



Business Intelligence bij Ahold Nederland:
Pallas, het Albert Heijn datawarehouse

Architectuurbeschrijving

Lidwine van As
Wouter van Aerle

Albert Heijn
oktober 2004

Inhoudsopgave

Inhoudsopgave.....	2
Inleiding.....	3
Werkwijze en documentindeling.....	3
Verantwoording van keuzes.....	3
Referentie.....	3
Over de auteurs.....	3
Rationale.....	4
Probleemstelling.....	4
Oplossing.....	4
Pallas in vogelvlucht.....	5
Functionaliteit.....	6
Uitgangspunt.....	6
Procesoriëntatie.....	6
Datakwaliteit.....	7
Rapportfunctionaliteit.....	8
Overige functionaliteit.....	8
Techniek.....	10
Infrastructuur.....	10
Architectuur.....	10
Overall architectuur.....	10
Transaction Repository.....	11
Datamarts.....	12
Data staging areas.....	12
Processing.....	13
Delta's.....	13
Heraggregatie.....	14
Interfaces en interactie met andere systemen.....	14
Aanpak.....	15
Organisatie.....	15
Bestaande services.....	15
Nieuwe services.....	15
Methoden, technieken en documentatie.....	16
Kosten en baten.....	17
Algemeen.....	17
Kosten.....	17
Baten.....	17
Toekomstperspectief.....	18

Inleiding

Deze architectuurbeschrijving vormt de inzending van het Albert Heijn Competence Center Business Intelligence voor het Nederlands Kampioenschap Architectuur 2004. Voor de beschrijving is de IEEE recommended practice voor architectuurbeschrijvingen, IEEE1471, als richtsnoer gebruikt.

Werkwijze en documentindeling

In lijn met IEEE1471 zijn eerst de stakeholders van het datawarehouse en hun concerns geïnventariseerd. De concerns zijn vervolgens vertaald naar en samengevat in 4 invalshoeken:

<i>Stakeholders</i>	<i>Concerns</i>	<i>Invalshoeken</i>
Gebruiker	- Welke processen worden ondersteund? - Welke functionaliteit wordt geboden? - Wat is de kwaliteit van de data?	Functionaliteit
Eigenaar	- Wat kost het datawarehouse? - Wat levert het op? - Wat is het toekomstperspectief?	Kosten & baten
Interne Accountantsdienst	- Wat is de kwaliteit van de data? - Hoe worden security en privacy gewaarborgd? - Hoe wordt traceability gewaarborgd?	Functionaliteit
IT-management	- Hoe beheersen en reduceren we de kosten?	Aanpak
DWH technisch team (ontwikkelteam, dwh-architect, beheerteam)	- Hoe zit het datawarehouse technisch in elkaar? - Hoe wordt de functionele vraag vertaald naar een technische oplossing?	Techniek, Aanpak
IT-Integratieteam	- Hoe past het datawarehouse in de AH IT-omgeving?	Techniek

In het hoofdstuk Rationale wordt de bestaansreden van het systeem toegelicht, waarna een globale systeembeschrijving gegeven wordt in het hoofdstuk Pallas in vogelvlucht. Vervolgens wordt deze in de hoofdstukken Functionaliteit, Techniek, Aanpak, en Kosten en baten vanuit de verschillende invalshoeken verder uitgediept.

Verantwoording van keuzes

In aanmerking genomen dat het in 2000 opgestelde oorspronkelijke architectuurconcept voor Pallas 73 pagina's besloeg, is het onderhavige document noodzakelijkerwijs beknopt, en kunnen niet alle onderwerpen even diepgaand besproken worden. Uit het aanbod aan informatie moet dan ook een selectie gemaakt worden.

Het belangrijkste uitgangspunt dat daarbij gehanteerd wordt, is dat *het unieke verkozen wordt boven het triviale*. Zo wordt weinig aandacht besteed aan de privacy- en beveiligingsaspecten van het datawarehouse, aangezien op dat gebied geen maatregelen getroffen zijn die wezenlijk afwijken van wat men voor een willekeurig systeem mag verwachten.

Referentie

- [1] "The Data Warehouse Life Cycle Toolkit", Ralph Kimball et al., ISBN 0-471-25547-5, Wiley, 1998
- [2] Kimball Design Tip #59: "The Surprising Value of Data Profiling", september 2004
- [3] "Corporate Information Factory, 2nd edition", W.H. Inmon et al, ISBN 0-471-39961-2, Wiley, 2000

Over de auteurs

Lidwine van As (las@grey-matter.nl) is zelfstandig IT-consultant op het gebied van software-ontwikkeling en business intelligence. Zij is vrijwel vanaf het begin als datawarehouse-architect bij het Pallas-project betrokken geweest.

Wouter van Aerle (wouter-van.aerle@ahold.com) kwam in 2001 in dienst van Albert Heijn bij het CC-BI. Als informatieanalist was hij betrokken bij de uitbreiding van het datawarehouse met bonuskaartgegevens en de realisatie van een aparte datamart voor bonuskaartanalyse. Daarnaast participeerde hij in vele andere nieuwe ontwikkelingen. Momenteel ligt zijn focus met name op het verder professionaliseren van het vakgebied informatieanalyse binnen het CC-BI.

Rationale

Probleemstelling

In 1999 werd vastgesteld dat de kwaliteit van de toenmalige management-informatievoorziening van Albert Heijn als sterk onvoldoende ervaren werd. De situatie werd gekenmerkt door slechte beheersbaarheid, onvolledigheid, elkaar tegensprekende cijfers, late beschikbaarheid en het niet-geïntegreerd benaderbaar zijn van uiteenlopende bronnen.

Los daarvan was verbetering gewenst in het licht van nieuwe ontwikkelingen in de business, zoals verdergaande differentiatie. Dit leidde tot een toenemende behoefte aan gegevens op een groter detailniveau (filiaal- en kassabonnniveau) dan destijds mogelijk was. De beschikbare management-informatiesystemen konden hier vanwege hun gesloten karakter niet in voorzien, of waren niet ingericht op het te verwachten datavolume.

Daarom ging in 1999 project Pallas van start met de volgende doelstelling:

“Creëer een Albert Heijn brede oplossing voor de huidige problematiek t.a.v. de informatievoorziening binnen de organisatie, breng deze op een hoger plan en leg het fundament voor verdere uitbreiding”

Een afgeleide doelstelling was het toevoegen van waarde door het afdekken en ondersteunen van de complete value chain en het faciliteren van integraal gebruik van de beschikbare informatie. Daarnaast moest de nieuwe oplossing een einde maken aan de Babylonische spraakverwarring die het gevolg was van het gebruik van meerdere, niet-geïntegreerde informatiebronnen naast elkaar. Door de inrichting van een centrale bron van historische gegevens, kon een gemeenschappelijk referentiekader ten aanzien van informatie en businessdefinities worden gewaarborgd: “one copy of the truth”.

Tenslotte streefde men ernaar, kostenvoordelen te behalen door hergebruik van de omgeving voor andere Ahold-werkmaatschappijen.

Oplossing

Om deze doelstellingen te realiseren, werd een algemene informatiebron ter ondersteuning van de besluitvormingsprocessen in de AH-organisatie ontwikkeld, waarin gegevens uit alle bedrijfsprocessen bij elkaar gebracht werden: Pallas, het enterprise datawarehouse.

Om de kans van slagen van de nieuwe omgeving te vergroten, werden de volgende uitgangspunten en randvoorwaarden voor ontwikkeling van de omgeving geformuleerd:

- het datawarehouse is business-driven: de voornaamste drivers voor ontwikkeling en uitbreiding van het datawarehouse zijn de informatiebehoeften van de business
- zonder de theorie uit het oog te verliezen, wordt een pragmatische aanpak gehanteerd, o.a. door aansluiting te zoeken bij bestaande best practices op gebied van datawarehousing
- think big, act small: het einddoel is een enterprise datawarehouse, maar het is een enorm karwei om dat in één keer neer te zetten. Leg in plaats daarvan eerst het fundament (de architectuur), en bouw daarop kleine incrementen
- er wordt bij voorkeur proven technology toegepast

Pallas in vogelvlucht

De Pallas-architectuur is opgebouwd rondom een centraal datawarehouse, de Transaction Repository (TR). De TR wordt gevuld vanuit de verschillende bronsystemen binnen Albert Heijn. Vanuit de TR worden datamarts gegenereerd die bevroegd kunnen worden door de eindgebruikers, en die zijn toegesneden op de ondersteuning van specifieke businessvragen en –processen.

Pallas heeft een vertraging van een dag ten opzichte van de bronsystemen. De brongegevens worden elke nacht aangeleverd, geladen en verwerkt in de TR en de datamarts, zodat ze de volgende dag beschikbaar zijn voor rapportage en analyse. Wekelijks worden zo'n 600 miljoen rijen in het datawarehouse geladen. De aanleverende systemen gaan geleidelijk aan over op real-time data-aanlevering, waardoor de reactietijd van het datawarehouse steeds korter wordt: inmiddels zijn rapportages beschikbaar die de omzet van tien minuten geleden laten zien.

Nadat in 2000 de basisarchitectuur was neergezet, is Pallas incrementeel verder ontwikkeld en uitgebreid. Hoewel uitbreidingen in eerste instantie getriggerd werden door, en uitgevoerd werden ten behoeve van het oplossen van concrete business issues, werd altijd gewerkt vanuit de integratiegedachte: de toegevoegde informatie kon weer worden geïntegreerd met de reeds beschikbare informatie, om op die manier extra inzicht in de ondersteunde businessprocessen te verschaffen. De beschikbare gegevens bestrijken vrijwel alle elementen van de retail value chain, van DC-leveringen tot aan kassatransacties. De gegevens worden gebruikt ter ondersteuning van vrijwel alle businessprocessen in de onderneming. Door de snelle beschikbaarheid van de gegevens en de hoge mate van detail kunnen zowel management- als operationele rapportages uit het datawarehouse geleverd worden.

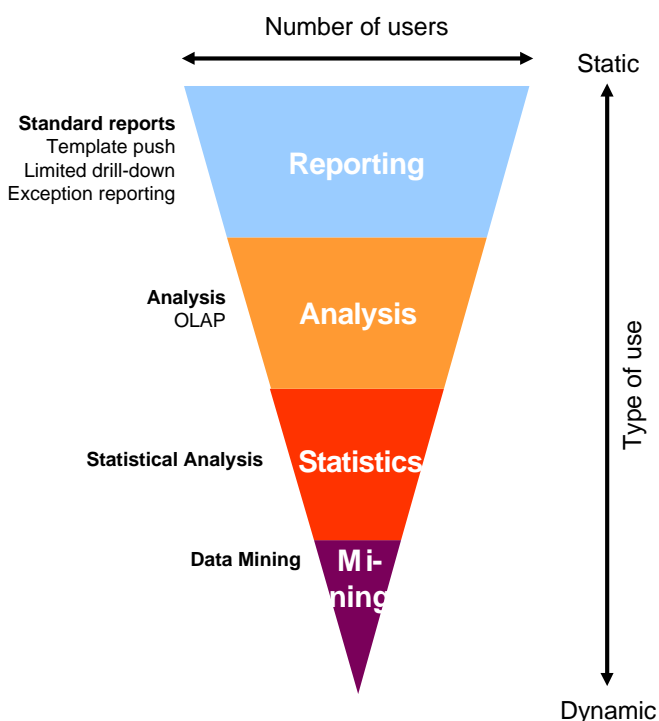
Sinds de start van het project is de Pallas-gebruikersgroep gestaag gegroeid. Wekelijks worden de rapportage- en analyse-omgevingen gebruikt door zo'n 1800 individuele gebruikers van hoog tot laag: van directieniveau tot aan de winkelvloer en het DC.

Functionaliteit

Uitgangspunt

Het doel van een business intelligence omgeving is eenvoudig: *verstrekking van de juiste informatie, op het juiste moment, aan de juiste persoon, op de juiste manier*. Daarmee is de basisfunctionaliteit van een BI-omgeving bepaald: het beschikbaar stellen van (management)informatie aan eindgebruikers. Functioneel gezien kan dit op verschillende manieren: van statische rapporten tot vergaande mogelijkheden om zelf in de data te zoeken. Technisch zijn er legio mogelijkheden om één en ander te realiseren.

De verschillende toepassingsmogelijkheden zijn geclassificeerd in een staalkaart:



Afbeelding 1 Staalkaart voor end-user functionaliteit

Alle end-user functionaliteit is volgens deze staalkaart geoperationaliseerd. Er is voor gekozen om ieder type functionaliteit (reporting, analyse e.d.) te realiseren met standaard front-end tools, die zowel out-of-the-box rapportagefunctionaliteit bieden als bouwstenen om zelf specifieke rapportagetypen (zoals scorecards) te bouwen. Door deze benadering werd a priori bepaald wat functioneel wel en niet mogelijk is: immers, de beschikbare functionele componenten en mogelijkheden van de standaardtools bepalen op welke wijze data gerapporteerd en beschikbaar gesteld kan worden. Daarnaast vormt de classificatie een goed communicatiemiddel bij de afstemming met de gebruikers over hun informatiebehoefte en de wijze van operationalisering daarvan.

Procesoriëntatie

Ieder type business proces (merchandising, fulfillment, replenishment etc.) kent zijn eigen unieke informatiebehoefte. Hierop is geanticipeerd door de architecturale keuzes die ten aanzien van de TR en datamarts zijn gemaakt. Relevante keuzes in dit kader zijn o.a. de toepassing van conforme dimensies, en het uitgangspunt van “één datamart per business proces”. (In sommige gevallen betekende dit principe dat precies dezelfde data in meerdere datamarts werd opgenomen. Zo zijn verkoopgegevens vereist voor het besturen van meerdere primaire processen: winkeloperatie, merchandising e.d.)

De gebruikte tooling ondersteunt het creëren van meerdere rapportageomgevingen die via één portal benaderbaar zijn. Hierdoor was het mogelijk om, als onderdeel van de gekozen tool, per ondersteund proces c.q. aandachtsgebied een eigen, dedicated rapportageomgeving te creëren, die kan beschikken over zijn eigen onderliggende data.

Onderstaande bus matrix [1] geeft een indruk van de primaire processen die door Pallas ontsloten en ondersteund worden, en de datamartdimensies die daaraan gerelateerd zijn:

	Tijd	Artikel	Filiaal	Verkoop- formule	Aktie	Distributie- center	Klant	Medewerker	Leverancier
Brand mgmt / CRM									
Format mgmt									
Merchandising									
Replenishment									
Fulfillment store									
Fulfillment logistics									

Afbeelding 2 Bus matrix: gerealiseerde sourcing en ondersteuning t.a.v. primaire processen

Over rapportageomgevingen heen zijn vorm- en navigatie-eigenschappen gestandaardiseerd. Hierdoor heeft ieder rapport en iedere omgeving dezelfde look-and-feel. Dit zorgt voor een uniforme en daarmee “rustige” presentatie naar de gebruiker, en creëert tevens herkenbaarheid ten aanzien van welke informatie wel en niet uit het datawarehouse afkomstig is.

Het beheer van datadefinities zorgt daarnaast voor uniformiteit van informatie over de omgevingen heen: zo is de benaming voor verkopen in geldwaarde in de ene omgeving “Omzet” maar in de andere omgeving “Klantomzet”, omdat de onderliggende definitie verschilt. De datadefinities zijn op een centrale plaats gedocumenteerd en toegankelijk voor de eindgebruiker, zodat deze de geboden informatie op de juiste wijze kan interpreteren. Informatie die qua definitie wel hetzelfde is, heet ook in iedere omgeving hetzelfde: zowel aan de back-end kant, op databaseniveau, als aan de front-end kant, op rapporten en kubussen.

Datakwaliteit

Een primaire driver voor de toegevoegde waarde en effectiviteit van een datawarehouse is de kwaliteit van de data die erin opgeslagen ligt¹. Het ontwerpen en ontwikkelen op datakwaliteit is van begin af aan een fundamenteel uitgangspunt geweest. Om dit te effectueren zijn de volgende maatregelen toegepast:

- Opstellen van KDD's² t.a.v. bronsystemen en brondata: aangezien een datawarehouse wordt gevoed vanuit bronsystemen, wordt de inhoudelijke kwaliteit in grote mate bepaald door de bronsystemen. De hiervoor vastgestelde KDD's behelzen het volgende:
 - Originele bron: data wordt alleen onttrokken aan de bron die de data creëert, en niet aan andere systemen die om efficiencyredenen ook over de betreffende data beschikken maar niet de bron ervan zijn.
 - Hoogst mogelijke detailniveau: gegevens worden op het hoogst mogelijke (beschikbare) detailniveau ontsloten en opgeslagen. Hierdoor worden de mogelijkheden van het datawarehouse niet onnodig ingeperkt door geaggregeerde data, terwijl in de toekomst misschien gedetailleerde data nodig zal zijn
 - Volledige bron: wanneer een bron(tabel) wordt ontsloten, wordt de hele bron(tabel) gesourced, en niet alleen de data waarvoor op dat moment requirements bestaan
- Bronnen worden verondersteld altijd vervuild te zijn. Derhalve wordt op iedere nieuwe bron een volledige technische en functionele bronanalyse uitgevoerd tot op attribuutniveau (data profiling, [2]). Hierdoor ontstaat inzicht in hoe een bron daadwerkelijk wordt gebruikt. Geconstateerde issues worden op één van de volgende manieren aangepakt:
 - De broneigenaar wordt verzocht het genoemde issue te verhelpen
 - In het ETL-proces wordt gecontroleerd op het voorkomen van het opgetreden issue. Ieder voorkomen wordt gelogd; zo kan bijvoorbeeld gerapporteerd worden over het aantal keren dat een

¹ Efficiency-aspecten worden eerder bepaald door de vorm van de gebruikte output. Gesteld zou kunnen worden dat zowel met een data-sheet dan wel een dashboard / exceptierapportage dezelfde beslissing of stuurmaatregel genomen zou worden, mits de onderliggende data hetzelfde is en van dezelfde kwaliteit is. Dit laatste wordt echter niet bepaald door de tools maar door de back-end van het datawarehouse.

² Key Design Decisions: fundamentele beslissingen t.a.v. functionaliteit en architectuur. Zie hoofdstuk Aanpak.

- specifiek issue is opgelost. Daarnaast vindt op het issue toegesneden follow-up³ plaats.
3. Doordat de TR de enige placeholder is voor alle data, vallen dubbelingen onmiddellijk op. Dezelfde soort informatie die in verschillende bronnen ontstaat wordt in de TR samengebracht, in één en dezelfde tabel indien van toepassing. Op dit punt bewijst de TR zijn toegevoegde waarde als integrator van data. Zo worden voorraadstanden van artikelen in distributiecentra en filialen apart opgeslagen in de operationele systemen; in de TR zijn mogelijkheden gecreëerd om deze data samen te voegen teneinde integraal te kunnen rapporteren over de totale voorraadpositie van een artikel.
 4. Referentiedata kent in iedere datamart dezelfde modellering en dezelfde inhoud. Dit principe van *conforme dimensies* [1] zorgt voor een uniforme bril over alle informatie heen. Bovendien bieden conforme dimensies de mogelijkheid om alle data die op hetzelfde subject (artikel, filiaal, klant e.d.) betrekking heeft, met elkaar te integreren.
Een uitdaging op dit punt is de omgang met informele indelingen van bijvoorbeeld artikelen of filialen. Er blijkt een veelheid aan alternatieve groeperingen te bestaan die niet worden ondersteund door een bronstelsel. Ten aanzien hiervan wordt een strikt beleid gevoerd: alleen wanneer voor een dergelijke indeling een formele definitie en een formele bron voorhanden is, wordt de data opgenomen in het datawarehouse.
 5. Wanneer eenmaal programmatuur is gebouwd om nieuwe data te ontsluiten vindt een *verificatie* plaats op de gegevensverwerking. Deze check controleert of alles wat wordt aangeleverd ook in het datawarehouse en succesvol de datamart terechtkomt.

Rapportfunctionaliteit

In overeenstemming met de staalkaart, vormt output in de vorm van rapportages op het datawarehouse (“Reporting”) de bulk van de functionaliteit die vanuit de Pallas-architectuur wordt geleverd. De mogelijkheden voor invulling van dit type functionaliteit lopen uiteen van werkbladen à la Excel met grote hoeveelheden detailgegevens, tot en met gepushte exceptierapportages, balanced scorecards en management dashboards met geaggregeerde informatie. Al deze mogelijkheden worden ondersteund door één en hetzelfde standaardtool. Om het aantal varianten qua rapporten beheersbaar te houden is een taxonomie opgesteld van rapporttypen; nieuw ontwikkelde rapporten moeten binnen deze taxonomie classificeerbaar zijn, anders worden ze niet ontwikkeld. Het is echter voornamelijk erg lastig om met deze standaardtypen in specifieke functionele requirements te voorzien. Niet zozeer om technologische, als wel om procesmatige redenen: standaardtypen rapportages vereisen tot op zekere hoogte gestandaardiseerde en gestructureerde processen – en processen zijn lang niet altijd gestandaardiseerd. Een positief effect is evenwel dat juist het gebruik van de BI-omgeving voor de business aanleiding is om dit onderwerp te adresseren.

Overige functionaliteit

Het tweede niveau in de staalkaart is Analyse. Binnen Pallas wordt hieronder verstaan: het ad-hoc stellen van een reeks vragen aan de database, waarbij iedere volgende vraag bepaald wordt door de antwoorden op de voorafgaande vragen. Dit vereist een uitstekende performance van de omgeving; de betreffende functionaliteit wordt binnen Pallas geoperationaliseerd met behulp van (M)OLAP-functionaliteit (multidimensionele kubussen). Doelgroep van dit type functionaliteit zijn kenniswerkers op ondersteunende afdelingen: Planning & Control, Marktonderzoek en dergelijke.

Daarnaast vindt op beperkte schaal statistische analyse plaats. Ook wordt een proof-of-concept uitgevoerd met data-mining technologie.

Twee vormen van functionaliteit die niet direct BI-gerelateerd zijn, maar wel door Pallas geleverd worden, verdienen kort de aandacht:

- **Gegevensaanlevering:** Pallas is niet alleen hét platform voor managementinformatie binnen Albert Heijn, het is tevens door de gekozen datawarehouse-architectuur de centrale opslagplaats van historische data. Deze eigenschap gecombineerd met het feit dat er sprake is van een hoge mate van datakwaliteit maakt Pallas de ideale leverancier van historische data. Dit is geëffectueerd in de vorm van interfaces terug naar operationele systemen. Er is dus sprake van *closing the loop*: operationele toepassingen die voor het uitvoeren van hun taak historische informatie nodig hebben, worden belevd door Pallas. Hierdoor konden o.a. het voorspellen van volumes ten behoeve van automatische beleving van winkels, het voorspellen van acties, en het ontwerpen van schappenplannen voor winkels worden ondersteund.
- **Operationele reporting:** omdat het geselecteerde standaardtool voor reporting ook te gebruiken is op operationele databases, is ervoor gekozen om alle reporting (niet alleen managementinformatie, maar

³ Reject, Ignore, Correct, Suspend

ook operationele rapportages) te standaardiseren op dit front-end tool. Dit maakt de afbouw mogelijk van vele varianten van andere reporting tools die vaak met standaardpakketten worden meegeleverd. Praktisch betekent dit het realiseren van een aanvullende rapportageomgeving, ditmaal echter op het operationele systeem in plaats van op het datawarehouse. Dergelijke omgevingen zijn o.a. gerealiseerd voor de logistieke systemen. Voor de gebruiker blijft een hoge mate van transparantie gewaarborgd: met dezelfde portal, tools en lay-out kan hij zowel zijn management- als zijn operationele rapportages opvragen.

Techniek

Infrastructuur

Het grootste deel van het Pallas datawarehouse draait op een IBM p670 met 16 1,45-GHz-processoren, onder AIX. Daarnaast draait er nog 1 performance-intensieve datamart op een IBM S85 met 8 750 MHz processoren, eveneens onder AIX. Alle data bevindt zich op een EMC schijvenkabinet ter grootte van ruim 11 TB.

Als RDBMS wordt Oracle gebruikt. Voor het transformeren en laden van de data in het datawarehouse wordt een ETL-tool gebruikt; beheersmatig verdiende dit de voorkeur boven handgecodeerde procedures, gezien de verwachte hoeveelheid processing jobs. Gekozen is voor Informatica Powercenter. De samenhang en scheduling van de Powercenter-workflows wordt vastgelegd en gemanaged in Control-M (BMC); geleidelijk aan wordt een deel van de scheduling en het workflow management verplaatst naar de Powercenter Workflow Manager.

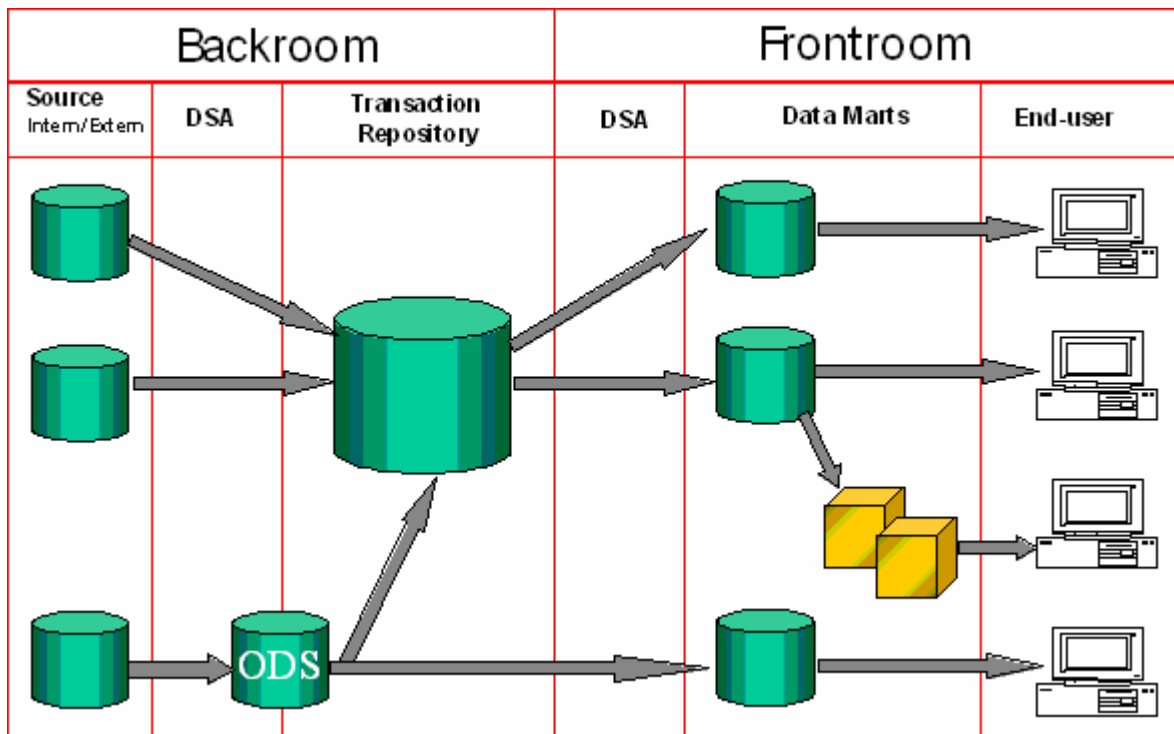
De eindgebruikers bevragen het datawarehouse niet rechtstreeks, maar met behulp van end-user tools. Voor standard reporting wordt Microstrategy gebruikt. De meeste gebruikers bekijken via Microstrategy Web hun rapporten in een webbrowser, of krijgen die in de vorm van PDF-files, Excelsheets of SMS'jes opgestuurd door Microstrategy Narrowcaster. Ter ondersteuning van Microstrategy Web zijn enkele webservers ingericht op basis van Microsoft IIS. Deze webservers kunnen onderling de load van gebruikersaanvragen balanceren.

Voor ad-hoc analyses is Hyperion Essbase beschikbaar, met als front-ends Hyperion Analyzer en Temtec Executive Viewer.

Architectuur

Overall architectuur

De Pallas-architectuur is opgebouwd rondom een centraal datawarehouse, de Transaction Repository (TR). Deze architectuur is feitelijk een hybride tussenvorm van enerzijds de Corporate Information Factory van Inmon [3], en anderzijds het op dimensionele datamarts gebaseerde datawarehouse van Kimball [1]. De TR wordt gevuld met brongegevens, deze worden aan het voorportaal (data staging area) afgeleverd en van daaruit verder verwerkt. De datamarts worden opgebouwd uit de TR. Uit de datamarts kunnen multidimensionele kubussen gegenereerd worden ten behoeve van ad-hoc analyse.

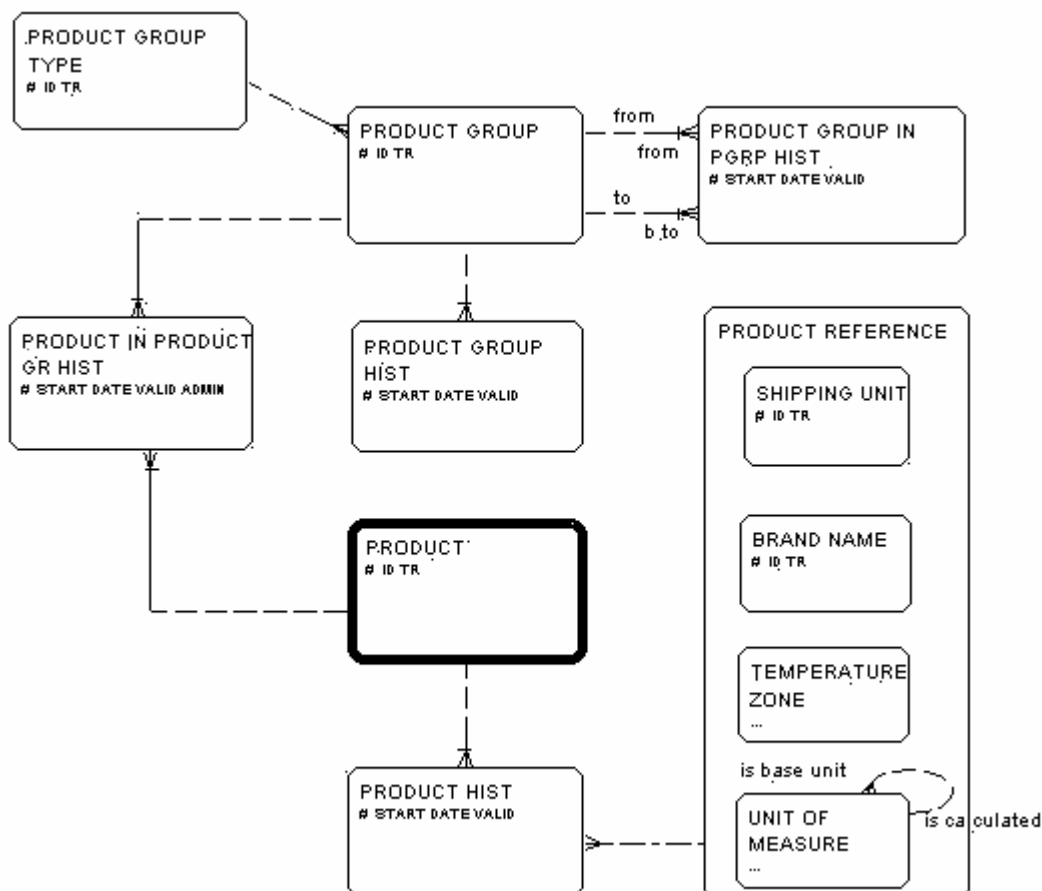


Afbeelding 3 Overall architectuur van Pallas

Transaction Repository

De Transaction Repository (TR) fungeert als de centrale “one copy of the truth”, de bewaarplaats van historische data op zo hoog mogelijk detailniveau. De TR is geoptimaliseerd voor bulk loading en queries op detailgegevens, en wordt in principe niet bevraagd door eindgebruikers. De uiteindelijke horizon voor de TR is 5 jaar, bij een databaseomvang van ongeveer 2,5 TB (raw data).

Met het oog op beheersbaarheid is het onwenselijk om het model van de TR vaak te wijzigen: de centrale bron voor historische data moet niet elk half jaar compleet opnieuw ontworpen, geïmplementeerd, en gemigreerd hoeven worden. Het TR-model is dan ook ontworpen met in het achterhoofd een zo groot mogelijke onafhankelijkheid van de bronsystemen enerzijds en de business requirements van het moment anderzijds. Voor de referentiegegevens wordt een generieke wijze van modelleren, gebaseerd op het relationele model, gehanteerd; referentiegegevens worden opgeslagen volgens het volgende patroon:



Afbeelding 4 Modellering PRODUCT subject area (vereenvoudigd)

De entiteit PRODUCT vormt het hart van de subject area PRODUCT: het domein in de TR waar alle referentiegegevens betreffende verkoopartikelen worden opgeslagen. Deze entiteit bevat alleen onveranderlijke attributen van PRODUCT; het is ook de plaats waar de mapping tussen operationele keys en surrogate keys plaatsvindt. Binnen het datawarehouse worden alleen de surrogate keys gebruikt. In PRODUCT_HIST worden alle veranderlijke attributen van PRODUCT opgeslagen. Deze tabel bevat als het ware een aantal snapshots van PRODUCT, voorzien van een tijdsaanduiding: voor ieder snapshot worden begin- en einddatum van de betreffende toestand opgeslagen. Hier worden ook relaties met referentiegegevens als unit of measure en brand name vastgelegd. Op deze manier wordt historie opgeslagen, ook als de bron dit niet zelf al doet.

Groeperingen van PRODUCT worden opgeslagen in PRODUCT_GROUP. Door middel van een link naar de tabel PRODUCT_GROUP_TYPE wordt vastgelegd wat voor soort groepering er opgeslagen wordt. De tabel PRODUCT_GROUP_IN_PGRP_HIST wordt gebruikt om groeperingen hiërarchisch aan elkaar te koppelen. Met de tabel PRODUCT_IN_PRODUCT_GR_HIST wordt een product in een hiërarchie gehangen. Ook de laatste twee tabellen bevatten weer een begin- en einddatum, omdat hiërarchieën in de tijd

kunnen veranderen. Door deze wijze van modelleren kunnen hiërarchieën en hiërarchieniveaus zonder gevolgen voor de modellering worden toegevoegd en uitgebreid.

Datamarts

Business requirements

Terwijl de TR een permanent karakter heeft, zijn de datamarts uitsluitend gericht op het vervullen van een informatiebehoefte uit de business, zonder aandacht voor eisen als toekomstvastheid, herbruikbaarheid etc. Het idee hierachter is dat de TR de onveranderlijke historische bron vormt, terwijl de manier waarop de eindgebruiker naar die gegevens kijkt, kan veranderen, afhankelijk van de ondersteuning die hij nodig heeft om zijn taak te vervullen. Als de gebruiker van de data verandert, kan dat betekenen dat een datamart volledig opnieuw gemodelleerd en gegenereerd moet worden. De datamarts zijn dus veel vluchtiger van karakter dan de TR. Zoals aangegeven in het hoofdstuk Functionaliteit, wordt er bij het modelleren naar gebruikersbehoefte naar gestreefd datamarts zodanig op te bouwen dat ze als geheel een bepaald bedrijfsproces ondersteunen.

Een ander verschil tussen de TR en datamarts is dat de datamarts meestal geen detailgegevens bevatten, maar vooral geaggregeerde data. De detailgegevens uit de TR dienen alleen als basis om er de views uit samen te stellen die de gebruiker nodig heeft voor ondersteuning van zijn werk. Zo dienen de details van kassatransacties die in de TR opgeslagen zijn, als basis voor het samenstellen van omzetgegevens per filiaal per kwartaal. De gebruiker is alleen geïnteresseerd in de geaggregeerde data, dus hoeft de datamart ook alleen de aggregaten te bevatten. (Een uitzondering is de datamart ten behoeve van bonuskaartanalyse, waar analyse op kassatransactieniveau plaatsvindt.)

Ook de tijdshorizon van iedere datamart wordt bepaald door de gebruikerswensen waarvoor hij ontwikkeld wordt. De datamart met de meeste historie omvat 3 jaar aan (geaggregeerde) data, maar er is ook een datamart met een historie van 6 weken. Daarnaast gelden voor kubussen ook technische beperkingen aan de hoeveelheid data die ze kunnen opslaan.

Toolbepaalde requirements; dimensionele modellering

Naast de businessdrive zijn voor de modellering van datamarts ook de eisen relevant die door het end-user tool opgelegd worden. De meeste end-user tools vragen een dimensionele modellering (sterschema's), dus wordt voor de datamarts aangesloten bij de dimensionele modelleerwijze van Kimball [1]. Daarnaast stelt Microstrategy een aantal specifieke eisen aan het dimensionele model. Zo gaat de performance van Microstrategy er significant op vooruit als de hiërarchieën van de dimensies in het model expliciet worden uitgemodelleerd ("snowflaking"). Op dit punt is afgeweken van de Kimball-modellering, die juist sterke denormalisatie voorschrijft. De performance van de snowflake wordt nog verder verbeterd door de sleutels uit de hogere niveaus te laten overerven naar lagere niveaus in de hiërarchie, zodat het aantal joins om een hiërarchieniveau te bereiken, geminimaliseerd wordt. Zo bevat de entiteit ARTIKEL de sleutels van alle bovenliggende niveaus. (Overigens geldt ook hier dat, als Microstrategy vervangen zou worden door een ander end-user tool, deze wellicht dusdanig andere eisen stelt aan de modellering van de datamarts dat deze opnieuw gemodelleerd en gegenereerd moeten worden. Door de scheiding tussen TR en datamarts, en de onafhankelijke modellering van de TR, kan dit zonder problemen binnen korte tijd verwezenlijkt worden.)

In de datamart kijken de gebruikers altijd door "de bril van vandaag" naar de feiten, of dat nu feiten van gisteren of van twee jaar geleden zijn. Dit betekent dat de feiten in de datamart altijd tegen de laatst bekende versie van de dimensies worden aangelegd (in Kimball-termen: er wordt een type 1 benadering van slowly changing dimensions gevolgd). Hoewel vanuit de TR andere "brillen" gegenereerd kunnen worden, is daar tot dusver nog geen business requirement voor geweest. In de datamarts wordt dus nog geen gebruik gemaakt van de in de TR beschikbare dimensiehistorie.

Het genereren van kubussen gebeurt op basis van de datamarts, en niet vanuit de TR, omdat de opbouw van de dimensies in de kubussen gelijk is aan die in de datamarts. Eenmaal gegenereerd in de datamarts, staan ze dus meteen klaar in het juiste formaat voor de kubussen.

Alle dimensies in de datamarts worden ontworpen als conforme dimensies, zodat de informatie in de verschillende sterschema's onderling vergelijkbaar en relateerbaar is ("drilling across").

Data staging areas

In de eerste DSA worden bronbestanden ontvangen, en wordt de brondata omgemodelleerd naar het TR-formaat. Hier vinden ook kwaliteitscontroles, schoning en verrijking van de data plaats. De initiële DSA bestaat zowel uit bestanden als uit tabellen. In de tweede DSA wordt bepaald welke feitgegevens doorgeladen moeten worden (zie paragraaf Delta's), en worden de conforme dimensies opgebouwd op basis

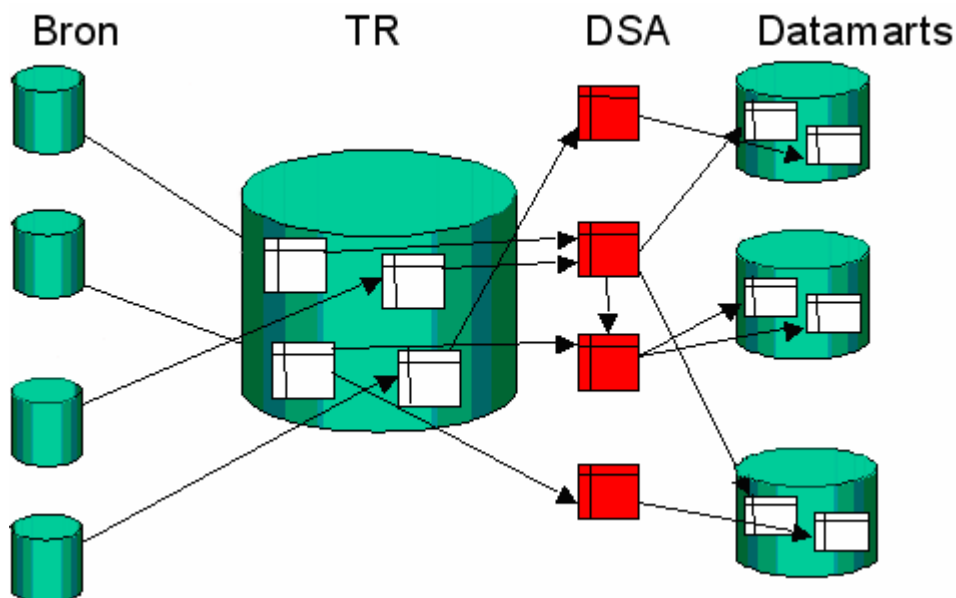
van de laatste stand van de referentiedata in de TR, en als geheel doorgezeten naar de datamarts. De DSA-tabellen maken fysiek onderdeel uit van de TR-database.

Processing

De verwerking van de gegevens uit de bronsystemen vindt grotendeels 's nachts batchgewijs plaats. Daardoor blijft de beschikbare processorcapaciteit overdag gereserveerd voor het draaien van rapporten op datamarts en het doen van analyses op kubussen; bovendien kunnen sommige gegevens pas na de dagafsluiting aangeleverd worden. De processing is onderverdeeld in een aantal hoofdstromen, die ieder een bepaald type data laden: kassatransacties, logistieke bewegingen, voorraadstanden, etc. De hoofdstromen zijn op hun beurt weer samengesteld uit in totaal zo'n 1500 Powercenter-jobs. Het managen van de onderlinge afhankelijkheden en het afstarten van deze stromen gebeurt in Control-M.

Gegevens kunnen in meerdere datamarts gebruikt worden; er is dus een m:n correspondentie tussen gegevensstromen en datamarts. In eerste instantie werden brongegevens in één stroom per bron doorgeladen van de bron via de TR naar de datamarts waarin ze nodig waren. Soms werden, om performance-redenen, gegevens geladen vanuit andere datamarts in plaats van vanuit de TR, omdat de benodigde tabellen al in de gewenste (basis)vorm in een andere datamart aanwezig waren. Hierdoor raakten de verschillende stromen steeds verder verweven, wat een ongunstig effect had op de onderlinge afhankelijkheden en de uitbreidbaarheid van de omgeving.

Daarom werd een nieuwe aanpak ontwikkeld, waarin gewerkt wordt met *halffabrikaten*. Dit houdt in dat de gegevensstromen in tweeën worden geknipt: eerst worden de gegevens van alle stromen verwerkt in de TR en de achterliggende DSA, waar er zgn. halffabrikaten van gemaakt worden. Deze halffabrikaten dienen vervolgens als basis om de afzonderlijke datamarts op te bouwen. Gemeenschappelijke basistabellen komen nu dus in de DSA terecht.



Afbeelding 5 ETL-verwerking middels halffabrikaten

Een traject ten behoeve van ontvlechting en gedeeltelijke herinrichting van de staande omgeving volgens het halffabrikaten-principe is inmiddels in gang gezet en zal nog tot eind 2004 doorlopen.

Delta's

Sommige bronnen kunnen wijzigingen (correcties) op reeds geladen feitgegevens aanleveren. Om dit correct te verwerken, wordt gebruikgemaakt van een "smart" deltamechanisme: er vindt een signalering plaats van "bewegingen" in de feitgegevens. Vastgesteld wordt welke gegevens sinds de vorige verwerkingslag gewijzigd of nieuw aangeleverd zijn. Op basis hiervan wordt een ToDo-list gecreëerd die aangeeft welke feiten moeten worden geactualiseerd. De ETL-verwerking bepaalt op basis van deze ToDo-list, welke gegevens vervolgens worden doorgeladen naar de datamarts. Gewijzigde gegevens worden eerst verwijderd uit de tabellen en vervolgens opnieuw ingevoegd. De feitentabellen worden dus incrementeel opgebouwd, in

plaats van steeds volledig opnieuw. Niet alleen vermindert dit de totale processing in vergelijking met “dom” bulkladen, ook wordt direct een administratie en logging bijgehouden van de gegevens die bewogen hebben.

Heraggregatie

Het gebruik van de bril van vandaag brengt een issue met zich mee bij geaggregeerde gegevens. Immers, deze zijn geaggregeerd op basis van de bril van het moment van aggregeren; feitegegevens die niet gewijzigd zijn, worden niet herladen, dus wijzigingen in de bijbehorende referentiedata (bijstelling van de “bril”) worden niet doorgevoerd in de aggregaten. Het beeld van de werkelijkheid wat de aggregaten geven, wordt dus hoe langer hoe minder accuraat. Dit geldt voor aggregaten die geaggregeerd worden langs dimensies waarin entiteiten een andere groepering kunnen krijgen; bijvoorbeeld bij de artikeldimensie, als artikelen van de ene assortimentsgroep naar de andere verplaatst worden.

Daarom moeten deze aggregaten met enige regelmaat opnieuw opgebouwd worden. Gezien de omvang van de beschikbare historie in sommige datamarts, is heraggregatie als onderdeel van de nachtelijke verwerking niet haalbaar: de heraggregatie zou simpelweg teveel tijd in beslag nemen. Wanneer het tijd is voor heraggregatie, wordt een kopie gemaakt van de basistabellen, en wordt het heraggregatieproces als verwerking in de achtergrond gestart. Als de heraggregatie voltooid is, worden de oude en de nieuwe aggregaat tabellen gewaapt.

Interfaces en interactie met andere systemen

Een datawarehouse, dat in het leven geroepen wordt om gegevens uit andere systemen te integreren, heeft per definitie koppelingen met andere systemen.

Bij het integreren van gegevens uit verschillende systemen had Pallas van begin af aan een belangrijk voordeel ten opzichte van vele van haar collega-datawarehouses: binnen Albert Heijn waren de bronsystemen reeds in hoge mate gestandaardiseerd en geüniformeerd, doordat werd gewerkt op basis van een corporate datamodel. Dat betekende dat voor ieder type brongegeven in principe maar één bronsysteem in gebruik was; aanlevering van gelijksoortige gegevens uit verschillende, qua modellering en datadefinitie ongelijksoortige bronsystemen was dus geen issue.

In 2000 werd besloten tot een hermodellering van het AH-applicatielandschap. Om beter te kunnen inspelen op business requirements ten aanzien van flexibiliteit en differentiatie, werd besloten tot een transitie naar een servicegeoriënteerde aanpak. Applicaties worden voortaan opgezet als onafhankelijke services die ieder hun eigen data store hebben, en asynchroon en event-driven met elkaar communiceren door middel van messages over een message broker.

Het is duidelijk dat dit ook zijn weerslag heeft op de aanlevering van brondata aan Pallas: hoewel het gros van de data nog als interfacebestand in batch aangeleverd wordt, leveren steeds meer applicaties, voor zover mogelijk en zinvol, hun brondata near-real-time aan in de vorm van messages. Dit stelt nieuwe eisen aan het ontvangst- en laadmechanisme van het datawarehouse, maar het biedt ook nieuwe mogelijkheden. Zo is Pallas inmiddels aangesloten op de berichtenstroom voor kassatransacties: iedere kassatransactie in de winkel wordt near-real-time doorgestuurd naar het hoofdkantoor, en daar verwerkt in diverse belanghebbende systemen, waaronder het datawarehouse. Het datawarehouse verzamelt de real-time gegevens in een *ODS* (operational data store), en laadt deze in kleine batches om de tien minuten door naar de TR. Dit wordt “trickle feed” genoemd: de gegevens “druppelen” als het ware het datawarehouse binnen. Niet alleen geeft dit Pallas meer ruimte in haar nachtelijk batchwindow – immers, hoe meer gegevens er overdag al binnendruppelen, hoe minder er 's nachts nog verwerkt hoeven worden –, ook kunnen gebruikers door middel van rapportages op het ODS near-real-time van actuele omzetinformatie voorzien worden. Aandachtspunt bij het verplaatsen van (een deel van) de verwerking naar overdag, is de balans tussen de tijdwinst door overdag laden, en de performance van de rapportages zoals ervaren door de gebruikers. Immers, als er overdag geladen wordt, kan dat ten koste van de processorcapaciteit gaan die beschikbaar is voor het runnen van rapporten.

Aanpak

Organisatie

Binnen Albert Heijn zijn de ontwikkeling en het onderhoud van Pallas belegd bij het Competence Center Business Intelligence (CC-BI). Dit is een onderdeel van AH IT, de IT-afdeling van Albert Heijn. Het CC-BI voert daarnaast ook BI-activiteiten uit voor diverse andere werkmaatschappijen van Ahold, op basis van de bestaande infrastructuur en werkwijze. Het CC-BI is verdeeld in twee onderafdelingen: Operations & Improvements (O&I) en Projecten. De uitgevoerde werkzaamheden omvatten alle voorkomende activiteiten, van DBA-werkzaamheden tot end-user support; alleen het beheer van het machinepark is elders belegd. Afhankelijk van het aantal lopende projecten, zijn op de afdeling 30 tot 50 man werkzaam. Het CC-BI levert *services* aan zijn gebruikers. Een service kan gezien worden als een logisch geheel van aangeboden informatie, inclusief de complete stroom erachter, van bron tot rapport.

Bestaande services

De levering van bestaande services wordt verzorgd door O&I. O&I is primair verantwoordelijk voor de dagelijkse operatie en het onderhoud van de staande architectuur en BI-oplossingen; ook het doorvoeren van kostenbesparende verbeteringen aan architectuur en infrastructuur valt daaronder. Om richting te geven aan het beheer, wordt met iedere gebruikersgroep een Service Level Agreement afgesproken voor de afgenomen services. Hierin worden afspraken vastgesteld over bijv. de verwachte tijd waarop de gevraagde rapportages elke ochtend beschikbaar zijn, en de maximale downtime van een service. Verder levert O&I support en opleiding aan de eindgebruikers van het datawarehouse, en voorziet in Quick Services: op aanvraag kan informatie worden geleverd die de gebruikers zelf niet via rapporten of kubussen kunnen benaderen, maar die wel in het datawarehouse opgeslagen ligt. Tenslotte worden binnen O&I de standaarden en richtlijnen ten aanzien van ontwikkeling van nieuwe oplossingen opgesteld en bewaakt, en wordt het Pallas Delivery Framework beheerd. In dit raamwerk zijn de ontwikkelprocessen en deliverables voor nieuwe projecten gedefinieerd.

Metadata

Om de omgeving goed te kunnen beheren, is *metadata* (gegevens over gegevens) over de operatie onontbeerlijk. Daarbij zijn twee typen metadata met name interessant. Ten eerste *dynamische* metadata, metadata over het proces: hoeveel rapporten worden er gedraaid, hoe verloopt de machinebelasting gedurende de dag, welk service level is gehaald etc. Het datawarehouse, als gestandaardiseerde omgeving voor rapportage, rapporteert zelf de informatie over zijn eigen operatie. Hiertoe worden met de eigen tooling feiten en dimensies in de TR geregistreerd. Deze worden vervolgens tot een aparte datamart verwerkt waaruit eveneens met standaardtools wordt gerapporteerd. Een tweede soort metadata ter ondersteuning van de operatie wordt gevormd door *statische* metadata: informatie over de structuur van het datawarehouse. Verschillende services kunnen bepaalde componenten (machines, databases, ETL-processen etc) gemeenschappelijk hebben, dus uitval van of een aanpassing aan zo'n component kan impact hebben op meerdere services. Het inzichtelijk maken van de afhankelijkheden tussen services onderling, en het beoordelen van de impact van wijzigingen en storingen op de verschillende services is nu nog voornamelijk handwerk. Het streven is om dit (verder) te automatiseren met behulp van een geïntegreerde metadata-oplossing.

Nieuwe services

De onderafdeling Projecten is georganiseerd in de vorm van projectteams die nieuwe services ontwikkelen, op basis van de bestaande omgeving en architectuur. Deze projectteams werken niet in isolatie: om de aansluiting van nieuw ontwikkelde oplossingen op de staande omgeving te waarborgen, zijn medewerkers van O&I als reviewers betrokken bij projecten. Uiteindelijk bepaalt O&I of en wanneer een nieuwe oplossing in productie genomen wordt. Het centrale raamwerk voor ieder projectteam is het Pallas Delivery Framework. Omdat doorgaans ieder project of release dezelfde basiselementen kent – datalogistiek en front-end functionaliteit – is een hoge mate van standaardisatie in de werkzaamheden haalbaar. Hoewel er veel ruimte is en moet zijn voor specifieke oplossingen, opereert iedereen vanuit hetzelfde referentiekader. Dit is ook nodig omdat feitelijk iedereen op dezelfde omgeving werkt.

Methoden, technieken en documentatie

- Fundamentele beslissingen ten aanzien van de functionaliteit en de architectuur van het datawarehouse worden vastgelegd als *Key Design Decisions (KDD's)*. Oorspronkelijk in het leven geroepen als medium voor het formaliseren van de besluitvorming, vormen ze een goede vastlegging en referentie voor de principes die aan Pallas ten grondslag liggen.
- Daarnaast worden richtlijnen en best practices geformuleerd en vastgelegd ten aanzien van de dagelijkse ontwikkelpraktijk. Na experimenten met diverse vormen van vastlegging, zijn voor dit doel recentelijk *design patterns* als middel ingevoerd. Zo zijn design patterns opgesteld voor de verwerking van feiten met onbekende referentie, de wijze van loggen van datakwaliteitsproblemen, en het verwerken van real-time aangeleverde berichten.
- Analyse: informatieanalyse voor BI-toepassingen blijkt lastige materie. Het ontwerpen van gewenste output in de vorm van rapportages is uiteindelijk niet zo moeilijk, als eenmaal is vastgesteld ten behoeve van welke besturings- of besluitvormingsprocessen informatie nodig is. Het analyseren, modelleren en beschrijven van dit laatste lijkt echter nog een braakliggend terrein. Binnen het CC-BI zijn eigen ideeën en ervaringen uitgewerkt tot best practices, waaronder vormen van prototyping, oplossingsscenario's en *menukaarten*. Een menukaart is een representatiewijze waarin een (operationeel) proces wordt geanalyseerd; per processtap wordt bepaald welke (stuur)informatie is vereist en welke bron deze informatie kan leveren.

In ieder geval is uit ervaring gebleken dat analysemethoden voor “normale” systeemontwikkeling ontoereikend zijn. Zo is vastgesteld dat Use Case Engineering (RUP/UML) nauwelijks houvast biedt, aangezien iedere use case zou neerkomen op “Print Report”. Ook ervaringen met functiepuntanalyse hebben aangetoond dat het karakter van informatie-analyse en functioneel ontwerp voor BI-toepassingen fundamenteel anders is dan voor transactionele systemen.

Kosten en baten

Algemeen

Is het maken van een kosten/baten afweging in het algemeen voor IT al lastig, voor BI-toepassingen is deze helemaal moeilijk te maken. Immers, hoe bepaal je de baten van een nieuwe, betere BI-omgeving? Uiteindelijk komt dit neer op het kwantificeren van het effect van betere besluitvorming en sturing; dit lijkt een onmogelijke opgave. Daar komt bij dat investeringen vaak tegen elkaar afgewogen moeten worden. Denk aan de afweging die het management van een onderneming moet maken om ofwel in logistiek ofwel in BI te investeren: voor het eerste valt wellicht eenvoudiger aannemelijk te maken dat er harde baten zijn dan voor het tweede.

Zoals in de rationale al werd aangegeven, hebben in de afweging om te investeren in BI bij Albert Heijn een aantal overwegingen een rol gespeeld:

- *Beheersbaarheid*: Er bestond een hoge mate van versnippering in de bestaande oplossingen: een veelheid aan technische oplossingen, die ieder door een andere IT-afdeling werden beheerd. Vanuit beheersoogpunt was het wenselijk een en ander te consolideren en te uniformeren. De kosten/batenafweging kon voor deze overweging worden uitgedrukt in termen van lagere beheerkosten.
- *Noodzaak*: iedere onderneming heeft functionaliteit nodig om in zijn management- en stuurinformatie te voorzien. Naarmate de complexiteit van de organisatie toeneemt, nemen de eisen ten aanzien van management- en stuurinformatie eveneens toe. Het was evident dat toekomstige ontwikkelingen binnen Albert Heijn (o.a. de differentiatie strategie) eisen zouden stellen die met de bestaande oplossingen niet meer gerealiseerd konden worden: iets nieuws was hoe dan ook nodig. In het kader van deze overweging was het met name van belang een acceptabel investeringsniveau vast te stellen.
- *Ideologie*: het geloof bestond – ook op directieniveau – dat het creëren van één geïntegreerde informatie omgeving vele voordelen voor het bedrijf zou opleveren in termen van datakwaliteit, integratie en combinatie van gegevens, beschikbaarheid van detaildata, het “one copy of the truth”-principe en dergelijke. Zonder dat deze in harde cijfers gekwantificeerd konden worden, was er de sterke overtuiging dat met deze oplossing de basis gelegd kon worden voor toekomstige baten. Met name deze laatste overweging heeft bijgedragen tot het besluit om daadwerkelijke diepte-investeringen te doen.

Het zal duidelijk zijn dat de combinatie van overwegingen doorslaggevend is geweest voor de uiteindelijke beslissing om te investeren.

Kosten

In het werkveld wordt als vuistregel aangehouden dat ruim 70% van de ontwikkelkosten in de back-end ontwikkeling gaat zitten, dus met name in de ontwikkeling van de ETL-verwerking (hardware- en licentiekosten niet meegerekend). Deze vuistregel ging ook voor Pallas op; om deze reden is veel sturing gelegd op dit deel van de systeemontwikkeling. Eén van de maatregelen die is toegepast om de kosten te beheersen, is om in het geval van bronontsluiting de hele bron(tabel) in te lezen, en niet slechts die attributen waarvoor specifieke requirements bestaan. De meerkosten om deze later alsnog toe voegen zijn substantieel hoger gebleken dan de extra kosten om ze bij de initiële ontwikkeling mee te nemen.

Verder is veel aandacht besteed aan datakwaliteitsanalyse, ETL-ontwerp en integratietesten. Onvermijdelijk is hier ook leergeld betaald, en zijn initieel op bepaalde dealerreinen te grote incrementen gekozen. In het algemeen heeft de praktijk echter uitgewezen dat de neergelegde basis solide en toekomstvast is.

Net als met ieder ander informatiesysteem, blijft Pallas zich continu uitbreiden en is er voortdurend vraag naar wijzigingen. Om wildgroei te voorkomen is een stuurgroep geïnstitutionaliseerd die niet alleen tot taak heeft de totale kosten van Pallas in de hand te houden, maar ook om toe te zien op de inhoudelijke kwaliteit en samenhang van de BI-omgeving als geheel.

Baten

Aan de batenkant worden de voordelen vooral gevonden in herbruikbaarheid van onderdelen van de datawarehouse-architectuur. Relevante onderdelen in dit kader zijn met name:

- kennis
- standaardtools
- aanpak en methodieken

Door het combineren van bovenstaande componenten zijn een aantal architectuurvarianten ontwikkeld – als spin-off van Pallas – waarmee andersoortige wensen op het gebied van managementinformatie bediend konden worden. “Andersoortig” wil zeggen: kleinschaliger, met een kleinere reikwijdte en lagere eisen ten aanzien van integratie van bedrijfsprocessen. Binnen het CC-BI wordt in dit verband gesproken over “de Ferrari en de Volkswagen”: het competence center heeft de architectuur, de kennis, de tools en de aanpak in huis om een Ferrari te bouwen, maar er is geen reden waarom die niet zouden kunnen worden ingezet om een Volkswagen te bouwen.

Deze alternatieve varianten zijn tot dusver ingezet voor andere werkmaatschappijen van Ahold, waaronder Gall&Gall, Ahold Vastgoed, albert.nl en de holding zelf. Zo is een datawarehouse-oplossing voor Gall&Gall gecreëerd tegen relatief zeer lage kosten. De beheerkosten zijn navenant laag, zeker in vergelijking met een scenario waarin Gall&Gall ontwikkeling en beheer in eigen hand gehouden zou hebben.

Strikt genomen vallen deze baten niet ten deel aan Albert Heijn, maar wel aan Ahold als concern. Dit heeft ook zeker in de overwegingen meegespeeld bij de initiële investeringsbeslissingen.

Voor Albert Heijn zelf heeft het integreren en combineren van gegevens nieuwe mogelijkheden geopend. Zo was er voorheen weinig tot geen inzicht in de financiële effecten van afboekingen in winkels: de gegevens die voor de bepaling daarvan relevant waren, waren verspreid over diverse systemen en konden, door hun onderling afwijkende datadefinities, niet gecombineerd worden. Nadat de gegevens in Pallas onder één datadefinitie gebracht en gecombineerd waren, kon voor het eerst structureel actie ondernomen worden om de afboekingen in te dammen.

Een andere belangrijke bate waar Albert Heijn rechtstreeks de vruchten van plukt, is de lead-time voor het aanleveren van gegevens op ad-hoc basis. Legio voorbeelden zijn te geven waarbij de geïntegreerde verzameling van gegevens, en de snelheid waarmee deze benaderbaar is, hun kracht hebben bewezen:

- Inzichtelijk maken van hamstergedrag na grote calamiteiten (11 september, oorlog Irak)
- Ondersteuning van het initiële prijsoffensief najaar 2003
- Tussentijdse evaluatie van specifieke acties

Daarnaast is gebleken dat het behoud van historie van zowel referentiële als feitdata veel toegevoegde waarde biedt. Niet alleen ontstaat hierdoor een beter inzicht in het historisch verloop van bijvoorbeeld acties, voorraadstanden en afzet, ook kan het historisch verloop van referentiedata, zoals de ontwikkeling van inkooprijzen door de jaren heen, geanalyseerd worden.

Toekomstperspectief

Nu eenmaal een groot deel van de in bronsystemen beschikbare gegevens in het datawarehouse voorhanden is, zal aan de aanbodkant de aandacht steeds meer verschuiven naar het “uitfilteren” van de golden nuggets uit de enorme hoeveelheid gegevens die voorhanden is: “minder is meer”. Gebruikers zullen bijvoorbeeld steeds vaker door middel van automatische alerts en proactieve exceptierapportages op basis van business rules geïnformeerd worden. Er wordt als het ware intelligentie toegevoegd aan de informatie.

Aan de vraagkant worden initiatieven opgestart om derden, zoals leveranciers, toegang tot onderdelen van het datawarehouse te geven. Binnenshuis gaat met name het integratie-aspect een grotere rol spelen: werden in de afgelopen jaren de meeste subject areas in eerste instantie ontsloten met als primair doel de ondersteuning van afzonderlijke business processen, in de toekomst zal de vraag naar integrale informatie over de gehele keten toenemen. Deze beweging is bijvoorbeeld zichtbaar aan de replenishment-kant, waar, na afzonderlijke business requirements voor ondersteuning van DC replenishment en winkelbevoorrading, nu de vraag opkomt naar informatie over voorraadbewegingen door de gehele keten heen.

Dit soort informatie is, door de gekozen aanpak en architectuur, in de meeste gevallen al geïntegreerd aanwezig in het enterprise datawarehouse. Ook op dit punt zullen de gemaakte keuzes ten aanzien van aanpak en architectuur dus in toenemende mate hun vruchten gaan afwerpen.