

Het echte operationele deel van de methodologie

Enterprise Information Management (4 en slot)

Mike Ferguson en Jan Henderyckx

In deel drie van deze serie over Enterprise Information Management (EIM) heeft u meer inzicht gekregen in de eerste twee fasen van de CIM (Continuous Improvement Methodology): de Define- en de Identify-stap. Waar deze eerste twee stappen veeleer het studie- en modelleerwerk behelzen, introduceren we ditmaal in de drie daaropvolgende stappen het echte operationele deel van de methodologie, namelijk de Assess-, de Integrate- en de Serve-stap. Hiermee is de CIM-cyclus helemaal doorlopen, waarna de cyclus van voor af aan kan worden herstart, weliswaar op een hoger informatie-governance- en maturiteitsniveau.

In de eerste twee stappen van de CIM zijn dan wel het model en de relaties tot het bestaande systeem vastgelegd, toch is dit niet meer dan een momentopname. Een continue toetsing van de bedrijfsgegevens in de Assess-fase dringt zich op. Hoe dan ook een belangrijke verantwoordelijkheid, die in handen wordt gegeven van de data steward. De data steward dient vooral blijk te geven van het nodige business inzicht, aangezien bewaking van de inhoud van de gegevens onlosmakelijk verbonden is met de business op zich. Het gepaste profiel van de data steward wordt vaak buiten de IT-afdeling gevonden en wel bij voorkeur in de medewerker die vanuit de business afdeling een bijdrage heeft geleverd tot de allereerste stap in de methodologie, de Define-stap.

Enterprise Information Management (4)

In deze vierde bijdrage en sluitstuk van de serie artikelen rond EIM introduceren Mike Ferguson en Jan Henderyckx de drie laatste stappen van hun Continuous Improvement Methodology. Hiermee is de CIM-cyclus helemaal doorlopen en is een hoger niveau van informatie governance en maturiteit bereikt. Uit hun conclusie blijkt dat dit meer is dan alleen een technologisch verhaal en dat er evenzeer behoefte is aan organisatorische bijsturingen en aan de introductie van adequate informatiestandaarden.

Assess

De Assess-fase is op zich een datakwaliteitsproces, waarbij de dialoog met de IT het beste zo open mogelijk blijft, aangezien oplossingen voor kwaliteitsproblemen vaak op IT-vlak kunnen worden gevonden, voor zoverre het probleem niet te wijten is aan inefficiënte bedrijfsprocessen. We verwijzen hierbij graag naar het tweede deel in DB/M2 2008 waarin Enterprise Data Quality ook al aan bod kwam. Enterprise Data Quality is echt fundamenteel voor compliance en performance management en het algehele vertrouwen in het bedrijf. Zonder betrouwbare data en een hoogwaardige kwaliteitsbewaking in de Assess-fase komt de rest van de methodologiecyclus in het gedrang. De vorige stappen in de cyclus hebben geen zin als niet wordt geïnvesteerd in goede datakwaliteitssoftware. De Assess-fase is het sluitstuk van de twee voorgaande stappen door een proces van continue toetsing. Er zijn nu voldoende metadata vergaard om over te gaan tot het operationele aspect, met name het ter beschikking stellen van de gegevens aan de processen in de Integrate-fase.

Integrate

Het proces van data-integratie omvat software tools en architectuurpatronen om de informatie op een consistente manier ter beschikking te stellen aan het volledige spectrum van informatiegebruikers. De Integrate-stap is als het ware de motor die de Enterprise Information Architecture aandrijft. Een zuiver gedefinieerd datamodel, in combinatie met integratiesoftware, maakt het mogelijk om de data op een universele manier, ongeacht de business context, te verwerken. De informatie wordt zodoende op hetzelfde niveau getild als de bedrijfsprocessen. Dit wordt ook wel eens aangeduid als 'Information on Demand'. In een On Demand-benadering kunnen componenten zoals Cleansing en Transformatie perfect in de business workflow worden ingeschakeld. Eventuele inconsistenties worden dan meteen aan de bron aangepakt, wat de kwaliteit van de informatie alleen maar ten goede komt. Daarenboven ligt het doel van near real-time Business Intelligence op die manier een stuk dichterbij, aangezien de grenzen van de data quality firewall opschuiven tot aan de bron van de gegevens.

Software tools

Technologisch gesproken was de markt voor data-integratietools tot voor kort versnipperd in een aantal deelmarkten, elk met een

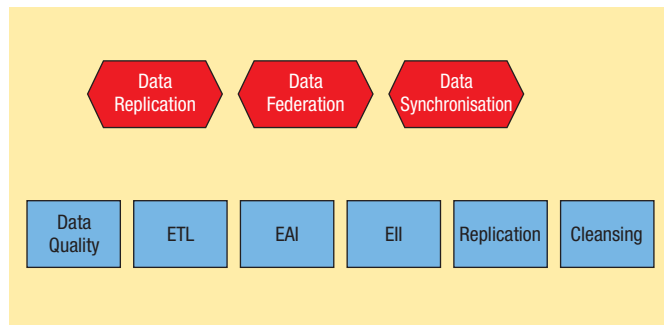
eigen technologische invalshoek. Vooral op de ETL-markt die een groot afzetgebied vond in Business Intelligence wordt nog steeds heftig gevochten om de knikkers. Verschuivingen in de behoeften van de BI-markt noopten leveranciers ertoe voortdurend bij te sturen en een kortere latentie van informatie-propagatie te bieden. De ETL-tools evolueerden steeds meer in de richting van integratiesoftware. Naast deze ETL-markt vormt de informatie-integratie (ook wel federatie genoemd, conform ons artikel Federatieve Databanken in DB/M1 2007) een belangrijk deelgebied. Hier heeft de mogelijkheid om naar een hoger abstractieniveau te gaan een belangrijke aanzet gegeven voor de algemene verspreiding als integratietool. De overige deelgebieden sluiten mooi aan bij de drie vorige stappen in de CIM, namelijk de datamodelleertools, de data quality tools en de data cleansing tools. Alle drie produceren ze metadata die noodzakelijk zijn om de integratiesoftware te laten functioneren. Op de versnipperde markt betekende dit dat de metadata vaak opnieuw dienden te worden ingevoerd. Wie dit wilde omzeilen kwam automatisch terecht bij *bridges* tussen de diverse tools.

Via een aantal technieken kan replicatie zelf worden ontwikkeld

Door de toenemende vraag naar geconsolideerde oplossingen bieden zich nu echter spelers aan die geïntegreerde oplossingen aanbieden. In Gartners Magic Quadrant van de data-integratie-producten, versie 10/2007, vinden we als leaders/visionaries dan ook niet geheel toevallig slechts twee namen terug: IBM en Informatica. Beide aanbieders zijn het verst in de samensmelting van de diverse deeloplossingen tot één coherent geheel.

Architectuurpatronen

De software mag dan wel onderhevig zijn aan een consolideringsgolf, binnen de organisatie op zich is consolidatie nog lang geen feit. De diverse afdelingen gebruiken elk eigen tools die zelden of nooit worden afgewogen in het licht van een optimale architecturale aanpak. In het versnipperde landschap met



Afbeelding 1: Architectuurpatronen.

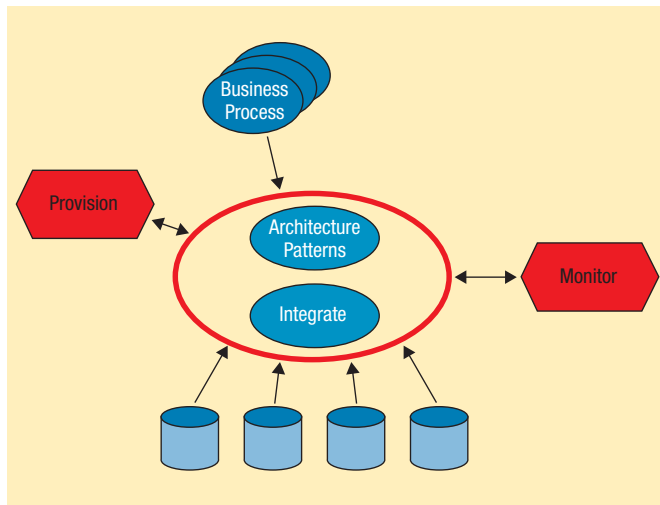
diverse architecturale oplossingen tekent zich overigens niet één duidelijke winnaar af. Het belang van een globaal overzicht is dermate beslissend dat hiervoor een informatiearchitect moet worden ingeroepen. De informatiearchitect moet kunnen beschikken over de nodige softwarecomponenten om de verschillende architecturale patronen te kunnen implementeren. De informatiearchitect moet op volstrekt objectieve manier het nodige inzicht aan de dag kunnen leggen, om in te schatten welke oplossing het beste kan worden toegepast voor welk probleem. Dit vertaalt zich naar een set van patterns met bijhorende tools, zie afbeelding 1.

Een eerste van deze patronen is *datareplicatie*. Totaal overbodig mochten alle gegevens slechts één keer worden bijgehouden. Doch zowel operationeel als functioneel is het niet altijd nodig of haalbaar om slechts één enkele kopie bij te houden. Bij meerdere kopieën is datareplicatie echt onontbeerlijk. Via een aantal technieken kan replicatie zelf worden ontwikkeld, zoals het aanduiden van de laatste wijziging, het gebruik van database triggers of applicatieve extensies. Aangezien replicatie vaak complexer uitvalt dan op het eerste gezicht lijkt, is het, met uitzondering van een *total refresh*, geen goed idee om de replicatie zelf te gaan bouwen. Desynchronisaties opsporen, de SLA bewaken, het herstarten van het replicatieproces; dit alles is niet zo eenvoudig zelf te bouwen. In de markt worden vele oplossingen aangeboden die voldoen aan alle voorwaarden en aankoop is dus een voor de hand liggende keuze.

Een tweede patroon is *datafederatie*. Er is veel voor te zeggen om gegevens in het bronsysteem te laten en de reële locatie van de gegevens te verbergen voor de applicatie. Een extra kopie is overbodig en de noodzaak tot replicatie valt dus meteen weg. Op elk moment is de meest actuele informatie voorhanden. Mocht de performance dan toch niet voldoen, kan nog altijd worden beslist toch over te gaan tot een extra kopie die via replicatie synchron wordt gehouden.

Datasynchronisatie

Een derde, meer complex, patroon is *datasynchronisatie*. Een tweewegsreplicatie is onontbeerlijk als beide kopieën in aanmerking komen voor wijziging. Deze geavanceerde vorm van replicatie wordt ook wel synchronisatie genoemd. De grotere mate van complexiteit komt volledig op conto van de logica, die nodig is voor het oplossen van conflicten die het gevolg zijn van het aanpassen van gegevens op meer dan één fysieke locatie. Er is een duidelijke neiging om dit soort problemen aan te pakken in de applicatielaag. Tekenend hiervoor is een bepaalde SOA-implementatie waarbij 30 procent van de services eigenlijk datareplicatie/synchronisatie-services bleken. Wij willen hoe dan ook pleiten om dit soort problemen in de informatie-architectuurlaag aan te pakken. Er is namelijk een fundamenteel verschil tussen het business event 'wijzigen van een adres' en het replicatie-event 'wijzigen van een adres'. Bij het business event ontstaat een volledige workflow om het event te verwerken, daar



Afbeelding 2: Monitoring.

waar bij de replicatie alleen sprake is van louter fysiek verplaatsen van gegevens. De informatiearchitect wenst bovendien de flexibiliteit om de informatiearchitectuur aan te passen los van de applicatiearchitectuur. De enige interface tussen applicatie en informatie bestaat uit de business objecten. Dit betekent dat de fysieke opslagvorm en locatie transparant zijn voor de applicatie.

Serve

Aangezien onze Enterprise Informatie Architectuur een centraal punt vormt voor de gehele Enterprise Architectuur, zijn fouten enorm kritiek en onmiddellijk voelbaar in de bovenliggende applicaties. Cruciaal is ervoor te waken dat de SLA van de applicaties op geen enkel ogenblik in het gedrang komt. Als ingeschat is hoeveel resources nodig zijn (provisioning), moet continu opgevolgd worden of de theoretische berekeningen de test van de praktijk doorstaan (monitoring).

Provision

Omdat de vraag vanuit de applicaties verre van constant is en daarenboven evolueert in de tijd, zijn voortdurende evaluaties aan de orde. Bij aankoop van een oplossing is het dan ook zaak rekening te houden met de schaalbaarheid van de oplossing en zo mogelijk meteen te opteren voor een federatieve oplossing, waarbij meerdere componenten met succes samenwerken, uiteraard met hergebruik van de metadata. De leveranciers spelen gretig in op deze vraag en bieden zelfs 'blade'-oplossingen aan: bij een toegenomen vraag kan gewoon een extra blade worden geactiveerd. De software doet de rest.

Monitor

Waar de Assess-fase tot doel heeft het model te bewaken hebben ook de architectuurpatronen behoefte aan bewaking. Zo zitten we in de Monitoring-fase, waarin men in kan schatten of het al dan niet nodig is de huidige implementatie aan te passen, bijvoorbeeld door extra kopieën van gegevens aan te maken. Het is raadzaam dat de integratielaag beschikt over virtualisatie-

mogelijkheden, zodat de informatiearchitect de fysieke implementatie kan aanpassen zonder aan de definitie te raken van de business objecten, zie afbeelding 2. In concreto dient een aantal zaken continu te worden gemeten, zoals de duurtijd van het aanleveren van informatie aan de processen en de efficiëntie van de replicatie/synchronisatie. De efficiëntie kan worden afgelezen uit de volumewijzigingen, vertragingen in de replicatie en het aantal conflicten.

Conclusie

We kunnen niet genoeg benadrukken dat IT zoveel meer is dan alleen processen, maar echt ook een zaak van informatie is. IT heet dan ook niet toevallig Information Technologie. Elk bedrijf behoeft dan ook absoluut een informatiearchitectuur. Organisaties die het er niet voldoende op laten aankomen riskeren in een situatie te verzeilen waarin compliance zo goed als onmogelijk is, waarin de kans op succes van een SOA erg klein wordt en waarin de bedrijfsprocessen gekortwiekt dreigen te worden.

Het spreekt voor zich dat het opzetten van een informatie-architectuur een stapsgewijs proces is

Kort samengevat valt of staat een adequate werkwijze voor een informatiearchitectuur met een goed inzicht in wat er voorhanden is en wat de onderlinge verbanden zijn. Vervolgens moeten nieuwe rollen worden gedefinieerd en ingevuld, zoals de data steward en de informatiearchitect. De organisatie moet wellicht worden aangepast, bestaande muren dient men af te breken en iedereen moet leren denken in een meer globale context. Natuurlijk dienen hiervoor aangepaste tools aangeschaft te worden en de nodige standaarden vastgesteld. Het spreekt voor zich dat het opzetten van een informatie-architectuur een stapsgewijs proces is, waarbij slechts geleidelijk aan vooruitgang wordt geboekt. De vooruitgang is echter direct meetbaar.

Mike Ferguson en Jan Henderyckx

Mike Ferguson is Managing Director van Intelligent Business Strategies Limited. Als analist en consultant is hij gespecialiseerd in Business Intelligence en enterprise business integration.

Jan Henderyckx is onafhankelijk consultant, spreker en auteur op het gebied van informatie-architectuur en databases.

Het copyright van de in deze serie beschreven methode voor Enterprise Information Management ligt bij de auteurs.