

Markt operationele BI-hulpmiddelen wordt volwassen

Real-time: de ontbrekende schakel in BI?

Jos van Dongen

Real-time, right time, event driven en operational BI, 'managing at the speed of thought', complex event processing, business activity monitoring; we kunnen zo nog wel even doorgaan met hype-termen. Inderdaad, we hebben het over het beschikbaar maken van informatie voor het nemen van beslissingen op het moment dat deze informatie ontstaat.

De aandacht voor dit fenomeen is niet nieuw. Er is in dit blad al regelmatig aandacht besteed aan het verkorten van de tijd tussen business events en de daaruit resulterende (management) acties. Diverse nieuwe tools als StreamBase, Progress Apama en recentelijk AltoSoft zijn de revue gepasseerd. Toch is er wel iets aan de hand. Was Apama in 2003 nog een interessante nieuwkomer met een stuk veelbelovende technologie, inmiddels is het een hoeksteen van het Progress SOA portfolio. Ook andere aanbieders zijn de startup-fase al lang voorbij en hebben hun eigen niche gevonden (StreamBase) of zijn overgenomen door een grotere partij (Celequest). Kortom, de markt begint trekjes van volwassenheid te vertonen en dat is alleen maar goed nieuws. De inzet van operationele (real-time) BI-hulpmiddelen kan namelijk zeer effectief zijn bij het monitoren en optimaliseren van uw bedrijfsprocessen. Dat er nu een ruime keus is aan degelijke tools is dan ook mooi meegenomen.

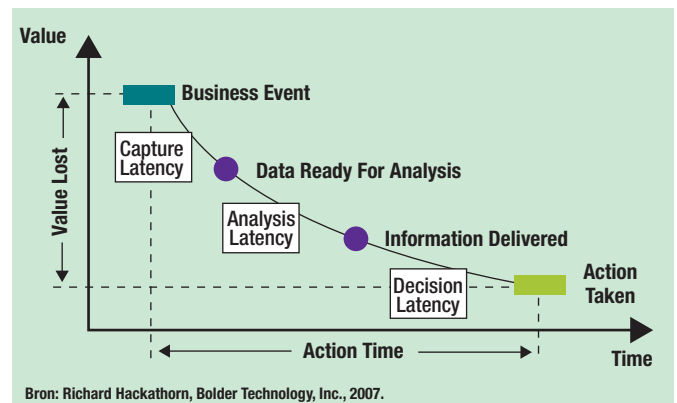
Achtergrond

Real-time BI, event driven BI en operational BI zijn drie termen die vaak door en naast elkaar gebruikt worden, waardoor het lijkt alsof het over hetzelfde gaat. Toch zijn er enkele subtiele verschillen. Operational BI is naast strategische en tactische BI de derde loot aan de BI-stam en heeft een paar afwijkende kenmerken. Zo is de business focus gericht op de operatie (vandaar de naam) en het optimaliseren van de primaire bedrijfsprocessen. Hier ook geen executives of analisten aan de knoppen, maar lijnmanagers en werknemers op de werkvloer. De data dienen actueel te zijn en hebben een beperktere horizon, vaak een dag of soms nog minder. En omdat deze data zo actueel mogelijk dienen te zijn spreken we bij de verwerking van de gegevens over *intraday*, near real-time of zelfs real-time. Real-time is dus meer een kenmerk van de data dan dat het een apart onderdeel van het BI-spectrum betreft, maar omdat real-time

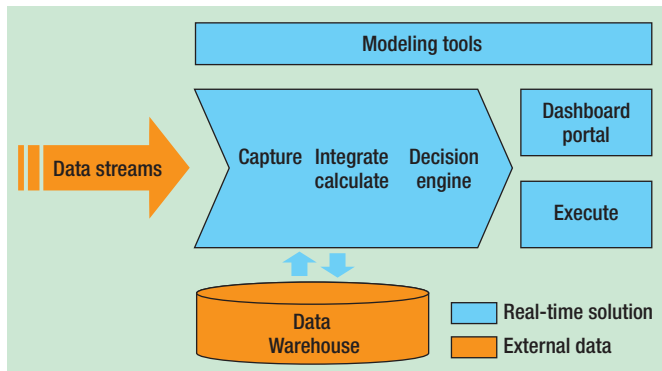
dataverwerking vrijwel altijd betrekking heeft op operationele BI (terwijl dit andersom niet zo hoeft te zijn) worden beide termen vaak als synoniemen gezien.

Een andere dimensie in dit verhaal is Event driven BI. Event driven BI wordt gekenmerkt door de wijze waarop het operationele of real-time systeem wordt aangestuurd. Gegevens worden hierbij niet opgehaald maar door het proces zelf gegenereerd en een 'event processor' ingestuurd. Wanneer het hierbij om veel, verschillende en/of gecompliceerde data gaat wordt ook wel gesproken over 'complex event processing'. Meestal gebeurt dit in real-time met als doel de operationele processen beter te monitoren en te besturen, waarmee de cirkel naar operationele BI weer rond is.

Van oudsher is BI ingezet voor tactische en strategische besluitvorming, en de data die hiervoor nodig zijn hoeven niet tot op de minuut (of zelfs seconde) actueel te zijn. Het is voldoende voor de 'klassieke' BI-toepassingen om data in een datawarehouse dagelijks te verversen. Voor operationele doeleinden



Afbeelding 1: Tijd versus waarde.



Afbeelding 2: Real-time engine.

gelden echter andere eisen die vooral betrekking hebben op het detailniveau en de actualiteit van de gegevens. In het algemeen zal gelden dat hoe meer tijd verstrijkt tussen een business event en de genomen actie, hoe meer waarde verloren gaat, zie afbeelding 1.

Voor sommige toepassingen zoals de verhandeling van aandelen en fraudedetectie is het evident dat de data tot op de (milli) seconde actueel dienen te zijn, voor andere ligt dit wellicht wat minder voor de hand. Denk bij dit laatste bijvoorbeeld aan een klantprofiel zoals dit gebruikt wordt voor (direct) marketingdoeleinden. Hoe actueel moet dit profiel zijn? Dat ligt er aan. Voor mailingdoeleinden kan prima met een (nacht)batch gewerkt worden. Voor e-commerce toepassingen ligt dit wat anders; het kan zijn dat de laatste transactie n t voldoende is om de klant in een ander segment te laten vallen en dat er een ander aanbod gedaan kan worden. Het is duidelijk dat dit behoorlijk wat eisen aan de achterliggende infrastructuur stelt. In het algemeen echter kan gesteld worden dat processen sneller wijzigen, en dat de druk op doorlooptijden ook steeds groter wordt. Denk maar eens aan het activeren van een mobiele telefoon. Waar dit tot voor enkele jaren terug nog enkele dagen in beslag nam loopt men inmiddels bellend de winkel uit. Of het bestellen van een 'built to order' PC: niemand meer neemt genoegen met een wachttijd van vier weken, maar men verwacht (eist) dat het apparaat de volgende dag al afgeleverd wordt. Dit betekent niet alleen dat alle processen optimaal ingericht moeten zijn om dit te kunnen ondersteunen, maar ook dat elke afwijking van dit proces onmiddellijk dient te leiden tot een alarm en bijbehorende actie. En dit is nu precies wat de nieuwe generatie operationele real-time BI-tools doet.

Toepassingen

Binnen het totale BI-werkveld is en blijft operationele BI een specifieke tak van sport, maar   n die zo snel groeit dat van een niche eigenlijk geen sprake meer is. De meest voor de hand liggende toepassing is de elektronische handel. Het kan hierbij gaan om aandelen en derivaten, maar ook om energie. Een andere veel voorkomende toepassing is het bewaken van netwerken. Dit kunnen computernetwerken zijn, maar bijvoorbeeld ook het spoor en telecom vallen in deze categorie. Eigenlijk zijn

alle activiteiten die een continue stroom van events genereren geschikt voor het analyseren met behulp van een event processing engine. Wat dacht u bijvoorbeeld van het monitoren van voertuigbewegingen via GPS, het volgen van goederen met behulp van RFID tags of het bewaken van hartslag, ademhaling en temperatuur van pati nten? Ook binnen de online gok- en game-wereld wordt gebruik gemaakt van event analyse, bijvoorbeeld om ongewenste vormen van samenwerking te traceren.

Wil de 'echte real-time' nu ontstaan?

Elke zichzelf respecterende BI-leverancier heeft op dit moment wel   n of meerdere 'real-time' oplossingen in het portfolio. Vooral de ETL-, database- en appliance-vendors schermen met het feit dat ze real-time, message based, data kunnen verwerken. Hoe waardevol dit misschien ook is, het is maar een deel van de oplossing. In afbeelding 1 is te zien dat hiermee het eerste deel van de vertraging wordt opgelost, namelijk de 'capture latency'. Dit deel van de *action time as* kunnen we aanduiden als real-time datawarehousing, en het eindresultaat is dat de data beschikbaar zijn voor analyse, niets meer en niets minder.

De volgende stap bestaat uit het analyseren van deze data, de tijd die dit kost wordt aangeduid met 'analysis latency'. Het real-time maken van deze fase in het proces betekent dat er regels vastgelegd dienen te worden op basis waarvan automatisch een bepaalde betekenis gegeven kan worden aan de beschikbare data. Dit lijkt moeilijker dan het is; vaak zult u al over dit soort regels beschikken en worden ze ook al gebruikt. Denk hierbij aan een management dashboard waarin de positie van een meter bepaald wordt door de waarde van de data, en de grenzen voor groen, geel en rood (de regels) een indicatie geven van de betekenis van deze waarde. Ook in een automatisch ververst top-10 rapport met de regel 'hoogste omzet' zit al iets van een analyse verstopt, en u kunt vast zelf nog de nodige voorbeelden bedenken.

Het real-time maken van de laatste fase en dus het wegnemen van de 'decision latency' is de meest verstrekkende. Het automatisch verwerken en analyseren van gegevens is   n ding, het daadwerkelijk actie ondernemen op basis van de resultaten van de analyse is iets waarvoor we in de meeste gevallen toch nog graag enige menselijke tussenkomst zien. Minimaal vinden we het prettig als deze beslissingen nog 'overruled' kunnen worden, of wanneer het systeem tijdelijk uitgeschakeld kan worden.



Afbeelding 3: Cognos Now! Appliance.

Maar wanneer is een systeem nu écht real-time? En bestaat dit wel? Strikt genomen natuurlijk niet; er zal altijd een vertraging zijn, zelfs als het maar enige nanoseconden is. Dit geldt zelfs bij het verwerken van streaming data vanuit message queues. Toch is het handig om daar het onderscheid te maken. Real-time BI heeft betrekking op streaming data of events, zodra de data worden opgeslagen voordat deze kunnen worden geanalyseerd is er hooguit sprake van 'near' real-time. Dit onderscheid betekent dat de meeste oplossingen zoals we deze nu kennen in deze laatste categorie vallen. De echte real-time systemen kenmerken zich door de volgende eigenschappen:

- event based capturing – afvangen business events en bijbehorende data zodra ze optreden, meestal met behulp van een message queue;
- in-memory analyse – in-memory filteren, aggregeren en berekenen van resultaten op basis van deze data;
- process execution – direct toepassen van regels en het uitvoeren van updates of andere acties.

Voor elk onderdeel van een dashboard kan een interval worden ingesteld

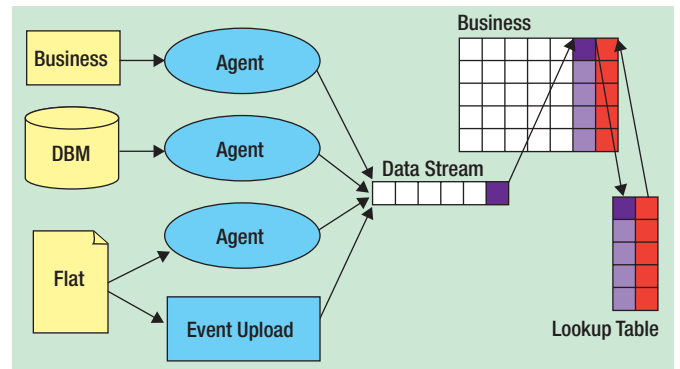
Daarnaast is er over het algemeen een workbench beschikbaar voor het modelleren van processen en regels, en beschikken de tools over een of andere vorm van visualisatie, meestal in de vorm van een portal. In afbeelding 2 is te zien hoe globaal gezien alle real-time of (complex) event processing oplossingen in elkaar zitten. Aan de linkerzijde van het diagram bevinden zich de data streams; vrijwel alle pakketten kunnen overweg met JMS, MQ, Tibco en Webservices, de zogenaamde infrastructuur adapters. Maar ook direct inprikken op Reuters, Wombat of het lezen van RFID- en GPS-data behoort al tot de mogelijkheden.

De modelleer-tools worden in de vorm van een workbench aangeboden en dienen niet alleen om business rules te definiëren maar ook om dashboards te bouwen. Vaak zijn dit verschillende tools met verschillende interfaces.

Het hart van het systeem wordt gevormd door de 'engine'. Dit is het werkpaard van de oplossing en zorgt voor extractie, filteren, berekenen en aggregeren van (event) data. Ook het toepassen van regels op binnengekomen events en het afvuren van vervolgacties gebeurt binnen deze engine. De meeste leveranciers hebben hiervoor een eigen, proprietary query taal ontwikkeld, omdat SQL niet geschikt is voor het bevragen van streams.

Bij Oracle is dit bijvoorbeeld CQL (Continuous Query Language) en Streambase beschikt voor dit doel over StreamSQL.

Aan de rechterzijde staan dashboards en uitvoering apart vermeld; een dashboard zorgt voor visualisatie ten behoeve van een eindgebruiker of analist en biedt extra mogelijkheden voor



Afbeelding 4: Eventstream processing.

analyses of het in gang zetten van vervolgacties. Met 'uitvoering' wordt bedoeld het direct starten van deze acties zonder menselijke tussenkomst, bijvoorbeeld door het bijwerken of afvuren van events naar andere systemen.

Om een en ander verder te verduidelijken is het altijd goed om een concreet voorbeeld te beschrijven. Nu zijn er meerdere tools die zich hiervoor lenen, maar een zoektocht in het DB/M archief leert dat er één product is dat een prominente positie in deze markt inneemt én waar vreemd genoeg nog nooit aandacht aan is besteed in dit blad. Het gaat hierbij om het voormalige Celequest, een product dat sinds alle overnames in de BI-markt als IBM Cognos Now! door het leven gaat.

Cognos Now!

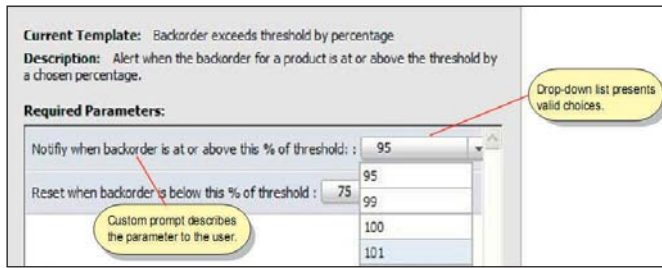
Het bijzondere aan Cognos Now! is dat we hier in eerste instantie niet te maken hebben met een software-oplossing, maar met een heuse appliance (zie afbeelding 3). Dat wil zeggen dat de complete oplossing in een stekkerklare server wordt geleverd en dat tijdrovende installatie- en configuratiewerkzaamheden niet nodig zijn. Afhankelijk van de omvang van de gebruikersgroep en de te analyseren data zijn er verschillende modellen beschikbaar met respectievelijk 8, 16 of 32 GB intern geheugen en 400 of 800 GB interne schijfcapaciteit.

De zwaarste uitvoering beschikt over twee quadcore processoren, wat uiteraard wel grenzen oplegt aan de maximale verwerkingscapaciteit van één machine. Naast de fysieke appliance levert Cognos het pakket ook als virtuele machine; hierbij worden de verschillende beschikbare hardware appliance-modellen afgebeeld in een virtuele VMWare ESX 3 machine. Nog steeds gelden dan de processor- en geheugenbeperkingen van de fysieke uitvoeringen, waarschijnlijk om redenen die met licentie-inkomsten te maken hebben. Het is namelijk wél mogelijk om de omgeving op te schalen door extra Cognos Now! servers toe te voegen. En wilt u helemaal geen extra infrastructuur in huis, dan biedt IBM de oplossing ook aan in een SaaS vorm.

Onder de motorkap

Het hart van het systeem wordt gevormd door een 64-bit processing engine met een zelf ontwikkelde in-memory database.

Ook de query-taal is een eigen ontwikkeling en heet C-SQL



Afbeelding 5: Voorbeeldregel.

(een gelukje voor Cognos dat Celequest ook met een C begint). C-SQL is een subset van ANSI-SQL99 en query's volgen dan ook het welbekende SELECT FROM WHERE stramien van reguliere SQL statements. De taal is vervolgens uitgebreid met specifieke elementen om met data streams en events overweg te kunnen. De meest in het oog springende toevoegingen hebben betrekking op het concept 'Windows' (nee, niet die van Microsoft). Een Window specificeert de set van rijen in een statement ten opzichte van de huidige rij in een specifieke event. Dit Window kan betrekking hebben op tijd (time-series window) en op events (event-series window). Een statement om de tien events te totaliseren, beginnend twaalf events terug ziet er als volgt uit:

```
SUM(Qty) OVER (EVENTS BETWEEN 11 PRECEDING AND  
2 PRECEDING) AS Total_Qty
```

Om hetzelfde totaal over een bepaald tijdsvenster (in dit geval van 18 uur en 15 minuten tot 45 seconden geleden) te berekenen dient een range opgegeven te worden:

```
OVER (RANGE BETWEEN INTERVAL '18:15' HOUR TO MINUTE  
PRECEDING AND INTERVAL '45' SECOND PRECEDING
```

Voordat deze query's kunnen worden uitgevoerd dient er uiteraard nog een en ander te worden ingericht. Zoals gezegd kan Cognos Now! met verschillende soorten databronnen en streams overweg. Data kunnen zowel vanuit bronnen worden opgehaald ('pull') als automatisch worden ingelezen ('push'). Beide manieren van data-acquisitie vormen intern een data stream die met behulp van zogenaamde Business Views kan worden benaderd.

Cognos Now! kan met verschillende soorten databronnen en streams overweg

De Business Views vormen het real-time datamodel en worden met behulp van de Cognos Now! Workbench gecreëerd. Het mooie van de Business Views is dat hierin zowel data uit streams als data uit bijvoorbeeld een datawarehouse kunnen worden verbonden. Zo is het dus mogelijk om binnenkomende waarden uit de streams te vergelijken met langjarige trends uit een DWH, en hoeven niet alle data beschikbaar gemaakt te worden in de

in-memory database. Afbeelding 4 verduidelijkt dit principe. De lookup table in het schema bevat data uit een 'statische' bron, de datastream data uit een 'dynamische' bron.

Alerts

Eén van de belangrijkste functies van een real-time oplossing is het genereren van 'alerts' op basis van business rules. Dit kunnen meldingen zijn van afwijkingen van een ingestelde grenswaarde of van het optreden van bepaalde events. Cognos Now! kent twee typen alerts: *stateless* en *statefull*. Bij een stateless alert wordt verder niet naar de toestand van een bepaalde waarde gekeken maar gewoon een bericht gestuurd met de melding van de afwijking. Een statefull alert houdt wél de toestand bij en is in staat om naast het geven van de melding deze ook weer in te trekken als de waarde weer onder (of boven) een ingestelde grens komt. Stel dat bijvoorbeeld de temperatuur van een processor te hoog oploopt. Hierdoor kan een systeem instabiel worden en verschijnt een alarm in beeld, maar wanneer de temperatuur weer genormaliseerd is verdwijnt het alarm weer. Denk bij alerts trouwens niet alleen aan meldingen in een dashboard, via e-mail of SMS, maar ook aan het aanroepen van een webservice. Op deze manier kan Cognos Now! ook ingezet worden om vervolprocessen aan te sturen.

Workbench

Het ontwikkelen van een Cognos Now! oplossing begint met de al eerder genoemde Workbench. Het spreekt voor zich dat eerst de benodigde dataconnecties dienen te worden aangelegd voordat er iets anders kan gebeuren. Het belangrijkste is echter het modelleren van de business data. Dit houdt in dat op basis van datastreams en lookup tables de business views worden gebouwd. Deze business views zijn modellen die een (deel van) een bepaalde business activiteit weergeven, en vormen de bouwstenen van zogenaamde scenario's. Een scenario bevat regels, alerts en report portlets (weergavecomponenten voor dashboards) die betrekking hebben op één specifiek proces of een verandering in data. Iets als 'Voorraadwijziging' is een voorbeeld van een scenario, maar ook 'Bestelling' of 'Annulering' zou een scenario kunnen zijn. Scenario's worden gegroepeerd in 'Business Activities' als Helpdesk of Verkoop zodat een logische onderverdeling van bedrijfsprocessen mogelijk is. Dit biedt tevens mogelijkheden om op zowel activiteit- als scenarioniveau autorisaties toe te passen zodat elke gebruikersgroep alleen toegang krijgt tot de voor hem of haar relevante gegevens. Het maken van rules en alerts is met de Workbench een fluitje van een cent. Mocht u wellicht afgeschrikt zijn door de beschrijving onder het kopje 'onder de motorkap', dan is dit geheel onterecht. Alle regels kunnen met een eenvoudige GUI worden opgebouwd. Ook hoeft er niet steeds opnieuw begonnen te worden: het is mogelijk om templates te maken (ook via de GUI) die het raamwerk voor een regel definiëren, waarbij vervolgens alleen nog maar de specifieke details hoeven te worden ingevuld. De template kan bijvoorbeeld bestaan uit de algemene

regel 'backorder-percentages overschrijdt grenswaarde', en de bijbehorende alerts specificeren wat er moet gebeuren als deze regel wordt afgevuurd. Op detailniveau hoeven vervolgens alleen de parameters te worden ingevuld, in dit geval alleen de grenswaarde of in het geval van een stateful alert ook de ondergrens, zie afbeelding 5.

Dashboards

Tot nu toe ging het voornamelijk over de 'achterkant' van Cognos Now! Dit is niet toevallig, aangezien al het marketingmateriaal en ook de online beschikbare demo voornamelijk over de 'voorkant', het Dashboard, gaat. Uiteindelijk is dit wel het deel waar de meeste gebruikers mee zullen werken, maar het is niet het onderdeel waarin het product uniek is. Er zijn meer op Flash technologie gebaseerde dashboardproducten op de markt. De kracht zit met name in de real-time data en rules engine, het dashboard verzorgt voornamelijk de presentatie, hoewel dat wellicht wat erg kort door de bocht is. Er zit namelijk wel enige overlap in hetgeen binnen de Workbench gebouwd kan worden en de functionaliteit die het dashboard biedt. Dashboard is eigenlijk ook een iets te beperkt begrip aangezien een gebruiker niet alleen dashboards kan ontwikkelen (door content elementen toe te voegen) maar ook in staat is om zelf regels, alerts en acties te definiëren. De basis hiervan blijven de business views en rule templates die de bouwstenen vormen voor de grafieken, indicatoren en tabellen waarmee een dashboard wordt opgebouwd, zoals in afbeelding 6 is te zien.

Voor elk onderdeel van een dashboard kan een interval worden ingesteld waarmee de gegevens worden ververs, zodat het geen kakofonie van knipperende lampjes en metertjes wordt als er elke milliseconde nieuwe data binnenkomen. Elk element kan ook worden aangepast naar de eigen behoefte, zowel qua content als qua layout, en de data kunnen op verschillende manieren worden gefilterd. Een filter kan ook een globaal karakter hebben en heet dan een 'perspective'. Dit betekent dat op dashboard niveau een selectie (doorsnijding) wordt toegepast die voor alle onderdelen van het dashboard geldt. Wanneer een grenswaarde overschreden wordt kunnen automatisch acties worden afgevuurd, maar dit kan ook resulteren in het toekennen van een bepaalde taak aan een gebruiker die tot doel heeft om het probleem te verhelpen. In de taaklijst kan elke gebruiker zien wat er nog voor hem of haar openstaat aan niet afgehandelde acties. Cognos Now! zorgt dus niet alleen voor de notificatie, maar biedt tevens ondersteuning voor de bijbehorende workflow.



Afbeelding 6: Dashboard.

Een complete BI-oplossing?

Eén vraag die vaak wordt gesteld over oplossingen à la Cognos Now! is of het voor een bedrijf volstaat om dit als enige BI-tool in huis te halen. Dit lijkt een aantrekkelijke optie, want zelfs zonder dat de specifieke real-time opties worden gebruikt kan een dergelijke tool ook prima met een datawarehouse overweg.

Een veel voorkomende toepassing is het bewaken van netwerken

Door het handig opzetten van de business views en door gebruik te maken van de mogelijkheden die de grafieken en tabellen bieden kan men ook een aardig eindje komen. Toch is het de vraag of dit verstandig is. Real-time of event processing oplossingen zijn ontworpen om meer grip te krijgen op primaire processen. De onderliggende ontwerpfilosofie is dan ook operationeel van aard, terwijl een 'klassieke' BI-oplossing meer gericht zal zijn op de ondersteuning van tactische en strategische besluitvorming. De middelen hiervoor overlappen deels (alerts en exception highlighting) maar deels ook niet. Real-time tools blinken niet uit in geavanceerde analysemogelijkheden als OLAP, datamining en forecasting, en ook het maken van complexe *pixel perfect* rapportages kunt u beter overlaten aan andere pakketten. IBM heeft dit natuurlijk zelf ook al bedacht en biedt daarom de mogelijkheid om vanuit de 'klassieke' Cognos 8 BI-tools informatie binnen de IBM Cognos Now! omgeving te benaderen.

Conclusie

Real-time BI-tools kunnen een nuttige aanvulling zijn op het bestaande portfolio aan BI-hulpmiddelen, maar u moet zich wel realiseren dat ze een heel specifiek doel dienen en nooit een volledige BI-oplossing kunnen bieden. Met name de gebruikersgroep zal anders samengesteld zijn en voor een groot deel bestaan uit medewerkers die de dagelijkse problemen op de werkvloer moeten zien op te lossen. Hiervoor krijgen ze echter met tools als Cognos Now! wel een eenvoudig te gebruiken, krachtig en mooi instrument in handen.

Literatuur

Building the Real-Time Enterprise, Colin White voor TDWI, 2003.
Managing the Real-Time Enterprise, Richard Hackathorn, 2003.
Are we there yet? Trends in business intelligence and beyond, seminar Claudia Imhoff, 2007.
Ebizq CEP site (www.ebizq.net/hot_topics/cep), 2008.
Complex Events site (<http://complexevents.com>), 2008.
Cognos Now! Manuals, 2008.

Jos van Dongen

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