

Zware fundamenten voor succes van de BI-voorziening

# De volgende generatie EDW (3) – Ontwikkeling

Ronald Damhof

**Dit derde en laatste artikel over Enterprise datawarehouses gaat in op de wijze waarop een Business Intelligence gerelateerde vraag beantwoord kan worden met behulp van een BI-basisvoorziening. Account management, beheer, exploitatie en ontwikkeling worden in dit kader beschouwd als de kernprocessen van de voortbrenging.**

Meer specifiek wordt in dit artikel ingegaan op de wijze van *ontwikkeling* van gegevens- en informatieproducten. Verder wordt een tweetal fundamentele ontwikkelvormen onderscheiden en wordt ingegaan op de concentratie en plaatsing van activiteiten in het ontwikkelproces. Tenslotte wordt kort ingegaan op een mogelijk ontwikkelmodel en de wijze waarop de verschillende vormen van ontwikkeling samen kunnen werken in een gecoördineerde context. De eerste twee delen zijn in 2008 gepubliceerd in DB/M 5 en 8.

## BICC

Eindgebruiker: "Ik heb dringend een overzicht nodig voor mijn manager".

BICC: "Kun je het zelf maken, of moeten wij het doen"?

Eindgebruiker: "Ik doe het wel".

BICC: "Heb je de gegevens tot je beschikking"?

## Schaal van de organisatie

Cruciaal met betrekking tot het *ontwikkelingsproces* en het (gestandaardiseerd) inrichten hiervan, is de *schaal* van de organisatie. Misschien wel het meest onderschatte attribuut bij het inrichten een EDW-voorziening. Schaal in termen van functiescheiding, complexiteit van werkprocessen, aantal medewerkers, hoeveelheid gegevens en gebruik van gegevens. Naarmate die schaal toeneemt wordt het belangrijker de voortbrenging van gegevens- en informatieproducten in een strakker regime te zetten en de governance daarop te bewaken. Wordt dit niet onderkend, dan is de duurzaamheid van de EDW-voorziening in gevaar. Het is echter een delicate balans, de menselijke maat mag niet uit het oog worden verloren. Want het is de menselijke maat die Business Intelligence succesvol maakt.

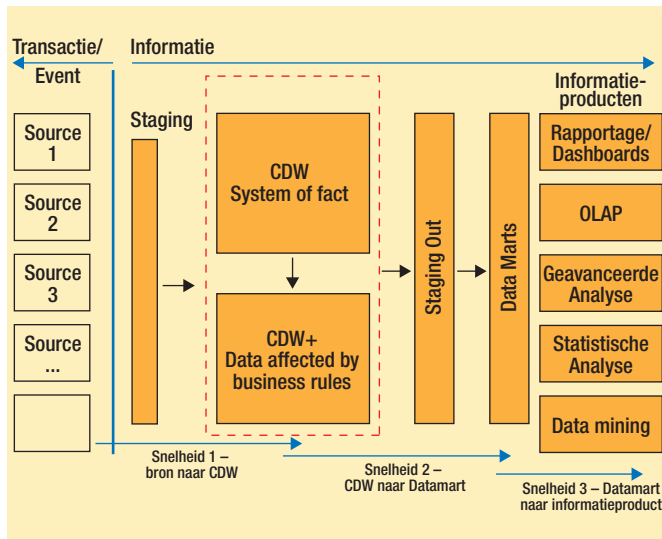
Eindgebruiker: "Voor het dringende deel wel, ik mis echter voor de volgende keer gegevens xyz".

BICC: "Die gegevens hebben we deels in het Centrale Data Warehouse al staan, wat we hebben staat eind deze week klaar, wat we niet hebben moeten we allereerst analyseren op herkomst".

Bovenstaande conversatie geeft aan dat het voortbrengen van informatieproducten in een Business Intelligence basisvoorziening verschillende snelheden kent. Bovendien kent de voortbrenging verschillende typeringen. Het initiële overzicht kan de eindgebruiker zelf maken – hij ontwikkelt zelf (lokaal) en waarschijnlijk test hij het rapport ook zelf of misschien wel samen met zijn manager. In het tweede en derde geval moeten gegevens worden voortgebracht. Het BICC neemt die taak op zich (centraal), maar kent een langere doorlooptijd. Het BICC heeft namelijk systeemvoortbrengingseisen die ze veilig wil stellen; er moet ordentelijk getest worden, capaciteitsimpact moet netjes worden ingeschat, versie- en configuratiebeheer worden ingericht voor de nieuwe datamart en de laadprocedures, alsmede moet voldaan zijn aan het beveiligingsbeleid. Voor het ontsluiten van een nieuwe bron zijn de systeemvoortbrengingseisen misschien nog wel een tikje zwaarder; er moet namelijk een koppelvlak met de OLTP-omgeving gerealiseerd worden.

## Ontwikkeling

De term 'ontwikkeling' wordt meestal automatisch geassocieerd met applicatieontwikkeling. Het resultaat van een applicatieontwikkelingsproject is direct relevant voor de business en als zodanig zeer herkenbaar. Andere ontwikkelgebieden, zoals infrastructuurontwikkeling, ontwikkelstraatontwikkeling en productie-instrumentariumontwikkeling zijn voor de business alleen indirect van belang. Budgettering is voor dat soort ontwikkelagenda's vaak een lastig verhaal. Vooral het ontwikkelen van productie-instrumentarium, bijvoorbeeld een applicatie ter ondersteuning van foutafhandeling of een applicatie voor het inzichtelijk maken van de ETL-productie, wordt gemakkelijk op de lange baan geschoven. Dit soort secundaire ontwikkelingen is vaak moeilijk te begrijpen en te verantwoorden, vooral omdat er sprake is van een ondoorzichtige mix van applicatie-, infrastructuur- en ontwikkelstraatontwikkeling. Resultaat is dan ook vaak dat een dergelijk instrumentarium pas tot stand komt als er problemen ontstaan in de exploitatiefase. Door schade en schande wordt dan duidelijk dat de productie onbeheersbaar is



**Afbeelding 1:** Verschillende snelheden van voortbrenging.

geworden omdat belangrijke instrumenten ontbreken. Voor een deel is dit de aard van de zaak: het is erg lastig te bepalen hoe het ondersteunende instrumentarium er precies uit moet komen te zien. Het is dan ook verstandig om een evolutionaire ontwikkelmethodiek toe te passen voor dit soort voorzieningen. In dit artikel zoomen we verder in op de verschillende vormen van applicatieontwikkeling.

## Verschillende snelheden van applicatie-ontwikkeling

Met de gegevens in het EDW worden verschillende gegevens- en informatieproducten voortgebracht. Gegevensproducten bestaan louter uit content. Dat kan zowel ruwe data betreffen, vanuit een authentieke bron (lees; Centrale Data Warehouse) of verrijkt (lees; datamarts). Informatieproducten zijn een combinatie van content, functionaliteit en presentatie. Bij informatieproducten is te denken aan rapportage, online analyse, data mining, enzovoort. De ontwikkeling van gegevens- en informatieproducten kent verschillende leveringssnelheden, zie afbeelding 1.

Belangrijk aspect in het voortbrengen van gegevens- en informatieproducten in een volgende generatie EDW is de onderkenning dat naarmate de Business Intelligence basisvoorziening langer draait, de snelheid van levering hoger wordt.

Het accent van applicatieontwikkeling zal in de startfase van de BI-basisvoorziening liggen op snelheid één – het ontsluiten van brongegevens naar het Centrale Data Warehouse. Naarmate het Centrale Data Warehouse meer wordt gevuld met gegevens kan steeds vaker worden volstaan met snelheid twee of drie.

Ten aanzien van informatieproducten zijn de eindgebruikers met name gebaat bij een (voor hun doel) zo groot mogelijke variatie in wendbaarheid. De eindgebruiker wil in staat zijn om producten snel aan te passen, te verbeteren of opnieuw te creëren. Daarbij komt ook nog dat verschillende groepen eindgebruikers verschillende wendbaarheidseisen kennen. Zo kent de bedrijfsvoering zeer waarschijnlijk een zwaarder testproces ten opzichte van bij-

voorbeeld marketing. Bovendien worden deze informatieproducten gekenmerkt door een grote individuele verscheidenheid (bijvoorbeeld eigen hiërarchieën, groepering, dimensionaliteit). Het ontwikkelen van gegevensproducten daarentegen wordt gekenmerkt door kwaliteitsattributen zoals leverbetrouwbaarheid, juistheid, tijdigheid, veiligheid, beschikbaarheid. Maar misschien nog wel belangrijker; gegevens worden getypeerd omdat ze voor de massa van de organisatie ter beschikking worden gesteld.

Wendbaarheid en individuele verscheidenheid voor informatieproducten versus betrouwbaarheid en massaal gebruik voor de gegevensproducten zijn kenmerken die op gespannen voet met elkaar staan. In de volgende generatie EDW wordt daarom onderscheid gemaakt tussen een 'Sovjet' achtige controle op de ontwikkeling van gegevensproducten<sup>1</sup> en een meer 'chaotische' ontwikkeling van informatieproducten<sup>2</sup>. Gegevensproducten moeten daarbij voldoen aan alle eisen die bekend zijn vanuit de systeemontwikkeling. Bij ontwikkelen van informatieproducten wordt echter een grote mate van vrijheid verleend aan de verschillende groepen eindgebruikers. Er wordt zelfs bewust rekening gehouden met een lichte vorm van 'chaos'. Het BICC speelt een cruciale rol in het neerzetten en handhaven van de spelregels waarbinnen die 'chaos' zich mag afspelen.

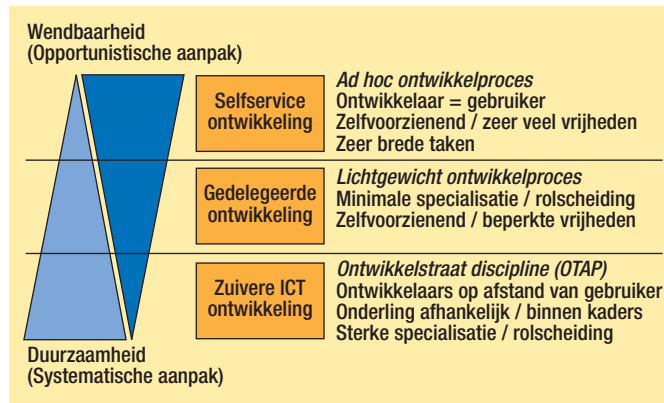
## Analyse van ontwikkelvormen

Applicatieontwikkeling laat zich vanuit twee invalshoeken karakteriseren: 1. de stijl van ontwikkelen; en 2. de concentratie en plaatsing van ontwikkelactiviteiten.

### Stijl van ontwikkelen.

De ontwikkeling kan een systematisch of een opportunistisch karakter hebben. Bij *systematische systeemontwikkeling* wordt de ontwikkeling gestuurd door een overkoepelende architectuurbeschrijving. Er wordt zo goed mogelijk rekening gehouden met lange termijn belangen: duurzaamheid is belangrijk. Er wordt veel aandacht besteed aan non-functionele eisen, zoals herbruikbaarheid, performance en de conceptuele consistentie van het totaal. De sterke samenhang en onderlinge afhankelijkheden die daarmee ontstaan maken het noodzakelijk het ontwikkelwerk centraal aan te sturen. Gevolg is dat de afstand tot de klant relatief groot is. Er moeten stabiele *requirements* opgesteld worden zodat formeel gecommuniceerd kan worden hoe het resultaat er uit moet komen te zien. Er wordt veel aandacht geschonken aan versie- en configuratiebeheer, testen, beheer en het opbouwen van documentatie ter ondersteuning van onderhoudswerkzaamheden. De voortbrenging van functionaliteit verloopt gefaseerd (Ontwikkeling, Test, Acceptatie, Productie) en de overdracht ligt in één hand. Systeemontwikkeling kan relatief grootschalig worden ingericht. Als omvangrijke complexe systemen moeten worden ontwikkeld dan is de systematische aanpak de enige werkbare.

Bij *opportunistische systeemontwikkeling* ligt het accent totaal anders. Doel is een snel resultaat en directe invloed van de



**Afbeelding 2:** Systematisch versus opportunistisch ontwikkelen.

gebruikers op dat resultaat. Als een oplossing niet voldoet wordt snel een aanpassing gedaan. Als dat niet kan dan wordt snel een nieuwe variant gemaakt. Er is geen behoefte aan een overkoepelende architectuur waar sturing vanuit gaat. De communicatie verloopt voor een belangrijk deel informeel, de ontwikkelaar heeft de meeste materiekennis paraat of directe toegang tot eindgebruikers gedurende ontwikkelwerkzaamheden. Er is geen behoefte aan een sterk gefaseerde voortbrenging (OTAP), dat werkt gewoon te traag. De praktijk leert dat grootschalige ontwikkeling bureaucratie in de hand werkt. De schaal van de ontwikkeling moet daarom relatief klein blijven om de communicatie- en coördinatie-overhead laag te houden. Ontwikkeling vindt meestal gewoon in de lijn plaats, projectmatig werken is geen noodzaak. Verregaande specialisatie van medewerkers is niet nodig, taken zijn relatief breed. Stabiele en formeel geaccordeerde requirements zijn niet nodig, de functionaliteit kan op basis van voortschrijdend inzicht enigszins organisch groeien. Systeemontwikkelaars zijn medewerkers die opgeleid zijn in de systematische ontwikkeling. Gebruikers die aan systeemontwikkeling doen gaan opportunistisch te werk. In de praktijk van de Business Intelligence ontstaat vaak een mix van de volgende ontwikkelvormen:

- Selfservice ontwikkeling: Medewerkers bouwen op een opportunistische manier oplossingen 'in place' voor zichzelf of collega's. De ontwikkelde resultaten zijn direct te gebruiken zonder enige vorm van fysieke overdracht. Er is geen rolscheiding tussen de ontwikkelaar en de gebruiker, iedereen kan aanpassingen doen. De omgeving wordt dus volledig gedeeld door iedereen. Deze werkwijze is gangbaar in de hoek van de kantoorautomatisering. Data-analisten en data miners werken meestal met selfservice. De mate van onvoorspelbaar middelenbeslag is voor deze vorm van ontwikkeling het grootst.
- Gedelegeerde ontwikkeling: Lokaal opererende medewerkers bouwen in een gescheiden omgeving oplossingen die worden gebruikt door directe collega's. Een door de ontwikkelaar direct aangestuurd overdrachtmechanisme zet resultaten vanuit de ontwikkelomgeving direct in productie. Er ontstaat een rolscheiding tussen de ontwikkelaar en gebruiker die in een lokaal verband samenwerken.

- IT-ontwikkeling: In de IT opgeleide medewerkers bouwen op een systematische manier oplossingen volgens de ontwikkel-disciplines die bekend zijn in de branche.

Het model in afbeelding 2 plaatst de geschetste ontwikkelvormen in een onderling verband.

### Concentratie en plaatsing van activiteiten.

Bij grote organisaties wordt het belangrijk de concentratie en plaatsing van ontwikkelwerkzaamheden goed te kiezen. Door mensen fysiek of in de vorm van een taakgroep bij elkaar te brengen ontstaat automatisch een proces van onderlinge aanpassing. De kosten van communicatie en coördinatie blijven laag, en handelingen worden min of meer vanzelf op elkaar afgestemd. Dus als het bijvoorbeeld de bedoeling is om een bedrijfsbrede orde af te dwingen, zoals bij het ontwerpen van een Enterprise Data Warehouse het geval is, dan moeten relevante activiteiten in één punt in de organisatie worden geconcentreerd. Als het de bedoeling is om maximaal in te spelen op lokale omstandigheden, zoals bijvoorbeeld bij de ontwikkeling van informatieproducten, dan is het beter ontwikkelactiviteiten te spreiden en zo dicht mogelijk naar de gebruiker toe te brengen. De architectuurambitie en organisatie-inrichting moeten goed op elkaar worden uitgelijnd. Gebeurt dat niet dan worden organisatieproblemen afgewenteld op de architectuur of vice versa. Op één locatie geconcentreerde front-end ontwikkeling is voor echt grootschalige organisaties niet goed voor te stellen. De kosten van communicatie en coördinatie worden zo hoog dat snel inspelen op lokale omstandigheden onmogelijk wordt. Snel kunnen inspelen op nieuwe vragen als gevolg van voortschrijdend inzicht is nu juist een essentieel kenmerk van BI. Er moet dus een zekere lokale autonomie bestaan. Dat werkt ook motiverend voor lokale medewerkers.

### Ontwikkelmodel

Afbeelding 3 toont een ontwikkelmodel in de context van de volgende generatie EDW:

- Centraal gecoördineerde Infrastructuurontwikkeling. Daaronder wordt verstaan het verzorgen van ingerichte platformcapaciteit en BI-systeemsoftware voor back-end en front-end. Tevens een portal-omgeving waarbinnen ontwikkelde webfunctionaliteit kan worden aangeboden aan afnemers.
- Lokaal gecoördineerde front-end ontwikkeling. Een lokale unit waarbinnen applicatiefunctie wordt gerealiseerd voor een deel van de organisatie. De units zitten zo dicht mogelijk tegen de primaire processen aan om 'customer intimacy' te waarborgen en efficiënte, doelmatige front-end ontwikkeling mogelijk te maken. Binnen een lokale unit bestaan twee ontwikkelvormen; selfservice ontwikkeling en gedelegeerde ontwikkeling.
- Centraal gecoördineerde EDW Ontwikkeling. Alle IT-nieuwbouwprojecten en applicatiebeheerwerkzaamheden die betrekking hebben op de ontsluiting van bronnen, het Centrale Data Warehouse en de daaruit opgebouwde datamarts.

## Ontwikkeling ‘in control’

De verschillende stijlen van ontwikkeling die zowel lokaal als centraal worden uitgevoerd stellen hoge eisen aan de wijze van samenwerking tussen de verschillende partijen. Standaardisatie van producten en processen en het toezicht daarop is hierbij een essentieel inrichtingsaspect.

*Standaardisatie in de front-end ontwikkeling (informatieproducten).* Door standaardisatie kan een bepaalde ordening worden afgedwongen. Een standaard kan betrekking hebben op producten of op processen. De lokaal gecoördineerde unit front-end ontwikkeling is zeer autonoom in haar voortbrenging van producten. Front-end producten laten zich daarmee niet goed standaardiseren. Het maximaal wenselijke is het standaardiseren van opmaak. Standaardisatiepogingen binnen de front-end omgeving zouden zich moeten richten op het uniformeren van het lokale ontwikkelproces en de daaraan ten grondslag liggende infrastructuurinrichting. Dat begint met het onderkennen van standaardrollen waaraan iedereen zich conformeert. In de tweede instantie is het van belang afspraken te maken over de wijze van testen en het beheer. Dit laatste wordt belangrijk omdat wijzigingen in de back-end door kunnen werken in de rapportages en dat veronderstelt een gecoördineerde actie waarbij het back-end team en meerdere front-end teams samen een change doorvoeren. Als iedereen een afwijkende werkwijze heeft zal dat niet lukken. Het vermogen zich aan te passen aan veranderingen in de back-end is een kritieke succesfactor. Zonder dat vermogen wordt het riskant iteratieve ontwikkeling in de data mart ontwikkeling te introduceren.

*Standaardisatie in de back-end ontwikkeling (gegevensproducten).* Bij de back-end ontwikkeling is standaardisatie van zowel producten als processen aan de orde. De toegepaste systematische ontwikkeling introduceert allerlei processtandaards en de datastructuren moeten zo uniform mogelijk worden ontworpen. Een belangrijke reden is dat de impact van fouten in de data-voortbrenging op de organisatie erg groot kan worden. De kosten van non-kwaliteit zijn erg hoog.

## Exploitatie ‘in control’

De ontwikkeling van producten wordt, op het lokale niveau, in gradaties vrijgelaten. Er is dus een zekere mate van onvoorspel-

baar middelenbeslag op de infrastructuur. Dit is in milde vorm aanwezig bij gedelegeerde ontwikkeling en in extreme vorm aanwezig bij selfservice ontwikkeling. Gevolg zou kunnen zijn dat de selfservice ontwikkelprocessen de performance van de gehele infrastructuur naar beneden trekken. Er zijn twee varianten in dit kader denkbaar:

- De selfservice ontwikkeling infrastructureel a priori te *compartimenteren*. Ofwel; selfservice ontwikkeling te isoleren op de infrastructuur<sup>3</sup>. Elke lokaal gecoördineerde unit heeft een service level afgesproken met de centraal ontwikkelde infrastructuur en architectuur. Hierin wordt bijvoorbeeld afgesproken wat het middelenbeslag is waarover de lokale unit vrijelijk mag beschikken. Gaan ze daar overheen, dan worden er nieuwe (prijs)afspraken gemaakt<sup>4</sup>.
- De mogelijkheid om het principe toe te passen van geleidelijke degradatie van prioriteit. Naarmate de gevraagde query langer loopt krijgt de betreffende query een steeds lagere prioriteit. In het laatste geval worden resources (met name CPU cycles) gedeeld en is het mogelijk een optimale verdeling van resources te bereiken. Het vereist echter veel van de infrastructurele inrichting.

De mogelijkheid om infrastructureel – al dan niet a priori – te compartimenteren is een lastige (technische) puzzel van BI-tooling (separatoren van analyse-services), besturingssysteemprocessen (workload management) en schaalbaarheidsaspecten (mocht er meer capaciteit nodig zijn dan moet dat technisch opschaalbaar zijn).

## Conclusie

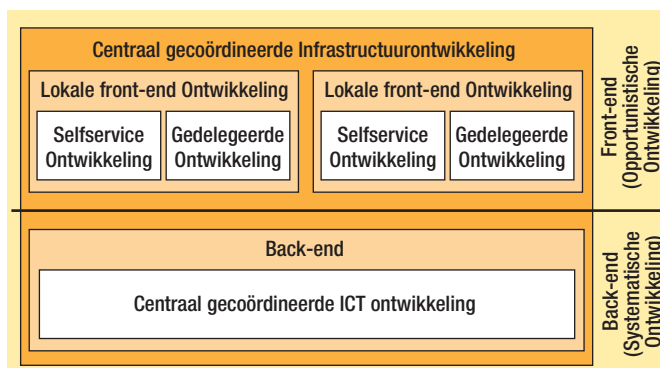
Een Business Intelligence voorziening moet de organisatie voorzien van (lever)betrouwbare gegevens in combinatie met een zeer vraaggestuurde – en dus wendbare – ontwikkeling van informatieproducten. Het is deze combinatie die ontworpen moet worden en wat ontwikkeling in de context van Business Intelligence voortbrenging zo uniek maakt. Een delicate puzzel van infrastructuurinrichting, architectuureisen, procesmodellering en organisatie-inrichting. Het is een opgave die – naarmate de schaal van de organisatie toeneemt – complex te noemen is en een lange adem vereist, maar die zware fundamenten zet onder het succes van de BI-basisvoorziening.

## Noten

1. Ook wel geduid als ‘Back-end’.
2. Ook wel geduid als ‘Front-end’.
3. In het tweede artikel van dit drieluik wordt dit bijvoorbeeld aangestipt door het gebruik van appliances voor onvoorspelbaar middelenbeslag door specifieke doelgroepen (bijvoorbeeld data mining).
4. Middelenbeslag moet meetbaar zijn en dus ook onderdeel zijn van het productie-instrumentarium van een BI-voorziening.

**Ronald Damhof** (ronald.damhof@prudenza.nl) is Informatie Management Architect en eigenaar van Prudenza.

Met dank aan Lidwine van As (lidwine@grey-matter.nl), sr. consultant op gebied van informatie-architectuur en -management bij Grey Matter.



**Afbeelding 3:** Ontwikkelmodel Business Intelligence basisvoorziening.