



Door virtualisatie lijkt voor gebruiker alle informatie lokaal opgeslagen

# Federatieve databanken

Jan Henderyckx

**Een databank kan in een applicatiearchitectuur meer bieden dan louter persistente informatie. Een federatieve databank kan heel wat bijdragen tot een efficiëntere informatiearchitectuur, al zijn er nog zwakke plekken.**

Ook alternatieven als Enterprise Applicatie Integratie (EAI) of Extract Transform en Load (ETL) software komen beter tot hun recht in combinatie met een federatieve databank. Voor u alvast de voor- en nadelen op een rij gezet. De conclusie is hoe dan ook dat met een doelmatige synergie meer kan worden bereikt dan met een individuele component.

### Conglomeraat

De hedendaagse business-omgeving stelt nieuwe eisen aan het opslaan van informatie. Steeds vaker blijkt de business operator een conglomeraat van meerdere bedrijven, elk met eigen informatiebronnen, afkomstig uit reeds bestaande toepassingen. Bedrijven met één enkele databank zijn eerder de uitzondering dan de regel. Bovendien is er een verschuiving van zelf ontwikkelde toepassingen naar pakketten voor ERP, CRM en andere. Deze pakketten hebben niet zelden een eigen implementatie van informatieopslag. Bedrijven met drie verschillende databankleveranciers zijn geen uitzondering. Voor de informatiezoeker lijkt het aantal informatiebronnen alvast onbeperkt.

Heel wat bedrijven hebben de omschakeling van de klassieke (EDI) Electronic Data Interchange-oplossingen naar een geavanceerde EAI-laag inmiddels achter de rug of er minstens de aanzet tot gegeven. De EAI software stack geeft bedrijven de mogelijkheid te evolueren naar een service oriented benadering, waarbij alle onderdelen van de software samenwerken teneinde het business-probleem op te lossen. Leveranciers van ETL bieden steeds krachtiger oplossingen die indien gewenst zo goed als

zonder vertraging informatie van de bron naar een kopie kunnen doorsturen.

Dankzij het toegenomen aanbod in functionaliteiten van zowel hard- als software is schaalbaarheid niet langer een probleem. Het consolideren van alle informatiebronnen binnen één (virtuele) machine is vanuit het oogpunt van schaalbaarheid een haalbaar objectief geworden.

### Alternatieven

Uiteraard bestaan er alternatieven voor federatieve databanken. Zo is het mogelijk om vanuit de applicatie de verschillende bronnen direct te benaderen. Nadeel hierbij is evenwel een verminderde productiviteit van de ontwikkelaars, een afgenomen efficiëntie in de uitvoering en een beperkte flexibiliteit. Vaak heeft elke bron een eigen opvragingstaal. De ontwikkelaar dient meerdere talen te beheersen, wat niet zelden leidt tot inboeten aan efficiëntie en productiviteit.

### Adapters voor niet-gestructureerde informatie zijn ontzettend waardevol

Het combineren van twee informatiebronnen verloopt via twee onafhankelijke opvragingen. De ontwikkelaar bepaalt de volgorde hiervan, helaas niet altijd even optimaal. Zo wordt voorbijgegaan aan meer dan twintig jaar evolutie in relationele databanken en dus ook optimizer-logica. Van alle alternatieven is voornamelijk het aanwenden van verschillende bronnen de minst efficiënte. In de context van één enkele informatiebron is een federatieve databank overbodig. Databankconsolidatie is dan op het eerste gezicht

Jan Henderyckx spreekt tijdens Database Systems op 20 maart 2007 over Enterprise Informatie Architectuur.

---

een alternatief. De ervaring heeft echter uitgewezen dat dit niet haalbaar is. Consolidatie mag nooit een doel op zich zijn.

In de praktijk is consolidatie veelal louter het gevolg van een applicatieve herstructurering.

Een ander alternatief is de replicatie van gegevens, wat eveneens de indruk wekt dat de gegevens slechts in één systeem aanwezig zijn. Replicatie heeft echter geen applicatieve interface. De applicatie maakt gebruik van de kopie van de gegevens die door de replicatie-infrastructuur wordt onderhouden. Bij replicatie is er meestal sprake van een vertraging tussen de wijziging van de informatie aan de bron en het tijdstip waarop de kopie wordt aangepast. Enkel in het geval van synchrone aanpassingen is er geen vertraging, maar dan is er eigenlijk niet echt sprake van replicatie. Het verkorten van de vertraging heeft niet zelden een impact op het resource-verbruik van de totale oplossing en wordt in de praktijk dan ook beperkt. Dit resource-verbruik staat in verhouding tot het aantal aanpassingen aan de bron en de gewenste vertraging. Replicatie blijkt dus aangewezen bij statische gegevens die vaak worden gelezen.

## Opbouw

Een federatieve databank vereist een aantal basiselementen, teneinde alle noodzakelijke functionaliteiten aan te bieden. Bedoeling is hoe dan ook steeds dat de virtualisatie de eigenschappen van het onderliggende systeem afschermt. Voor de gebruiker lijkt het alsof alle informatie gewoon lokaal is opgeslagen. Elke federatieve databank omvat hierdoor minstens twee componenten meer dan een gewone databank: een back-end adapter en een laag voor functiecompensatie. De rol van de back-end adapter is vrij eenvoudig. Een opvraging wordt door de federatieve databank vertaald naar een opvraging in de bron-databank. De federatieve databank is voor de brondatabank niets anders dan een client-toepassing. Niet alle federatieve databanken scoren trouwens even goed op dit vlak. Hoe beter het inzicht van de federatieve databank in de functionering van de brondatabank, hoe groter de kans op performance en schaalbaarheid van de oplossing. De adapter dient vaak niet alleen voor het uitvoeren van de vraag, maar fungeert daarnaast ook als doorgeefluik van metadata: informatie over gedefinieerde indexen, de frequentie van bepaalde waarden. Een gespecialiseerde adapter verdient daarom veelal de voorkeur boven meer generieke oplossingen als ODBC of OLE-DB drivers.

Het samenbrengen van gestructureerde en ongestructureerde informatie leidt vaak tot bijkomende business-inzichten. Adapters voor niet-gestructureerde informatie zijn dan ook ontzettend waardevol. Recentelijk zien we daarom een toegenomen aanbod van mogelijkheden om ongestructureerde informatie via een gestructureerde interface te benaderen.

Vaak zijn bronnen niet te benaderen op het niveau van de tabellen maar wel via de applicatielaag. Een adapter die webservices kan aanroepen is daarom ook essentieel voor het opzetten van een federatie met complexe bronnen. Het is daarbij wenselijk dat

de adapters van een eventuele EAI-oplossing kunnen worden ingezet.

Een wezenlijk onderdeel van de federatieve oplossing is de laag voor functiecompensatie. De vraag van de gebruiker wordt uit elkaar getrokken per gegevensbron. De syntax van de gegevensbron is niet altijd even uitgebreid als die van de federatieve server. Dan dient de federatieve databank een deel van de vraag voor zijn rekening te nemen. Het uitvoeren van een GROUP BY CUBE (COL1, COL2) query kan op die manier uiteenvallen in een eerste deel GROUP BY COL1, COL2, uitgevoerd door de bron, waarbij CUBE op niveau van de functiecompensatielaag ligt.

## Interfaces

Indien de gegevens aan de bron veelal ongewijzigd blijven, is het interessant niet de gegevens op zich te bevragen, maar de informatie op niveau van de federatieve databank op te slaan in een lokale cache. Het onderhouden van een lokale persistentielaag kan op deze manier de performance en schaalbaarheid sterk verbeteren. Werken met een lokale cache veronderstelt evenwel enerzijds het onderhouden van de informatie in de cache en anderzijds een doelmatig herschrijven van de opvraging naar de informatie in de cache.

## Bij gebruik van een lokale cache in de federatieve server neemt de complexiteit sterk toe

Het onderhouden van de cache kan gebeuren door het periodiek uitvoeren van een query op de bron die alle gegevens opvraagt en de cache vervangt. Dit is alleen mogelijk als de frequentie van refresh niet te hoog ligt en de gegevens niet te omvangrijk zijn. Bij grotere hoeveelheden gegevens of een hogere frequentie van wijzigen, is het mogelijk een replicatieoplossing in te schakelen voor het bijwerken van de cache. Het herschrijven van de query vraagt de nodige zorg en veronderstelt een perfecte kennis van de optimalisatiecomponent van de relatie tussen alle bronnen. Niet zelden dient referentiële integriteit te worden toegevoegd om op een correcte manier joins tussen gegevens aan de bron en informatie in de cache uit te voeren. Het is daarom belangrijk bij de eventuele keuze van een product niet alleen te kijken of een lokale cache mogelijk is, maar ook de efficiëntie van deze laatste te evalueren.

Een bijkomende differentiator bestaat uit de aangeboden client interfaces. Klassieke talen als SQL zijn een mogelijkheid, maar beperken de inzetbaarheid en in zekere mate ook de mogelijkheid om uit te groeien tot een informatielaag. Het ondersteunen van bijkomende interfaces als XPATH of XQUERY is dus zeker een sterke troef. Ander alternatief is het beschikbaar stellen van de informatie als een webservice. Via deze weg kan de federatieve

databank uitgroeien van een persistentielaag tot een applicatie-component en zo het traject aanzienlijk inkorten naar een Service Oriented Architecture.

## Beperkingen

De kracht van de virtualisatie is tevens de zwakke plek. Niet zelden hebben de bronnen andere niet-functionele eigenschappen. Zo zal de databank voor thuisbankieren bijvoorbeeld een grotere beschikbaarheid hebben dan de klanteninformatie. Een opvraging die informatie uit het thuisbankieren dient te koppelen aan klantengegevens zal alleen werken als beide informatiebronnen beschikbaar zijn. Aangezien alle informatiebronnen ogenschijnlijk gelijkwaardig dienen te zijn, is het risico niet denkbeeldig over te blijven met de kleinste gemene deler.

Een EII-oplossing gedraagt zich in de meeste gevallen als een databank die we ondervragen via een synchrone interface.

De client-applicatie wacht op het antwoord tijdens de eigenlijke verwerking. Bij consultatie van een bron via een asynchrone adapter ontstaan problemen, aangezien er geen garantie is met betrekking tot de antwoordtijd van de asynchrone call. Opnieuw is hier sprake van de zwakke plek van virtualisatie: het geheel is maar zo sterk als de zwakste schakel.

## De syntax van de gegevensbron is niet altijd even uitgebreid als die van de federatieve server

De beveiliging van informatie is soms alleen mogelijk via een business-logica die complexer is dan het louter uitvoeren van een SQL-query. Zo kan binnen een bepaalde toepassing een beveiligingsregel afdwingen dat de rekeninginformatie voor alle werknemers alleen toegankelijk is als de rekeningstand lager uitvalt dan 15.000 euro, tenzij de klant een product bestelde dat door de afdeling van de informatiegebruiker wordt beheerd. Deze business-logica zit vaak niet verrat in de databank, maar wel in de applicatielaag. Directe toegang tot de onderliggende gegevens is niet toegestaan om de beveiliging niet in het gedrang te brengen. De federatieve oplossing kan in dit geval niet anders dan de applicatielaag aanspreken, bijvoorbeeld via een webservice. Ook wanneer het datamodel te complex blijkt of ongekend is om bruikbaar te zijn, duiken er problemen op. Zo kunnen bijvoorbeeld de tabellen van een SAP implementatie niet worden gelezen. Opnieuw vormen webservices de oplossing om informatie te ontsluiten.

Niet alle sleutels en velden in de databanken zijn op dezelfde manier gedefinieerd, wat complexe transformaties noodzakelijk maakt teneinde verschillende informatiebronnen te combineren. Om real-time integratie en virtualisatie mogelijk te maken moet de federatieve databank in staat zijn complexe transformaties uit

te voeren. Deze eigenschap vinden we overigens vaak terug in de wereld van EAI, aangezien het uitvoeren van transformatie en de versleuteling een basisvereiste is van EAI. EII en EAI vullen elkaar goed aan op dit vlak.

## Implementatie

Een federatieve laag is in zijn basisvorm niet veel meer dan een hoeveelheid metadata met verwijzingen naar brongegevens. Een eerste stap naar een federatieve databank is dan ook niet meer dan het vervangen van de gegevenstoegangen van het bronstelsel naar de federatieve databank. Indien de syntax van de brondatabank compatibel is met de syntax van de federatieve databank, kan dit zonder enige aanpassing aan de applicatie. Door meerdere bronnen samen te brengen in één federatieve databank wordt de eigenlijke meerwaarde gecreëerd. Nieuwe toepassingen kunnen dan gebruik maken van de nieuwe databank. Aangezien het alleen over metadata gaat, kunnen de beschikbaarheid en schaalbaarheid worden gegarandeerd door meerdere kopiën van de federatieve databank te creëren. In tweede instantie kunnen de performance en schaalbaarheid worden verbeterd door het toevoegen van lokale caches.

## Een blik op de toekomst

Federatieve databanken bevinden zich nog in een fase van volledige evolutie. Met de recente aankondiging van de 'Information Server' heeft IBM de grenzen van federatieve databanken verschoven. In het model van de Information Server is er een verschuiving van een eenvoudige persistentielaag naar een informatieservice, die intern over alle componenten beschikt die nodig zijn voor het beheer van informatie: zo ook transformatie, federatie, datakwaliteit.

In de steeds veranderende business-context dient dus ook de federatieve databank continu te evolueren. In de industrie bestaat er echter geen duidelijke standaard voor het vastleggen van de integriteitsregels bij de consultatie van informatiebronnen. Een applicatie die bewust gebruik maakt van meerdere informatiebronnen is zelf verantwoordelijk voor de onderlinge integriteit van de bronnen. Als we deze bronnen nu virtualiseren en achter een federatieve dataserver plaatsen, verschuift de integriteitsproblematiek naar de federatieve server. Zonder het gebruik van een lokale cache in de federatieve server is er op zich nog geen probleem: de problematiek blijft immers dezelfde. Bij gebruik van een lokale cache in de federatieve server neemt de complexiteit echter sterk toe. Stel dat de lokale cache een vertraging van 30 seconden heeft ten opzichte van de eigenlijke informatiebron. Kan de toepassing met deze vertraging overeind blijven of niet? Momenteel bevat geen enkele API deze informatie. De federatieve databank zal dus bijkomende metadata moeten afdwingen teneinde intelligent gebruik te maken van de lokale cache. In de huidige systemen is hier vaak nog een gebrek aan granulariteit: de cache is of actief of uitgeschakeld, maar zelden op basis van de gebruikte objecten.

## Update

### Conclusie

Het is waarschijnlijk mogelijk om alle business-problemen op te lossen zonder gebruik te maken van EII, maar een federatieve oplossing kan een wezenlijk verschil maken in het totaalbeeld van de informatiearchitectuur. Het is een gegeven dat elk systeem sterke en zwakke punten heeft en het is dus zaak om zoveel mogelijk de sterke punten te valoriseren door middel van een adequate architectuurkeuze. Einddoel is steeds een informatie-architectuur die de gegevens ter beschikking stelt van de gebruikers en waarbij de onderliggende implementatie niet relevant is. Deze implementatie is onderhevig aan wijzigingen als het gedrag van de applicatielaag verandert. De informatiearchitect dient dus gebruik te maken van de drie technologieën om de informatielaag te bouwen: ETL voor het verplaatsen en onderhouden van de lokale cache en extra kopieën, EAI met het oog op complexe transformaties en de rijkheid van de adapters en EII omwille van de efficiëntie en de flexibiliteit.

#### Jan Henderyckx

Jan Henderyckx (jan\_henderyckx@brainware.be) is directeur van Brainware, een adviesbureau op het gebied van IBM DB2.

### Nederlandse BI-pionier Anago realiseert omzetgroei

Business intelligence-specialist Anago uit Houten wist zijn klantenbestand sterk uit te breiden en realiseerde daarmee een omzetgroei van 50 procent over 2006. Ook deed Anago succesvol zijn intrede in twee nieuwe specialismen, call centers en activity-based costing.

Anago is gespecialiseerd in het aansturen van complexe bedrijfsactiviteiten met oplossingen waarin rekenen een belangrijke rol speelt, zoals capaciteitsplanning, workload forecasting, what-if analyse en dashboarding. Met het rekenplatform van Anago zijn de daarvoor benodigde rekenkundige modellen snel en eenvoudig te vertalen in gebruiksvriendelijke applicaties. Dat hieraan grote behoefte is blijkt uit de voortdurende samenwerking met instellingen als ABN AMRO en het UWV. Van beide organisaties ontving Anago in 2006 nieuwe opdrachten, na eerder andere projecten succesvol te hebben afgerond. Daarnaast wist Anago in 2006 meer dan tien nieuwe grote klanten op te tekenen, waaronder hypotheekdienstverlener Stater en callcenterbedrijf HCN. Tot de klantenkring van Anago behoren onder meer financiële dienstverleners, (semi)overheidsorganisaties, outsourcing providers, venture capitalists en call centers.



# BI-ware

## De harde en de zachte kant van Business Intelligence

BI-initiatieven mislukken nog veel vaker dan andere projecten.

De BI-initiatieven moeten van de harde en de zachte kant komen. En als de harde kant van BI al praat met de zachte, spreken ze niet elkaars taal.

Het boek BI-ware is een boek voor ICT'ers en voor managers en vertelt in gewoon Nederlands wat er allemaal fout kan gaan en wat daaraan kan worden gedaan.

BI-ware bevat een bundeling van artikelen van Karien Verhagen en is een nieuwe uitgave in de reeks van DB/M Essays. De artikelen zijn gepubliceerd in de periode 2002 – 2006.

Wilt u weten hoe u Business Intelligence kunt laten slagen?

Dan kunt u niet zonder deze uitgave!

Ga snel naar [www.array.nl](http://www.array.nl) en bestel BI-ware!

Deze uitgave is mogelijk gemaakt door: **Getronics** **PinkRocade**

**DB/M**

**Array** PUBLICATIONS