



Informatica kiest voor stabiliteit en duurzame ontwikkeling

# Universal Data Services voorziet in behoefte

Martin Misseyer

**Informatica heeft Universal Data Services gelanceerd als antwoord op de hevige competitie in de werelden van BI en datawarehousing. Martin Misseyer onderzoekt of Informatica zich met UDS serieus heeft voorbereid op de toekomst of dat het een slim marketing-verhaal is om nieuwe markten aan te boren.**

De laatste jaren zien we een haast onuitputtelijke stroom aan nieuwe concepten en innovaties, zowel op inhoudelijk als op technologisch vlak, zeker binnen de werelden van Business Intelligence (van onder meer recht-toe recht-aan reporting via OLAP naar portal en dashboard) en van datawarehousing (onder meer profiling, cleansing, en data- en metadata-integratie). Hier is de competitie hevig en staan er grote belangen op het spel. Als antwoord op de concurrentie heeft Informatica de afgelopen jaren verschillende producten uitgerold, van PowerCenter (data-integratie) tot PowerExchange (data-uitwisseling) en van PowerAnalyzer (informatiepresentatie) tot en met SuperGlue (metadata-integratie en -presentatie). Onlangs heeft Informatica haar Universal Data Services gelanceerd. Universal Data Services, kortweg UDS, is een architectuur die Informatica's visie op – en strategie voor – de bedrijfs- en technologie-ontwikkelingen reflecteert. UDS is de komende jaren de belangrijke rode draad in Informatica's productontwikkeling. Hans Lamboo heeft in DB/M nummer 4 reeds verslag gedaan van UDS, dat op de begin mei gehouden gebruikersconferentie Informatica World 2004 uitgebreid aan bod is gekomen.

## ETL-hulpmiddel: PowerCenter

De roots van Informatica liggen bij data-integratie; vanaf 1993 tot de introductie in 1996 is er ontwikkeld aan PowerCenter (de oorspronkelijke naam is PowerMart). PowerCenter is het hulpmiddel voor het ontwikkelen van ETL-processen. PowerCenter is een Transformation Engine, zie afbeelding 1. Dit betekent dat alle processen die met dit hulpmiddel worden gemodelleerd, als metadata worden vastgelegd in een database. Processen worden via de server-componenten in het geheugen uitgevoerd. Dit in tegenstelling tot het hulpmiddel Code Generator dat compileerbare code genereert, die zonder het hulpmiddel is uit te voeren, en het hulpmiddel Code Engine, die weliswaar code likt, maar waarbij

de DBMS engine de code tijdens het uitvoeren interpreteert. Een Code Generator genereert bijvoorbeeld C++ of JAVA-code en een Code Engine bijvoorbeeld PL/SQL.

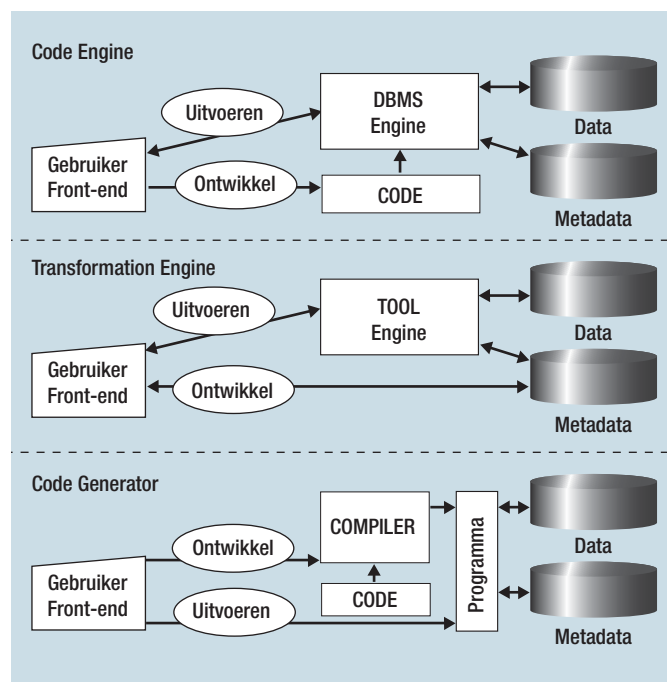
## Interface middleware: PowerConnect en PowerExchange

Naarmate er met meer verschillende typen bronnen diende te worden geïntegreerd, ontwikkelde Informatica en haar partners meer 'interfaces' of 'adapters'. In eerste instantie waren dit de PowerConnects, die als 'add-on' dienen te worden gebruikt op PowerCenter om bijvoorbeeld data uit specifieke standaardpakketten te ontsluiten. Dit voorjaar heeft Informatica ook PowerExchange uitgebracht. In tegenstelling tot de PowerConnects, is PowerExchange een op zichzelf staande productlijn. PowerExchange behelst het batch-gewijs, wijzigingsgewijs of real-time ontsluiten van data op basis van 'Change Data Capture'-technologie. Deze technologie maakt deels gebruik van bestaande mechanismen voor het vastleggen van wijzigingen (bijvoorbeeld via de DBMS transactie-log), en deels heeft Informatica deze technologie zelf ontwikkeld (bijvoorbeeld in het geval van VSAM data).

Vanzelfsprekend kunnen PowerExchange-producten ook als 'add-on' worden gebruikt in combinatie met PowerCenter. Enkele PowerConnects alsook een deel van de in PowerExchange ondergebrachte technologie is afkomstig van Striva, de technology partner, die in 2003 werd overgenomen.

## BI-hulpmiddel: PowerAnalyzer

Het Informatica-product PowerAnalyzer is een hybride hulpmiddel. Met PowerAnalyzer kunnen zowel standaardrapportages als real-time dashboards (BAM) worden ontwikkeld en alles wat daar tussen zit. PowerAnalyzer beschikt over OLAP- en portal-functio-



**Afbeelding 1:** Drie typen ETL-tools.

naliteit. Een bijzondere feature is onder meer de 'analytical workflow', waarmee rapportages op basis van inhoud kunnen worden gekoppeld tot een logische analytische keten. Daarnaast zijn er uitgebreide faciliteiten voor administratie en beheer beschikbaar, inclusief scheduling; zowel de eindgebruiker als de beheerder moeten het doen met één en dezelfde web-interface, die is gebaseerd op Java-technologie. Onderliggend dient er een Java applicatie-server te draaien om PowerAnalyzer te 'hosten'. Evenals PowerCenter is PowerAnalyzer gestoeld op repository-technologie: in de PowerAnalyzer metadata repository worden alle (definities van) objecten van onder meer dashboards, rapportages, en parameters, bewaard.

## Metadata hulpmiddel: SuperGlue

SuperGlue is het Informatica-product voor metadata-integratie en -presentatie. SuperGlue is bedoeld om metadata toegankelijk en bruikbaar te maken. Toegankelijk omdat metadata veelal is opgeslagen in afgeschermdes omgevingen of in niet gangbare formaten; bruikbaar omdat het samenbrengen van metadata het inzicht significant verbetert. SuperGlue biedt als het ware een kijkglas op allerlei metadata in (en buiten) de organisatie. SuperGlue is gebaseerd op het centrale repository-principe (zie afbeelding 2). Dit houdt in dat metadata uit andere hulpmiddelen, van BI tot en met DBMS en van een datamodelleerhulpmiddel tot en met een standaardpakket, periodiek fysiek in de centrale repository dienen te worden gerepliceerd, voordat er rapportages kunnen worden gegenereerd en er analyses op kunnen worden uitgevoerd. Afgezien van het feit dat alle metadata 'onder één dak' beschikbaar is (Informatica spreekt toepasselijk van een 'metadata warehouse') is het bijzonder aantrekkelijk om bijvoorbeeld in de metadata te kunnen zoeken, te kunnen herleiden waar bepaalde

data 'vandaan komen', en/of te kunnen bepalen wat de impact is van een specifieke wijziging (bijvoorbeeld de wijziging van een tabelstructuur of attribuutnaam).

## Partners

Hoewel partners van Informatica gebruik konden maken van de beperkte beschikbare programmeerbare interfaces (API's via de SDK) van bijvoorbeeld PowerCenter, waren tot een jaar of twee geleden de mogelijkheden voor het ontwikkelen van complementaire producten door hen beperkt. De afgelopen twee jaar is dit aanzienlijk verbeterd met de komst van de nieuwe SDK's voor PowerCenter en SuperGlue.

Tot voor kort was de 'integratie' van de Informatica-producten geregeld via de 'bovenzijde' (reporting) en de 'onderzijde' (repository). Zo wordt PowerAnalyzer als reporting-hulpmiddel (de 'metadata reporter') meegeleverd met zowel PowerCenter als met SuperGlue voor specifiek metadata reporting. Aan de 'onderzijde' fungeert, vanzelfsprekend, SuperGlue als (optionele) centrale metadata repository. Voorgaande integratie is tamelijk oppervlakkig en daarmee niet ook voldoende duurzaam. Informatica was genoodzaakt om met een betere, meer strategische, visie te komen, waarin theorie, techniek en toepassing elkaar naadloos kunnen vinden. Aldus ontstaat Informatica's Universal Data Services architectuur.

## Universal Data Services architectuur

De kern van UDS wordt gevormd door de 'data server'. Deze data server is het kroonjuweel van Informatica. De data server is de toolbox die beschikbaar is om alle gewenste data te kunnen ontsluiten voor elke willekeurige toepassing, ongeacht bovenliggend toepassingsgebied en onderliggende technologie.

De data server is opgebouwd uit: een aantal fundamentele services; open repository ('metadata gedreven'); een virtual machine ('schaalbaarheid' en 'onafhankelijkheid'); een uitbreidbaar metamodel.

## Tussen het dataportal en het datawarehouse bevindt zich nog de staging area

Bovenop de data server zijn generieke, dat wil zeggen herbruikbare, services gedefinieerd. Het voorgaande dient zodanig te zijn gedefinieerd dat enerzijds derden extensies kunnen toevoegen die Informatica niet biedt, omdat er bijvoorbeeld zeer specifieke eisen en wensen kunnen bestaan in een bepaalde situatie, en anderzijds dat er voldoende mogelijkheden zijn voor integratie van de data server/services in een bestaande omgeving, waarin zich ook andere hulpmiddelen, applicaties en dergelijke, bevinden waarmee moet worden gecommuniceerd en opdrachten en data moeten worden uitgewisseld.

Uit het groeiende klantenbestand van Informatica is, in de tijd gezien, op te maken dat er uit steeds meer en zeer uiteenlopende applicaties data worden ontsloten. UDS geeft aan dat het feitelijk niet uitmaakt om welk type applicatie het gaat: er moet eenvoudigweg een mogelijkheid zijn om data uit te wisselen, of dit nu een legacy-applicatie betreft in een host-omgeving, een berichtverwerkende transactieomgeving of een exotisch DBMS. Er moet een weg zijn, liefst meerdere!

Door de jaren heen is door verschillende instanties onderzoek gedaan naar de kosten en baten van een datawarehouse. In veel gevallen bleek dit te duur en niet aan de verwachtingen te voldoen. Informatica onderbouwt UDS mede doordat ze heeft gesignaleerd dat de afgelopen jaren steeds meer behoefte is ontstaan aan:

- hergebruik van data die is opgeslagen in datawarehouses (data hub en datasynchronisatie-initiatieven);
- nieuwe toepassingen voor datawarehouses (Business Activity Monitoring);
- hergebruik voor ingezette hulpmiddelen (datamigratie).

Het voorgaande geeft overigens ook aan dat de functie van het 'klassieke' datawarehouse kan worden verruimd (bredere inzet van de data en het datawarehouse), maar ook dat de onderliggende technieken en technologie niet meer 'exclusief' voor het datawarehouse hoeven te worden aangeschaft.

De hulpmiddelen van Informatica zijn reeds beknopt behandeld. Het is zinvol om te benadrukken dat Informatica, getuige UDS, er alles aan doet om haar producten 'zo open mogelijk te maken', om enerzijds de integratiegraad met andere applicaties te verhogen en anderzijds om de ontwikkeling van allerlei extensies te stimuleren.

Voor een product als SuperGlue is dit onontbeerlijk, en is het logisch dat ze een implementatie is van OMG's CWM (metamodelvastlegging) standaard, inclusief MOF (metamodel/integratie) en XMI (metadata-uitwisseling). Het is immers wenselijk om metadata van alle gangbare hulpmiddelen te integreren, zodat een klant een allesomvattend metadata-beeld kan worden voorgeschoteld. Op termijn zou dit ook zover moeten gaan dat metadata efficiënt, dat wil zeggen op één plaats, moet kunnen worden beheerd, hetgeen impliceert dat het metamodel tweerichtingsverkeer (opvragen én wijzigen) toestaat.

## Nieuwe applicaties, nieuwe ontwikkelingen

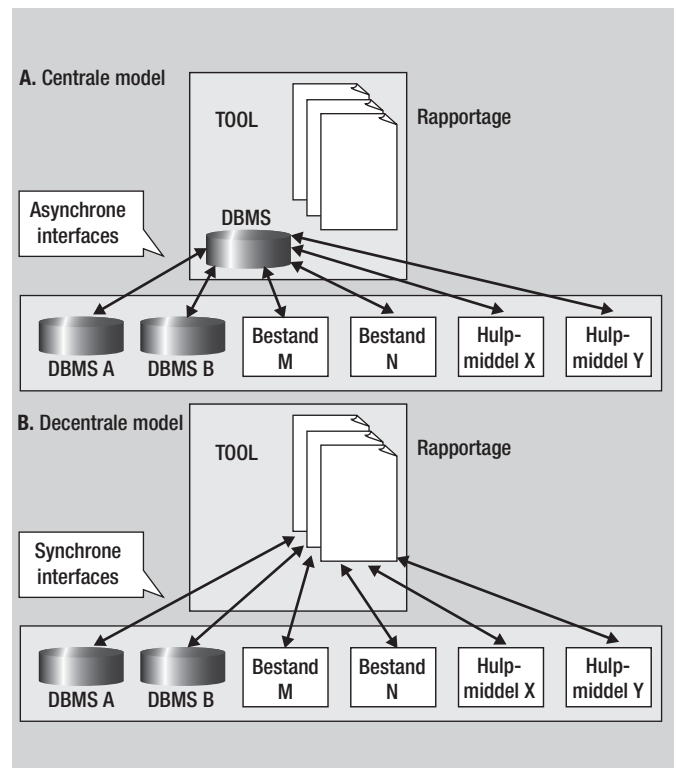
Een van de belangrijkste ontwikkelingen die Informatica voorziet, is die van de vorming van de Integration Competency Centers. Om slagvaardig te zijn is een efficiënte organisatie noodzakelijk waarin kennis en kunde bij elkaar zijn gebracht om effectief te kunnen ontwikkelen en uitrollen, en te kunnen beheren en onderhouden. Een Integration Competency Center draagt zorg voor de noodzakelijke kennisborging van applicaties, architecturen, toegepaste methoden en technieken, gebruikte infrastructuur en dergelijke. Gegeven de beschrijving van de eerder genoemde ontwikkelingen is dit pure noodzaak.

In de loop van het vorige decennium dienden zich ERP-systemen aan, gevolgd door SCM- en HRM-systemen. Hoewel voor de eindgebruiker de oplossing, introduceerden deze ontwikkelingen problemen voor de IT-organisatie. Een van de problemen van deze nieuwe systemen was dat ze feitelijk een black box waren met, zeker in het begin, matig programmeerbare interfaces. In de tweede helft van het vorige decennium kwamen hier onder meer de CRM-systemen bij. Aan de eindgebruikerszijde kwamen de BI- en OLAP-tools beschikbaar, evenals de eerste balanced scorecard (BSC) en web-toepassingen.

In het huidige decennium is het als gevolg van de kortere 'time-to-market' en product-levenscycli, noodzakelijk dat men steeds sneller reageert. Informatie moet sneller, actueler en betrouwbaarder beschikbaar zijn; er is geen ruimte voor bedenktijd.

Concepten als de Real Time Enterprise (RTE) en Business Activity Monitoring (BAM) die de afgelopen laatste jaren zijn ontwikkeld, spelen hier op in.

Recentelijk is steeds meer wet- en regelgeving ontwikkeld die verplicht om gedetailleerde informatie aan te houden, dan wel te publiceren op het gebied van de financiële (IFRS) en niet-financiële (SOX) verantwoording. Daarnaast is specifieke wet- en regelgeving (Basel II) vernieuwd voor de rapportage van kredietrisico's door voor financiële instellingen. Deze ontwikkeling alleen al heeft een enorme stimulans gegeven aan de (her)ontwikkeling van datawarehouses, eenvoudig omdat financiële instellingen meer en langer gedetailleerde historie moeten bewaren van het kredietrisico dat ze lopen.



Afbeelding 2: Metadata repository-technologie.

Een belangrijkste ontwikkeling is die van de Service Oriented Architecture (SOA). Verschillende applicaties dienen zodanig open te zijn dat eenmaal gebouwde componenten benaderbaar (herbruikbaar) zijn via generieke services. Dit voorkomt logge applicaties, redundantie, overmatig onderhoud enzovoort. Dit is niet alleen van toepassing op interne applicaties. SOA impliceert dat er generieke middleware, een applicatiebus, wordt geïmplementeerd. Via een dergelijke applicatiebus vindt gestandaardiseerde communicatie plaats, worden verzoeken gedaan, transacties uitgevoerd en (meta)data verstuurd. Waar in het verleden vaak kopieën van datasets werden bijgehouden in verschillende applicaties, is het tegenwoordig slimmer om slechts één master dataset bij te houden en deze toegankelijk te maken voor de verschillende applicaties. Hoewel de integratie van data in het kader van SOA cruciaal is, spreken we nu vanuit de applicatie, waardoor het onder de noemer Enterprise Application Integration (EAI) wordt gevat. Vanzelfsprekend bestaat er een behoefte aan een gateway tussen de primaire en de niet-primaire informatiesystemen. Daarnaast wordt SOA in één adem genoemd met web services en onderliggende technologie.

## Van datawarehouse tot dataportal

Een flink aantal datawarehouses heeft naast de generatie van relevante rapportages voor een bepaalde business, ook de mogelijkheid gekregen om data uit te leveren voor andere afdelingen in de organisatie. Naarmate dit serieuzere vormen begint aan te

nemen, kunnen we spreken van het 'dataportal' (in Informatica termen een 'data hub'). Naarmate er meer bronnen worden ontsloten in het datawarehouse, ontstaat binnen de organisatie de behoefte om de ontsloten brondata ook te gebruiken voor andere doeleinden. Het heeft immers geen zin om de data dubbel te ontsluiten voor andere toepassingen. Het bewandelen van de 'koninklijke weg', dat wil zeggen het betrekken van de brondata via het datawarehouse is vaak ook geen optie, omdat bijvoorbeeld de opdrachtgever in een ander bedrijfsonderdeel zit dat in hergebruik geen belang ziet (of geen belang heeft, het leidt immers tot vertraging van het bereiken van de eigen doelen). Bovendien is de weg die data aflegt naar het datawarehouse en/of de datamart vaak zo lang en voorzien van allerlei extra tussenstappen, dat de drempel te hoog is.

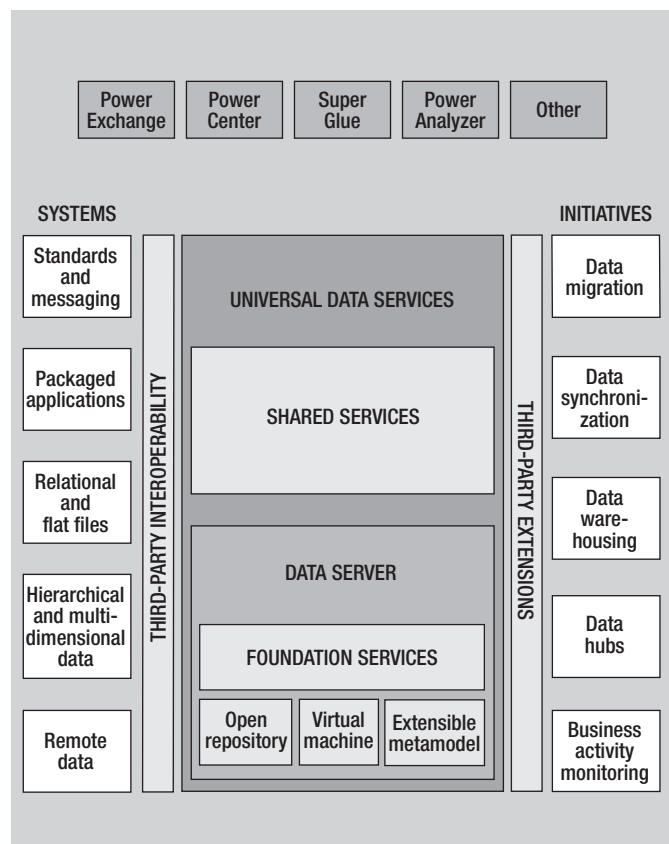
## Bovenop de data server zijn generieke, herbruikbare, services gedefinieerd

In sommige gevallen ontstaat dan een stap tussen de bronsystemen en het datawarehouse: het dataportal. Het dataportal is de omgeving waarin de brondata worden verzameld, gearchiveerd, tot op zekere hoogte gecontroleerd, maar ook uitgestuurd. Tussen het dataportal en het datawarehouse bevindt zich nog de staging area. Wanneer data van het dataportal dienen te worden verwerkt in het datawarehouse, dan vindt dit via de staging area plaats. Tevens geldt dit voor data die uit het datawarehouse worden geëxtraheerd en dienen te worden verzonden naar afnemers. Deze data komen via de staging area in het dataportal terecht. Het dient uitdrukkelijk te worden vermeld dat het dataportal niet noodzakelijkerwijs fysiek (bestanden, database) hoeft te zijn geïmplementeerd. Anders gezegd, het dataportal is een concept dat doorgaans op een hybride wijze wordt geïmplementeerd (tot en met een eigen beheerorganisatie aan toe)!

## Datakwaliteit

De beschreven ontwikkelingen dwingen organisaties ertoe om steeds meer aandacht te besteden aan de kwaliteit van hun informatievoorziening. In veel gevallen is de oorzaak te vinden in allerlei (zelfgebouwde) legacy-applicaties die verantwoordelijk zijn voor een matige kwaliteit van de data. Herbouw brengt conversie- en migratieproblemen met zich mee. Methoden en technieken op het gebied van data-analyse, zoals profiling, dienen structureel te worden ingezet om de datakwaliteit te verbeteren. Ook de beschikbaarheid en de kwaliteit van metadata, bijvoorbeeld logische (business) en technische (implementatie) begrippen, definities en formaten zijn hier onontbeerlijk.

De eerder genoemde nieuwe wet- en regelgeving kunnen grote consequenties hebben voor organisaties, omdat organisaties, hun



Afbeelding 3: Algemeen overzicht UDS.

bewindvoerders en management (persoonlijk) aansprakelijk kunnen worden gesteld voor niet alleen ondeugdelijk bestuur, maar ook voor de kwaliteit van de informatie en de kwaliteit van de onderliggende (controle) processen en mechanismen waarmee deze informatie tot stand was gekomen. Dit maakt dat niet alleen de vastlegging van data secuur dient te gebeuren, maar dat er tevens uitgebreid aandacht dient te worden besteed aan de vastlegging van alle controle- en bewerkingsstappen, zowel uit oogpunt van volledigheid als uit oogpunt van herleidbaarheid.

Elk van de bovengenoemde ontwikkelingen is met Universal Data Services te positioneren, en voor elk is de technologie van Informatica te combineren tot een oplossing. Ook valt te concluderen dat elk van de Informatica-technologieën hybride kan worden ingezet. Zo kan PowerCenter niet alleen worden toegepast voor het klassiek laden van een datawarehouse, ook in de data-migratie- en datasynchronisatie-initiatieven is een belangrijke rol weggelegd voor PowerCenter. Een vergelijkbaar betoog is te houden voor een product als PowerAnalyzer, dat zowel rapportages en dashboards voor de eindgebruiker (met als focus: content-gerichte data en metadata) als voor de beheerder (met als focus: proces-gerichte data en metadata) kan produceren.

## Is UDS volledig?

In het voorgaande is geprobeerd aan te geven dat UDS als architectuur en de Informatica-producten als implementatie van deze architectuur, potentieel een tamelijk breed gebied kunnen afdekken. Toch is er op een aantal punten nog een en ander onvoldoende beschikbaar. Voor een deel heeft Informatica zelf reeds aangegeven dat er nog het nodige op handen is. Aan de drie volgende ontwikkelingen valt te denken:

- Generatie van data-integratieprocessen (ontwerp & definitie). Op basis van de technische datamodellen van de staging-omgeving en het datawarehouse, moet het mogelijk zijn om niet al te complexe dataverwerkingsprocessen te genereren. In dit geval spreekt men van een meer workbench-achtige benadering: specificeren van functionaliteit en genereren van 'code' (lees: metadata die processen definiëren). Met concepten als Model Driven Architecture (MDA) en CWM zou dit op termijn mogelijk moeten zijn. De eerstvolgende stap is door aan de Informatica-producten meer intelligentie in de vorm van wizards toe te voegen;
- Automatische configuratie en tuning (gebruik & beheer). Een flinke evolutionaire stap kan worden gemaakt door UDS shared services te definiëren voor performance-optimalisatie en -tuning. Op basis van infrastructuur-configuraties, performance-metingen en (vuist)regels kunnen automatisch verbeteringen in de configuratie van de verschillende producten worden doorgevoerd. Tevens kunnen dergelijke 'cases' ook als 'kennisbank' worden opgenomen als metadata;
- Data server versus database server: 'its all about integrity'. Een derde ontwikkeling waarover in het kader van evolutie dient te worden nagedacht is de rolverdeling tussen de data server en

## Middleware

In een organisatie schrikt men doorgaans als de kosten worden genoemd van middleware. Meestal luidt dan het antwoord dat (a) men al verschillende varianten in gebruik heeft ("we betalen ons een ongeluk"), (b) de IT-expertise rondom de middleware die in gebruik is, toch al beperkt is ("dit kan IT er niet bij hebben"), en (c) de business case niet rond te krijgen is ("onze klanten vragen hier niet om" of "we hebben geen budget").

Veel organisaties vertrouwen nog op klassieke vormen van data-extractie (genereren extractiebestand) en datatransport (fysiek bestandsverkeer). Doorgaans is het beheer rondom klassiek datatransport matig geregeld, dat wil zeggen dat extractie- en transportfouten niet worden gelogd, of als ze al worden gelogd, toch niet worden opgemerkt. Gebruik van meer geavanceerde middleware wordt soms ook niet aangemoedigd, doordat bijvoorbeeld een (DBMS) gateway wordt geïmplementeerd op een piepkleine server. Een objectieve kosten- en batenanalyse, dient ook indirecte kosten (als gevolg van incidenten en problemen) en risico's (vertaald in onder andere maatregelen) mee te nemen. Serieuze verstoringen, onnodig herdraaien van processen, en dubbel en/of excessief datatransport moeten ook worden meegewogen. Dit kan het investeringsplaatje en met name de terugverdientijd danig beïnvloeden.

de database server. Tot nu toe worden constraints vaak impliciet (database server) én expliciet (data server) gecontroleerd. Het zou mogelijk moeten zijn om (shared) services te ontwikkelen waarmee gedrag van respectievelijk de data server en/of de database server is te beïnvloeden, om te voorkomen dat er onnodig constraints worden afgedwongen.

## Conclusie

Informatica heeft met Universal Data Services een duidelijk statement gemaakt naar haar klanten én de markt, en beide een heldere richting aangegeven. Informatica heeft met UDS gekozen voor stabiliteit en duurzame ontwikkeling, waar in het algemeen de markt, en niet in de laatste plaats de klant, profijt van zullen hebben. Immers, UDS is op basis van duidelijke architectuur-principes en -concepten ontwikkeld, en voorziet in een duidelijke behoefte.

De implementatie van UDS met Informatica's technologie is vergevorderd. Overigens, in de tijd gezien is niets compleet te noemen, omdat er altijd wel wat te verbeteren, uit te breiden of gewoon te klagen valt. De ontwikkelingen van Informatica's UDS en technologie geven vertrouwen in de toekomst.

### Martin Misseyer

Dr. Martin P. Misseyer (martin.misseyer@ordina.nl) is Profession Leader bij Ordina VisionWorks.