

Ab Initio: een gebruiker vertelt!

De tip van de sluier verder opgelicht

Dick Schievels

In februari van dit jaar (DB/M 1, 2011) schreef mijn voorganger Hans Lamboo onder de titel 'Tipje van de sluier' een enthousiast artikel over Ab Initio (latijn voor 'vanaf het begin'), een leverancier van een 'groot strategisch datamanagementplatform in stilte', aldus de chapeau boven het artikel. Nu, in dit laatste nummer van het jaar, een gesprek met Michael Hansen van Vodafone, een gebruiker!

Een prikkelende column in dit blad over de wolk van mystiek die over Ab Initio hangt, leidde bijna een jaar geleden tot een door FourPoints (de Nederlandse vertegenwoordiger van Ab Initio) georganiseerde demo voor ex-hoofdredacteur Hans Lamboo en onze columnist Rick van der Lans, die in de woorden van mijn voorganger als 'indrukwekkend' werd ervaren.

Die wolk van mystiek rond Ab Initio is terug te voeren op een onorthodoxe marketingstrategie die, en ik citeer nogmaals Hans Lamboo, 'niet gebaseerd is op advertenties, folders & brochures, whitepapers, sponsorships en magic quadranten, maar puur op kwaliteit van het product, op POC's en recommandatie'. Tot die onorthodoxe marketingstrategie behoort ook, zo viel te lezen, dat klanten gevraagd wordt niets over Ab Initio naar buiten te brengen, tenzij er sprake is van een recommandatie. Lamboo kon begin dit jaar daarom nog geen enkele naam van een Ab Initio-klant noemen, maar ging als rechtgeaarde vakjournalist toch op jacht naar een interessante klantencase. Mede door het masserende werk van FourPoints richting Ab Initio lukte het hem vlak voor zijn afscheid in september die case binnen te halen.

Het gaat om Vodafone Nederland waar nu sinds een jaar of acht, op voorspraak van Vodafone Engeland, met Ab Initio wordt gewerkt. Dat althans vertelt ons Michael Hansen, tegenwoordig manager Business Intelligence Competence Center bij Vodafone Nederland en destijds betrokken bij de aanschaf van het betreffende datamanagementplatform.

Michael Hansen is een ervaren BI-specialist. Voordat hij in 2001 in dienst trad bij Libertel – een jaar later ingelijfd door Vodafone – werkte hij acht jaar bij bank/verzekeraar KBC in België. "Bij KBC heb ik mij vanaf het begin beziggehouden met BI-activiteiten, hoewel men dat in die tijd meestal nog 'decision support' noemde. In ieder geval heb ik altijd ondersteunend

gewerkt voor mensen die beslissingen moeten nemen, en dat gefundeerd willen doen op basis van data-informatie. Ik heb daar zowel een designer- als een analistenrol gehad."

Eenmaal bij Vodafone ging Hansen als BI-projectmanager aan de slag. Vervolgens groeide hij door naar een architectenrol en tegenwoordig is hij verantwoordelijk voor strategie, architectuur en data-provisioning. "We hebben ons datawarehouse opgedeeld in twee grote blokken: informatie-exploitatie, alles wat we produceren, en de data-provisioning, alles wat nodig is om tot informatie te komen. Daarboven hangt het strategiestuk en de architectuur: hoe zijn we ingeregeld, wat voor tooling en wat voor methodieken gebruiken we."

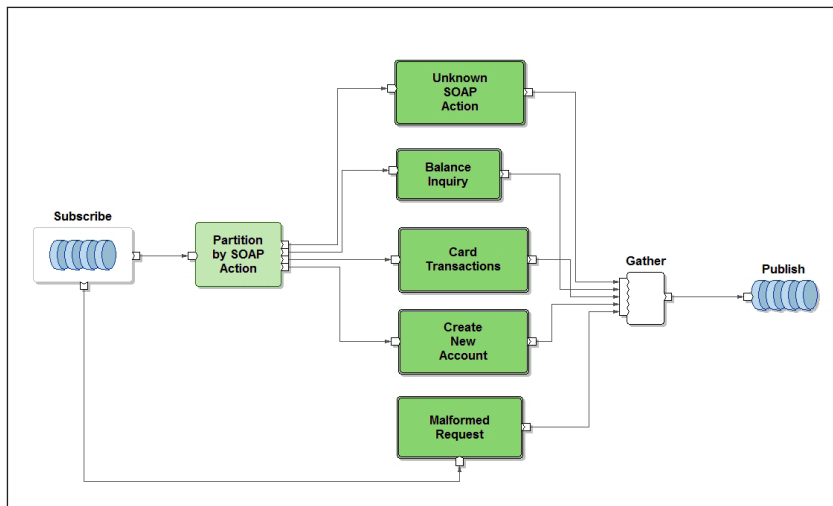
Recent heeft Vodafone Nederland een BI Competence Center (BICC) in het leven geroepen waar rond de 45 mensen werkzaam zijn. Michael Hansen: "In juni zijn we daarmee gestart. Daarbinnen hebben we de end-to-end verantwoordelijkheid voor BI. Persoonlijk ben ik als een van de BICC-managers met name verantwoordelijk voor de strategie: waar gaan we naar toe qua tooling, maar ook conceptueel, waar willen we heen met onze informatiehuishouding."

Hoe zijn jullie met Ab Initio in contact gekomen?

"Via onze Engelse collega's, want we kenden ze zelf niet. Het was een beetje een sprong in het diepe, maar vanaf dag één ben ik tevreden over hun accountmanagement, over hun technische mensen en over het product. Ik weet dat uit eigen ondervinding, want ik heb er in het begin zelf ook nog mee gewerkt. Het is een intelligent product, maar ook weer niet zo gemakkelijk dat je er direct vernuftige dingen mee kunt doen. Het heeft een steile leercurve. Je moet er dus wat tijd in willen investeren. Het vereist een enigszins andere manier van denken over hoe je met data omgaat."

Kunt u iets nader uitleggen wat dat precies inhoudt?

"Wat je je altijd moet realiseren als je met Ab Initio werkt, is dat je telkens met één record werkt, met één data-elementje uit een groot bestand. Daarbij moet je er eerst goed over nadenken, wat je met dat bestandje wilt doen. Achteraf als je het klaar hebt, hang je er dan misschien duizenden of miljoenen records aan. Ab Initio gaat heel snel hele kleine elementjes lezen en daar al iets mee doen. Doordat ze hele kleine elementjes nemen die



Afbeelding 1: Real-time dataprocessing met Ab Initio Continuous>Flows.

heel weinig processorcracht nodig hebben voor de verwerking, kunnen ze ook als het om duizend of een miljoen van die elementjes gaat, die nog heel snel verwerken."

Dat is heel wat anders dan dat je bijvoorbeeld een heel datablok in geheugen hebt?

"Ja. Want dan kan je bijvoorbeeld kijken naar record 1 en record 2, een vergelijking doen en iets wegschrijven. Als je als developer maar één record in geheugen hebt gelezen, hoe moet je dan vergelijken met het eerste of het laatste? Dat soort dingen kunnen wel in Ab Initio, maar je moet even onder de knie krijgen – met look-up tables of anders – hoe je dat dan doet. Dat vereist een verandering van mindset. Je moet leren op een andere manier te werken."

Waar gebruiken jullie Ab Initio voor?

"Wij gebruiken het om alle data die we ter beschikking hebben en nuttig vinden om in ons datawarehouse te plaatsen, van bron- naar doelbestand te converteren. Dus de klassieke ETL: Extract, Transform en Load. We extraheren data uit databases of uit bestanden die we aangereikt krijgen. We transformeren die, dat wil zeggen: we converteren codes, combineren databronnen met elkaar of zuiveren en verrijken de data. En uiteindelijk laden we de getransformeerde data in een datamodel, de Load."

Wat moet ik me bij die bronbestanden voorstellen?

"Onze belangrijkste bronbestanden bestaan uit de zogeheten CDR's, Call Detail Records. Van elk telefoontje wordt een record gegenereerd. Daarin worden dingen opgeslagen als: persoon A heeft gebeld met persoon B op tijdstip C. Een andere bron is winkelinformatie: data over wat mensen kopen, online of fysiek via onze Vodafone-winkels. Ons billingsysteem, dus het systeem dat de facturen genereert, is een derde belangrijke databron. Op dit moment hebben we 22 bronsystemen die informatie aanleveren over onze klanten en wat die doen, of over wat Vodafone zelf doet. Die data worden geïntegreerd in ons datawarehouse,

waarna we er, uiteraard met inachtneming van de bestaande privacyregels, analyses op doen en rapportages over genereren. We zijn beursgenoteerd, dus we zijn verplicht bepaalde zaken te rapporteren. Verder willen we uit de vele systemen die we hebben proberen één beeld te vormen van de klant, van de Vodafone-business als geheel, of van een bepaalde business unit waar we in werken, met als doel al die zaken goed aan te kunnen sturen."

Jullie gebruiken Ab Initio alleen voor het ETL-stuk, de data-provisieering, en niet voor de productiekant, de data-exploitatie?

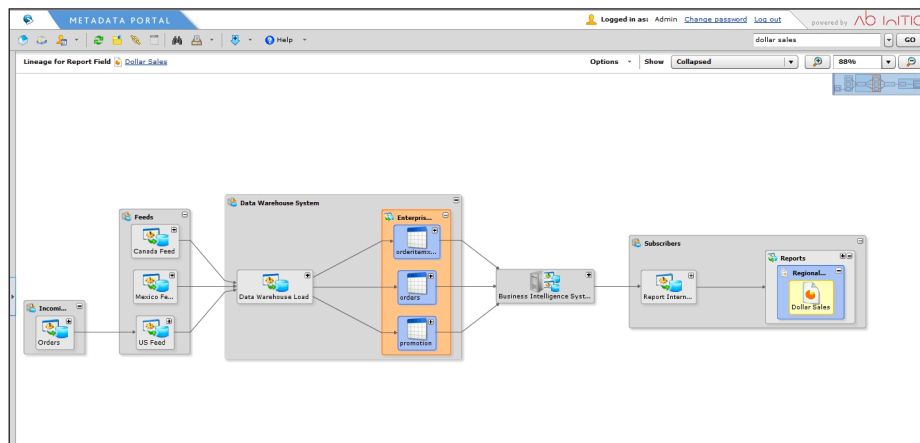
"Nee, nog niet. Er loopt nu wel een groot project, een nieuw billinginitiatief, om te onderzoeken wat de mogelijkheden zijn van bijvoorbeeld wat ze bij Ab Initio noemen 'indexed compressed flat files' voor bepaalde toepassingen. Maar dat hebben wij op dit moment nog niet geïmplementeerd."

Want het idee is dan dat je soortgelijke dingen kan doen als met een DBMS?

"Ja, klopt! Ik denk dat er toepassingen voor te vinden zijn. We kijken bijvoorbeeld naar een toepassing voor informatie die we moeten kunnen leveren aan instanties. Stel, bijvoorbeeld dat we van één bepaalde ingang, een 06-nummer of zo, informatie moeten kunnen leveren aan de politie. Je zou dan kunnen overwegen zo'n bestand 'index compressed' op te zetten, met de index op het 06-nummer, zodat je daar heel snel bij kunt. Afhankelijk van de toepassing zie ik dus mogelijkheden voor het gebruik van Ab Initio's manier van dataopslag om ook de exploitatiekant te ondersteunen. Maar dat doen we op dit moment nog niet."

Ab Initio beschikt ook over functionaliteit om gegevens in real-time te kunnen verwerken. Maken jullie daar gebruik van of is alles nog batch-georiënteerd?

"Dat is nu nog altijd batch-georiënteerd. Maar ook daarmee zijn we in het nieuwe billingproject bezig. Daarin hebben we dat



Afbeelding 2: Ab Initio's metadata managementoplossing toont end-to-end data-lineage.

geïntroduceerd om ook tussentijds, intra-batch, informatie te kunnen verkrijgen, of, indien de bronsystemen de data kunnen aanleveren, data continu te kunnen lezen. Ab Initio heeft daarvoor een add-on in de vorm van Continuous>Flows."

Wat voor belangrijke voordelen biedt Ab Initio naar uw mening?

"Een van hun key-voordelen die ik meerdere keren zelf heb ervaren, is dat Ab Initio in de praktijk 100 procent schaalbaar blijkt. Stel je moet iets sneller kunnen doen, of het datavolume neemt toe, maar de window om het te verwerken blijft hetzelfde. Eén optie is dan je hardware uitbreiden. Ab Initio blijkt dan effectief lineair schaalbaar. Bij andere oplossingen die dat ook claimen, is er op zijn minst altijd een beetje verlies. Je doet een dubbele hardware investering en je krijgt in plaats van 100 procent performancewinst 98 of 95.

Als je bij Ab Initio de CPU's verdubbelt, is je verwerkingstijd echt de helft. Dat heb ik effectief al twee of drie keer gezien toen we de hardware moesten upgraden en vooraf en achteraf gingen meten."

Dat betekent dus dat Ab Initio de hardware heel nauwkeurig weet aan te sturen.

"Het grootste geheim van Ab Initio ligt ook precies in dat aspect: hoe gebruik je met je software zo efficiënt mogelijk de hardware-componenten die je tot je beschikking hebt. Ze kunnen daarbij draaien op een mix van platformen."

Daar zit MPP onder, Massive Parallel Processing?

"Ja, dat zit gewoon ingebakken in hun software. Daarbij wordt echter heel slim gekeken naar welke zaken wel en welke zaken geen baat hebben bij MPP. Dus ze weten precies wat je beter wel en niet parallel kunt verwerken."

Hoeveel mensen werken met Ab Initio bij Vodafone?

"Ik weet het niet precies. We hebben al het ontwikkelwerk inmiddels outsourced naar HP. Ik schat dat er daar misschien zo'n 25 man mee werken. Intern hebben we nog twee à drie mensen met Ab Initio-kennis."

Kunt u eens een wat uitgebreider beeld schetsen van een traject waarin Ab Initio gebruikt wordt?

"Stel de marketingafdeling wil op bepaald moment data uit sociale netwerken kunnen analyseren. Dan luidt vraag één: van waar betrekken we die data en in welke datamodelstructuur ga ik ze opslaan in mijn datawarehouse. Als dat is bepaald, komt vervolgens de Ab Initio-developer in beeld. Na een eerste analyse van de brondata en een blik op het model waar hij naar toe moet schrijven, gaat deze in gesprek met de businessanalist over het design voor de transformatiefase; wat voor formules moeten er precies inzitten, et cetera. Als hij dat weet, heeft hij eigenlijk al zijn ingrediënten om aan de slag te gaan. Vervolgens komt de ETL.

Voor het E- en het L-stuk, voor de extractie en de load, hebben we bij Vodafone op basis van Ab Initio generieke componenten gemaakt."

De transformatie, dáár gaat het om?

"Het is inderdaad het middenstuk, de transformatie, waar het om draait. Dat is echt het stuk waar je je bedrijfsregels inzet. Daar zeg je: van bron A, B, C ga ik omvormen naar tabel A, B, C, D. Dat is het typische gebruik van Ab Initio bij ons: het ontsluiten van een nieuwe bron in het datawarehouse. Waarbij je kijkt, welke componenten zijn al generiek en welke zijn specifiek. Naar die specifieke componenten, dat T-stuk, zijn we nu overigens ook aan het kijken hoe we dat met behulp van Ab Initio's metadata programming nog generieker kunnen maken. Om ook daar zo weinig mogelijk wijzigingen te hoeven aanbrengen mocht het bron- of doelsysteem veranderen. Het is volgens mij een utopie dat dat ooit volledig intuïtief wordt, maar dat zo dicht mogelijk benaderen, zodat je geen of weinig onderhoud hebt, dat is een nobel streven."

Kunt u de mogelijkheden van dat metadatagedreven programmeren nog iets nader toelichten?

"Heel vroeger ging je van Assembler programmeren naar ontwikkelen in C, Cobol en de 2-, 3- en 4-GL's. Als je Ab Initio dan ziet als hogere programmeertaal, dan is de volgende stap dat je

met nog minder programmeurs, nog minder codegenererende maatregelen je code kan blijven runnen, ook al verandert de wereld daaromheen.

Stel, je hebt bijvoorbeeld een bepaalde databron in Ab Initio gemaakt en vervolgens ingelezen en opgeslagen in een data-model. En stel je datamodel verandert, of je bron verandert. Hoe kan je er dan voor zorgen dat je met zo weinig mogelijk aanpassingen aan de betrokken Ab Initio-componenten toch hetzelfde werk kan blijven doen. Dat bereik je door middel van metadata programming. Ze zijn daarin ver gevorderd. Ze hebben door middel van stuurtabellen, zoals ik het maar even noem, een flexibiliteit in hun product weten te bouwen, die kan omgaan met wisselende zaken zowel aan de voorkant als aan de achterkant. Dat heeft even geduurd, maar ze hebben dat nu live en we zijn dat product nu ook aan het implementeren. Dat vereist overigens opnieuw een verandering van mindset voor de developers die langzamerhand gewend zijn om 'gewoon' in Ab Initio te werken, en nu met de metadata-driven ontwikkeling conceptueel weer totaal anders moeten denken. Want je ontwikkelt niet meer blokje voor blokje, maar moet steeds bedenken bij de ontwikkeling van die blokjes wat je moet inbouwen voor als er straks op een bepaald punt een verandering optreedt en je alleen op basis van het aanpassen van parameters je code wilt kunnen blijven draaien."

Is het ontwikkelgemak de belangrijkste factor geweest om Ab Initio te blijven gebruiken, of is het vooral de snelheidswinst die telt?

"De developers waarmee ik gewerkt heb zijn vooral enthousiast over het feit dat ze snel tot een resultaat komen. Dat is aantrekkelijk voor een developer. Je hoeft niet, zoals vroeger bij Cobol, eerst heel veel te programmeren, te compileren en te debuggen. In Ab Initio neem je een klein testbestandje, zet wat componenten bij elkaar en je gaat letterlijk lijntjes trekken tussen de componenten. Vervolgens druk je op start en je ziet de records quasi door die flow lopen."

Dus je bent op een hoger niveau bezig als een soort businessanalist en de code wordt als het ware onder water gegenereerd?

"Het is iets complexer, maar het ziet er wel zo uit. Als je een developer van een afstandje bezig ziet, dan zie je iemand die vooral lijntjes tussen blokjes trekt. Maar je moet heel goed weten wat elke component doet en daar de juiste kenmerken aanhangen. Daarvoor moet je developer zijn. Je moet de vertaalslag kunnen maken van de requirements enerzijds naar de technische realisatie anderzijds."

Nu werken nog alleen developers met die metadata portal en geen businessanalisten?

"Klopt!"

Het liefst zou je natuurlijk de businessmensen zelf rechtstreeks achter de knoppen willen zetten.

"Ja, maar dan moet je dus als businessanalist snappen wat je aan

het veranderen bent. Als je een parameter wijzigt, wat doet dat dan onderliggend? En een businessanalist wil niet teveel in de details verzeild raken. Ik denk wel dat er misschien een nieuw soort functie in bedrijven gaat ontstaan, die tussen de businessanalist en de developer in ligt. Mensen die zowel de noodzakelijke technische als voldoende businesskennis in huis hebben om dit soort werk te kunnen doen."

Zou jullie onlangs opgerichte BI Competence Center niet bij uitstek de plek zijn om dit soort mensen onder te brengen?

"Dat is ook precies het streven. Het BICC als kweekvijver van een nieuw ras medewerkers in het gebied tussen business en techniek."

Want het BICC functioneert met name ook als intermediair tussen de ontwikkelaars bij HP enerzijds en de business anderzijds?

"Exact! Als de business bepaalde informatie verlangt, dan komen ze eerst bij mijn collega. Die bekijkt of de daarvoor benodigde data beschikbaar zijn. Zo ja, dan doet hij een analyse, draait hij een rapport uit of bouwt hij een dashboard. Heeft hij die data niet, dan komt hij bij mij en kijken we, samen met HP, uit welke bron we die data zouden kunnen halen en wat we daarvoor door middel van Ab Initio moeten ontwikkelen. Vervolgens zetten we die data in het datawarehouse, waarna mijn collega het weer verder kan afhandelen. Dat is wat er in ons BICC gebeurt."

Tot slot, waarom is Ab Initio volgens u minder bekend dan andere leveranciers van datamanagementoplossingen?

"Een van de redenen daarvoor is volgens mij dat op scholen en universiteiten Ab Initio als tool niet wordt aangeboden. Het is ook niet te downloaden van het internet. En als je gaat zoeken op het web vind je een website met één of twee pagina's. Er wordt veel geheim gehouden rond het bedrijf en de precieze reden daarvoor is enigszins in nevelen gehuld."

Heb je dan wel goed zicht op de plannen van het bedrijf?

"Nee, niet echt. Dat is ook iets dat in hun nadeel kan gaan werken. Kijk, dat er wat geheimhouding is rond de interne zaken van het bedrijf, maakt mij niet uit. Dat vind je waarschijnlijk bij elk bedrijf dat in private handen is. Als niet-beursgenoteerd bedrijf hoeft je tenslotte niet je cijfers te publiceren. Maar waar ik wel zeer geïnteresseerd in ben, is in hun roadmap: waar gaan ze naar toe? De manier waarop ETL nu op de wereld is gezet, hoe mensen daarmee werken en wat je ermee kan, dat evolueert. Maar waar gaat Ab Initio straks naar toe wat betreft dingen als datavirtualisatie? Je moet het altijd maar afwachten. Hebben ze daar straks tooling in? Ik heb gehoord van wel. Maar wat doen ze wel en niet in de cloud? Al die nieuwe hypes die opkomen, gaan ze daarin mee of niet? Dus als klant weet ik vaak niet voldoende waar Ab Initio naar toe gaat en wanneer ze met iets komen. Maar ik weet wel, als ze met iets komen dan werkt het ook!"

Dick Schievels is hoofdredacteur van Database Magazine.