

ETL-tool of Query Server?

Dossier Semi ETL-tools

Freek Kamst

De negatieve ervaringen die in het recente verleden zijn opgedaan met de ontwikkeling van datawarehouse-omgevingen, hebben ertoe geleid dat er is gezocht naar alternatieve toepassingen die dezelfde resultaten kunnen leveren tegen minder inspanningen en kosten. De kosten van een datawarehouse-omgeving hebben voornamelijk betrekking op de ontwikkeling van de extractie-, transformatie- en laadprocessen (ETL).

ETL-tools reduceren deze kosten aanzienlijk, maar een investering in een dergelijke tool is voor veel organisaties nog steeds geen haalbare situatie. De leveranciers van

Business Intelligence-oplossingen bieden daarom vaak hun query server-toepassing aan als alternatief voor een ETL-tool en er zijn inmiddels ook leveranciers die door

De verleiding om snel met een query server van start te gaan is hiermee gecreëerd

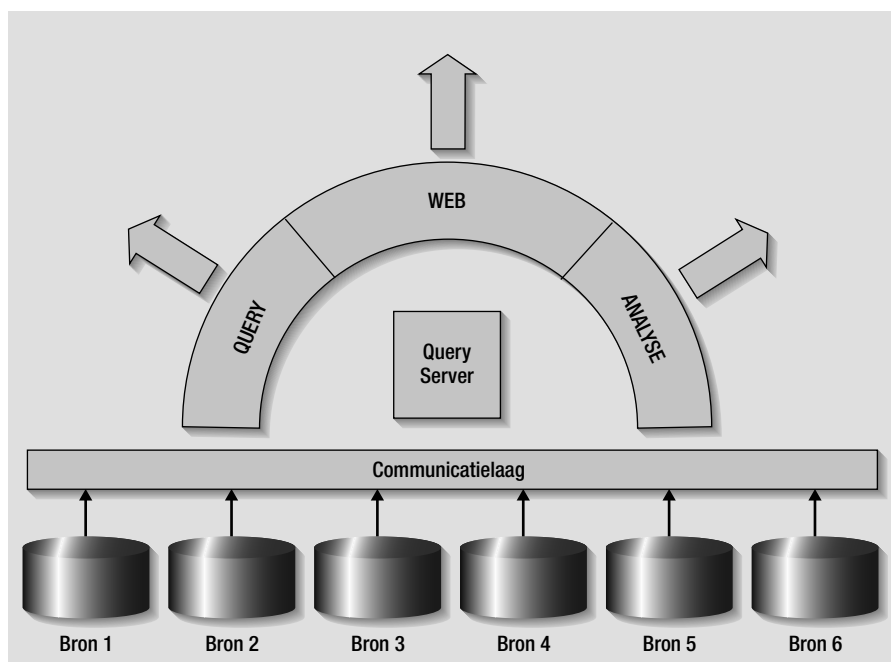
middel van een generieke query-tool een goedkope semi ETL-oplossing aanbieden. Welke oplossing de meeste kans van slagen heeft en wanneer een datawarehouse-omgeving of een query server-toepas-

sing moet worden gekozen is een dilemma waar vandaag de dag nog steeds veel organisaties mee worstelen. In dit artikel wordt dieper ingegaan op de mogelijkheden met een query server-toepassing en aansluitend wordt een overzicht gegeven van de huidige marktontwikkelingen rondom Business Intelligence, die van invloed zijn op ETL-tools en query server-omgevingen.

DE QUERY SERVER

De query server bestaat uit een semantische communicatielaag bovenop één of meerdere brondatabases en zorgt ervoor dat de generieke gegevens uit deze databases kunnen worden gelezen. Vervolgens worden de gegevens getransformeerd en doorgegeven aan een rapportage- of analyse-module. Ook kunnen de gegevens in een bepaald formaat, zoals Excel of XML, worden opgeslagen. In de metagegevenslaag van de query server worden de gegevens ook nog eens uniform gepresenteerd, zodat de gebruikers geen kennis nodig hebben van de onderliggende gegevensstructuren in de brondatabases.

Met de query server komt informatie beschikbaar zonder een ingewikkelde datawarehouse- of Business Intelligence-toepassing. Hierdoor worden deze query servers dan ook vaak verkocht met het argument dat er geen datawarehouse of ETL-tool benodigd is. Soms terecht, maar ook vaak worden deze beslissingen genomen op basis van kostenbesparing en wordt er geen rekening gehouden met de informatiebehoefte op langere termijn.



FIGUUR 1: DE QUERY SERVER.

Leverancier	Product	Website	Modulaire open structuur	Analytische applicaties beschikbaar	Rapportage applicaties beschikbaar	Open database interface	Open Standaard metagegevens	Onderdeel van EBIS-structuur
Cognos	Cognos Query	www.cognos.com	•	•	•	•		•
Business Objects	Business Query	www.businessobjects.com	•	•	•	•	•	•
SAS	SAS/Access	www.sas.com		•	•	•	•	•
Crystal Decisions	Crystal Enterprise	www.crystaldecisions.com	•	•	•	•		
Hyperion	Hyperion Enterprise	www.hyperion.com		•	•	•	•	
Actuate	Actuate Server	www.actuate.com		•	•	•		
Information Builders	WebFOCUS	www.informationbuilders.com		•	•	•		•
MicroStrategy	Intelligence Server	www.microstrategy.com		•	•	•		
SAP	MySAP Enterprise Portal	www.sap.com		•	•	•	•	•
Oracle	E-Business Suite	www.oracle.com	•	•	•	•	•	•
Brio Software	Brio Intelligence	www.brio.com	•	•	•	•		
Hummingbird	BI Server	www.hummingbird.com		•	•	•		•
Inforay	Inforay Server	www.inforay.com		•	•	•		
Noetix Corporation	Query Server	www.noetix.com	•			•		
Sagent	Sagent Data Access Server	www.sagent.com		•	•	•		•
Informatica	PowerConnect	www.informatica.com		•	•	•	•	•
Keper	DBExplorer	www.keper.nl				•		
InformStrategy	Database Analyzer	www.informstrategy.nl				•		
Aruna	Aruna Logic Server	www.arunasoftware.com				•		

FIGUUR 2: LEVERANCIERS VAN QUERY SERVERS EN HUN KENMERKEN.

In veel praktijkgevallen is er door een projectgroep een onderzoek uitgevoerd, waarin wordt onderbouwd dat een datawarehouse omgeving op de lange termijn noodzakelijk is en worden de kosten hiervoor in het voorstel opgenomen. De verleiding om snel met een query server van start te gaan is hiermee gecreëerd, omdat op korte termijn hetzelfde resultaat kan worden verkregen en de initiële kosten aanzienlijk lager liggen.

Heel belangrijk is dat de keuze van een query server of een datawarehouse-omgeving onafhankelijk is van de informatiebehoefte en de manier waarop deze moet worden ingevuld. De transformaties van de gegevens blijven in beide gevallen hetzelfde. Echter bij een datawarehouse-omgeving zullen de transformaties eenmalig worden uitgevoerd en vastgelegd, terwijl een query server de resultaten telkens opnieuw zal moeten berekenen. De keuze om te starten met een query server moet dan ook weloverwogen worden, zodat een schaalbare groei naar een generiek informatieproces mogelijk is.

Het gebruik van een query server op de bedrijfskritische operationele systemen

bevat echter ook een aantal nadelen. De performance van deze systemen is vaak cruciaal ten aanzien van de beschikbaarheid van bedrijfsprocessen binnen een organisatie. Ook zijn de operationele systemen optimaal ontworpen om de afhandeling van transacties zo efficiënt mogelijk uit te voeren en wanneer op de databases van deze systemen omvangrijke queries

De relatief korte geschiedenis van ETL-tools heeft nu drie generaties voortgebracht

worden uitgevoerd, dan heeft dit zeker consequenties voor de beschikbaarheid. Een ander nadeel is dat de gegevensstructuren van de operationele systemen genormaliseerd zijn ontworpen, waardoor er geen of weinig mogelijkheden zijn om de gegevens vanuit diverse invalshoeken te analyseren. Ook de hoeveelheid historische gegevens binnen dergelijke systemen is meestal beperkt, zodat analyse over een langere periode niet altijd mogelijk is.

De voordelen van query servers zijn onder andere de mogelijkheden om op korte termijn aan een informatiebehoefte tegemoet te komen, waarbij de initiële kosten vele malen lager zijn dan een start met een ETL-tool en een datawarehouse. Eveneens is het direct raadplegen van actuele informatie vanuit de bronsystemen een voordeel dat niet altijd binnen datawarehouse-omgevingen mogelijk is.

Ondanks dat de investering minimaal lijkt, is het aan te bevelen om de keuze van een query server-omgeving zorgvuldig uit te voeren. Schaalbaarheid om daardoor mee te kunnen groeien met de organisatie en haar informatiebehoefte is uitermate belangrijk. Overige belangrijke aandachtspunten in een keuzetraject van een query server zijn:

- Is het product modulaair opgezet?
- Is er een rapportage- en analysecomponent aanwezig?
- Is de query server-database onafhankelijk?
- Voldoen de metagegevens aan een open standaard?
- Is de query server onderdeel van een Enterprise Business Intelligence Suite?

Het gaat te ver om alle query servers uitgebreid in dit artikel te benoemen, maar in figuur 1 wordt een aantal producten en hun voornaamste kenmerken in tabelvorm vermeld.

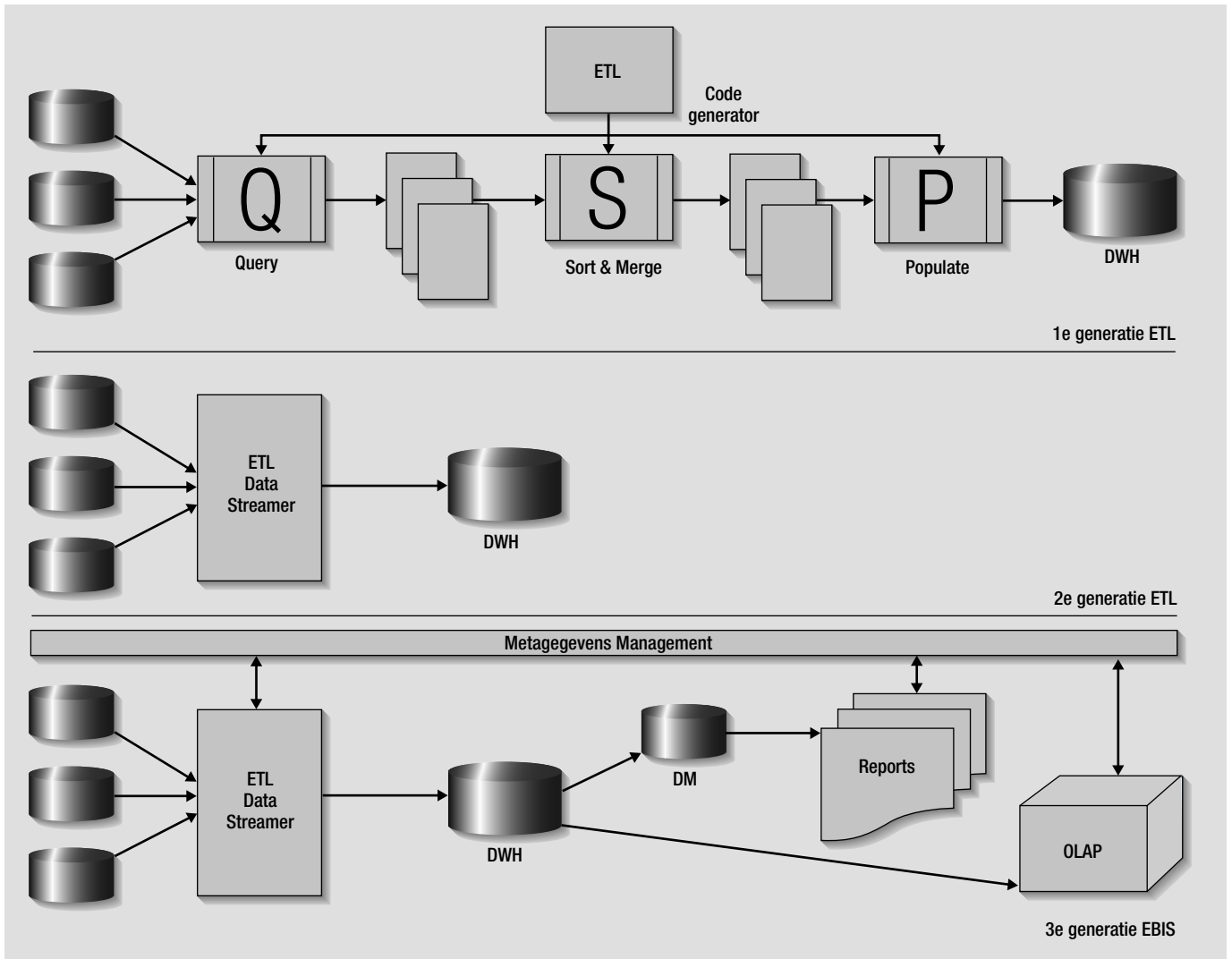
MARKTONTWIKKELINGEN IN BUSINESS INTELLIGENCE

Het is een feit dat de komende jaren organisaties de uitgaven ten aanzien van hun informatietechnologie scherp zullen blijven bewaken. Hierdoor zullen er meer korte en overzichtelijke projecten met meetbare resultaten worden uitgevoerd en ook het implementeren van Business Intelligence-omgevingen zal aan deze eisen moeten voldoen. Daarnaast zijn er ook nieuwe trends zichtbaar die organisaties kunnen helpen bij de ontwikkeling en

marketing van hun diensten en producten. Business Area Monitoring en Corporate Performance Management zijn onder andere nieuwe ontwikkelingen waar zeker de komende jaren aandacht voor zal worden gevraagd. Business Area Monitoring richt zich volledig op het direct waarnemen van de performance van een organisatie, terwijl bij Corporate Performance Management op een gestructureerde manier de businessplanning wordt uitgevoerd en gecontroleerd. Business Intelligence-omgevingen zijn een essentieel onderdeel van deze nieuwe ontwikkelingen en de verwachting is dat veel leveranciers de komende periode hier de aandacht op zullen gaan vestigen.

Binnen de omgeving van Business Area Monitoring zal ook real-time informatie een belangrijke rol gaan spelen.

Organisatiebrede real-time toepassingen zullen nog wel even op zich laten wachten, maar voor sommige specifieke informatiebehoeften is nu eenmaal direct informatie benodigd uit bronsystemen. Een voorbeeld van real-time informatie zijn de aandelenkoersen die constant aan verandering onderhevig zijn en die in de traditionele Business Intelligence omgeving niet direct kunnen worden opgevraagd. Immers datawarehouses worden met een minimale frequentie van één dag bijgeladen. Wanneer er bijvoorbeeld een actuele koers moet worden opgevraagd om direct de waarde van een aandelenportefeuille te bepalen, kunnen de portefeuillegegevens uit het datawarehouse worden opgevraagd en moeten de koersen direct uit het beurssystem worden gehaald, zodat de actuele waarde van de portefeuille on-line berekend kan



FIGUUR 3: DE DRIE GENERATIES ETL-TOOLS.

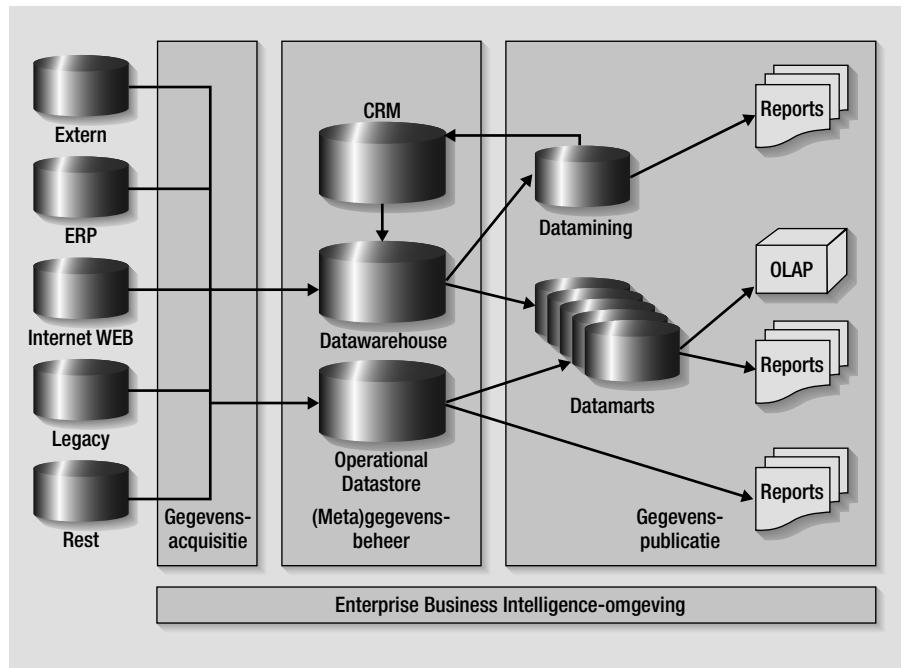
worden. Het produceren van real-time informatie is zeker een toepassing die door query servers kan worden uitgevoerd.

CONVERGENTIE

Eveneens een marktontwikkeling is dat nu ook Microsoft, PeopleSoft en SAP veel investeren in de ontwikkeling van Business Intelligence-toepassingen en deze vervolgens als add-on modules bij hun bestaande producten aanbieden. De concurrentiestrijd die hiermee gepaard gaat, noodzaakt de traditionele Business Intelligence-leveranciers om hun producten te innoveren. Een van de trends die uit deze verschuiving ontstaat, is de convergentie van Business Intelligence-toepassingen en ETL-tools. Waren deze twee werelden kort geleden nog gescheiden, binnen niet al te lange tijd zullen de ETL-tools opgaan in de zogenaamde Enterprise Business Intelligence Suites (EBIS). Binnen een Enterprise Business Intelligence Suite worden vanaf het bronsysteem tot aan proces van presentatie de toepassingen op elkaar afgestemd en kunnen de meta-gegevens uniform worden gebruikt. Een voorbeeld is Cognos dat een aantal jaren geleden de ETL-tool DecisionStream geïntegreerd heeft met haar bestaande producten Impromptu en Powerplay. Inmiddels wordt de gehele Cognos-productlijn als een Enterprise Business Intelligence Suite aangeboden. Ook Business Objects heeft onlangs de firma ACTA overgenomen en hiermee zullen ook de huidige Business Objects producten worden geïntegreerd met de ETL-tool ActaWorks. De specifieke ETL-tool leveranciers, zoals Informatica en Sagent, ontwikkelen daarentegen diverse Business Intelligence-toepassingen die in het verlengde van de ETL-tool worden geïntegreerd. Rapportage en OLAP-functionaliteit zijn inmiddels ook bij deze leveranciers standaard aanwezig.

ETL-TOOL HISTORIE

De Enterprise Business Intelligence Suites worden momenteel beschouwd als de derde generatie ETL-tools. Hierin zijn de



FIGUUR 4: DE ENTERPRISE BUSINESS INTELLIGENCE-OMGEVING.

ETL-toepassing en Business Intelligence-producten volledig met elkaar geïntegreerd door het eenduidig gebruik van de metagegevens. De relatief korte geschiedenis van de ETL-tools heeft dus nu inmiddels drie generaties voortgebracht.

De verwachting is dat er voorlopig geen nieuwe spelers meer zullen verschijnen

De eerste generatie ETL-tools bestaat voornamelijk uit programmageneratoren die query-, sort-, merge- en populate-programmatuur genereren en deze vervolgens laten uitvoeren op de bron- en doelsystemen.

De tweede generatie ETL-tools bestaat uit zogenaamde gegevensstreamers die op een specifieke server via het netwerk de gegevens naar zich toe trekken (extraheren), daarna deze gegevens op de server bewerken (transformeren) en als laatste de informatie in een database opslaan (laden). Beide generaties beperken zich tot het transformeren en laden van gegevens in een datawarehouse-omgeving.

De derde generatie ETL-tools is onderdeel van een complete suite aan programmatuur gericht op een Business Intelligence-omgeving. Hierbij is de ETL-tool een vaste component van het gehele informatieproces geworden.

Het resultaat van een EBIS-implementatie is een specifieke informatie architectuur die een Enterprise Business Intelligence-omgeving kan worden genoemd en de organisatie voorziet van actuele en kwalitatief hoogwaardige informatie. Deze architectuur zal ook de basis zijn voor de genoemde ontwikkelingen als Business Area Monitoring en Corporate Performance Management en zal uiteindelijk evalueren naar een bedrijfskritisch proces binnen een organisatie.

GROEISCENARIO NAAR EBIS

Omdat het niet aan te bevelen is om een complete Enterprise Business Intelligence-omgeving in één big-bang project te implementeren, is het advies om een strategisch groeiscenario te plannen. Hierdoor kan gefaseerd aan de informatiebehoefte van de organisatie tegemoet worden gekomen en worden ongecontroleerde investeringen voorkomen. Een door het manage-

ment ondersteunde informatiestrategie moet dan ook de basis zijn voor een dergelijk groeiscenario.

Een van de scenario's om van een query server naar een Enterprise Business Intelligence-omgeving te groeien is in de volgende vier stappen weergegeven:

1. Veelal wordt er gestart met een query server-toepassing op een beperkt aantal bronsystemen ten behoeve van een rapportage- of analysebehoefte. De analyse- en rapportagetoepassing is vaak geïntegreerd.
2. Na verloop van tijd ontstaat er een behoefte om meerdere bronsystemen op de toepassing aan te sluiten of om historische gegevens voor analytische doeleinden op te slaan. De gegevens moeten dan periodiek worden opgeslagen in een database, zodat er een geïntegreerde gegevensverzameling ontstaat. De basis voor een datawarehouse is dan gecreëerd.
3. Door de gegevens ook nog eens in een gedenormaliseerde structuur op te slaan, kunnen deze flexibel en vanuit meerdere invalshoeken geanalyseerd worden. Belangrijk moment in deze fase is de aansluiting van de query server omgeving op het datawarehouse. De gedane investeringen in rapporta-

Deze real-time informatie wordt in een aparte database opgeslagen en is meestal een kopie van de operationele bronsystemen

ges en analyses is hiermee dan gewaarborgd en de reeds gerealiseerde informatievoorziening wordt dan tevens gecontinueerd. Voor het frequent bijladen van het datawarehouse kan een ETL-tool worden ingezet en wanneer het geheel onderdeel uitmaakt van een Enterprise Business Intelligence Suite kunnen de meta-gegevens worden geïntegreerd.

4. Het ontsluiten van eventuele real-time informatie is een laatste fase om tot een

enterprise Business Intelligence-omgeving te komen. Deze real-time informatie moet in een aparte database worden opgeslagen en is meestal een kopie van de operationele bronsystemen. Deze real-time informatie database wordt ook wel Operational Data Store (ODS) genoemd. Voorbeelden van informatie in een ODS zijn openstaande transacties of offerte-informatie die gebruikt kan worden om financiële voorspellingen te doen. De frequentie van het bijladen in het ODS ligt vele malen hoger dan die van het datawarehouse.

TENSLOTTE

Het zal nog wel een aantal jaren duren voordat organisaties hun informatievoorzieningsproces optimaal zullen hebben ingericht, maar een feit is dat de leveranciers hun productstrategie momenteel steeds beter afstemmen op de actuele informatiebehoefte die binnen organisaties aanwezig is.

Ontwikkelingen in de informatietechnologie, zoals Web Services en de toenemende integratie van processen door middel van CRM, ERP en SCM, zorgen ervoor dat de vraag naar Business Intelligence-toepassingen zal toenemen. Organisaties blijven nu eenmaal de behoefte houden om zich te onderscheiden in hun branche en zullen daarom blijven zoeken naar mogelijkheden om sneller beslissingen te nemen. Daarnaast spelen er zeker voor de komende jaren nog de onvoorziene economische factoren. Kosten moeten constant bewaakt blijven en de voorkeur zal uitgaan naar kortere en overzichtelijke projecten met meetbare resultaten. De consolidatietrend (alle producten van één leverancier) zorgt er tevens voor dat de Business Intelligence-leveranciers hun product-assortiment compleet afstemmen op het vergaren, transformeren, analyseren en publiceren van informatie. Daarom is de verwachting dat er voorlopig geen nieuwe spelers meer op deze markt zullen verschijnen, met uitzondering van een aantal nieuwkomers uit de arena's van ERP, CRM en DBMS.

CONCLUSIE

De eindconclusie is dan ook, dat naast Business Intelligence-leveranciers ook de leveranciers voor CRM, ERP en DBMS zich concentreren op het leveren van complete productsuites waarmee enterprise Business Intelligence-omgevingen kunnen

Gebruik van een query server op bedrijfskritische operationele systemen bevat een aantal nadelen

worden ingericht. Echter de keuze en de manier waarop en hoe zo'n omgeving moet worden ingericht, blijft afhankelijk van de strategische informatiebehoeften binnen organisaties. Informatie is nu eenmaal naast kapitaal, middelen en mensen, één van de vier steunpilaren van een organisatie. De inrichting en de wijze waarop deze tot stand komt, moet altijd vanuit de strategische doelstellingen van de organisatie worden bepaald en niet door trends in informatietechnologie. ●

LITERATUUR

Building the Customer-Centric Enterprise. Claudia Imhoff, Lisa Loftis, Jonathan G. Geiger.
Uitgever: John Wiley & Sons Inc, 2001.

Freek Kamst (freek.kamst@newcom.net) is consultant bij Newcom Information Systems