



Sunopsis levert complete gegevens-integratiesuite

Betaalbare oplossing voor 'no-nonsense' ETL

Freek Kamst

ETL-toepassingen zijn de laatste tijd niet echt innovatief te noemen. Na de eerste generatie ETL-tools kwam de tweede generatie data streamers en vervolgens was het wel een tijdje stil in ETL-land. Wel hebben er veel overnames plaatsgevonden van BI-leveranciers die graag een ETL-tool aan hun productassortiment wilden toevoegen om zo een compleet platform te leveren. Hiermee werd vooral de indruk gewekt dat er geen plaats was voor nieuwe spelers in de markt van ETL-toepassingen. Er komen gelukkig steeds meer nieuwe spelers in de BI-branche en het ETL-segment.

In dit artikel wordt Sunopsis onder de loep genomen en deze organisatie typeert zich als een van de nieuwkomers in de ETL-markt. Sunopsis levert een omgeving waarin de migratie en integratie van gegevens centraal staat. De organisatie is opgericht in 1998 en introduceerde haar eerste product, Sunopsis ETL, in het jaar daarop. Ondanks het onvriendelijke investeringsklimaat van de laatste jaren, wist Sunopsis een product te ontwikkelen dat kan wedijveren met haar concurrenten die al veel langer in het ETL-segment actief zijn. Sunopsis is een Franse onderneming met het hoofdkantoor in Lyon, maar heeft inmiddels ook vestigingen in de Verenigde Staten, Duitsland, Italië, Singapore en Engeland. Sinds kort is er ook een vertegenwoordiging in de Benelux. Wereldwijd zijn er ruim vierhonderd implementaties met zo'n 1500 eindgebruikers.

Derde-generatie

Sunopsis positioneert zichzelf als leverancier van een derde-generatie ETL-tool. De eerste-generatie ETL-tools kenmerken zich als toepassingen, die specifieke query- en laadprogrammatuur genereren aan de hand van mappings en transformaties, welke via een grafische gebruikersomgeving worden ingevoerd.

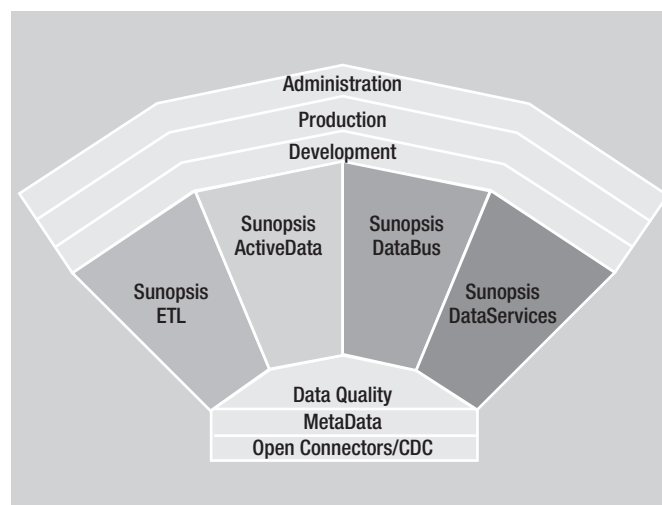
Tweede-generatie ETL-tools typeren zichzelf als zogenaamde data streamers, die gegevens via een specifieke ETL-server direct uit de bronsystemen extraheren en deze gegevens vervolgens direct laden in het doelsysteem.

Volgens Sunopsis kenmerkt een derde-generatie ETL-tool zichzelf als een platformafhankelijke toepassing. Het maakt hierbij niet uit op welk systeem extracties, transformaties en laadprocessen worden uitgevoerd. Deze kunnen vooraf worden gedefinieerd op elke willekeurige plaats in het netwerk, maar bij eventuele knelpunten in de performance kunnen deze processen snel en effectief worden overgezet op systemen die minder belast worden. Alle ETL-functionaliteit wordt in database-specifieke SQL code

gegenereerd. Hiermee worden de databases effectief ingezet voor migraties en transformaties. De huidige database-systemen bieden namelijk steeds meer standaard BI-functionaliteit. Sunopsis zorgt er hiermee voor dat kracht van de database *engine* optimaal wordt benut.

Sunopsis Integration Suite

Met de Sunopsis Integration Suite zijn alle gegevensintegratiewerkzaamheden, zoals datawarehousing, Enterprise Application Integration (EAI) en migraties, vanuit één omgeving te realiseren. Dankzij de gelaagde structuur van het platform (zie afbeelding 1) kunnen migratie-, ETL- en EAI-processen snel en flexibel binnen elke organisatie worden ingevoerd. De Sunopsis Integration Suite is een modulaire omgeving met één uniforme eindgebruikers-interface en metadata-omgeving. Ook de functionaliteit rondom versiebeheer, gegevenskwaliteit en beveiliging zijn hetzelfde voor



Afbeelding 1: Sunopsis Integration Suite.

alle modules. De Sunopsis software is operationeel op alle platformen die Java ondersteunen. Dit betekent dat de Sunopsis Integration Suite bijna overal kan worden geïnstalleerd (Windows, Unix, Linux, etcetera).

De Sunopsis Integration Suite bestaat uit de volgende vier modules:

- Sunopsis ETL is de basis ETL-toepassing van Sunopsis. Dit is het product waarmee Sunopsis zes jaar geleden op de markt kwam;
- Sunopsis Databus is gebaseerd op de Sunopsis MQ (Message Oriented Middleware) omgeving en kan worden ingezet voor directe communicatie tussen applicaties (EAI);
- Sunopsis ActiveData zorgt ervoor dat gegevens op real-time basis kunnen worden gerepliceerd naar een database-omgeving (real-time datawarehousing);
- Sunopsis DataServer is een module waarmee bedrijfsgegevens binnen een SOA (Services Oriented Architectuur) beschikbaar kunnen worden gemaakt.

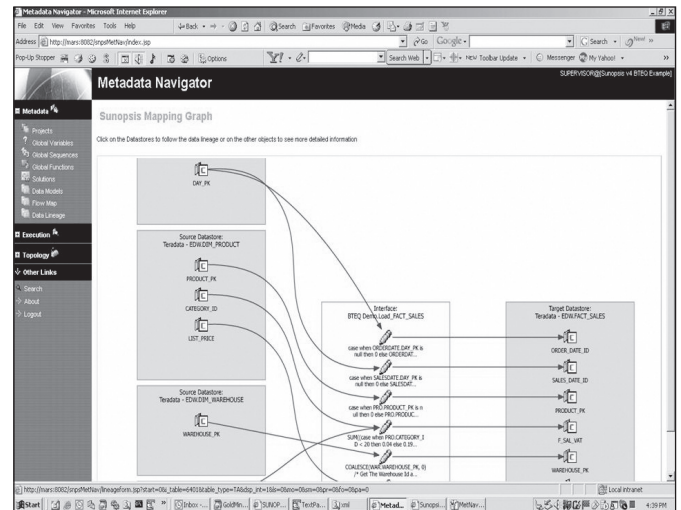
Metadata

Het gebruik van metadata wordt met de toenemende complexiteit van de informatieverwerkende systemen steeds belangrijker. De grootste drijfveer achter metadata zijn momenteel de compliance-regels, die vooral in de Verenigde Staten een paranoïde vorm aannemen. De hoeveelheid compliance-gegevens binnen een datawarehouse-omgeving overschrijdt daar qua omvang de eigenlijke gegevens. Het gevolg is echter wel dat ETL- en EAI-leveranciers meer aandacht gaan besteden aan metadata.

Een exportfunctie is aanwezig om metagegevens te exporteren naar een CWM-omgeving

De Sunopsis metadata-omgeving die gebruikt wordt voor de ontwikkeling en besturing van het ETL-proces, bestaat uit één of meerdere databases die op een willekeurige plaats binnen het netwerk kunnen worden geïnstalleerd. In deze metadata-omgeving worden de mappings, transformaties en gegevenskwaliteitsregels opgeslagen, waarbij de keuze van de database geheel productonafhankelijk is. Dit wil zeggen dat de database-omgeving op MS Access, Oracle, SQLServer, etcetera kan worden geconfigureerd.

Binnen de metadata-omgeving is één zogenaamde *master repository* aanwezig waarin de toegangsbeveiliging voor de ontwerpers en ontwikkelaars is vastgelegd. Hiërarchisch onder deze master repository kunnen zich een of meerdere working repository's bevinden. In de working repository's bevindt zich de procesinformatie, zoals scheduling parameters, mappings, transformaties, etcetera. Met de inzet van deze working



Afbeelding 2: Sunopsis Metadata Navigator.

repository's zijn dus bijvoorbeeld ook de creatie van multi-ontwikkel-omgevingen mogelijk, zoals ontwikkeling-acceptatie en -productie. Ten behoeve van het versiebeheer wordt gebruik gemaakt van de bekende check-in/check-out faciliteiten. Functies ten behoeve van impact-analyses, afhankelijkheidsdiagrammen en zogenaamde where-used functies zijn eveneens standaard aanwezig. De Sunopsis metagegevens zijn niet conform het standaard Common Warehouse Model gegevensmodel opgeslagen. Deze standaard is een aantal jaren geleden in het leven geroepen om metagegevens volgens een uniform standaard model op te slaan. Wel is een exportfunctie aanwezig om de Sunopsis metagegevens te exporteren naar een Common Warehouse Model-omgeving.

Knowledge Modules

De verbindingen met de diverse bron- en doelsystemen worden gerealiseerd door de zogenaamde *Knowledge Modules*. Het idee achter deze knowledge modules is dat elke module specifieke kenmerken bezit, waarmee SQL kan worden gegenereerd om de gewenste gegevens in de bron- en doelsystemen te benaderen. Bijvoorbeeld de Sybase IQ knowledge module bevat alle specifieke SQL-extensies van een Sybase IQ-omgeving. Ook zijn er knowledge modules voor standaard transformaties en datawarehouse-functies. Zo zijn er knowledge modules voor het laden van type-één of type-twee dimensies in een multi-dimensionele datawarehouse-omgeving.

Er zijn momenteel ongeveer tachtig standaard knowledge modules beschikbaar voor onder andere Oracle, DB2, Sybase ASE, Sybase IQ, SQLServer, Teradata, Hyperion en Netezza, maar ook voor toepassingen als SAP, Peoplesoft, Oracle Applications, JD Edwards, Siebel, Business Objects en nog veel meer. Wanneer een andere database-technologie wordt gebruikt, dan zijn natuurlijk ook JDBC- en ODBC-verbindingen mogelijk. Bovendien is het mogelijk om zelf knowledge modules te ontwikkelen. Aanvullend op de standaard set van knowledge modules levert Sunopsis ook standaard Change Data Capture-functionaliteit. Change Data Capture biedt de mogelijkheid om wijzigingen uit

de bronsystemen te detecteren volgens een aantal standaard functies en/of methodieken.

Sunopsis ETL

Sunopsis beschrijft haar ETL-component als een derde-generatie ETL-toepassing. Binnen een derde-generatie tool is de plek waar men de extractie, transformatie of het laadproces uitvoert niet meer relevant. Dit kan in principe op elk systeem binnen een netwerk worden uitgevoerd. Het is natuurlijk nog steeds mogelijk om een specifiek systeem voor het ETL-proces te definiëren, maar dit is niet noodzakelijk. Ook in-memory systemen kunnen binnen de architectuur worden gebruikt, die vooral bij complexe staging-omgevingen veel performance winst kunnen opleveren.

Sunopsis ETL genereert relevante SQL ten behoeve van de bronsystemen waar de gegevens uit geëxtraheerd moeten worden. Zoals eerder in dit artikel beschreven levert Sunopsis hiermee een functionaliteit die effectief gebruik maakt van de aanwezige systemen en zorgt er vooral voor dat er zo min mogelijk in extra hardware moet worden geïnvesteerd.

Ook Sunopsis ETL maakt gebruik van de standaard front-end omgeving. Voor het ontwerpen van het ETL-proces wordt de Sunopsis Designer-omgeving aangeroepen, waarin op een conventionele manier de mappings en transformaties kunnen worden ontwikkeld met behulp van *drag and drop*-faciliteiten. Het toevoegen van commentaar is op procesniveau mogelijk en functies voor het loggen van foutmeldingen zijn aanwezig.

Het automatisch nemen van beslissingen aan de hand van fouten in het transformatieproces is een standaard faciliteit, evenals het versturen van berichten per e-mail bijvoorbeeld wanneer handmatige interventie in het proces nodig is.

Het traceren van fouten binnen het ETL-proces is binnen Sunopsis geen standaard functionaliteit. Functies als *step-by-step running*, het definiëren van *breakpoints* en het tussendoor bekijken van specifieke variabelen is niet mogelijk en moet met behulp van SQL zelf worden ontwikkeld. Ook standaard ETL-functies als *splitting*, *multiple targets*, *union* en *pivoting* zijn

niet aanwezig. Hiervoor zal er binnen de Sunopsis ETL Designer gebruik moeten worden gemaakt van de beschikbare SQL-instructies en logica om het gewenste resultaat te behalen.

Functionaliteit

Sunopsis Databus is volledig operationeel rondom Sunopsis MQ, welke het message queuing-systeem is van Sunopsis. Met Sunopsis DataBus is het mogelijk om specifieke transacties direct te vertalen naar een eigen formaat-transactie. Sunopsis DataBus is hiermee het applicatie-integratie (Enterprise Application Integration) component van de Sunopsis Integration Suite.

Het is mogelijk om gegevens-structuren beschikbaar te stellen als webservices

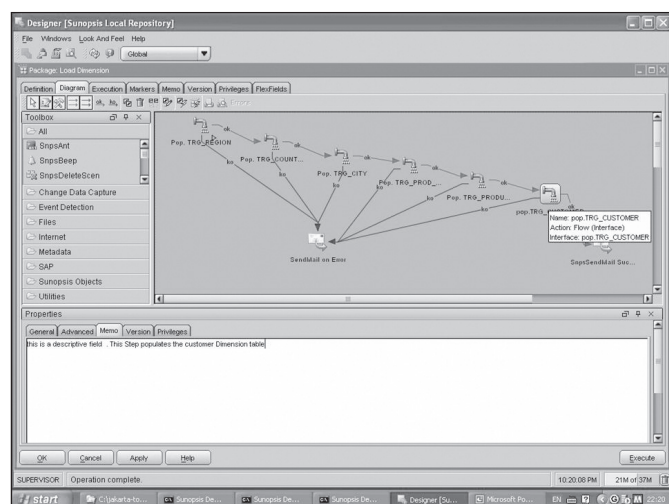
Sunopsis ActiveData levert meer functionaliteit voor BI- en datawarehouse-omgevingen die direct gegevens vanuit de bronsystemen willen repliceren. De vraag voor zogenaamde real-time datawarehouse-omgevingen neemt steeds toe, omdat datawarehouses meer en meer informatie leveren ter ondersteuning van operationele processen. Deze processen willen tot op de minuut nauwkeurig actuele gegevens beschikbaar hebben, waardoor BI-omgevingen een constante stroom transacties vanuit de bronsystemen vereisen. Sunopsis ActiveData kan worden ingezet voor diverse doeleinden, zoals het vullen van een ODS (Operational Data Store) of het vullen van een datamart-omgeving ter ondersteuning van BAM (Business Activity Monitoring). Samenvattend biedt deze module een architectuurbenadering die is gebaseerd op een Active Data Hub, waarin de gegevens in een standaard formaat en persistente werkwijze real-time beschikbaar worden gesteld aan de verschillende systemen.

Met Sunopsis DataServices is het mogelijk om gegevensstructuren beschikbaar te stellen als webservices. Het is in principe de brug tussen de gegevens en de webservices binnen een Service Oriented Architectuur (SOA).

Interbank

In Nederland heeft men momenteel een succesvolle implementatie van Sunopsis achter de rug bij Interbank. Interbank is onderdeel van het ABN/AMRO-concern en is de grootste verstrekker van consumentenleningen in Nederland.

Bij Interbank in Amsterdam heeft men voor Sunopsis gekozen, omdat er behoefte was aan een integratie van de gegevens binnen de bank om meer inzicht te verkrijgen in de risicoprofielen van haar klanten. Het oude datawarehouse-systeem voldeed hiervoor niet meer aan de eisen en een nieuwe opzet met zogenaamde scorecards was noodzakelijk voor de risico-analyses. Het nieuwe datawarehouse is op een DB2-omgeving gebaseerd en met behulp van Sunopsis was na drie maanden ontwikkeling het eerste



Afbeelding 3: Sunopsis Designer-scherm.

increment van het datawarehouse operationeel. De productiviteit van de ontwikkelteams is met 50 procent verbeterd ten opzichte van het ontwikkelen van ETL-processen met SQL. Naast de productiviteit is de performance van het laadproces ook sterk verbeterd. Doordat Sunopsis effectief gebruik maakt van de DB2-functionaliteit is het laadproces van acht uur naar twee uur gegaan, waarbij de hoeveelheid gegevens is toegenomen en de complexiteit van de transformaties hoger is. Met de komst van de Sunopsis Integration Suite kijkt Interbank nu ook naar de mogelijkheden van real-time informatie in het datawarehouse.

Conclusies

Sunopsis heeft met de ontwikkeling van de Sunopsis Integration Suite een belangrijke positie ingenomen binnen de ETL-markt. Het product heeft inmiddels ook de volledige aandacht van Gartner en andere analisten en ontbreekt niet in de diverse reviews en analyses. Het product oogt goed en biedt voldoende functionaliteit om migraties en ETL-processen te ontwikkelen. Het feit dat de omgeving op alle platformen beschikbaar is, creëert mogelijkheden om het product vooral binnen grote datawarehouse-projecten in te zetten. Een sterk punt van het product is de performance, omdat optimaal gebruik wordt gemaakt van de aanwezige systemen en databases binnen een netwerk. De kennis

die benodigd is voor de ontwikkeling en beheer van ETL-processen met Sunopsis is beperkt tot SQL. Een training van maximaal twee dagen is dan ook voldoende om het product te implementeren en ermee aan de slag te gaan.

Sunopsis levert een betaalbaar ETL-product dat effectief kan worden ingezet bij integratie- en migratie-trajecten. Met de komst van de Sunopsis Integration Suite is het product uitgegroeid tot een omgeving waarin gegevensintegratie op allerlei manieren kan worden uitgevoerd. Echter op het gebied van ETL moet Sunopsis nog wel de nodige functionaliteit ontwikkelen om op gelijk niveau te acteren met haar concurrenten. Derhalve oogt het product compleet en bevat zeker een aantal unieke eigenschappen binnen het ETL-segment. Het is voor organisaties dan ook zeker aan te bevelen om het product op de shortlist te zetten, wanneer men erover denkt om in een ETL-product te investeren.

Literatuur

Philip Howard. Sunopsis Integration Suite Evaluation. Bloor Research 2005.

Freek Kamst

Freek Kamst (Freek.Kamst@newcom.nl) is General Manager bij Newcom Information Systems.