

Bronnen en uiteindelijke producten blijven hetzelfde

KPN richt nieuw datawarehouse in

Teus Molenaar

KPN stelt voortaan de klant centraal in zijn bedrijfsprocessen en niet meer de applicaties. Dat heeft gevolgen voor de inrichting van het information warehouse. Bij de overgang naar het nieuwe platform is afgesproken de bronnen en output hetzelfde te houden, om het project beheersbaar te laten zijn.

Voorheen was de informatievoorziening bij het telecommunicatiebedrijf ingericht rond de diverse diensten die het aanbiedt. Als iemand, of een bedrijf, bijvoorbeeld de vaste telefonie en mobiele telefonie van deze aanbieder betrok, dan was dat genesteld in aparte systemen en kwam hij dus meerdere malen voor. Tegenwoordig is de klant het centrum en zijn de diensten aan hem gekoppeld in plaats van andersom. De verschillende bedrijfs-onderdelen (zoals verkoop en marketing) zijn alle geïnteresseerd

in de klantgegevens. Het is de taak van de afdeling Information Services om die gegevens zo accuraat en snel mogelijk te leveren. Tom Konings is architect/consultant bij Information Services en heeft de supervisie over het project van data-integratie. Hij weet zich daarbij gesteund door Thomas Reichel, principal consultant bij Getronics PinkRoccade. Reichel is verantwoordelijk voor het beheer van het information warehouse, dat KPN dus heeft uitbesteed aan deze ICT-dienstverlener.

Artemis-project

De basis voor het data-integratie project bij KPN Mobile vormt het Artemis-project. Dit is in 2005 afgerond en behelst een migratie van het hardware-platform. De onderneming had een reeks HP True 64 Unix servers staan, maar die werden niet meer ondersteund door Informatica. Daarom is KPN overgestapt op de HP Integrity servers met de Intel Itanium-processoren.

Tegelijkertijd is gekozen voor de 64-bit versie van Powercenter (het enterprise data-integratie platform van Informatica) om meer geheugen te kunnen alloceren. In de 32-bit versie kwam dat uit op maximaal 1,8 GB, dat bleek veel te weinig. "Dat is essentieel voor KPN", legt Thomas Reichel van Getronics PinkRoccade uit. "Er zijn ongeveer zes miljoen aansluitingen. Als je dat afzet tegen het model van het information warehouse dat multidimensionaal is ingericht, waarbij 'aansluitingen' één dimensie is, dan zie je dat je met heel grote dimensies te maken krijgt. Die moet je gaan cachen om snelheidswinst te halen. Als je ze op schijf zet, omdat de cache te klein is, dan gaat de performance drastisch naar beneden."

Met PowerCenter is het gelukt de tijd om call detail records te verwerken met de helft te verminderen, evenals de kosten voor het beheer van het datawarehouse."

Visie

Reichel is al langer betrokken bij de ontwikkeling van het information warehouse dat ooit is begonnen als een applicatie die MIS heette, Markt Informatie Systeem. Dat was opgezet met Oracle; veel maatwerk en dus erg duur. Het was geen datawarehouse, maar had wel de kenmerken daarvan zegt Reichel. In die tijd was het aantal 'verkeersfeiten' (de call detail records) beperkt tot 10 à 15 miljoen per dag; tegenwoordig ligt dat aantal op ruim 40 miljoen. "Die omgeving kenmerkte zich door enorme volumes en complexiteit", zegt Reichel. "En dynamiek. Er lopen altijd veel verschillende projecten tegelijk. Daarom zijn we in 2000 een programma gestart voor de bouw van een information warehouse", vult Konings aan. "De markt staat onder druk en dat we snel informatie moeten kunnen leveren die accuraat is, spreekt voor zich."

Eén van de deelprojecten betreft data-integratie. Er was een behoorlijk lange lijst met requirements (onder andere: stabiliteit, voldoende performance, bewezen technologie, lokale ondersteuning) waaraan de ETL-tool moest voldoen. Oracle Warehouse Builder was toen nog niet volwassen, meent Reichel. Ook Ascential (nu IBM WebSphere Information Integration) viel af. De keus viel op Informatica, dat Reichel toen al koploper op het vlak van ETL vond. "Het is op de data-integratiemarkt één van de



Tom Konings (l) is architect/consultant bij Information Services van KPN, Thomas Reichel (r) is principal consultant bij Getronics PinkRocade.

weinige bedrijven die nog zelfstandig zijn. En ik had en heb vooral veel waardering voor de visie van het bedrijf. Kijk maar naar een concept als Informatica-on-demand, waarbij data-integratie gewoon een service is die je kunt aanroepen vanuit elke denkbare applicatie", verklaart Reichel. Hij verwijst naar Salesforce.com dat een verkoopapplicatie als service aanbiedt. Informatica heeft koppelingen gemaakt tussen het Salesforce.com-platform en de bedrijfsapplicaties. "Dit soort ontwikkelingen, daar lopen ze wel bij voorop ten opzichte van andere leveranciers. Als je puur kijkt naar de ETL-functionaliteit, dan ontlopen de meeste pakketten elkaar niet veel meer."

Informatiekloof

KPN heeft veel geïnvesteerd in Informatica de afgelopen jaren. Niet alleen in licenties, maar ook in kennis van de producten en applicaties die zijn gebouwd. "Je stapt niet zomaar over van de ene naar de andere ETL-tool", is Reichels boodschap. "Want dan zou je helemaal opnieuw moeten beginnen en de componenten die je ooit hebt gebouwd, kun je dan niet meer gebruiken." Vorig jaar is KPN het Artemis-project begonnen (zie kader) omdat het niet meer lukte de gevraagde data binnen de in SLA's afgesproken tijd te integreren en de gewenste informatie te

leveren. Bij ETL is de I/O één van de belangrijkste aspecten, dat moet snel gebeuren. Daarom is overgestapt naar de 64-bit versie van Powercenter 7 op HP Integrity servers. "We willen de

In de prijzen

KPN Mobiel is het afgelopen jaar met zijn datawarehouse flink in de prijzen gevallen. In maart ontving het de Business Intelligence Award 2006. "Beweeglijkheid en klantoriëntatie zijn voor de jury doorslaggevend geweest om aan KPN Mobiel de eerste prijs toe te kennen in de categorie profit. Deze organisatie is al jarenlang in veranderingsprocessen gewikkeld en heeft na het overwinnen van de nodige moeilijkheden kans gezien om een sterke oriëntatie op de klant te realiseren", aldus juryvoorzitter Leo Kerklaan, vennoot van de Holland Consulting Group.

Dit jaar ook kreeg de mobiele telco de Innovation Award 2006 van Informatica. "KPN Mobiel heeft de klanttevredenheid en winstgevendheid vergroot via een doordacht programma voor de levenscyclus van klanten, dat een verbetering oplevert voor het zicht op de klant, klantensegmentatie en -analyse.

Pure data-integratie

Wat Informatica zo aantrekkelijk maakt, is volgens Reichel de gerichte aandacht voor data-integratie. "Ze hebben een paar jaar geleden wel een uitstapje gemaakt naar het analyseren van gegevens met PowerAnalyser, maar daarmee begaven ze zich op het terrein van bijvoorbeeld Cognos en Business Objects en ze hebben ingezien dat ze daar niet thuis horen. Data-integratie is hun core business, waarbij kwaliteit van data voorop staat", aldus Reichel. Wat dat betreft, is de wereld volgens hem een goed speelveld. "Je ziet dat overal consolidatieprojecten plaatsvinden. Dat betekent dat er veel data moeten worden gemigreerd. Daar kun je een ETL-tool voor gebruiken." Overigens gaat dat niet altijd van een leien dakje. Zo kreeg KPN Powercenter 7 maar niet geïnstalleerd op de Integrity Servers. Dit had te maken met problemen rond een Unix patch, waarvan de oplossing uiteindelijk op een Japanse site is gevonden.

informatiekloof overbruggen. De hoeveelheid informatie neemt alleen maar toe, terwijl de tijd om die te verwerken beperkt blijft." "Op basis van de toegenomen snelheid hebben wij ook de KPI's kunnen vaststellen", vertelt Konings. De afspraken met Getronics PinkRocade zijn in SLA's vastgelegd, waaraan een bonus/malus-regeling is gekoppeld. De dienstverlener is gehouden accurate informatie binnen de gestelde tijd te leveren. Hoe weet je of de geleverde kwaliteit van de data goed is? "Informatica heeft zelf een aantal tools ingebakken dat de kwaliteit controleert", antwoordt Reichel. "De processen leveren informatie op die uiteindelijk weer door een tool wordt gebruikt om de kwaliteit te bepalen om er analyses op te kunnen plegen." "Maar ook daar zijn we nog in ontwikkeling", vult Konings aan. "We zijn onlangs een project gestart om het IWH opnieuw in te delen. We gaan er een modulaire structuur van maken. We hebben nu nog te maken met de bekende 'spaghetti-structuur'. Dat is organisatorisch zo gegroeid. Dat heeft te maken met de snelheid waarmee je soms producten moet opleveren voor de business. Dan ga je wel eens wat te kort door de bocht. Waar we nu tegenaan lopen is dat de voorspelbaarheid van de ontwikkelingen, zeg maar de doorlooptijd, niet aan de verwachtingen voldoet. Daarom hebben we ervoor gekozen om een modulaire structuur in te brengen, waarbij je de halffabrikaten uit de back room kunt hergebruiken. Dat biedt tijdswinst. Daarnaast zijn we aan het kijken – nu we de doos toch open hebben liggen – of we de kwaliteit kunnen verbeteren. Daarbij kijken we naar Powercenter 8; voornamelijk omdat de performance daarmee sterk verbetert. En je wilt gewoon ook technisch up to date zijn."

Generieke opzet

We hebben een paar aannames gemaakt voor de aanpassing van het IWH. We houden de bronnen gelijk, evenals de eindproducten die wij opleveren. Het voorkomt dat je al te veel impact op de business hebt. Natuurlijk gaan we wel dingen veranderen aan

de definities, dus daar richten we een klankbordgroep voor in", vertelt Konings.

De kern van het information warehouse wordt herontworpen. Er moet dus worden nagedacht over het inrichten van generieke datamodellen in het datawarehouse. "We willen voorbereid zijn op de integratie van vaste en mobiele telefonie. Ons information warehouse is nu gebouwd rond de klanten van mobiele telefonie, maar diezelfde klant kan ook een vaste aansluiting hebben. Dat betekent dat we die gegevens ook moeten opnemen in ons datawarehouse. Op het moment dat ons model generiek is opgezet, dus onafhankelijk van de bronnen, dan kunnen we ook aan de wens van de business voldoen die gewoon analyses wil uitvoeren op alle gegevens van een klant. Dat betekent wel dat we de huidige structuur moeten herontwerpen en een nieuw datawarehouse moeten gaan opbouwen."

Kun je dat doen terwijl de winkel doordraait? "Ja", klinkt het volmondig. "We houden het oude datawarehouse nog open naast het nieuwe. We ontwikkelen ook nog in het oude datawarehouse, hoewel we natuurlijk heel slim gaan kijken of een bepaalde wens meteen moet worden vervuld of dat het wel even kan wachten. Ook gaan we na of we al gebruik kunnen maken van het nieuwe datawarehouse bij een bepaalde ontwikkeling."

"We werken wel incrementeel, waarbij we steeds met deelopleveringen werken waar je meteen gebruik van kunt maken. Dat is vooral om het beheersbaar te houden, maar ook om er sneller de vruchten van te kunnen plukken. Als je alles in één keer wilt opleveren en eerst een jaar onder water gaat, dan is het nog maar de vraag waar je weer boven water komt. Onlangs is de investeringsbeslissing genomen om dit project te starten."

Toekomstvast

Wat de overgang naar een opnieuw gestructureerd datawarehouse vergemakkelijkt, is dat er een actuele replica bestaat. Er is dus altijd een achterwacht.

"Het speelt nu nog niet", zegt Reichel, "maar Powercenter 8 is helemaal gebaseerd op een Service Oriented Architecture. Daar is KPN nu volop mee bezig, zodat alles wat nu wordt ondernomen toekomstvast is. In de toekomst zullen daar namelijk ook koppelingen gaan ontstaan. Het is nu nog niet in de plannen meegenomen, maar wellicht in de nabije toekomst."

Reichel wil nog wel kwijt dat de overgang naar Powercenter 7 en het nieuwe hardware-platform niet alleen een performance-kwestie is geweest, maar evenzeer een geldkwestie. "De beheerkosten zijn gehalveerd", zegt hij. Alles is twee keer zo snel geworden. Voor de toekomst ziet Reichel een situatie ontstaan dat de informatie niet eerst naar een datawarehouse gaat, alvorens ze beschikbaar is voor analyses, maar dat de gegevens beschikbaar moeten worden gesteld op het moment dat ze ontstaan. Dat is de enige mogelijkheid om de honger naar snel beschikbare en accurate gegevens te stillen. Data-integratie is, zo blijkt, een zich steeds verder ontwikkelend proces.

Teus Molenaar is freelance journalist.