



Pervasive gaat in de markt van data-integratie

# BI-paraplu wordt steeds breder

Paul van der Linden

**Data-integratie heeft betrekking op meer dan alleen ETL. Alle vormen van data-integratie, niet alleen die te maken hebben met datawarehousing, maken er onderdeel van uit. Denk daarbij bijvoorbeeld aan real-time activity monitoring, data-migratie van legacy-systemen, data-synchronisatie, replicatie van data en uitwisseling van data tussen meerdere applicaties.**

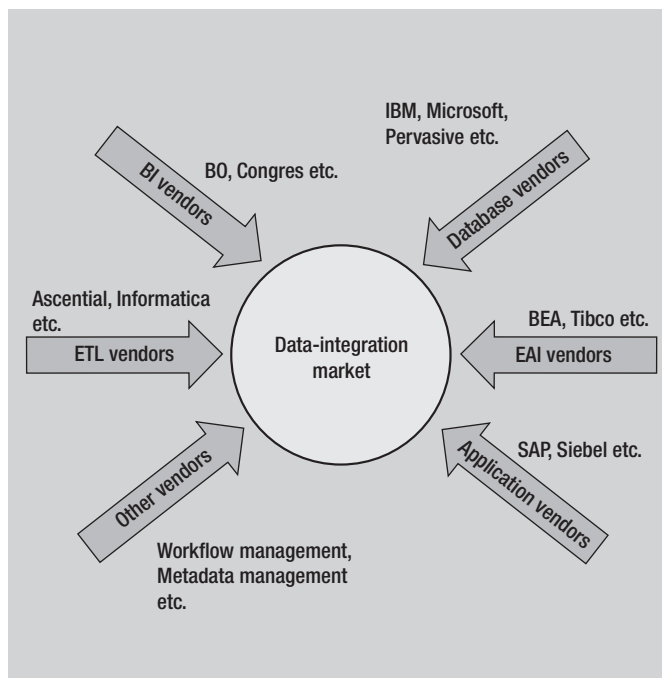
Steeds meer onderwerpen worden onder Business Intelligence geschaard, vertelde Dan Vesset, Research Director Analytics and Data Warehousing bij IDC, tijdens een conferentie in mei van dit jaar. Hij stelde vast dat onderwerpen zoals business analytics, Corporate Performance Management, enterprise reporting, portals, statistics, ETL, data visualization en scorecards, om er maar een paar te noemen, onder de steeds bredere BI-paraplu vallen.<sup>1</sup> Interessant is om vast te stellen waarom dit zo is. Is de (IT-centrische) definitie van Business Intelligence aan het veranderen? Groeien al deze onderwerpen naar elkaar toe? Of

gaat het om een meer triviale reden zoals: er valt goed geld te verdienen in BI, dus voorzien leveranciers hun software tegenwoordig van een BI-label – ongeacht waar het vandaan komt en waarvoor het oorspronkelijk bedoeld is? Waarschijnlijk spelen al deze redenen een rol.

## ETL-operatie

Een goed voorbeeld van hetgeen Vesset vaststelde is wat nu gebeurt op de markt voor data-integratie. Wie bekend is met de wereld van datawarehousing weet dat voordat data in het datawarehouse wordt opgenomen, eerst een – meestal heftige – filter- en messageslag moet plaatsvinden. Data afkomstig uit verschillende operationele systemen moeten opgepoetst, aangevuld en gecombineerd worden. In jargon: hier vindt een ETL-operatie plaats (extractie, transformatie en laden van data). De meesten (maar niet allen) onder ons zullen ETL en het datawarehouse zien als onderdeel van BI. Maar hoe zit het met data-integratie? Het zal niemand ontgaan zijn dat zelfs de zuivere ETL-leveranciers zoals Informatica en Ascential in de afgelopen periode bijzonder hun best hebben gedaan om van dit label af te komen. Zij houden zich niet langer bezig met ETL: zij houden zich bezig met data-infrastructuur. Het gaat hier daadwerkelijk om een andere boodschap.

Volgens een IDC-studie zullen ondernemingen in de komende twee jaar gemiddeld te maken krijgen met 25 data-intensieve initiatieven. Informatica heeft deze studie aangegrepen om haar Universal Data Services (UDS) concept te lanceren. Dat er momenteel een data-integratiemarkt ontstaat waar software beschikbaar is om al die verschillende data-integratie issues mee aan te pakken, is inmiddels duidelijk. Wat ook duidelijk zal zijn is dat als data-integratie in deze betekenis ook onder BI wordt geschaard, BI hiermee een betekenisloze 'catch all' term is geworden.



**Afbeelding 1:** De data-integratie-markt wordt door een grote hoeveelheid aan leveranciers bestormd.

## Eisen aan data-integratie-software

De data-integratie-markt is ontstaan vanuit het besef dat het point-to-point realiseren van datastromen elke keer dat de behoefte zich voordoet alleen in de meest eenvoudige gevallen de optimale oplossing biedt. Indien er sprake is van vele en/of veranderende datastromen (en dat is zo volgens de eerder aangehaalde IDC-studie), is een hub-and-spoke-oplossing te prefereren. Hier kan weer worden verwezen naar de ontwikkelingen die in de afgelopen jaren in ETL hebben plaatsgevonden. Ook daar hebben hub-and-spoke-oplossingen (hier genoemd: data transformation engines) het pleit gewonnen van de eerdere en eerste generatie code-generatoren. De voordelen van een hub-and-spoke aanpak zijn evident: onder andere een lager kostenniveau doordat minder interfaces onderhouden en beheerd hoeven te worden. Dit betekent meer hergebruik, minder opleidingskosten en een verbeterde kwaliteit (want minder componenten die vaker worden gebruikt). De flexibiliteit en aanpasbaarheid van een dergelijke oplossing is ook groter, doordat vanuit een centraal punt zaken worden beheerd en een beter overzicht aanwezig is.

Wat zijn de vereisten waaraan software moet voldoen om als data-integratie-software aangemerkt te kunnen worden? Op de eerste plaats moeten grote hoeveelheden data verwerkt kunnen worden. De combinatie van meer historische gegevens, een lager detailleringniveau en meerdere onderwerpen, betekent per definitie dat er meer data binnen de organisatie worden opgeslagen. Uit een survey<sup>2</sup> van TDWI (The Data Warehousing Institute) van vorig jaar blijkt dat de gemiddelde data load binnen de komende anderhalf jaar beduidend toeneemt. Had het merendeel te maken met een data load van minder dan 1 GB, de verwachting is dat 18 maanden later gemiddeld genomen 50 procent een data load tussen 1 en 500 GB heeft.

## Pervasive kent twee productlijnen: database en data-integratie.

Een tweede vereiste is het aankunnen van veel verschillende databronnen. Uit het eerder aangehaald onderzoek blijkt dat organisaties gemiddeld genomen uit maar liefst 12 databronnen hun data betrekken. Hierbij geldt dat Excel en Access als belangrijke (zelfs kritische) databronnen worden genoemd. Tegenover de toename van data staat het kleiner worden van het beschikbare batch window waarin de data betrokken en verwerkt moeten worden. Een beschikbaarheid van 24\*7 betekent dat er geen system downtime meer is. Een ander probleem hierbij is dat naarmate het aantal relevante databronnen toeneemt, de kans groter is dat de update cycles van deze systemen moeilijk verenigbaar zijn, waardoor het nog moeilijker wordt om tijdens

	Eisen aan data-integratie software
1	Grote hoeveelheden data
2	Verschillende databronnen
3	Krimpende batch windows
4	Operationeel BI
5	Datakwaliteit add-ons beschikbaar
6	Metadata management
7	Bieden van packaged solutions

**Afbeelding 2:** De zeven aandachtspunten bij data-integratie, volgens TDWI.

de al kleiner wordende batch windows de gedefinieerde jobs te laten lopen.

De vierde eis is dat BI steeds meer een operationeel karakter heeft gekregen. De business is veeleisender geworden als het gaat om de 'versheid' van de aangeboden informatie. Het TDWI-onderzoek wijst uit dat het percentage bijna real-time informatie/meerdere keren per dag verversen waarschijnlijk oploopt van de huidige 21 procent (cijfers 2003) naar een vermoedelijke 49 procent in 2005. Als vijfde eis geven organisaties aan dat ze behoefte hebben aan datakwaliteits-tools en dat ETL-tools een centrale rol moeten spelen in metadata management (eis zes). Tenslotte bestaat de behoefte aan packaged solutions: het standaardiseren op een enkel infrastructuurproduct waar het gaat om alle data-integratiebehoefte. Zie hier de zeven punten waaraan data-integratie-software idealiter voldoet (afbeelding 2).

## Pervasive

De wortels van het huidige Pervasive gaan terug tot 1982 toen SoftCraft werd opgericht. SoftCraft is het bedrijf dat de Btrieve database ontwikkeld heeft. Btrieve is een veel toegepaste database die vaak embedded met applicaties wordt meegeleverd. Vanwege het embedded karakter ervan is de bekendheid van Btrieve minder groot dan zou passen bij het grote aantal toepassingen. Binnen de database-wereld heeft Btrieve echter zonder meer een goede naam. Vijf jaar na oprichting werd SoftCraft overgenomen door Novell, die ook het bedrijf Btrieve noemde en vervolgens zelfstandig in de markt zette. Sinds 1996 heet het bedrijf Pervasive. De opvolger van Btrieve, Pervasive.SQL, werd in 1999 geïntroduceerd. Tot vorig jaar was Pervasive een database only company. Met de overname van Data Junction Corporation (augustus 2003) is daar een einde aan gekomen. Het nieuwe Pervasive kent twee productlijnen: database en data-integratie.

De combinatie van deze twee lijnen moet Pervasive als data-infrastructure company positioneren. Intelligent Enterprise noemt Pervasive als een van de bedrijven op hun 'companies to watch'-lijst.

## Producten

De producten die Pervasive biedt vallen uiteen in database software (voortkomend uit de oorspronkelijke wortels van SoftCraft) en data-integratie-software (verworven via de overname van Data Junction). Binnen deze twee productlijnen gaat het om de volgende producten.

Database Software:

- Pervasive.SQL: nieuwste versie van de Btrieve database. Gericht op kleine tot middelgrote organisaties, die wel de behoefte hebben aan een database, maar niet aan een DBA. Vaak toegepast als embedded database;
- Pervasive DataExchange: voor data-replicatie en disaster recovery;
- Pervasive AuditMaster: mogelijkheid tot real-time monitoring van data.

Data-integratie Software:

- Pervasive Business Integrator is bedoeld om meerdere applicaties real-time en event-gestuurd samen te laten werken. Deze applicaties kunnen zich binnen hetzelfde bedrijf of bij verschillende bedrijven bevinden. Om dit te bewerkstelligen staan grafische tools ter beschikking om de processtromen (workflow) te ontwerpen. Daarnaast zijn er ontwerp-tools aanwezig om documentschema's te ontwerpen (ter ondersteuning van standaards zoals EDI, FIX, SWIFT, EDIFACT, HL7, HIPAA, Sarbanes-Oxley en vele op XML gebaseerde standaarden) en is er ondersteuning voor IBM MQ-series, MSMQ en JMS-compatibele wachtrijen. Tot de beschikbare add-ons behoren de EDI Connector bundle (EDI, EDIFACT, TRADACOMS), een Health Care en een Finance Connector bundle alsook een Application Connector bundle waarin alle ERP-, CRM- en andere applicatie-connectoren zitten. De Business Integrator vormt de applicatie-integratiecomponent van het aanbod van Pervasive;
- Pervasive Data Integrator richt zich op het samenvoegen van data. Hiertoe zijn zo'n 150 directe data-connectoren beschikbaar alsmede tools voor het visueel mappen van data-transformaties. Overigens wordt native support voor MS SQL Server, Oracle, Sybase en DB2 gezien als een add-on, die apart moet worden aangeschaft (zitten in de zogenaamde Enterprise Connection bundle). Data Integrator is de ETL-tool in het data-integratie aanbod. Data Integrator kent een Process Designer, Map Designer en Schema Designer. De Data Integrator draait onder Windows2000/XP, Linux, SUN Solaris, HP-UX en IBM AIX;
- Data Junction Migration Toolkit is bedoeld om data tussen oude en nieuwe applicaties of datastores te migreren. Het beschikt over connectiviteit met meer dan 80 verschillende soorten

databronnen en -doelen. Visuele mapping-tools moeten gebruikers in staat stellen om data snel te mappen en te transformeren van bron naar doel en een structuur te creëren voor data waarvoor geen gedefinieerde schema's bestaan.

De Migration Toolkit draait alleen onder Windows 2000/XP.

## Concurrentie

Met de overname van Data Junction vorig jaar en de introductie van deze producten dit jaar, stort Pervasive zich in de nieuw ontstane data-integratiemarkt. Deze markt wordt gedomineerd door de ETL-kampioenen Ascential en Informatica. Beide bedrijven worden dan ook door Gartner geplaatst in het leiderssegment in het zogenaamde Magic Quadrant. De ETL-markt zal waarschijnlijk opgaan in de nieuwe data-integratiemarkt. Het lijkt niet logisch dat er een aparte ETL-markt blijft bestaan naast de data-integratiemarkt – zeker niet als de zwaargewichten Informatica en Ascential de centrale rol hierin bezetten. Ascential heeft met versie 7 van DataStage Real Time Integration Services geïntroduceerd. Informatica biedt met PowerCenter 7 de mogelijkheid om van objecten in een dataflow een webservice te maken. De UDS data server wordt gebruikt om te zorgen voor de performance, schaalbaarheid, beschikbaarheid, beveiliging, scheduling en de workflow. Dat de emoties hoog kunnen oplopen bleek afgelopen april, toen in METAspectrum naast Informatica en Ascential ook opeens SAS en Pervasive als leiders in data-integratie werden gepresenteerd.

## De business is veeleisender geworden als het gaat om de 'versheid' van informatie

Pervasive richt zich op bedrijven vanaf middelgroot tot groot. Positief is dat Mike Hoskins, overgekomen van Data Junction, vice-president en general manager van Pervasive's Integration Products Division is. Hiermee is gezorgd voor een stuk stabiliteit en eenheid in visie.

Kijkend naar de zeven aandachtspunten van TDWI, kan geconcludeerd worden dat het nu ontstane aanbod van Pervasive zeker gezien mag worden. Met name de ETL-kant ervan alsook de Enterprise Application Integration (EAI) zijn goed ingevuld. Aspecten die wellicht versterking kunnen gebruiken hebben betrekking op datakwaliteit en metadata management. Hiervoor biedt Pervasive wel wat handvatten, maar geen echt doorontwikkelde visie en software.

## Conclusie

Vakbladen als Intelligent Enterprise, Infoworld en Data Management Magazine hebben Pervasive al meermaals betiteld

---

als 'promising', 'innovative' en als 'company to watch'. De data-integratiemarkt is een nieuw ontdekte groene weide waarop steeds meer leveranciers hun ogen hebben laten vallen.

Leveranciers uit de ETL-hoek (zoals Ascential en Informatica), maar zeker ook uit de EAI-wereld (zoals BEA en Tibco) hebben uitstekende mogelijkheden om hun oorspronkelijke sterkten aan

## Zeker bij middelgrote ondernemingen zou Pervasive een belangrijk deel van de markt moeten kunnen afdekken

te vullen met functionaliteit die nu nog ontbreekt om zich met recht data-integratie vendor te mogen noemen. Gezien de veelheid van eisen die aan data-integratie wordt gesteld (zie de zeven punten zoals genoemd door TDWI) zullen nog vele nieuwkomers en oude bekenden zich als data-integratieleverancier (gaan) opwerpen.

De kaarten liggen voor Pervasive niet slecht. Met de vorig jaar verworven en onlangs geïntroduceerde data-integratieproducten heeft het bedrijf zeer zeker de mogelijkheid om een eigen plek in deze zich ontwikkelende markt in te nemen. In het verlengde van de Btrieve database horen daarbij woorden als lage TCO, snelle terugverdiendtijd (en hoge ROI), eenvoud, simpliciteit en minimaal onderhoud en beheer. Dit zijn ook precies de termen die naar de markt worden gecommuniceerd. Onderbouwing met harde cijfers, bij voorkeur in *head-to-heads* met concurrerende producten, zou een en ander nog overtuigender maken. Aandachtspunten liggen vooral op het gebied van datakwaliteit en metadata management. Maar zeker bij middelgrote ondernemingen zou Pervasive een belangrijk deel van de markt moeten kunnen afdekken.

Of ze daarbij zo *pervasive* worden als met Btrieve in embedded data-bases valt echter te bezien.

1 'From Data Integration to Information Delivery: Latest Trends in Business Analytics', Dan Vesset, IDC - 2004.

2 'Evaluating ETL and Data Integration Platforms', Wayne Eckerson en Colin White, TDWI - 2003.

**Paul van der Linden** (Paul.PFH.vanderLinden@AtosOrigin.com) is senior consultant Data Warehousing/BI bij Atos Origin en geeft leiding aan Data Warehousing Cost & Life cycle Management.

---