

ETL-software wordt centraal component van informatiemanagement

# Trends in ETL-tools

Freek Kamst

**M**ees kritiek in een datawarehouse-project is en blijft de ontwikkeling van processen voor extractie, transformatie en laden (ETL). Vast staat dat bij meer dan zeven -tijdrovende-ETL-processen binnen een organisatie, een ETL-tool al een zinvolle investering is. De betreffende softwareproducten zijn de laatste jaren uitgegroeid tot toepassingen voor algemene gegevens-integratie. Ze worden inmiddels ook buiten datawarehouse-omgevingen ingezet in trajecten van migratie en integratie. Freek Kamst kijkt naar heden en toekomst in de ETL-markt.

Werden extractieprocessen voorheen nog wel ontwikkeld met behulp van traditionele programmeertalen, tegenwoordig is het ETL-tool een begrip geworden binnen elke datawarehouse-omgeving. Ook de toch nog omvangrijke investering die dergelijke software met zich meebrengt, vormt vandaag de dag geen discussiepunt meer. Zij verdient zichzelf snel terug door de aanzienlijke tijdbesparingen in ontwikkeling en onderhoud.

Databaseleveranciers hebben ETL-tooling inmiddels in hun assortiment opgenomen en leveren deze als extra functionaliteit bij hun database-engine. Ook de opkomst van nieuwe technologieën, zoals ERP, CRM en e-commerce, vragen om geïntegreerde informatievoorziening. Veel softwareleveranciers blijven bovendien vasthouden aan hun eigen standaards, waardoor de klant ETL-tools steeds vaker moet inzetten om een uniform geïntegreerd beeld te realiseren.

Redenen genoeg om eens stil te staan bij de huidige situatie in de markt voor ETL-tools en de ontwikkelingen en trends naar de toekomst nader te beschouwen. Wereldwijd groeit deze markt momenteel met 11% per jaar. De verwachting is dat de toename de komende jaren gelijk zal blijven, als het huidige economische klimaat aanhoudt.

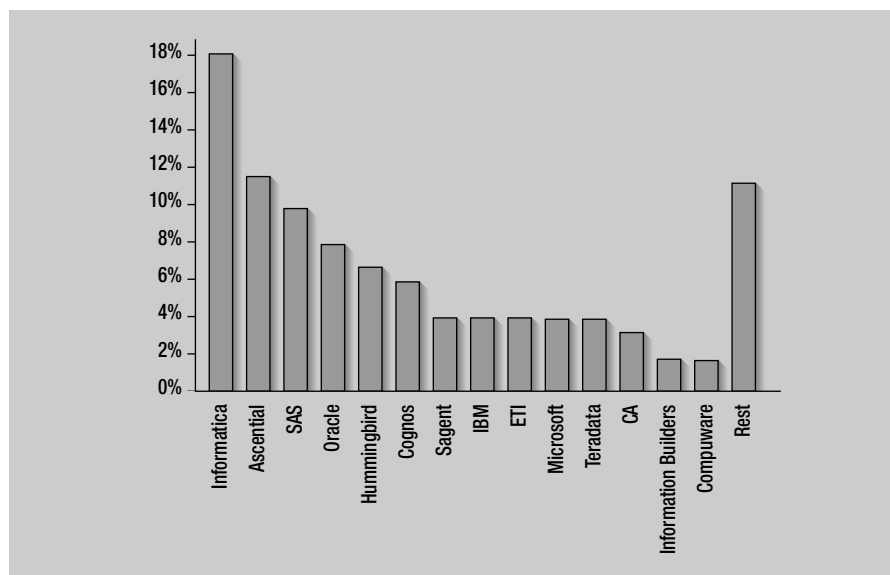
Onbetwiste marktleider is nog steeds Informatica, maar de snelste groeiers zijn op dit moment Microsoft en IBM. Het marktaandeel van Microsoft is in 2001 gestegen van 2% naar 6%, terwijl IBM met Data Warehouse Manager haar aandeel zag verdubbelen van 2% naar 4%. Reden hiervoor is dat databaseleveranciers steeds meer ETL-technologie zonder meerprijs meeleveren met de onderliggende database.

De trends die van invloed zijn op de ontwikkeling van ETL-tools liggen mede ten grondslag aan de visie en strategie van de leveranciers ervan. De belangrijkste trends die opgeld doen zijn:

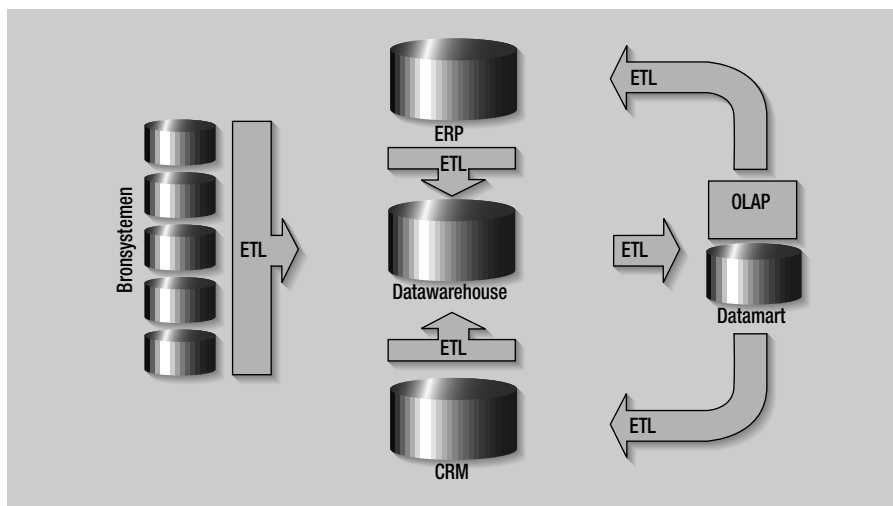
- het meeleveren van ETL-technologie met de onderliggende database;
- integratie van gegevens met ERP- en CRM-toepassingen;
- gebruik van de XML-standaard;
- integratie en communicatie met message brokers;
- datakwaliteit wordt steeds belangrijker;
- aantal gegevens neemt toe.

## TREND I: MEELEVEREN

De 'meelever'-trend is in 1999 in gang gezet door Microsoft. Zij kwam toen met



FIGUUR 1: DE VERDELING VAN DE MARKT VOOR ETL-TOOLS OVER DE DIVERSE LEVERANCIERS.



FIGUUR 2: INZET VAN ETL-TOOLS BIJ ERP- EN CRM-TOEPASSINGEN.

haar Data Transformation Services (DTS), een combinatie van extra functionaliteit voor datawarehouses, gebaseerd op en als onderdeel van SQL Server. Meegeleverd worden een relationeel schema, een OLAP-kubus, ETL-toolroutines en interfaces voor metadata-beheer. Hoewel ze niet uitgebreid is, voldoet de functionaliteit voor de bouw van redelijke business intelligence-toepassingen en dient het dikwijls als basis voor een professionele datawarehouse-omgeving.

Microsoft had succes en kreeg dan ook al snel, in 2000, navolging. IBM doopte haar Visual Warehouse om in Data Warehouse Center en leverde het gratis mee met de Windows NT-versie van DB2. De functionaliteit is echter beperkt ten opzichte van de andere platforms waarop DB2 draait. Wie een niet-NT-versie, met meer functionaliteit, wil aanschaffen, moet in licenties investeren.

Oracle heeft de productstrategie gebaseerd op integratie van ETL-, OLAP-en datamining-functionaliteit. De nieuwe versie Oracle 9i bevat standaard al beperkte ETL-functionaliteit gebaseerd op PL/SQL en ETL-SQL. Wie volledige functionaliteit wenst, kan dat krijgen door aanvullende modules te kopen, zoals Oracle Warehouse Builder en Data Warehouse Manager.

De grote drie hebben de ontwikkeling van hun ETL-functionaliteit alle gegroepeerd rondom de eigen database-engine. Dat blijft natuurlijk een risico, evenals het gegeven dat vroeg of laat de licentiestructuur zal wijzigen. En veranderingen vereisen meestal extra investeringen.

### TREND 2: ERP, CRM

De tweede trend stoelt op de toename van het aantal ERP- en CRM-systemen. Beschikbaarheid van historische klant- en productiegegevens is essentieel voor een juiste werking van deze systemen. Door analyses uit te voeren op deze historische gegevens kan men aannamen en patronen terugkoppelen naar de ERP- en CRM-systemen. Bij de eerste kan dit leiden tot bijsturing van het productieproces, terwijl bij CRM onder meer klantwaarden kunnen worden bepaald die ondersteunend zijn aan het verkoopproces.

Uiteindelijk heeft alles natuurlijk te maken met de behoefte gegevens geïntegreerd beschikbaar te maken, zodat meer inzicht ontstaat in de bedrijfsprocessen. Niet alleen bij ERP- en CRM-systemen wordt de relevantie van het informatie-component onderschat, maar ook bij vele andere standaard applicaties geniet het proces zelf veelal de aandacht. Een ETL-tool biedt dan uitkomst. Het helpt op een flexibele manier de gewenste informatie zichtbaar te maken.

### TREND 3: XML

XML is inmiddels uitgegroeid tot een universele uitwisselingstaal voor gestructu-

reerde en ongestructureerde gegevens. Veel softwareleveranciers ondersteunen reeds de mogelijkheid van dataverwerking en -aanmaak in XML.

XML-inzet door ETL-tools kan op twee manieren. De eerste optie is gegevensstromen vanuit de bronsystemen in XML-formaat te verwerken. Daarnaast zijn metagegevens in XML-formaat worden op te slaan, zodat deze beschikbaar kunnen komen voor andere toepassingen, zoals ontwikkel- en metadata-beheersystemen.

### TREND 4: MESSAGE BROKERS

Message brokers verzorgen op recordbasis realtime communicatie tussen applicaties en worden vooral toegepast binnen projecten voor enterprise application integration (EAI). Message brokering-toepassingen vallen onder de categorie middleware; een van de bekendste standaards is momenteel MQ Series. Het gebruik ervan heeft als voordeel dat een informatiecomponent -bijvoorbeeld naam- binnen één toepassing is te wijzigen en dat deze wijziging vervolgens via een *message queue* wordt doorgegeven aan alle overige relevante

Leverancier	Product
Informatica	PowerMart/Center
Ascential	DataStage XE
SAS	Warehouse Administrator
Oracle	Oracle Warehouse Builder
Hummingbird	Genio Designer/Engine
Cognos	DecisionStream
Sagent	Sagent Technology Suite
IBM	DataWarehouse Manager
ETI	ETI*Extract
Microsoft	DTS
Teradata	Warehouse Builder
Computer Associates	Infopump/Data Transformer
Information Builders	IWay/Copy Manager
Compuware	File-Aid ETL
Acta	Actaworks
DataJunction	DataJunction Integration
I-flex	Flexcube ETL
Embarcadero	DT Studio

FUNCTIONALITEIT BINNEN EEN GROOT AANTAL ETL-TOE

toepassingen. De message bestaat uit het gewijzigde record en een identificatie van de toepassing.

Voor een ETL-tool zijn deze message brokers niet meer dan een bron- en doel-object, en bij gebruik hiervan wordt slechts één record per gebeurtenis verwerkt.

De combinatie van ETL en EAI is tot op zekere hoogte in de praktijk toepasbaar. Terwijl EAI-omgevingen altijd transactie-georiënteerd zijn, zullen ETL-omgevingen zich meer richten op query-intensieve omgevingen, waarbij grote hoeveelheden gegevens moeten worden getransporteerd.

### TREND 5: DATAKWALITEIT

Met de toenemende behoefte aan geïntegreerde klantinformatie die via diverse contactkanalen tot stand wordt gebracht, is de kwaliteit van de klantgegevens essentieel. Bijvoorbeeld als klantgegevens via een callcenter, via Internet en via directe verkoop worden bepaald, legt men deze gegevens meestal op uiteenlopende manieren vast (bijvoorbeeld een klantnaam kan worden genoteerd als: Dhr. Jansen, De heer K. Jansen en Jansen K.).

Standaardisatie van de naam naar één uniforme naam kan met speciale tools voor datakwaliteit worden gerealiseerd. Een trend is dat dergelijke functionaliteit nu ook binnen ETL-tools steeds meer als standaard wordt geleverd.

### TREND 6: AANTAL DATA

De omvang van de gegevensstroom neemt binnen alle organisaties exponentieel toe. Oorzaak is vooral de invoering van veel nieuwe technologieën, zoals de al genoemde CRM, ERP en e-commerce. Ook moeten gegevens frequenter worden geladen in het datawarehouse, zodat de organisatie een actueler beeld krijgt waarop het management kan sturen. Dataverwerking wordt dus een steeds intensiever en langduriger proces.

De concurrentie op de ETL-toolmarkt wordt veelal uitgevochten met performancecijfers. De leveranciers maken hierbij gebruik van parallelle laadtechnieken. Hierbij worden meerdere ETL-processen op multiprocessorsystemen ingezet, waardoor meer gegevens in een kortere periode zijn te verwerken. Vendors voeren momen-

teel allerlei benchmarks uit, maar ieder gebruikt zijn zelf ontworpen benchmark. Zo is behoefte ontstaan aan een standaard benchmark voor ETL-technologie, ontworpen door een onafhankelijke derde partij.

### ONTWIKKELINGEN

Alle leveranciers van ETL-tools reageren over het algemeen direct op de genoemde trends. De database-vendors richten zich op integratie met hun bestaande producten, de onafhankelijke ETL-toolleveranciers leggen zich voornamelijk toe op integratie met open standaards en marktgerichte 'totaaloplossingen'.

Deze toepassingen worden ook wel analytische applicaties genoemd. Ze bestaan uit standaard sterschema-datamodellen voor specifieke industrieën en processen, zoals financiën, telecommunicatie, marketing, e-commerce en voorraadbeheer. Ook standaard ERP- en CRM-gebaseerde gegevensmodellen van specifieke leveranciers (SAP, PeopleSoft, Siebel en andere) worden steeds meer als optie aangeboden. De verwachting is dat deze analytische applicaties in omvang zullen toe-

Website	Database-leverancier	Analytische applicaties beschikbaar	XML-gegevens-transport	XML-metadata	Message broker ondersteuning	Datakwaliteit functies	Parallele laadopties
www.informatica.com		X	X	X	X	X	X
www.ascentialssoftware.com			X		X	X	X
www.sas.com	X				X		X
www.oracle.com	X		X	X		X	
www.hummingbird.com			X		X		
www.cognos.com		X					
www.sagent.com		X	X	X	X	X	X
www.ibm.com	X						
www.eti.com		X	X	X	X		
www.microsoft.com	X		X	X	X		
www.teradata.com	X						X
www.ca.com							
www.informationbuilders.com					X		
www.compuware.com							
www.acta.com		X					
www.datajunction.com							
www.iflexsolutions.com							
www.embarcadero.com				X			

nemen, maar uiteindelijk geen kernactiviteit van de ETL-vendors worden. De winst moet uiteindelijk toch komen van het aantal licenties en de dienstverlening rondom het ETL-tool. Wie de hoop mocht koesteren dat aanschaf van een analytische applicatie een integratietraject snel en efficiënt maakt, zal al snel tot de conclusie komen dat de noodzakelijke aanpassingen en implementatie van het standaard data-

model soms evenveel inspanning vergen als het ontwerpen van een gegevensmodel op maat.

Volledige samensmelting van ETL- en EAI-systemen zal voorlopig een punt van discussie blijven tussen ETL-leveranciers en hun klanten. Nu processoren steeds krachtiger worden en de applicaties flexibeler zijn voor het aanleveren van één record per wijziging, komen de mogelijk-

heden voor realtime datawarehouses binnen handbereik. ETL/EAI-integratie vraagt echter om een tegenstrijdige kijk op de strategie voor gegevenstransport binnen één organisatie. De belangrijkste knelpunten die de komende tijd moeten worden opgelost zijn de volgende.

- De huidige systemen bevatten tal van legacy databases, waardoor één standaard benadering (ETL-bulkload of per transactie) niet mogelijk is.
- Wensen en eisen voor metagegevens worden nog niet ondersteund door de huidige leveranciers van message brokering-systemen.

**STANDAARDS**

Zoals uit de trends blijkt, neemt de behoefte aan een onafhankelijke benchmark toe. Naar verwachting zal een onafhankelijk instituut -zoals de Transaction Processing Performance Council- binnen een jaar een standaard benchmark voor ETL-tools aanbieden.

Hoewel standaardisatieprocessen nooit een gesignaleerde behoefte kunnen bijhouden, heeft de Object Management Group (OMG) inmiddels het Common Warehouse Metamodel (CWM) opgesteld. Deze standaard specificatie voor een metadata-model is bovendien al zo breed geaccepteerd, dat de meeste leveranciers haar implementeren in hun ETL-product. XML en het bijbehorende uitwisselingsformaat XMI zijn de standards binnen het CWM. Een van de productstrategieën van de ETL-leveranciers is de metagegevens conform het CWM op te zetten en de communicatie met andere metadata-applicaties te laten plaatsvinden via XMI. Met als uiteindelijk resultaat meer inzicht voor de eindgebruiker en dus een hogere productiviteit binnen de organisatie.

**INFORMATIEMANAGEMENT**

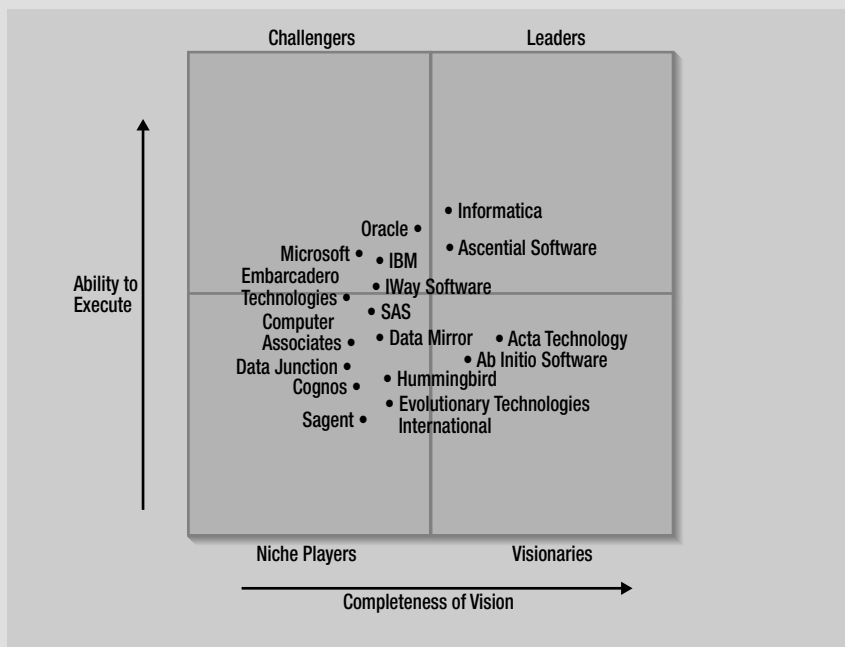
Samenvattend stellen we vast dat de ETL-toolmarkt gestaag groeit en dat die groei de komende jaren blijft aanhouden. Er is een migratie aan de gang van een markt voor datatransportapplicaties naar een markt voor een groep applicaties die gege-

**Gartner over ETL-markt**

Informatica is door Gartner Group uitgeroepen tot leider in de markt voor ETL-tools. In het vorige maand uitgekomen rapport *ETL Magic Quadrant Update: a market in evolution* beschrijven de analisten de zeventien spelers en de trends in de ETL-markt.

Gartners "ETL Magic Quadrant" ziet er momenteel uit als in figuur 3. 'Leiders' zijn volgens Gartner leveranciers die goed presteren, een duidelijke visie op marktontwikkelingen hebben en actief hun competenties uitbouwen om hun positie in de kop van het peloton vast te houden. Het draait dus niet alleen om marktaandeel.

"Door vendors in dit Magic Quadrant te plaatsen, benadrukken wij de levensvatbaarheid van de onderneming, allianties, prijsmodellen en productcriteria," aldus het Gartner-rapport. "Verder wegen de plannen en mogelijkheden van de vendors trends te zetten in de data-integratiemarkt, zwaar in het klassemment."



**FIGUUR 3: ETL MAGIC QUADRANT 2002 (BRON: GARTNER GROUP).**

De ETL-markt blijft volgens Gartner groeien en veranderen. De huidige spelers hebben diverse achtergronden, van ETL als *core business* tot business intelligence- en dbms-vendors. Informatica is begonnen met data-integratie als kernactiviteit -met haar ETL-product PowerCenter- en heeft inmiddels ook analytische applicaties ontwikkeld. Het rapport van Gartner is te vinden op [www.gartner.com/reprints/informatica/106602.html](http://www.gartner.com/reprints/informatica/106602.html)

vensintegratie, analytische applicaties, software-ontwikkeling en gegevensbeheer als belangrijkste processen onderkennen. Het ETL-tool wordt dus een belangrijk component van het informatiemanagement binnen een organisatie. Hierbij is informatie, naast mensen, middelen en financiën, de vierde belangrijke pijler binnen een organisatie.

De huidige trends zijn er oorzaak van dat de ETL-leveranciers zich meer richten op het leveren van analytische applicaties. Sommige vendors doen dit door het leveren van standaard sterschema's, andere daarentegen leveren speciale portals als onderdeel van een totaal geïntegreerde informatiebenadering. Feit blijft dat naast ETL steeds meer business intelligence-functionaliteit wordt toegevoegd.

XML is en blijft de standaard voor de opslag van van metagegevens. Omdat metadata-beheer strategisch is voor de

informatievoorziening binnen organisaties, is het dan ook van belang om ETL-tools te conformeren aan deze standaard. Het Common Warehouse Metamodel van de Object Management Group is een specificatie die compleet genoeg voor implementatie door ETL-leveranciers.

Over het algemeen zijn ETL-tools ontwikkeld om frequent grote hoeveelheden gegevens te transporteren en transformeren naar query-intensieve gegevensbanken, zoals datawarehouses en datamarts. EAI-applicaties zijn meer toepasbaar in een transactionele omgeving waarin realtime synchronisatie van operationele systemen actueel is. Volledige integratie van ETL- en EAI-toepassingen zal voorlopig dan ook niet tot de mogelijkheden behoren.

Gezien de ontwikkelingen in de markt voor ETL-tools is er momenteel niet echt één leverancier die met kop en schouders

boven de rest uitsteekt. De selectie van een ETL-tool moet altijd zijn gebaseerd op de behoefte aan integratie van informatie binnen een organisatie. Het is daarom van groot belang zorgvuldig de wensen en eisen met betrekking tot de geïntegreerde informatiebehoefte in kaart te brengen en vervolgens een selectietraject te beginnen. Kwalificatie en selectie van een ETL-tool is nog steeds een lange-termijngebeurtenis, waarbij de huidige en toekomstige trends zeker moeten worden meegenomen. ●

**Referentie:**

Lou Agosta: *Market Overview Update: ETL*. Giga Information Group, Inc, 2002.

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

**U P D A T E**

**INFORMATION BUILDERS:  
TEMPLATES VOOR VERTICALE  
MARKTEN**

De Vertical Business Reporting Templates -of *intelligence foundations*, zoals Information Builders ze noemt- vormen het fundament voor een rapportage-omgeving die binnen enkele weken op maat gebouwd kan worden en die is afgestemd op de branche en de organisatie.

De intelligence foundations omvatten standaard reporting templates voor het analyseren van bedrijfstakspecifieke meetwaarden.

Meer informatie op [www.informationbuilders.nl](http://www.informationbuilders.nl).

**SAS BREIDT DATAMINING-  
SOFTWARE UIT MET TEXT  
MINER**

SAS kondigde onlangs de beschikbaarheid van Text Miner aan. Hiermee kan kennis worden gehaald uit een scala aan tekst-documenten, zoals rapporten, verslagen, medische files, websites, patenten, garantieverklaringen en e-mails.

De software kan tekstgegevens bereiken in

Word-documenten, PDF's, html en vele andere files opgeslagen in diverse typen databases en systemen.

Meer informatie op [www.sas.com](http://www.sas.com)

**FRAMEWORK VOOR  
CORPORATE PERFORMANCE  
MANAGEMENT**

Cognos claimt de eerste leverancier te zijn die een compleet framework op de markt brengt voor *corporate performance management* (CPM), de strategische toepassing van business intelligence. Het framework faciliteert de volledige managementcyclus van plannen, budgetteren, uitvoeren, rapporteren en analyseren.

Voor meer informatie [www.cognos.com](http://www.cognos.com)

**TEMTEC BENUT DRILL-  
THROUGH MOGELIJKHEDEN  
VAN ESSBASE**

Temtec lanceert een nieuwe versie van Executive Viewer, haar analytische frontend-toepassing voor Essbase en andere OLAP-tools. Met versie 4.2 van Executive Viewer kunnen

gebruikers diep doordringen in geabstraheerde data in een OLAP-kubus.

Tegelijk met deze nieuwe release brengt Temtec een API voor Executive Viewer uit, waarmee Executive Viewer kan worden geïntegreerd met andere rapportage-omgevingen en standaardproducten, zoals Microsoft PowerPoint.

Mede door Executive Viewer 4.2 verwierf Temtec de Platinum Technology Partner-status van Hyperion.

Meer informatie op [www.temtec.com](http://www.temtec.com)

**COGNOS KOMT MET  
CMG EN ORACLE VERSTERKEN  
SAMENWERKING**

CMG en Oracle Nederland gaan hun samenwerking intensiveren op het gebied van kennisoverdracht, training, ondersteuning en wederverkoop. De overeenkomst versterkt CMG's positie als system integrator, waarbinnen ook consultancy en implementatie van Oracles 9i database, Application Server en de volledige E-Business Suite vallen. CMG is Certified Advantage Partner van Oracle sinds 1996.

Voor meer informatie: [www.cmg.com](http://www.cmg.com) en [www.oracle.nl](http://www.oracle.nl)