

Vier complementaire views op BI en datawarehousing

Generiek A&I-raamwerk bewijst zijn nut

Pim Aarts en Martin Misseyer

Al vele jaren wordt Business Intelligence toegepast binnen organisaties. Business Intelligence komt in allerlei soorten en maten voor; we hebben het allang niet meer alleen over rapporteren. Tegenwoordig is het zelf analyseren van data en het zelf samenstellen van informatie binnen handbereik. Steeds meer medewerkers krijgen beschikking over één of meer scorecards of dashboards voor het eigen team of de eigen afdeling. Het tempo waarin de business nieuwe functionaliteit wil, of beter gezegd eist, neemt nog altijd toe.

De meest recente ontwikkelingen als BI 2.0 gooien alleen maar meer olie op het vuur. Echter, veel organisaties zijn nog steeds niet goed in staat BI om adequaat te faciliteren. Het onvoldoende specificeren van uitgangspunten en randvoorwaarden, eisen en wensen, kan leiden tot verkeerde architectuur- en/of infrastructuurkeuzes. Onder druk van de business vergeet men projecten goed te leiden, test men onvoldoende en laat men beheer aan zijn lot over. Hoewel datakwaliteit en metadata al jaren als belangrijk worden beschouwd en onder de aandacht worden gebracht, is het resultaat doorgaans zwaar ondermaats. Datakwaliteit en metadata worden snel 'vergeten' als hierdoor het 'succesvol', dat wil zeggen tijdig, afronden van een project in gevaar komt. Vervolgens is het te hopen dat wanneer het project live gaat, dat het met de datakwaliteit wel meevalt en tijdens de gebruikerstraining wordt uitgelegd wat de rapportages precies betekenen en hoe deze tot stand komen.

In dit artikel wordt een generiek architectuur- en infrastructuur-raamwerk (A&I-raamwerk) voor BI en datawarehousing gepresenteerd. Met dit generieke raamwerk kunnen alle hiervoor genoemde maar ook andere belangrijke problemen communiceerbaar en adresseerbaar worden gemaakt.

Klassieke beschrijving BI-omgeving

In gangbare applicatieontwikkeltrajecten worden onder andere 'use cases' en 'node diagrammen' gebruikt om de werking van een systeem uit te leggen. Hoewel deze beschrijvingen ook voor datawarehousing gebruikt kunnen worden, leveren ze niet de informatie die voor datawarehousing gewenst en specifiek zijn in een eenvoudig toegankelijke vorm.

Het is nutteloos om use cases op te stellen voor gebruikers die eerst inloggen en vervolgens rapportages bekijken. Het zou wat nuttiger kunnen worden wanneer de ervaren gebruiker bijvoorbeeld eerst een aantal handelingen verricht alvorens hij de

rapportage bekijkt of print. Echter, het gebruik van moderne rapportage- en analysehulpmiddelen brengt talloze mogelijkheden met zich mee. Het is onmogelijk om al deze standaard functionaliteiten apart te beschrijven. Dit in tegenstelling tot het op deze wijze beschrijven van de functionaliteit van een reguliere applicatie.

In gangbare architectuurmethoden die worden gehanteerd voor applicatieontwikkeling, wordt een *project start architectuur* opgesteld. Ook bij aanvang van een BI-project wordt er vaak een dergelijk document opgesteld. Hoewel dit op zich prima is, is een project start architectuur een specifieke component en kunnen hier lang niet alle vragen mee worden beantwoord. Soms dijt een project start architectuur document uit tot een complex en niet te onderhouden document. De invulling en interpretatie van een project start architectuur verschilt nogal per organisatie. Waar de ene organisatie meer de logische of functionele zijde beschrijft, neigt de andere meer naar de technische of zelfs infrastructurale kant. Waar het aan ontbreekt is een kapstok of raamwerk waaraan een beoogde BI-oplossing aan dient te voldoen.

Naar een generiek A&I-raamwerk

Uit onze ervaringen hebben wij vier complementaire 'views' gedestilleerd. Afbeelding 1 toont deze vier views, te weten de functionele en technische architectuur views (bovenste laag) en de functionele en technische infrastructuur views (onderste laag). De *logische architectuur* beschrijft de beoogde BI-omgeving. Ter illustratie zijn de belangrijkste onderscheiden logische architectuurcomponenten (bron, e.d.) genoemd. Deze view ligt het dichtst bij de (beoogde) gebruiker: de applicatie wordt beschreven, inclusief de te genereren informatieproducten voor de gebruikers. De *technische architectuur* beschrijft de technische aspecten van de logische architectuur. Er is slechts een handjevol technische architectuur-componenten: interface (welk type koppeling bestaat

er tussen componenten), database (welk type database), proces (welk type proces) en presentatie (welk type presentatievorm). De *logische infrastructuur* beschrijft het proces van ontwikkeling binnen de gedefinieerde logische architectuur. De logische infrastructuur is geconcentreerd rondom de (logische) OTAP-inrichting. OTAP staat voor Ontwikkel, Test, Acceptatie en Productie. Een OTAP-inrichting is voorwaarde voor een gecontroleerde, betrouwbare en veilige ontwikkeling van door de business gevraagde functionaliteit. Met OTAP worden vooral de methoden, technieken en hulpmiddelen bedoeld, alsook de processen en procedures voor promotie en deployment van wat is gerealiseerd. De gekozen OTAP-inrichting bepaalt hoe vaak de functionele architectuur dient te worden weergegeven in de logische infrastructuur.

De *technische infrastructuur* beschrijft de fysiek ingerichte systemen, van hardware tot netwerk en van bronsysteem tot infrastructuur voor de eindgebruikers. Op de technische infrastructuur dienen zowel de logische infrastructuur (OTAP-inrichting) als de technische architectuur te worden geprojecteerd. Op deze manier kan worden vastgelegd in welke mate de (logische) architectuurcomponenten gescheiden worden ingericht.

In afbeelding 1 is door middel van pijlen aangegeven wat de relatie tussen de onderscheiden views is. Een technische architectuur volgt uit een logische architectuur. Vervolgens wordt de technische architectuur geïmplementeerd op de technische infrastructuur. Met behulp van de logische infrastructuur wordt hetgeen dat als de logische architectuur is ontworpen, op een gecontroleerde wijze ontwikkeld, getest en geaccepteerd en vervolgens in productie gegeven. Waar de logische infrastructuur het te doorlopen ontwikkelproces weergeeft, toont de technische infrastructuur waar dit proces op wordt geïmplementeerd.

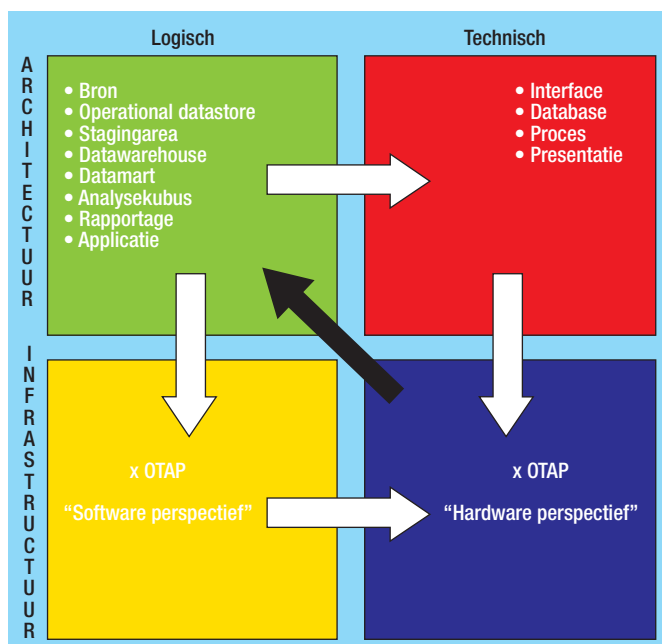
Tenslotte geeft de zwarte pijl aan dat na oplevering van het project

de gevraagde functionaliteit (logische architectuur) in productie is genomen (technische infrastructuur). Gezamenlijk vormen deze vier views het fundament voor het generieke A&I-raamwerk voor BI en datawarehousing.

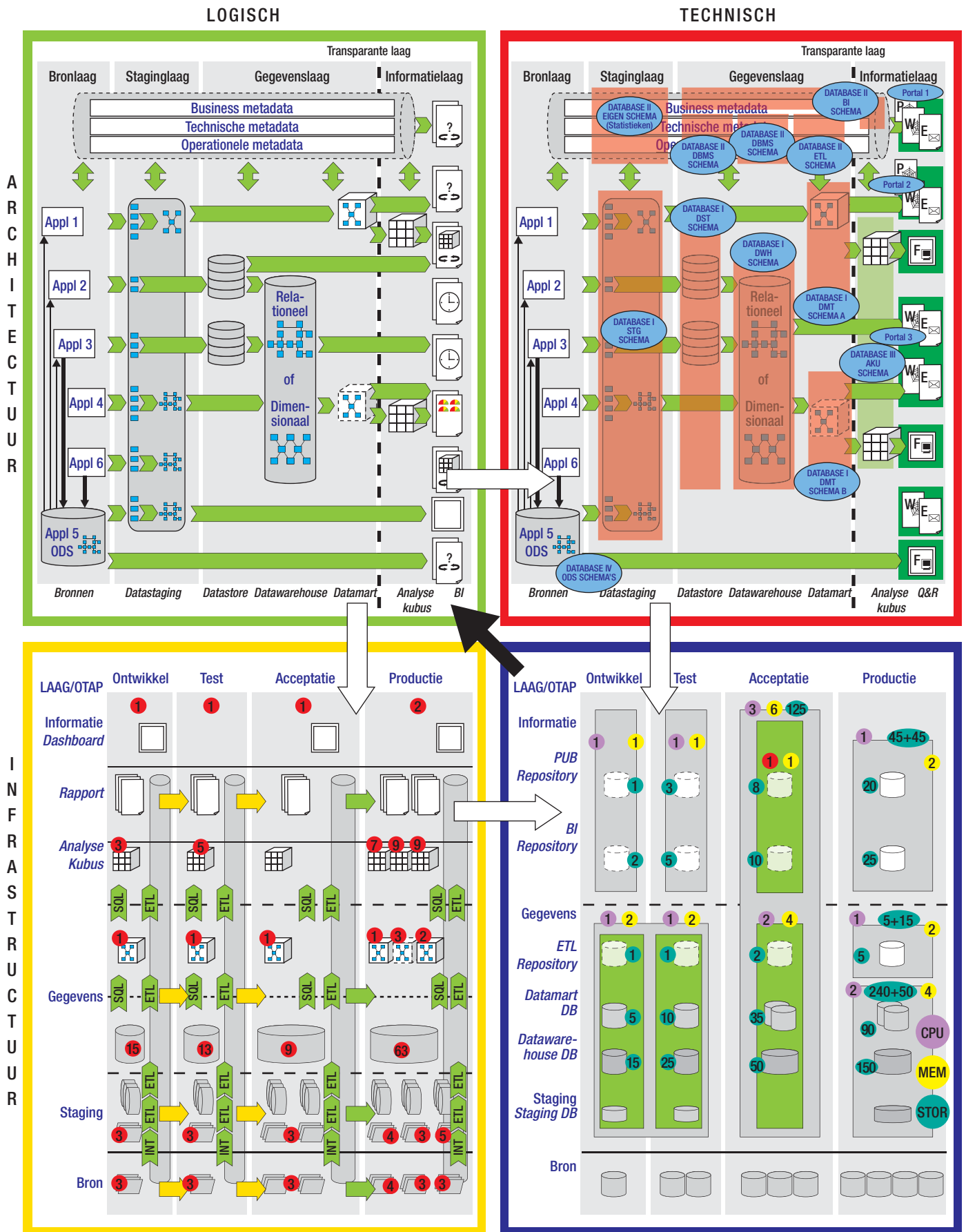
Logische architectuur view

Aan de hand van requirements wordt de logische architectuur ontworpen. Ten grondslag van een logische architectuur ligt doorgaans een business architectuur (met de business en informatiedomeinen) en architectuurprincipes. Het blok links bovenaan in afbeelding 2 (zie pagina 15) toont een tekentechniek (visualisatiwijze), voor de logische architectuur. De afbeelding toont *geen* specifiek ontworpen doelarchitectuur. De logische architectuur komt tot stand na vaststelling van de functionele requirements. Denk hierbij aan requirements zoals onder meer betrouwbaarheid, volledigheid, accuraatheid en gedetailleerdheid. Uit deze functionele requirements worden onder meer ontwerpbeslissingen gedestilleerd. Vervolgens worden deze vertaald in het wel of niet nodig hebben van specifieke architectuurcomponenten. In de tekentechniek is een vaste, gedefinieerde set architectuurcomponenten opgenomen, elk gerepresenteerd door een symbool. Elk van deze architectuurcomponenten is helder gedefinieerd en beschreven aan de hand van een aantal kenmerken. Daarnaast is per architectuurcomponent aangegeven welke vrijheidsgraden er zijn gespecificeerd. Voorbeeld: een staging area kan uit één geheel bestaan of bijvoorbeeld zijn opgeknipt in een 1-op-1 voorportaal voor de dataontvangst en een integratiegebied waarin de ontvangen data zijn gecontroleerd en (voor)geïntegreerd en klaar staan om in, bijvoorbeeld, het datawarehouse te worden gepubliceerd.

Een belangrijk onderdeel van de logische architectuur is de detaillering van de op te leveren informatieproducten. In de getoonde tekentechniek wordt een onderscheid gemaakt naar de informatieproducten standaardrapportages (klokje), flexibele rapportages (vraagteken), scorecard/dashboard (metertjes), analyses (kubus) en maatwerkapplicaties (rechthoek). Een ander aspect betreft het al dan niet fysiek of virtueel aanwezig zijn van architectuurcomponenten. Door de omlijning van bepaalde objecten gestippeld te tekenen kan men aangeven dat de betreffende objecten alleen virtueel bestaan, zoals bijvoorbeeld een datamart. Wat in het betreffende blok in afbeelding 2 ontbreekt, het is immers een tekentechniek en geen specifiek ontworpen logische architectuur, is de naamgeving van de verschillende objecten. Wanneer een logische architectuur wordt opgesteld is het vanzelfsprekend dat elke architectuurcomponent (of bijvoorbeeld per set informatieproducten) wordt voorzien van een omschrijvend label. Dit kan men doen door de naam in of naast het object zelf te plaatsen of door hiervoor een apart symbool, bijvoorbeeld een doorschijnende ovaal, op de betreffende architectuurcomponent te plaatsen. Voor een beperkt aantal architectuurcomponenten kan bijvoorbeeld ook additionele informatie toegevoegd worden, zoals het aantal objecten in het datawarehouse, het aantal kubussen of



Afbeelding 1: Contouren van een generiek A&I-raamwerk.



Afbeelding 2: Het generieke A&I-raamwerk.

het aantal rapporten. De logische architectuur wordt opgesteld door, en is eigendom van, de architect. Deze is bedoeld om een adequate governance te kunnen voeren op het ontwerp en de realisatie en anderzijds om een goed (functioneel) beheer te kunnen voeren over de te realiseren BI-omgeving. De logische architectuur wordt vertaald in een technische architectuur.

Technische architectuur view

Het blok rechts bovenaan in afbeelding 2 toont de technische architectuur view. De technische architectuur is een vertaling van de logische architectuur naar systeemcomponenten. Waar in de logische architectuur gesproken wordt over een flink aantal te onderscheiden architectuurcomponenten (zoals 'bron', 'datawarehouse' en 'kubus'), is het aantal systeemcomponenten in de technische architectuur beperkt tot '(bron)interface', 'database', 'proces' en 'presentatie'. De meeste architectuurcomponenten in de logische architectuur worden in de technische architectuur gerepresenteerd als database; deze systeemcomponenten zijn onderling verbonden door middel van interfaces. Langs deze interfaces vinden processen plaats. Naast de systeemcomponent database bestaat er ook de presentatiecomponent. Waar in de logische architectuur over informatieproducten wordt gesproken, hebben we het in de technische architectuur over presentatiecomponenten.

Wanneer we de interface met de bronapplicatie beschouwen dan kunnen we per bron specificeren of de interface met de bronapplicatie is gebaseerd op bestandsverkeer (via bestandssysteem),

op databaseverkeer (via databasekoppelingen) of op berichtenverkeer (via berichtenbus). Hier wordt niet alleen gesproken over louter data-extractie. Er kunnen functionele eisen zijn gedefinieerd voor het publiceren van (verrijkte) data uit de de BI-omgeving in een bronapplicatie. Alle koppelingen tussen database systeemcomponenten noemen we proces. Ook wanneer er binnen een database de data in meer dan één toestand bestaan, noemen we de verandering van de ene toestand in een andere een proces. In alle gevallen worden deze processen aangegeven door middel van pijlen. Functioneel zijn processen gedefinieerd op basis van het ETL-principe of het ELT-principe. Technisch zijn processen gebaseerd op statische of dynamische SQL, gecompileerde of geïnterpreteerde programmeertaalcode of op een engine.

De systeem componentpresentatie toont de informatieproducten uit de informatielaag. Presentatie omvat ook distributie. Mogelijke presentatie- en distributievormen zijn bijvoorbeeld 'Full' (informatieproducten via dedicated desktop software), 'Portal' (een informatieproduct via een webportal), 'Web' (een informatieproduct via een desktop webclient) en 'E-mail' (informatieproduct via een e-mail client). Recentelijk is hier de 'Service' aan toegevoegd, waarbij wel moet worden gesteld dat deze doorgaans via Portal of Web wordt afgenomen.

De technische architectuur wordt opgesteld door de architect in samenwerking met de database administrator. In een goed ontworpen technische architectuur wordt rekening gehouden met de combinatie c.q. de separatie van data, de gebruiksfrequentie, het



Stilstand is achteruitgang.
Hoe staat het met *jouw* carrière?

Join FourPoints !!

Dé specialist in Data Warehousing & Business Intelligence

volume, de mutatiegraad, de vergankelijkheid en de eisen ten aanzien van beveiliging etcetera. Zo is het bijvoorbeeld wenselijk om voor vergankelijke data (staging), historische data en metadata gescheiden databases in te richten, of in ieder geval voor te bereiden (gescheiden schema's in één database). Op deze manier is het in een later stadium vrij eenvoudig om alsnog een scheiding te realiseren. De technische architectuur is bedoeld om applicatie-beheer en technisch beheer in controle te kunnen laten zijn en voldoende instrumenten te geven om de BI-omgeving goed presterend, stabiel en schaalbaar te houden. De technische architectuur wordt gemapt op de technische infrastructuur.

Logische infrastructuur view

Veel organisaties, ook organisaties die al jaren BI inzetten, worstelen met de kwaliteit van het ontwikkelen van hun BI-omgeving. Bij een professionele ontwikkeling, beheer en onderhoud van BI hoort een adequate OTAP-inrichting. De OTAP-inrichting scheidt ontwikkeling van functioneel testen, functioneel testen van gebruikers(acceptatie)testen en gebruikers-(acceptatie) testen van productie. In de meeste gevallen volstaat deze (logisch) gescheiden vierdeling. De mate van scheiding wordt bepaald door functionele (en non-functionele) requirements. Functioneel bijvoorbeeld vanwege een geëiste functiescheiding, vertrouwelijkheid van productiedata en het zuiver kunnen testen en accepteren. Non-functioneel bijvoorbeeld vanwege eisen rondom performance, stabiliteit en beschikbaarheid.

Het blok links onderaan toont de tekentechniek voor de logische infrastructuur. De logische infrastructuur is een vertaling van de logische architectuur naar de gekozen OTAP-inrichting. Indien de OTAP-inrichting daadwerkelijk uit vier omgevingen bestaat, dan wordt de logische architectuur vier keer getekend. Het betreffende blok in afbeelding 2 illustreert de status van elke omgeving in de OTAP-inrichting. Met andere woorden, de logische infrastructuur wordt regelmatig ververs om aan te geven welke incrementen, projecten of modules van de beoogde c.q. bestaande BI-omgeving in ontwikkeling, respectievelijk onderhoud, zijn. De betreffende informatie is aan de logische infrastructuur door middel van een gekozen symbool (ovaal) toe te voegen.

Het afgebeelde blok toont tevens het promotie- en deployment-proces dat de ontwikkelde architectuurcomponenten doorlopen. In dit verband is het essentieel om alle specifieke aspecten aangaande dit proces goed af te stemmen op de toegepaste software (infrastructuur) en hieromtrent duidelijke procedurebeschrijvingen op te stellen. Het moge duidelijk zijn dat er in dit verband ook een roldefinitie (ontwikkelaar, tester, gebruiker, beheer) met bijbehorende definitie van autorisatie niet mag ontbreken. De logische infrastructuur wordt opgesteld door de architect in overleg met ontwikkelaars, testers en beheerders binnen de organisatie.

Technische infrastructuur view

Het blok rechts onderaan in afbeelding 2 toont de tekentechniek voor de technische infrastructuur. In de technische infrastructuur

komen de technische architectuur en de logische infrastructuur samen. Dit wil zoveel zeggen dat de OTAP-inrichting uit de logische infrastructuur wordt gecombineerd met de belangrijkste systeemcomponenten uit de technische architectuur.

De OTAP-inrichting wordt vertaald in de benodigde infrastructuur (netwerk, disk, systemen, processoren, geheugen). Hierin zijn non-functionele requirements en ICT-standaarden van de organisatie leidend. Zo kan een organisatie hebben gekozen voor het virtualisatieprincipe, waarbij de verschillende OTAP-omgevingen 'slechts' virtuele partities zijn op één of enkele grote systemen. Of heeft de organisatie gekozen voor centrale disk-opslag in de vorm van een SAN en wordt elk systeem hieraan gekoppeld. Verder kunnen op basis van bijvoorbeeld non-functionele requirements de functies *opvragen* (informatieproducten, metadata), *verwerken* (processing) en *bewaren* (databases) worden vertaald in een fysiek onderscheid van systemen. In de logische infrastructuur view is de productieomgeving op basis van deze drie-indeling ingericht.

Fysieke én virtuele servers dienen te worden voorzien van belangrijke informatie over de dimensionering, zoals het aantal processoren, de omvang van intern geheugen en de toegewezen diskruimte. Ook zou de wijze waarop systemen onderling zijn verbonden (bijvoorbeeld de dikte van de getekende lijn tussen de systemen staat voor de bandbreedte) kunnen worden toegevoegd. In de praktijk worden voor het netwerk (LAN, WAN) doorgaans aparte tekeningen opgesteld. Deze kunnen als bijlage bij het besproken raamwerk worden bijgevoegd.

De technische infrastructuur is doorgaans het domein van de infrastructuurspecialisten van een organisatie. Veel infrastructuurspecialisten in organisaties hebben geen of beperkte kennis van applicaties, laat staan van BI-omgevingen met al hun specifieke beheer. Denk bijvoorbeeld aan het feit dat technisch beheerders/infrastructuurspecialisten gewend zijn om applicaties en onderliggende query's die urenlang draaien en onnodig veel processor-capaciteit verbruiken, te beëindigen. Daarom is het wenselijk dat de (technisch) architect en de DBA op de hoogte zijn van de geïmplementeerde technische infrastructuur alsook de wijzigingen die hier in de loop der tijd op worden gepleegd (denk bijvoorbeeld aan de consolidatie van diskruimte naar een SAN). Dit om allerlei problemen rondom de performance en de stabiliteit van de BI-omgeving voor te kunnen zijn.

Toepassing

Het beschreven A&I-raamwerk is al enkele jaren in gebruik en heeft z'n nut inmiddels ruimschoots bewezen. In plaats van te gokken hoe een omgeving er mogelijkzins uit zou kunnen zien, is het absoluut aan te raden om dit zeker te weten. Dit vereist discipline ten aanzien van de te voeren governance en het onderhouden van een accurate en actuele documentatie.

Pim Aarts en Martin Misseyer

Dr. Martin Misseyer is partner bij Ordina VisionWorks, Ir. Pim Aarts is senior consultant (architect) bij Ordina VisionWorks.