



Entrepôt

Data Warehouse Consultancy B.V.

Data Warehouse Architectuur

Versie januari 2002

Adres: Lovensestraat 41, 5014 DM Tilburg
Telefoon: 06-51657583
Email: Pieter.Priems@entrepot.nl
Internetsite: www.entrepot.nl

Inhoudsopgave

1	Inleiding.....	3
2	Conceptuele architectuur	3
2.1	Aanleiding	3
2.2	Conceptuele architectuur van een data warehouse	3
2.3	Filosofie van een data warehouse	3
2.3.1	Leveranties.....	4
2.3.2	Tussenhandel	4
2.3.3	Afname door consumenten.....	4
2.3.4	Beheer	5
2.3.5	Besturing.....	5
2.3.6	Verandering en aanpak	5
2.4	De functionele componenten en gegevensverzamelingen van een data warehouse	6
2.4.1	Structuur	6
2.4.2	Kennis en taakverdeling	7
2.4.3	Componenten.....	7
2.5	Flexibiliteit.....	8
2.6	Eenrichtingsverkeer	8
3	Functionele Architectuur	9
3.1	Componenten Functionele Architectuur.....	9
3.2	Huidige situatie.....	9
3.3	Gewenste situatie	9
3.4	Van de huidige situatie naar de gewenste situatie	9
4	Technische architectuur	10
4.1	Samenhang functionele en technische componenten.....	10
4.1.1	Aanlevercomponent.....	10
4.1.2	Aanleververwerkingscomponent	10
4.1.3	pakhuis.....	10
4.1.4	winkelbevoorradingcomponent en winkelverwerkingscomponent	10
4.1.5	informatiewinkels	10
4.1.6	winkeltoepassingscomponent	10
4.1.7	beheercomponent.....	10
4.2	Componenten Technische Architectuur	10
4.3	Huidige situatie.....	11
4.4	Gewenste situatie	11
4.5	Van de huidige situatie naar de gewenste situatie	11

1 Inleiding

Een architectuur beschrijft een systeem op drie verschillende niveaus:

1. Conceptueel, de **onderdelen** (componenten)
2. Functioneel, het **wat**, de inhoud
3. Technisch, het **hoe**, de techniek.

Een data warehouse is in elke organisatie anders ingericht. Dit is afhankelijk van de aanwezige systemen en de informatiebehoefte (inhoud). Conceptueel en technisch zijn er sterke overeenkomsten tussen de data warehouses in organisaties. Dit document gaat vooral in op de gemeenschappelijke conceptuele architectuur. De functionele en technische architectuur zijn in hoofdlijnen weergegeven. Details staat in de projectdocumentatie.

2 Conceptuele architectuur

2.1 Aanleiding

Een van de grootste problemen bij het introduceren van nieuwe zaken is de communicatie over het onderwerp. Ieder heeft zijn eigen beeld van het onderwerp en de begrippen zijn niet gedefinieerd. Dat kan tot gevolg hebben dat men langs elkaar heen praat. Een oeroplossing voor dit probleem is het gebruik van een metafoor. Hierbij de metafoor die de architectuur van een Data Warehouse (DWH) weergeeft en die wordt gebruikt voor het communiceren over het onderwerp.

2.2 Conceptuele architectuur van een data warehouse

Dit hoofdstuk beschrijft een conceptuele architectuur van een DWH. Het gaat in op de filosofie van een DWH, geeft een beschrijving van de functionele componenten, de gegevensverzamelingen van een DWH en de organisatie.

De architectuurbeschrijving is een *conceptuele* beschrijving. Er is bewust voor gekozen afstand te nemen van de technologie of de specifieke toepassingssituatie. Op deze manier fungeert de architectuur als een ideaalmodel op basis waarvan in een later stadium ontwerpbeslissingen kunnen worden afgeleid c.q. getoetst. Met dit stuk wordt in vogelvlucht een overzicht gegeven van de functie die een DWH vervult en de componenten waaruit een DWH bestaat.

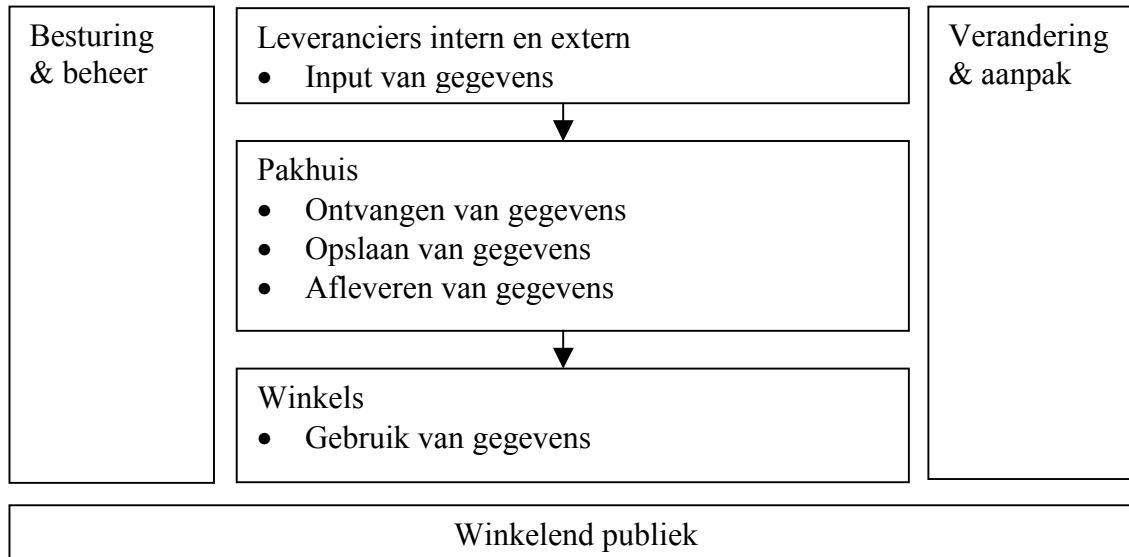
Paragraaf 2.3, de filosofie van een data warehouse, geeft een globale beschrijving van een DWH voor degenen die niet geïnteresseerd zijn in de verdere details.

In paragraaf 2.4 worden de acht functionele componenten en gegevensverzamelingen van een DWH kort toegelicht. Deze vormen de basis voor de techniek en is vooral bedoeld voor beheerders en ontwikkelaars.

2.3 Filosofie van een data warehouse

De naam "data warehouse" geeft al aan dat centraal in de architectuur van een DWH een gegevenspakhuis staat. Deze metafoor kan worden gebruikt om de overige functies van een DW te kunnen afleiden en te kunnen begrijpen. De filosofie van een DWH wordt daarom beschreven aan de hand van een vergelijking met de goederenvoorziening in de detailhandel.

In het verleden kochten de consumenten direct goederen in bij de fabriek. Het nadeel hiervan is dat de leveranciers in moeten spelen op de wensen van individuele consumenten en dat de consumenten bij alle leveranciers apart moesten inkopen. Kortom een onoverzichtelijke situatie. Dit gat in de markt wordt gevuld door de tussenhandel. Door het inrichten van een centraal pakhuis kunnen de leveranciers zich richten op massaproductie (lean en mean) voor de grote klanten en is er voor individuele consumenten één loket met een hoog serviceniveau: snelle (na)levering van goederen, een compleet en overzichtelijk assortiment en een gegarandeerde kwaliteit. Hoe ziet die goederen stroom er uit?



2.3.1 Leveranties

Grootwinkelbedrijven hebben één *pakhuis* waar goederen (hier: *gegevens*) door leveranciers (hier: *gegevensleveranciers of transactiesystemen*) worden afgeleverd.

2.3.2 Tussenhandel

Leveranties gaan niet "zomaar" het pakhuis in; ze worden eerst geschikt gemaakt voor opslag. Voorbeelden van dergelijke bewerkingen zijn: het van een pallet halen (hier: *inlezen*), het verwijderen van onnodig transportmateriaal (hier: *strippen van onnodige gegevenselementen*), het overpakken in andere eenheden (hier: *converteren*), etc. Vervolgens worden de goederen opgeslagen in de schappen (hier: *databases of directories*) van het pakhuis.

2.3.3 Afname door consumenten

Consumenten (hier: *eindgebruikers*) winkelen niet graag in het pakhuis zelf. De wijze waarop en de eenheden waarin de goederen zijn opgestapeld, de beperkte manier waarop de goederen liggen uitgestald en het rondlopende personeel, zijn ontworpen om een zo efficiënt mogelijk pakhuis te krijgen. Voor het winkelend publiek is deze wijze echter onhandig en werpt zij onnodige drempels op eens lekker te gaan winkelen (hier: *gegevens te gebruiken*). Vandaar dat een pakhuis haar goederen uitlevert aan winkels (hier: *informatiewinkels*). Winkels bestaan er in soorten en maten, variërend van warenhuizen (breed assortiment, maar geen specialiteiten) tot speciaalzaken (met uitgebreid advies rondom het toegesneden assortiment). De winkel bepaalt welk soort goederen (hier: *gegevenssoorten*), in welke hoeveelheden (hier: *mate van historie*), variëteit (hier: *detailgraad*) en versheid (hier: *actualiteit*) hij voert; afhankelijk van het specifieke publiek dat de winkel als doelgroep beschouwt.

De winkel wordt bevoorradt vanuit het centrale pakhuis. Er zijn meerdere bestel- en afleverstrategieën denkbaar (bijvoorbeeld vaste bestellingen, naleveringen, incidentele bestellingen, etc.). Het pakhuis beschikt dan ook over een *uitleveringsfunctie* waarmee op bestelling vanuit de winkel, de juiste levering te kunnen samenstellen en versturen. Ook hier zal het pakhuis bewerkingen voor zijn rekening nemen.

2.3.4 Beheer

De magazijnmeester (hier: *operationeel pakhuisbeheerder*) is verantwoordelijk voor de operationele logistieke stroom via het pakhuis. Hij houdt hiervoor de voorraad in het pakhuis in de gaten, waar de goederen vandaan zijn gekomen, hoe leveranciers goederen verpakken, welke winkels welk assortiment hebben, wat winkels besteld hebben, hoe leveringen verpakt moeten worden voor bepaalde winkels, etc. Tevens is hij het aanspreekpunt als er onverhoopt een kink in de logistieke kabel komt (leveringen die foutief verpakt zijn, niet compleet zijn, te laat afgeleverd worden, etc.) en bij interne herallocaties van goederen (zoals het opruimen van winkeldochters, het vooraan zetten van veelgevraagde goederen, etc.).

Naast het operationeel afstemmen van toe- en afleveringen, zullen ook op het gebied van assortiment vraag en aanbod in evenwicht moeten zijn. In het grootwinkelbedrijf zijn hiervoor de inkopers verantwoordelijk (hier: *functioneel pakhuisbeheerders*). Zij houden voeling met de smaak van het winkelend publiek om de juiste gegevens te kunnen inkopen. Uiteraard zijn zij hierbij gebonden door economische randvoorwaarden zoals prijs/opbrengst, technologische randvoorwaarden zoals beschikbare pakhuiscapaciteit en wettelijke/beleidsrandvoorwaarden (hier: *wie mag welke gegevens hebben/zien*).

Bij een groot aantal winkels maakt men gebruik van winkelbeheerders. Deze winkelbeheerders nemen voor hun winkel een aantal taken over van de inkopers en magazijnmeesters.

2.3.5 Besturing

De raad van bestuur bepaald de randvoorwaarden, het serviceniveau van de winkels en veranderingen in het logistieke proces.

Het serviceniveau van de winkel: assortiment, versheid van goederen, snelheid van levering, kwaliteit van de goederen en uitbreidingsmogelijkheden van het assortiment wordt bepaald door het serviceniveau van het pakhuis. Voor de bepaling van het serviceniveau van het pakhuis is het nodig te weten waar het winkelend publiek behoefte aan heeft.

Het pakhuis vervult primair een *intermediairfunctie*: zij streeft ernaar om zo min mogelijk goederen op voorhand in voorraad te houden. Pas bij concrete bestellingen gaat zij de inkoopmarkt op. Er is een duidelijke afhankelijkheid tussen de winkels, het pakhuis en de behoefte van het publiek.

2.3.6 Verandering en aanpak

Grote veranderingen in het logistieke proces: nieuwe winkels, andere of nieuwe leveranciers, uitbreiding of aanpassing van het pakhuis, nieuwe wegen en vrachtvervoer worden aangepakt in de vorm van projecten. Nadat de verandering is doorgevoerd worden de resultaten overgedragen aan beheer.

Samengevat, een data warehouse:

- **ontvangt gegevens van verschillende leveranciers;**
- **slaat deze hanteerbaar en efficiënt op;**
- **houdt overzicht over wat er ligt, hoe oud het is, waar het naar toe mag, enz.;**
- **bevoorradt informatiewinkels volgens hun specificaties;**
- **verzorgt service aan de afnemers van gegevens.**

2.4 De functionele componenten en gegevensverzamelingen van een data warehouse

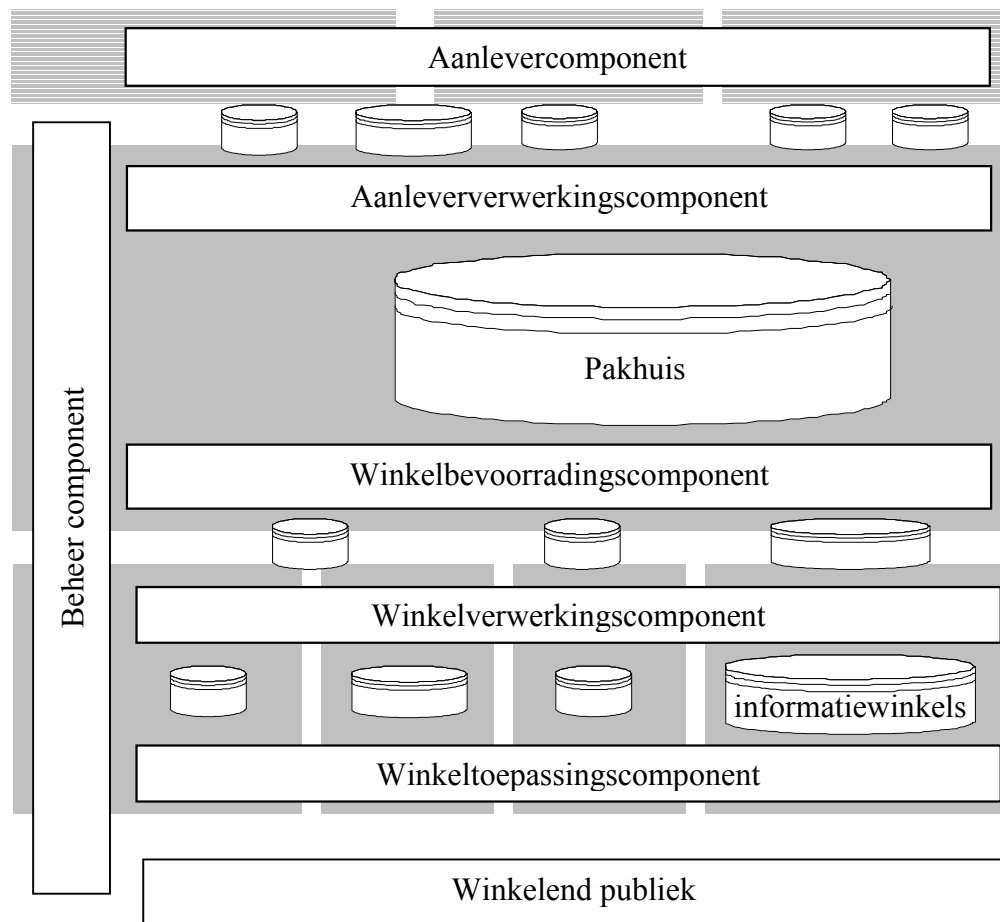
In deze paragraaf wordt de acht onderkende functionele componenten en gegevensverzamelingen kort beschreven.

2.4.1 Structuur

Onderstaand figuur geeft, gebaseerd op de pakhuismetafoor, in vogelvlucht de basisarchitectuur van een DWH. In de figuur zijn de verschillende functionele componenten en gegevensverzameling te onderkennen.

Inherent aan de functie van de basisarchitectuur, ligt de nadruk op de *structuur* van de functionaliteit en niet zozeer op specifieke invulling van deze functionaliteit voor een specifiek toepassingsgebied binnen een organisatie. Dit heeft tot gevolg dat de beschrijving van de functionaliteit een abstract karakter heeft.

De architecturele keuzen en de achterliggende overwegingen, die bij de nadere invulling van de componenten een rol spelen, worden in de functionele en technische architectuur uitgewerkt. Ook besturing en verandering (de organisatorische items) worden daar verder ingevuld.



2.4.2 Kennis en taakverdeling

Bij het toekennen van taken gaan we uit van het idee dat de interpretatie en bewerking van gegevens *daar geschiedt waar de kennis aanwezig is*. Er zijn drie mogelijke partijen die kennis hebben van gegevens: transactiesysteembeheerders, pakhuisbeheerders en winkelbeheerders. Beheerders van transactiesystemen hebben kennis van hun systemen. Samen met de pakhuisbeheerders, die kennis hebben van het gehele assortiment, realiseren zij de aanlevergegevens. Winkelbeheerders bezitten kennis van hun specifieke winkel. Zij stellen samen met de pakhuisbeheerders het winkelassortiment vast.

2.4.3 Componenten

Zoals uit de figuur valt op te maken is een DWH, net als de logistieke keten van de detailhandel, gelaagd opgebouwd. Elke component heeft een eigen kenmerkende functie. In een zevental stappen bereiken de gegevens vanuit de transactiesystemen de eindgebruiker, de beheercomponent is voor het beheer van de andere componenten.

2.4.3.1 Aanlevercomponent

De ruwe transactie- en referentiegegevens worden vanuit de transactiesystemen en externe systemen klaargezet in een door het DWH te verwerken formaat, de aanlevergegevens. Het maken van de aanlevergegevens is een taak van de *aanlevercomponent* en valt onder het beheer van de transactiesystemen.

2.4.3.2 Aanleververwerkingscomponent

Vervolgens worden deze aanlevergegevens getransformeerd naar een uniforme structuur en verwerkt in het gegevenspakhuis. Het is de taak van de *aanleververwerkingscomponent* deze processen uit te voeren. Deze taak is de verantwoordelijkheid van de pakhuisbeheerders.

2.4.3.3 Pakhuis

De gegevens die liggen opgeslagen in het pakhuis worden onder verantwoordelijkheid van het pakhuisbeheer onderhouden. Hierbij moet men denken aan het verwijderen/archiveren van oude gegevens of juist het overhalen van gegevens vanuit archief naar de werkruimte van het pakhuis, het aanmaken van gedenormaliseerde structuren of indexen ten behoeve van ontsluiting en het uitvoeren van verdichtingoperaties (bijvoorbeeld het maken van tellingen per maand). Het is de taak van het *pakhuis*, onder verantwoordelijkheid van de pakhuisbeheerders, deze processen uit te voeren.

2.4.3.4 Winkelbevoorradingcomponent

Vanuit het gegevenspakhuis worden de winkels bediend. Als gebruikers gegevens vragen, dan worden de benodigde gegevens geselecteerd, gegroepeerd en geconverteerd tot het vereiste formaat, de winkelbevoorradinggegevens. Hierbij kan men denken aan het klaarzetten van een database voor queryepakketten, het voeden van een eindgebruikersomgeving met Acces-bestand, of het uitvoeren van voorgedefinieerde queries. Het is de taak van de *winkelbevoorradingcomponent* deze processen uit te voeren. Deze taak valt onder verantwoordelijkheid van de pakhuisbeheerders.

2.4.3.5 Winkelverwerkingscomponent

De winkelbevoorradinggegevens toevoegen aan de informatiewinkels is een taak van de van de *winkelverwerkingscomponent*. Deze component valt onder de verantwoordelijkheid van de winkelbeheerders. De winkelbeheerder geeft aan wanneer een winkel open is en of de goederen in de winkel de juiste houdbaarheidsdatum bezitten

2.4.3.6 Informatiewinkels

De eindbestemming van de gegevens zijn de informatiewinkels. Belangrijke constatering is dat er verschillende soorten winkels bestaan: verschillend omdat het assortiment verschilt (de gegevensonderwerpen), en verschillend qua winkelformule (de wijze waarop een gebruiker door de gegevens kan snuffelen op zoek naar iets van zijn gading). Het verwerken van de gegevens uit de winkelbevoorrading is een taak van de winkelbeheerder.

2.4.3.7 Winkeltoepassingscomponent

De gegevens in de informatiewinkels worden *gebruikt* middels de *winkeltoepassingscomponent*. Voor winkeltoepassingen kunnen verschillende soorten pakketten worden gebruikt. Winkeltoepassingen staan functioneel onder beheer van de winkelbeheerder.

2.4.3.8 Beheercomponent

"Linking pin" in een DWH is de *beheercomponent* van de DWH-processen: het tijdig (laten) uitvoeren van gedefinieerde processen, het (laten) aanpassen van DW-processen conform nieuwe wensen en het toezien op de operationele gang van zaken in en rondom het pakhuis. Beheer vergt het opslaan van gegevens over de aangeleverde, opgeslagen en uit te leveren gegevens (de metagegevens). Ook definities van DWH-processen (op basis waarvan bijbehorende modules worden gegenereerd) behoren tot de metagegevens. Het is de taak van de beheercomponent deze processen uit te voeren. De pakhuisbeheerders zijn hiervoor verantwoordelijk.

2.5 Flexibiliteit

Bovenstaande stappen zijn met elkaar verbonden. De reden waarom de stappen onderkend worden, is dezelfde reden waarom stappen in een normale logistieke keten worden onderkend: elke stap heeft zijn eigen functie en schermt bepaalde problemen af van hogergelegen lagen. Op deze manier ontstaat een stabiele structuur, waarbinnen flexibel kan worden ingespeeld op veranderend aanbod en veranderende vraag.

2.6 Eenrichtingsverkeer

Benadrukt wordt dat de gegevensstroom in bovenstaande figuur slechts van transactiesystemen naar de winkels loopt en niet terug (*one-way replication*). Dit wil zeggen dat mutaties in afgeleide gegevens *niet* teruggevoerd worden naar de oorspronkelijke transactiebestanden. Mutaties zullen via de transactiesystemen aan het data warehouse worden aangeboden.

Samengevat, de data warehouse componenten zijn:

- een aanlevercomponent die gegevens in het juiste formaat levert;
- een aanleververwerkingscomponent die gegevens in ontvangst neemt en in het pakhuis plaatst;
- een pakhuis dat gegevens in bewaring heeft en deze in goede staat houdt;
- een winkelbevoorradingscomponent die in het juiste formaat uitgeeft aan de informatiewinkels;
- een winkelverwerkingscomponent die de gegevens in de informatiewinkels plaatst;
- informatiewinkels die gegevens beschikbaar stellen voor de eindgebruiker;
- een winkeltoepassingscomponent die de toegangspoort vormt voor de eindgebruiker;
- een beheercomponent die zorg draagt voor het correct opereren van de overige componenten.

3 Functionele Architectuur

Hierbij de functionele architectuur op hoofdlijnen. Het doel is het overzicht te krijgen.

3.1 Componenten Functionele Architectuur

De componenten bestaan uit de bronssystemen, het pakhuis en de winkels van de organisatie.

3.2 Huidige situatie

3.3 Gewenste situatie

3.4 Van de huidige situatie naar de gewenste situatie

Plan van aanpak

4 Technische architectuur

4.1 Samenhang functionele en technische componenten

Hieronder een vertaling van functionele componenten uit de conceptuele architectuur naar technische componenten. Het centrale punt waar de documentatie van alle componenten is opgeslagen is de repository.

4.1.1 Aanlevercomponent

De aanlevercomponent maakt onderdeel uit van het bronsysteem en ook de techniek is de verantwoordelijkheid van het bronsysteem. Het uiteindelijk opgeleverde aanleverbestand is de input voor het data warehouse. De aanleverbestanden kunnen in verschillende formaten worden geleverd.

4.1.2 Aanleververwerkingscomponent

Voor het in ontvangst nemen en in het pakhuis plaatsen van de gegevens wordt gebruik gemaakt van een Extractie, Transformatie en Laad Tool, kortweg ETL-tool.

4.1.3 pakhuis

Voor de opslag van gegevens wordt gebruik gemaakt van een database management systeem.

4.1.4 winkelbevoorradingcomponent en winkelverwerkingscomponent

Winkelbevoorrading en winkelverwerking geschied met hetzelfde ETL-tool als dat gebruikt wordt voor de aanleververwerkingscomponent.

4.1.5 informatiewinkels

Informatie winkels bevatten gegevens en maken net als het pakhuis gebruik van een database management systeem.

4.1.6 winkeltoepassingscomponent

De winkeltoepassingscomponent maakt gebruik van een Queriepakket dat deels werkt op de werkstations van de eindgebruikers.

4.1.7 beheercomponent

Voor beheer wordt gebruik gemaakt van een repository die in verbinding staat met het database management systeem, het ETL-tool en het Queriepakket.

4.2 Componenten Technische Architectuur

Hier de volgende onderdelen:

- **Database management systeem;**
- **Repository;**
- **ETL-tool;**
- **Queriepakket;**
- **Hardware en operating system.**

4.3 Huidige situatie

4.4 Gewenste situatie

4.5 Van de huidige situatie naar de gewenste situatie

Plan van aanpak