

seeMore Virtual Database biedt real-time toegang

# Alle enterprise data bereikbaar

Bram Dons

**Ondernemingen zetten voor hun bedrijfsvoering talrijke nieuwe complexe applicaties voor Business Intelligence, CRM, ERP, SCM en e-Commerce in. Om hiervan effectief gebruik te kunnen maken, is het wenselijk dat deze krachtige applicaties toegang hebben tot alle aanwezige data binnen de onderneming.**

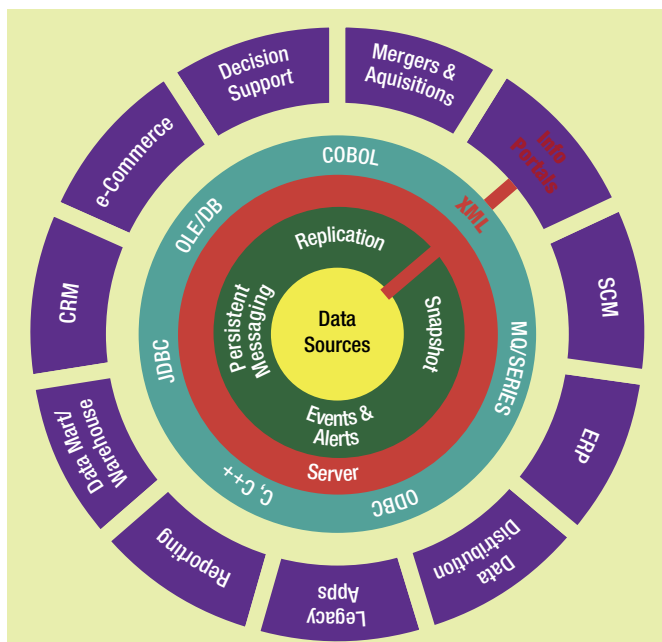
In IT-omgevingen met meerdere typen databases is het ondersteunen van een dergelijke functionaliteit aan de eindgebruiker voor de IT-afdeling een lastige en kostbare zaak. Niet alleen is door de ontwikkeling van de database-technologie het aantal verschillende typen databases de laatste tien jaar enorm toegenomen, maar ook als gevolg van het toenemend aantal gefuseerde ondernemingen. Volgens onderzoek heeft een doorsnee onderneming de data al in gemiddeld veertien databases opgeslagen (Gartner Group), moeten data bidirectioneel tussen 49 verschillende systemen of formats worden uitgewisseld (META Group), is 80 procent van de informatie 'gevangen' in een niet-standaard text format en wordt 35 procent van de IT-investeringen gespendeerd aan de integratie van bedrijfsinformatie (META Group).

Er bestaan al van oudsher talrijke tools en oplossingen om data uit verschillende databases te onttrekken ten behoeve van applicaties, waaronder 'adapters' of 'frameworks', maar deze vragen allemaal de nodige programmering. De vraag is dus op welke wijze ondernemingen vanuit één applicatie, zonder de noodzaak te moeten (her)programmeren, tegelijkertijd toegang kunnen krijgen tot alle in de onderneming, in diverse databases opgeslagen, aanwezige informatie. Om ondernemingen in staat te stellen om op een strategische manier van alle enterprise-data gebruik te kunnen maken is een applicatie nodig die toegang tot alle data heeft, ongeacht waar deze zich bevinden en ongeacht het data format. Bovendien moet een dergelijke applicatie de mogelijkheid bieden dat de verschillende bestaande applicaties (Report Writers, Business Intelligent Tools, Data Auditing & Cleansing Tools, Applications Servers en Web Services) op elk moment van deze data gebruik kunnen maken.

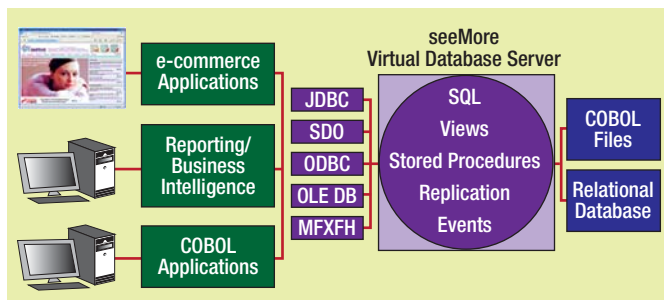
De firma seeMore heeft daartoe een zogenaamde 'Virtual Database' ontwikkeld die de 'look and feel' heeft van een relationele SQL database en bovendien gelijktijdige toegang verschaft tot een heterogene database-omgeving. Voor deze gevirtualiseerde database-omgeving is geen enkele programmeerinspanning nodig en hij is gebaseerd op open standaarden.

## Virtual

De seeMore Virtual Database is vanaf het begin ontworpen om deze er net zo uit te laten zien als een standaard SQL database, zodat de leertijd van de IT-beheerder kort kan worden gehouden en er ook van de bestaande ervaring op SQL-gebied binnen de onderneming gebruik kan worden gemaakt. seeMore biedt een Virtual Database Server die beheer, toegang, view, bundeling, wijziging, replicatie en gebruik van alle binnen de onderneming aanwezige informatie ondersteunt, ongeacht locatie, data type, of format. Corporate data kunnen daarmee vanaf verschillende systemen zonder vertraging worden geraadpleegd, zonder dat er dure programmering van verschillende type data-interfaces aan te pas moet komen. Informatie is real-time te wijzigen en te repliceren voor het consistent houden van meerdere remote databases. Het verschaft gebruikers of bedrijfsapplicaties real-time toegang tot alle beschikbare databronnen, zonder te weten waar de data fysiek aanwezig zijn. De seeMore server kan naast relationele databases ook bestaande niet-relationele databronnen als flat files, DISAM/CISAM en COBOL naar de virtuele



Afbeelding 1: seeMore Data Sources.



**Afbeelding 2:** seeMore Virtual Database Server.

relationele database omzetten. De meeste bekende databronnen worden ondersteund, waaronder: Oracle, DB2, Informix, Ingres, SQL Server, Sybase, MySQL, Postgres, Informix, Adabas, Micro Focus Cobol, Fujitsu NET Cobol, XML en Flat files (CSV en Text).

Nieuwe databases zijn transparant met bestaande applicaties te integreren of oude databases met nieuwe applicaties te gebruiken, zonder de noodzaak om data-integratiecode te moeten schrijven. Creatie van nieuwe business logic dat toegang heeft tot data en bestaande business logic gebeurt zonder dat de gebruiker hoeft te weten waar deze fysiek zijn gehuisvest. Alle enterprise data komen daarmee beschikbaar voor 'decision-making' doeleinden, waarbij data kunnen worden gebundeld, een toepassing die voorheen met de traditionele methoden niet mogelijk was. Voordeel voor zowel de beheerder als gebruiker is dat de op meerdere databronnen opgeslagen corporate data vanuit een enkel beheerpunt beschikbaar zijn. Het biedt ondernemingen de mogelijkheid om hun bestaande legacy-systemen te behouden en de rapportage- en BI-applicaties toch te kunnen gebruiken, zonder dat men van bepaalde data-extractietechnieken gebruik moet maken.

## Kenmerken

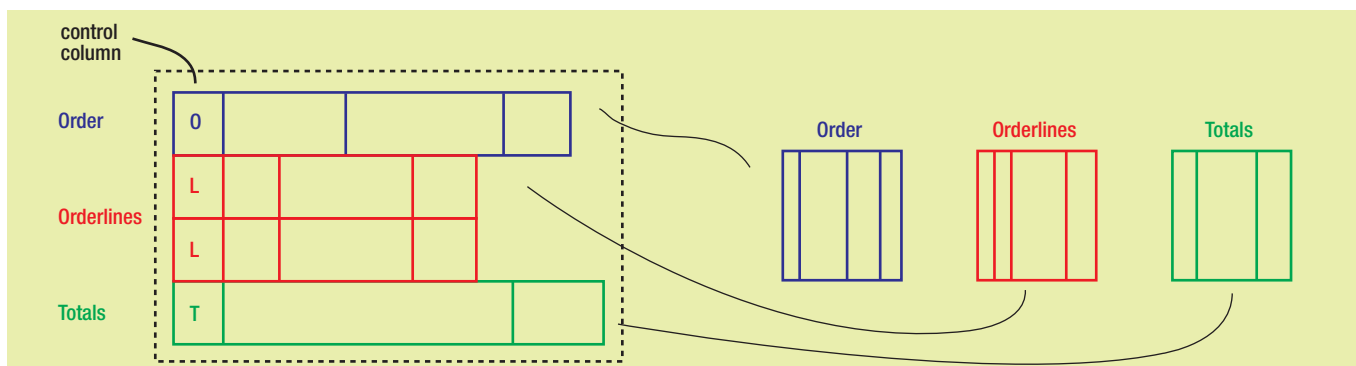
Nadat alle enterprise data binnen de Virtual Database zijn gecatalogiseerd, kunnen applicaties real-time door middel van client API's als ODBC, OLE DB, JDBC, Micro Focus/Fujitsu COBOL of SDO/SOA transparant en direct de gewenste informatie opvragen. De in de databronnen aanwezige data worden niet in de virtuele database opgeslagen, maar alleen de metadata

die beschrijven waar en in welke format alle tabellen en views fysiek zijn opgeslagen. Op het moment dat een applicatie via de seeMore Server gegevens opvraagt, wordt seeMore's Meta data dictionary geraadpleegd. De seeMore server zoekt vervolgens de juiste locatie van de gevraagde data op en op basis van de hiervoor genoemde industrie applicatie-interfaces vraagt het de data in de fysieke database(s) op en stuurt de verzamelde data vervolgens in het juiste format naar de eindgebruiker of applicatie toe (zie afbeelding 2).

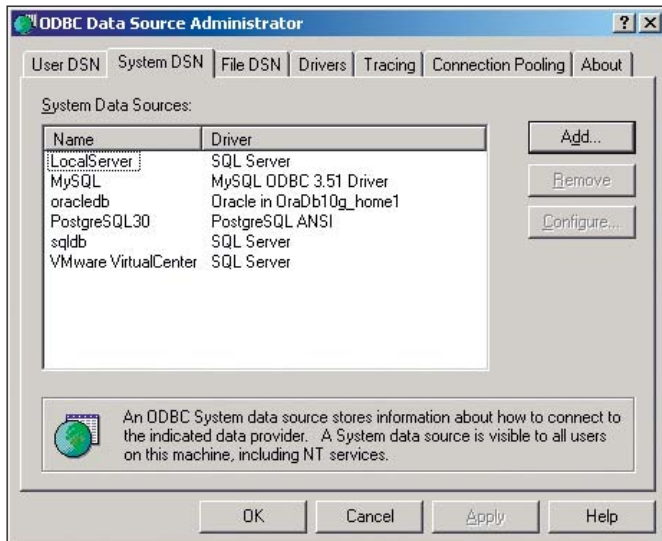
De kern van seeMore's Virtual Database wordt gevormd door de Meta Data Dictionary. Het is een door de database-beheerder geconfigureerde logische representatie van de fysieke informatie van alle in de onderneming aanwezige databronnen. De Data Dictionary bevat talrijke informatie, waaronder: users, access lists, schema's, stored procedures, events, tables, definitions en columns, indexes, relations, views, enzovoort.

Binnen de virtuele databases kunnen tabellen worden gedefinieerd met meerdere definities die kolommen met array's en classes bevatten. Dit stelt seeMore in staat om complexe data-structuren als COBOL of objectgeoriënteerde databases te kunnen ondersteunen. seeMore ondersteunt meerdere schema's waarmee database-objecten logisch opgesplitst kunnen worden in een virtuele database; dit is met name zinvol wanneer de namen van tabellen, views of stored procedures conflicteren met gelijke namen van tabellen in meerdere databronnen. De toegang van de host tot de server is beveiligd en naast de bestaande database-beveiligingen kunnen aan gebruikers van seeMore verschillende toegangsprivileges worden toegekend, zelfs tot op kolomniveau! De toegang tot gevirtualiseerde databronnen kan op basis van read-only of read/write worden ingesteld.

seeMore maakt van 'instances' gebruik om informatie te partitioneren en te repliceren over meerdere bestandssystemen, computersystemen, sites of locaties. Een scheduled replicatie kan gebeuren op basis van tabellen of views per partitie. seeMore heeft de ANSI standaard (SQLJ) geïmplementeerd, dat specificeert op welke wijze statische Java methodes vanuit SQL als procedures of functies kunnen worden aangeroepen. Met behulp van SQLJ zijn database-onafhankelijke Java stored procedures te



**Afbeelding 3:** COBOL conversie naar SQL.



**Afbeelding 4:** ODBC Data Source Administrator menu.

creëren, voor de koppeling met SQL, XML, Java, J2EE en Web services.

seeMore maakt van services gebruik voor de communicatie met andere seeMore Servers op andere systemen. Het door een service gebruikte communicatieprotocol kan zijn gebaseerd op TCP/IP of IBM's Websphere MQ (voormalig MQ/Series) messaging middleware; MQ ondersteunt een gegarandeerde data delivery, seeMore ondersteunt dit zelf niet.

Elke seeMore database heeft tenminste één partitie die, naast de replicatietoepassing, kan worden gebruikt voor de mapping van verschillende schema's voor meta-tabellen. Als een database is gedistribueerd over meer dan een site, dan is er op elke site tenminste één instance aanwezig. Instances zijn in een drie-lagen-boomhiërarchie gestructureerd, waarbij elke blad binnen de boomstructuur een exacte kopie bevat van de metadata dictionary; de data op verschillende locaties kunnen verschillen, bijvoorbeeld elke locatie kan een bepaald product bestrijken. Verder biedt seeMore een snapshot-voorziening om binnen de seeMore server een snapshot-replicatie te kunnen plannen, wat een snellere toegang tot de data biedt zonder dat van een transactiesysteem gebruik hoeft te worden gemaakt; seeMore ondersteunt incrementele, gesynchroniseerde en gespiegelde snapshots. seeMore heeft de kern SQL99 ANSI standaard voor haar SQL engine geïmplementeerd. Deze ondersteunt de uitvoering van heterogene 'joins' dat een eenduidige 'view' van meerdere databronnen creëert.

## Database-tabellen

De database zelf kan in schema's worden onderverdeeld; een logische opsplitsing van de database. Definities (columns) worden in seeMore gebruikt om de tabelkolommen te beschrijven. Doorgaans hebben tabellen in een relationele database slechts een enkele definitie die de kolommen bevatten, maar in sommige databronnen (bijvoorbeeld COBOL) kunnen in een

enkele tabel de kolommen vele definities bevatten. seeMore beschrijft dit als een 'complex table'. In afbeelding 3 zien we de COBOL-tabel 'Order' die verschillende record lay-outs bevat: header records, informatie over de order (zoals klantendetails en orderdatum); detail records: informatie over de bestelde items; trailer records, de totalen.

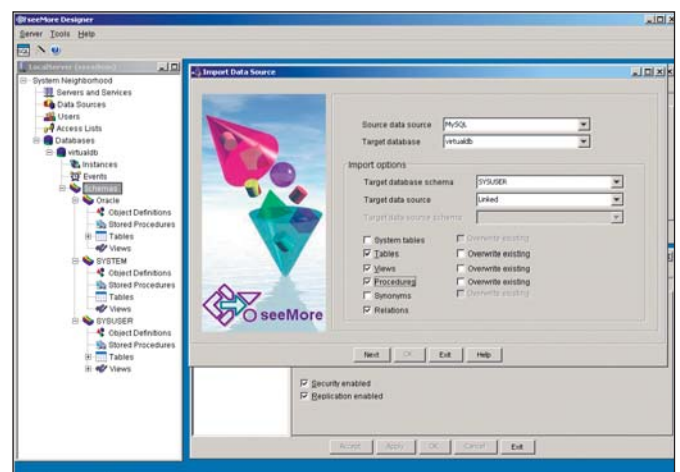
De fysieke tabel 'orders' bevat drie verschillende sets met rijen, elk met een eigen kolom lay-out. Door middel van 'table definitions' ziet seeMore deze complexe tabel als drie aparte logische tabellen. seeMore bezit dus de mogelijkheid om drie definities als drie aparte tabellen te zien en ze daarmee geschikt te maken voor SQL-query's.

Op deze manier hebben eindgebruikers via hun favoriete applicaties toch toegang tot COBOL databestanden, in het geloof dat ze een relationele tabel in een SQL-gebaseerde database raadplegen. Via ODBC, OLE DB en JDBC-drivers hebben eindgebruikers van, bijvoorbeeld populaire applicaties als Excel en Crystal Reports, toegang tot COBOL databases. Met behulp van de seeMore's Designer zijn relationele views en JAVA-gebaseerde stored procedures te creëren.

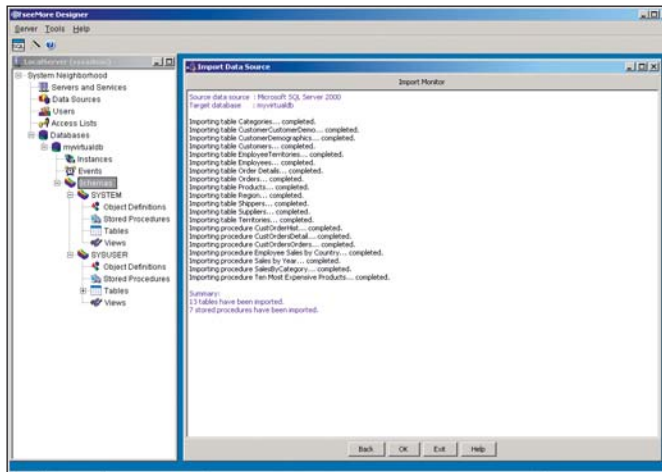
## Planning voor ontwerp

Met behulp van de seeMore Designer wordt de metadata dictionary ontworpen die de virtuele database beschrijft. Op enkele definities na, maakt de Designer van dezelfde definities gebruik als een relationele database. De definities worden op hiërarchische wijze gecreëerd. Bovenaan staan de definities die aan het systeem zelf zijn gerelateerd, daaronder die betrekking hebben op elke virtuele database en daar weer onder voor elke databases definities die betrekking hebben op elk object binnen de virtuele database.

Voordat wordt begonnen met een ontwerp met behulp van de Designer is het verstandig dat de database-beheerder een plan opstelt waarin hij de wijze aangeeft waarop elke virtuele database gebruikt gaan worden: de gebruikte data sources; de logische opdeling van de virtuele databases met schema's, welke tabellen, views en stored procedures beschikbaar zijn voor appli-



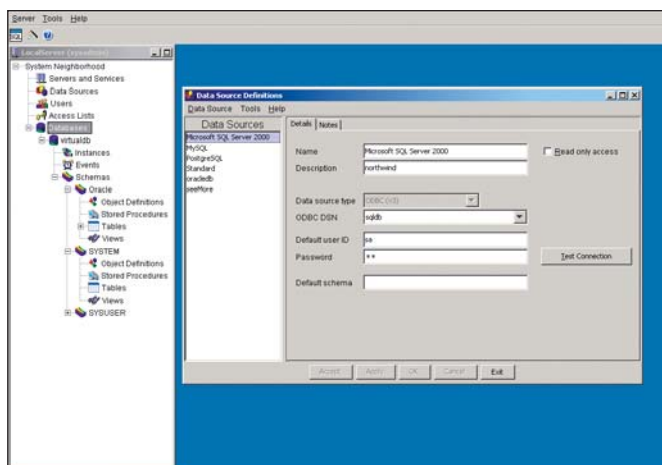
**Afbeelding 5:** Import Data Sources.



**Afbeelding 6:** Import SQL Server database.

caties; de locatie en organisatie van de database wanneer deze over meerdere sites is gedistribueerd; welke gebruikers toegang hebben tot welke data sources en met welke bevoegdheden.

Het complete ontwerp en onderhoud van de Virtual Database gebeurt vanuit de seeMore Designer GUI. De Designer is zodanig ontworpen dat het er als een 'database' uitziet. Het is een intuïtieve en eenvoudig te gebruiken grafische interface zodat bestaande vaardigheden en bronnen snel zijn toe te passen. Meer dan 90 procent van seeMore's functionaliteit zal onmiddellijk door gebruikers worden herkend, zodat een leertijd van enkele dagen en weken beperkt kan worden tot enkele uren of dagen. Voordeel van dit innovatieve ontwerpconcept is een verminderde onderhoud- en beheerkostenpost (TCO), met inbegrip van verminderde trainingkosten, snellere beschikbaarheid en ontwikkeling van projecten en daarom een verbeterde terugverdientijd (ROI). De Designer biedt de IT-ontwikkelafdeling een tool met een aantal 'wizards' waarmee snel tabellen, views en stored procedures van bestaande databronnen zijn te importeren. Deze geïmporteerde database-objecten vormen de basis van seeMore's eigen tabellen, views en



**Afbeelding 7:** Geïmporteerde Data Sources.

procedures. Zodra deze tabellen zijn geïmporteerd dan kunnen met de Designer de tabellen, views en procedures naar eigen inzicht worden ingericht voor verticale, horizontale of procesgeoriënteerde oplossingen ten behoeve van een eenvoudig gebruik van elke business-applicatie.

## Implementatie heterogene database-omgeving

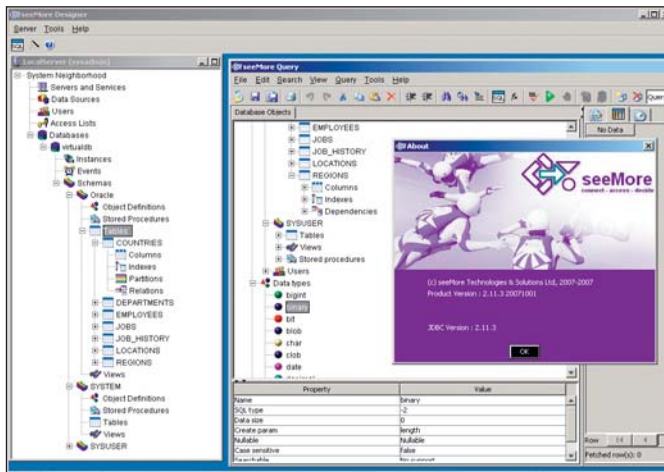
Om te zien of seeMore de belofte van een virtuele database kan waarmaken, creëren we een virtuele database op basis van een Microsoft SQL Server 2000, Oracle 10g, PostgreSQL 8.2 en MySQL Server 4.1 database. Een doorsnee entry-level installatie kan bestaan uit de seeMore Virtual Information Server (dat default ook de License Server bevat) op een UNIX (Solaris, HP-UX en Aix), Linux (Redhat of Suse) of Windows Server platform (Windows Server, XP of Vista). De seeMore Designer kan op een PC en een of meer clients worden geïnstalleerd met ondersteuning van ODBC, OLE DB of JDBC connectiviteit. Wij installeren de 'complete' installatie, bestaande uit de seeMore Virtual Information Server en de Designer op een Windows 2003 R2 32-bit server (64-bit wordt nog niet ondersteund); een 'compact' installatie bestaat alleen uit de client tools en connectiviteit (ODBC, JDBC en OLD DB drivers).

De configuratie van een database gebeurt in een viertal basisstappen: creatie van een databron, gebruikers, virtuele database, de creatie en de import van de fysieke database-tabellen. De eerste stap is dus de creatie van de diverse 'data sources'. Daartoe creëren we voor elk van de hiervoor genoemde fysieke databases een 'System DSN' ODBC-verbinding, met behulp van het standaard Windows 'ODBC Data Source Administrator' menu.

Zodra de System ODBC databron bestaat, is op basis daarvan een seeMore databron te creëren door 'Data Sources' te selecteren in de 'System Neighbourhood' tree in de Designer. De creatie van een Data Source voor een relationele database, zoals Oracle en SQL Server, is relatief eenvoudig. In de meeste gevallen is alleen een user ID en wachtwoord nodig, maar voor bijvoorbeeld de Focus of Fujitsu COBOL databron is het directory path nodig dat de databestanden bevat en de bestanden die de nodige importinformatie bevatten bij gebruik van de 'import wizard'. Vervolgens dienen er seeMore gebruikers te worden gecreëerd; standaard zijn de 'sysadmin' en 'guest' gebruikers al aangemaakt.

Na de creatie van de diverse databronnen is er een virtuele database nodig voordat de fysieke database-tabellen handmatig of automatisch kunnen worden geïmporteerd. Vanuit de 'System Neighbourhood' klikken we 'Databases' waarna een Databases window verschijnt, van waaruit zoveel als nodig databases worden gecreëerd en iedere database aan een databron wordt verbonden. De hiervoor gecreëerde databases dienen default als bron voor het importeren van de tabellen in de virtuele database. De eenvoudigste manier is om dit via de Import Data Source Wizard 'automatisch' te doen. De beheerder kan individuele tabellen, allemaal, of op basis van een filter selecteren.

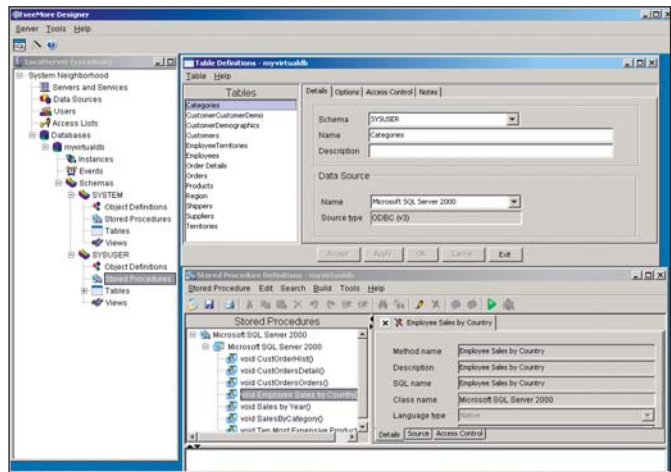




Afbeelding 8: Query tool Designer.

## Tools

In het 'Tools Menu' zijn verder 'Directory Aliases' en 'Pseudo Columns' aan te maken. Alle path-definities binnen de seeMore Designer zijn geen absolute paden. Ze moeten beginnen met een directory alias, een logische naam voor een directory. Voor elke 'replication instance' wordt de directory alias 'ge-mapped' naar een fysieke locatie door gebruik te maken van de Directory Aliases tab binnen het 'Instances Definitions' venster. Directory aliases zijn nodig zodat de definities van een virtuele database en de daarmee geassocieerde objecten door verschillende replication instances gebruikt kunnen worden. Want er is geen garantie dat databases of tabellen in identieke directory's zijn geplaatst op systemen met verschillende replication instances. seeMore ondersteunt ook het gebruik van pseudo columns waarmee data van verschillende databronnen zijn te bundelen in geval men niet van de kolommen van deze databronnen gebruik kan maken voor data-identificatie. Er zijn bijvoorbeeld applicaties van third party leveranciers die toestaan om applicaties voor meerdere ondernemingen binnen een organisatie te gebruiken. Daartoe zou men op een bepaalde onderneming moeten inloggen om toegang tot de data te krijgen. Voor rapportage en updating van data, buiten de applicatie om, is gelijktijdige toegang tot alle ondernemingen dan niet mogelijk omdat elke onderneming zijn eigen database heeft en de tabellen daarin geen kolommen bevatten die de ondernemingen kunnen identificeren die de data vertegenwoordigen. Met seeMore zijn pseudo kolommen te creëren die binnen seeMore's metadata dictionary bestaan, maar niet in de databron zelf. De kolom kan vervolgens aan de tabellen worden toegewezen om een 'ondernemingwaarde' te vertegenwoordigen. Daarbij wordt de constraint value van de pseudo kolom gebruikt die als een constante waarde fungeert. Na dit alles is de database in principe klaar voor gebruik, maar de beheerder zal meestal nog de diverse beveiligingen willen creëren. Via de 'access list' zijn verschillende soorten toegang (bijvoorbeeld lezen, schrijven of uitvoeren) aan gebruikers toe te kennen. Een access list kan daarna aan een database-object,



Afbeelding 9: Geïmporteerde tabellen SQL Server.

zoals een table, view, kolom of stored procedure worden gekoppeld. Verder staan de beheerder diverse tools ter beschikking, waaronder de SQL Query Explorer, Import en Export Wizard, Deployment Wizard, Properties en System Log. De SQL Query Explorer is een visuele query tool die eenvoudig in het gebruik is en waarmee SQL query's op elk van de aanwezige databases kunnen worden uitgevoerd. De SQL Query Explorer is een subset van het stand-alone SQL Query Product. Met de 'Properties' tool kan de gebruiker zijn persoonlijke voorkeur, zoals de 'look and feel' configureren voor de SeeMore Designer en tools.

## Samenvatting

seeMore's Virtual Database Server biedt vanuit een enkel toegangspunt toegang tot de meeste bekende databases in de enterprise-omgeving. De seeMore server is niet een echte database waarin echte gebruikers en data worden opgeslagen, maar fungeert slechts als opslagplaats van metadata die betrekking hebben op de fysieke locatie en eigenschappen van database-tabellen. Met behulp van de Designer configureert de gebruiker de virtuele database van waaruit hij met een SQL GUI data de gegevens uit de database(s) kan opvragen. seeMore's SQL Query Explorer fungeert als een gewone SQL GUI waarmee database-beheerders al bekend zijn. De kracht van seeMore's virtuele database-oplossing is het feit dat het toegang biedt tot een heterogene database-omgeving, waarbij de gebruiker vanuit één interface toegang heeft tot alle data in de enterprise-omgeving. De noodzaak om vanuit verschillende applicaties via adapters of frameworks toegang te krijgen, komt daarmee te vervallen.

Bestaande applicaties voor Business Intelligence, CRM, ERP, SCM en e-Commerce hebben via de seeMore virtuele database server niet alleen toegang tot de meeste bekende heterogene relationele databases, maar ook tot COBOL databases, XML en flat files (CSV en Text).

**Bram Dons** is onafhankelijk IT-consultant.