



Prof. Jacques Vandenbulcke over vier evoluties in de markt

Databasefederation-based information integration

Hans Lamboo

Prof. Jacques Vandenbulcke constateert dat geen van de bekende database-vendors recent hun producten totaal van scratch opnieuw heeft ontwikkeld. Naar zijn mening is er in de database-markt dan ook alleen sprake van evoluties. Information Integration eist daarin volgens hem de hoofdrol op. Een interview.

“Hoewel er geen sprake is van revoluties op de database-markt, zijn de evoluties aanzienlijk. Als eerste noem ik de evolutie die het voortbrengen van informatie en kennis stimuleert, ongeacht waar en hoe die informatie is opgeslagen. Als tweede evolutie zie ik de toenemende ondersteuning van de productiviteit van de ontwikkeling van database-applicaties; als derde de evolutie die de kosten van deployment en management van de exploitatie van databases moeten terugdringen; en als vierde een evolutie die voorziet in een robuuste on-demand infrastructuur voor databases. Ik zal deze vier evoluties stuk voor stuk verfijnen.”

De eerste evolutie

“We zien een groeiende belangstelling voor integratie in het algemeen. Dat is een gevolg van de snelle technologische ontwikkelingen en het toenemend aantal korte-termijn beslissingen en politieke processen. We zitten vast aan ‘eiland-automatisering’ – hoewel we dat eigenlijk helemaal niet willen, maar dat niet anders kan door de manier waarop bedrijven zijn georganiseerd en de toewijzing van budgetten. Bovendien zitten bedrijven door het breedschalig gebruik van pakketten met applicatie-silo’s.

We leven in een wereld van schaalvergroting en globalisering. De toenemende interesse voor integratie vindt zijn belangrijkste reden dan ook in het toenemende aantal mergers en acquisities. Het effect daarvan is immers, dat bedrijven opgescheept zitten met een applicatieportfolio die heterogeen is, in belangrijke mate redundant en veelal uiterst inconsistent. Als gevolg hiervan zijn bedrijven niet in staat om snel processen aan te passen als de marktomstandigheden dat verlangen. Dat vereist een korte time-to-market. Daarvoor is integratie nodig, men beseft meer dan ooit dat de huidige incompatibele systemen één moeten zijn.

Dit fenomeen wordt nog versterkt door de visie dat een modern bedrijf niet monolithisch meer is, maar een schakel in een waarde-omloopproces. Kortom: Business Integration is een key focus voor

alle bedrijven. Helaas heb ik de indruk dat men het dan vooral heeft over integratie van processen en applicaties, en dat men informatie-integratie vergeet.”

“Informatie-integratie heeft tot doel om data, ongeacht formaat, locatie, device of applicatie, over alle organisatorische eenheden van een bedrijf en over bedrijven die als partners samenwerken heen, te integreren. We zien als eerste belangrijke beweging op de database-markt dat vele vendors enorme inspanningen doen op het vlak van Information Integration. En er bestaan verschillende manieren om die integratie te bewerkstelligen. Ik noem Programmed Integration, maar ik vind dat absoluut geen goede methode. Het betekent immers dat de integratie *hard coded* in programma’s wordt vastgelegd en dat deugt niet. Een tweede manier is Portal-based information integration, maar ik pleit meer voor een Databasefederation-based information integration of een Model-driven information integration. Tot mijn vreugde zie dat een aantal belangrijke vendors die kant opgaat.”

Drie dimensies

“Information Integration heeft drie grote dimensies. De eerste die we zullen moeten overwinnen is de heterogeniteit van de data. We willen absoluut gestructureerde, semi-gestructureerde en niet-gestructureerde data transparant kunnen integreren. We willen één view kunnen bezitten over alle data-stores heen. Daar zit de grote uitdaging. Of een gegeven nu aanwezig is in een traditioneel bestand, in een relationele database, in een andere database, in een XML-document of een e-mail, en noem maar op, dat maakt niet uit. We willen daar op een transparante manier naartoe kunnen gaan en als het nodig is gegevens uit de verschillende stores integreren. Meerdere vendors hebben daar inmiddels behoorlijke oplossingen voor.

Een totaal andere dimensie van Information Integration is de

verspreiding van de gegevens. En hier is de uitdaging het integreren van gegevens, ongeacht de plaats waar ze voorkomen en het medium waarop ze zijn vastgelegd. Er is een aantal oplossingen op de markt voor synchrone informatie-integratie die niet allemaal maar grotendeels gebaseerd zijn op webservices, waarbij databases als informatie-provider of -requester kunnen fungeren. Er is ook een aantal oplossingen op de markt voor asynchrone informatie-integratie gebaseerd op het gebruik van message queuing, die volledig in de database wordt meegenomen. En dan de derde dimensie: het consolideren en analyseren van data. Het gaat hierbij om het integreren van business value uit informatie afkomstig uit datawarehouses (denk aan OLAP en data-mining) en informatie gehaald uit andere databronnen, tekst en noem maar op. Hier is een aantal producten op de markt die eigenlijk gegevens- en kennis-management naar elkaar brengen. We zien een aantal vendors met Relational Extractions for OLAP en met OLAP Webservices.

Information Integration is uiterst belangrijk en de vendors pakken het naar mijn mening redelijk goed aan. Er is al heel wat vooruitgang geboekt. Maar ik zie ook een groot minpunt:

om informatie modelmatig goed te kunnen integreren, spelen meta-gegevens een heel belangrijke rol. En ik vind nog altijd, dat men de problematiek van het modelleren van metadata, ondersteund met repository's waar die metadata worden opgeslagen, in de praktijk onderschat en veel te weinig aandacht geeft. Metadata zijn misschien niet 'sexy', maar uiterst belangrijk. Hier hebben ook de vendors nog flink wat werk te verrichten."

De tweede evolutie

"Dan kom ik aan wat ik de tweede grote evolutie zou willen noemen: ondersteuning van de productiviteit van de ontwikkeling van database-applicaties. Het is in die markt een tijdje stil geweest, maar inmiddels komen er – *gelukkig* – veel betere, geautomatiseerde en visueel ondersteunende database modellerings-tools. Ik zie ook steeds meer database-applicatie development plug-ins voor IDE's op de markt komen. Wat ik een goede ontwikkeling vind, is dat we meer en meer gaan

Consolidatie

"Als ik 'de vendors' noem, spreek ik vooral over IBM (DB2v8), Oracle (Oracle 10g), Microsoft (SQL Server Yukon) en Sybase (verschillende producten). De traditionele vendors dus. Daarnaast zien we (bijvoorbeeld) InterSystems die met Caché een voortreffelijk product in de markt heeft, maar die (waarschijnlijk) moet toezien dat er een reeks van producten in gebruik is die door bedrijven niet van de ene op de andere dag wordt weggedaan. Er bestaat een legacy waar men op voortbouwt die zo in het bedrijf is ingebed, dat het voor de betreffende bedrijven onmogelijk is om dat te verlaten. We zien niet veel nieuwe vendors, er zijn er alleen maar verdwenen. Er is dus sprake van flinke consolidatie op de database-markt."



Prof. Jacques Vandenbulcke: "Webservices zijn onvoorstelbaar veranderd, maar komen nu wel tot ontplooiing".

naar wat ik wil noemen Service Oriented Development. Hierbij vertrekt men vanuit de diensten aan de cliënten; deze worden vertaald naar objecttypen die men nodig heeft om de dienst te bewerkstelligen, en die weer worden vertaald naar (bijvoorbeeld) webservices in een context van SOA. Als de dienst eventueel wordt gewijzigd weet men direct welke component moet worden aangepast.

De Java-support voor databases wordt vele malen beter, zo is er nu bijvoorbeeld goede support voor de Java Development Kit. En bijna alle vendors bieden support voor .Net Microsoft Visual Studio, voor de creatie van tabellen, indexen, views en triggers, en voor .Net Microsoft Common Language, voor het creëren, uitvoeren en debuggen van onder meer stored procedures. Tussen .Net en Java is de samenwerking steeds beter. Als bij een bedrijf voor één van de twee is gekozen, sluit dat de ander niet langer uit. Vendors hebben daar goed op ingespeeld en ook vanuit beide kampen heeft men de zaken vrij goed geregeld. Ze zullen begrepen hebben dat ze naast en met elkaar zullen moeten leven, en dat interoperabiliteit een must is. We zien nieuwe drivers voor

OLE, JDBC en ODBC, bij .Net noemt men ze Data Providers, allemaal voor het sneller benaderen van gegevens. En er gebeurt nog meer rond de persoon die de applicatie-ontwikkeling moet uitvoeren en tegelijk gegevens nodig heeft vanuit de database.

Er zijn nog andere exponenten van deze evolutiedimensie: de ondersteuning van user defined types en user defined functions, native XML-datatypes voor opslag en management van XML-documenten en referentiële integriteits-checking tussen XML en SQL. Bovendien zien we steeds betere multimedia-support."

De derde evolutie

"Als derde evolutie zie ik de komst van producten die de deployment en het management van databases vereenvoudigen en zo dus de exploitatiekosten terugdringen. Een heel belangrijk attentiepunt in dit kader is Autonomic Computing. Er wordt steeds meer geïnvesteerd in de ontwikkeling van functionaliteit die het mogelijk maakt dat een database server zichzelf instelt en optimaliseert. Een en andere betekent een aanzienlijk betere ondersteuning voor de DBA. Daarvan zijn heel wat exponenten. Ik noem daarvan bijvoorbeeld Design Advisors; administratieve tools voor het evalueren van de belasting van database servers en voor het aanbevelen van wijzigingen in database-ontwerp die de performance van uitvoering van programma's moeten verbeteren. Een ander voorbeeld is Automated Statistics Collectors en Activity Monitors die door middel van intelligente agents zelf kunnen ingrijpen (waardoor je een beetje komt tot wat Rick van der Lans noemt: carefree database servers). Dan de Advanced Maintenance faciliteiten, dat zijn wizards voor het bepalen, wijzigen en eventueel *disabelen* van automatische onderhoudswerkzaamheden (self-tuning backups, automatic setting of number of buffers, buffer size, enzovoort). Ik noem Partitionary Advisors, die bepalen hoe de partitionering van de data het beste kan zijn, welke enterprise class-applicaties waar actief zijn en waar intra-query parallelisme heel belangrijk is. En tot slot: security enhancements, vele, vele verbeteringen op het gebied van informatiebeveiliging."

De vierde evolutie

"Dan is er de vierde evolutie, die voorziet in een robuuste on-demand infrastructuur voor databases, een infrastructuur die kan groeien of krimpen met de behoefte. Daar is één belangrijk niet te missen attentiepunt: Grid Computing. Er bestaan vele opinies en definities van Grid Computing; ik beschouw Grid Computing als een evolutie van distributed computing. Het betreft het gecoördineerd, transparant en beveiligd delen van IT resources over geografisch en/of administratief gedistribueerde locaties. We combineren eigenlijk een aantal onafhankelijke, heterogene systemen tot één virtueel systeem. Database-vendors spelen hier handig op in met producten zoals Automatic Storage Management (ook automatic mirroring, automatic I/O tuning), Automatic Shared Memory Management en volledige ondersteuning van NAS en SAN. Allemaal functionaliteiten die vooral performance, beschikbaarheid en betrouwbaarheid ten goede komen.

Verdwenen

"Niet alleen wat wij in de markt zien is belangrijk, ook wat we er niet meer zien. Waar zijn bijvoorbeeld de Object Oriented databases? Die zijn weg. Men treft ze hooguit nog ergens in een blackbox, waarvan niemand weet dat er zo'n database inzit. Maar voor zichtbaar gebruik zijn ze compleet verdwenen. Native XML-databases? Een groot vraagteken. Ik denk niet dat die kunnen overleven nu de grote vendors native XML-ondersteuning in de relationele databases binnenbrengen. En waar is Parallelele database software gebleven waarover men een tijdje de mond vol van had? Verdwenen."

Ik denk dat dit een enkele belangrijke evoluties zijn. Als u me zou vragen: hoe zit het met de Open Source-beweging en database software? Dan zou ik zeggen: ik treed niet in een discussie over Open Source op het database-vlak. Ik bekijk Open source liever als onderdeel van de gehele software-markt. Maar het is duidelijk dat producten als MySQL aan een inhaalslag bezig zijn; de gap wordt steeds kleiner."

"Webservices zijn onvoorstelbaar veranderd, maar komen nu wel tot ontplooiing. Het zal nog wel enige tijd duren voor er bedrijfskritische afhandelingen mee worden gedaan, want de beveiliging is het grootste probleem, dat niet eenvoudig is op te lossen. Hoewel webservices een vorm van informatie-integratie zijn, zal het gebruik van datawarehouses er niet door afnemen. Ik heb er moeite mee als een Datawarehouse wordt gezien als vorm van informatie-integratie. Ik zie het liever als een database, waarin door middel van een ETL-proces veredelde informatie bijeen gebracht wordt, en waar met OLAP en data-mining kennis voor met name het hogere management wordt gedistilleerd. Webservices spelen zich op een ander niveau af, over een aantal processen heen: de database kan zowel requestor als provider zijn.

De enige bedreiging voor het datawarehouse is de komst van real-time OLAP. Dan wordt de informatie immers direct uit de operationele datastore gehaald en na gebruik weggegooid.

Er zou een nieuwe Codd moeten opstaan, die een nieuw model naar voren schuift, gebaseerd op het relationele model, maar dan zonder de onvolmaaktheden. Tot nu toe is er niemand in geslaagd een op wiskunde gebaseerd en wetenschappelijk onderbouwd Object-Relationele database gestalte te geven. Met de bovengenoemde evoluties voor ogen, lijkt dit een hele opgave. Toch ben ik overtuigd dat dit haalbaar én noodzakelijk is."

Hans Lamboo is hoofdredacteur van Database Magazine.

Jacques Vandenbulcke is hoogleraar Informatiekunde aan de Universiteit van Leuven, België.