

**Data Warehouse Release 3**  
**Dossier softwarearchitectuur (SAD)**  
**Versie 1.2**

## Historiek van de aanpassingen

Datum	Versie	Beschrijving	Auteur
18/11/2010	0.1	Eerste versie voor interne revisie BB	Romuald Dumonceaux
21/11/2010	0.2	Correcties volgens opmerkingen van Michael Delire	Romuald Dumonceaux
15/12/2010	0.3	Aanpassingen na revisie door FOD Financiën	Romuald Dumonceaux
07/01/2011	1.0	Aanpassingen na tweede revisie door FOD Financiën en Logica	Romuald Dumonceaux
12/01/2011	1.1	Verduidelijkingen met betrekking tot het gebruik van MS SQL Server voor de optionele opslag van data marts (6.4 en 7.3.3.1).	Romuald Dumonceaux
01/03/2011	1.2	Introducties van de principes van auditing/logging en correcties na tactisch comité.	Romuald Dumonceaux

# Inhoudstafel

<b>HISTORIEK VAN DE AANPASSINGEN</b> .....	<b>2</b>
<b>INHOUDSTAFEL</b> .....	<b>ERREUR ! SIGNET NON DEFINI.</b>
<b>1. PRESENTATIE VAN HET DOCUMENT</b> .....	<b>6</b>
1.1 VOORWERP VAN HET DOCUMENT .....	6
1.2 REFERENTIEDOCUMENTEN .....	<b>ERREUR ! SIGNET NON DEFINI.</b>
<b>2. INLEIDING</b> .....	<b>ERREUR ! SIGNET NON DEFINI.</b>
2.1 SCOPE VAN HET DOCUMENT IN ZIJN HUIDIGE VERSIE .....	7
2.2 BEPERKINGEN .....	<b>ERREUR ! SIGNET NON DEFINI.</b>
2.3 DEFINITIES EN AFKORTINGEN .....	<b>ERREUR ! SIGNET NON DEFINI.</b>
<b>3. DOELSTELLINGEN, VEREISTEN EN MECHANISMEN VAN DE ARCHITECTUUR</b> .....	<b>8</b>
3.1 NALEVING VAN DE STANDAARDEN .....	<b>ERREUR ! SIGNET NON DEFINI.</b>
3.2 ERGONOMIE .....	9
3.3 PRESTATIEVERMOGEN .....	<b>ERREUR ! SIGNET NON DEFINI.</b>
3.4 CAPACITEIT .....	9
3.5 BESCHIKBAARHEID EN BETROUWBAARHEID .....	10
3.6 VEILIGHEID .....	<b>ERREUR ! SIGNET NON DEFINI.</b>
3.6.1 <i>Identificatie</i> .....	10
3.6.2 <i>Authenticatie</i> .....	11
3.6.3 <i>Autorisatie</i> .....	11
3.6.4 <i>Vertrouwelijkheid van de persoonsgegevens</i> .....	11
3.7 BEHEER .....	<b>ERREUR ! SIGNET NON DEFINI.</b>
<b>4. VIEW GEBRUIKSCASE</b> .....	<b>13</b>
<b>5. LOGISCHE VIEW</b> .....	<b>ERREUR ! SIGNET NON DEFINI.</b>
5.1 LOGISCHE ARCHITECTUUR .....	14
<b>6. VIEW IMPLEMENTERING</b> .....	<b>16</b>
6.1 GLOBAAL OVERZICHT – UML-DIAGRAM VAN DE COMPONENTEN .....	16
6.2 LAGEN – UML-DIAGRAM VAN DE COMPONENTEN .....	17
6.3 STRUCTURERING IN PACKAGES .....	18
6.4 PACKAGING / STRUCTURERING IN EENHEDEN VOOR INGEBRUIKNEMING .....	19
6.5 REALISATIES VAN STRUCTURERENDE GEBRUIKSCASES – UML-DIAGRAMMEN VAN CATEGORIEËN EN INTERACTIES (VERPLICHT) .....	20

<b>7.</b>	<b>VIEW GEGEVENS .....</b>	<b>21</b>
7.1	LOGISCHE GEGEVENSARCHITECTUUR .....	21
7.2	TECHNISCHE ORGANISATIE VAN DE GEGEVENS.....	<b>ERREUR ! SIGNET NON DEFINI.</b>
7.2.1	Bronsystemen .....	<i>Erreur ! Signet non défini.</i>
7.2.2	Bestandssysteem.....	<i>Erreur ! Signet non défini.</i>
7.2.3	Database .....	<i>Erreur ! Signet non défini.</i>
7.3	E/R-DIAGRAM VAN DE BLIJVENDE GEGEVENS.....	25
7.3.1	ODS.....	25
7.3.2	Data warehouse .....	26
7.3.3	Data marts.....	27
7.3.4	Operationele metagegevens .....	<i>Erreur ! Signet non défini.</i>
7.4	BACK-UPS .....	27
7.4.1	Service voor integratie van gegevens.....	28
7.4.2	BI-service .....	28
7.4.3	Bestandssysteem.....	<i>Erreur ! Signet non défini.</i>
7.4.4	Database .....	<i>Erreur ! Signet non défini.</i>
<b>8.</b>	<b>VIEW PROCESSEN .....</b>	<b>ERREUR ! SIGNET NON DEFINI.</b>
8.1	STANDAARDSTROOM .....	<b>ERREUR ! SIGNET NON DEFINI.</b>
8.1.1	Stap 0: Voorbereiding van de gegevens - Selectie .....	31
8.1.2	Stap 1: Formattering (Transformatie) .....	31
8.1.3	Stap 2: Laden van de ODS .....	32
8.1.4	Stap 3: Selectie en validering van de DWH-gegevens .....	32
8.1.5	Stap 4: Gegevenstransformatie .....	33
8.1.6	Stap 5: Uploaden van de DWH .....	34
8.1.7	Stappen 6,7 en 8: Opbouw van de Data Marts .....	34
8.1.8	Stap 9: Post-processing.....	35
8.2	STRATEGIE VOOR UITVOERING VAN DE VERWERKINGEN.....	35
8.3	UML-SEQUENTIEDIAGRAM .....	36
<b>9.</b>	<b>VIEW INGEBRUIKNEMING .....</b>	<b>37</b>
9.1	DIAGRAM VOOR INGEBRUIKNEMING .....	37
<b>10.</b>	<b>OMVANG EN PRESTATIES – FEEDBACK EN SUGGESTIES.....</b>	<b>38</b>
10.1	STROOM VOOR GEGEVENSINTEGRATIE .....	38
10.1.1	Prestatievermogen.....	<i>Erreur ! Signet non défini.</i>
10.2	DATABASE.....	<b>ERREUR ! SIGNET NON DEFINI.</b>
10.2.1	Prestatievermogen.....	<i>Erreur ! Signet non défini.</i>
10.3	BI-SERVICES .....	38

10.3.1	<i>Prestatievermogen</i> .....	<i>Erreur ! Signet non défini.</i>
<b>11.</b>	<b>KWALITEIT</b> .....	<b>ERREUR ! SIGNET NON DEFINI.</b>
11.1	UITBREIDBAARHEID / ONDERHOUD .....	39
11.1.1	<i>Beschrijving</i> .....	<i>Erreur ! Signet non défini.</i>
11.1.2	<i>Oplossing</i> .....	<i>Erreur ! Signet non défini.</i>
11.2	OVERDRAAGBAARHEID .....	<b>ERREUR ! SIGNET NON DEFINI.</b>
11.2.1	<i>Beschrijving</i> .....	<i>Erreur ! Signet non défini.</i>
11.2.2	<i>Oplossing</i> .....	<i>Erreur ! Signet non défini.</i>
11.3	MONITORING .....	40
11.3.1	<i>Beschrijving</i> .....	<i>Erreur ! Signet non défini.</i>
11.3.2	<i>Oplossing</i> .....	<i>Erreur ! Signet non défini.</i>
11.4	LOGGING .....	40
11.4.1	<i>Beschrijving</i> .....	<i>Erreur ! Signet non défini.</i>
11.4.2	<i>Oplossing</i> .....	<i>Erreur ! Signet non défini.</i>
<b>12.</b>	<b>TOEPASSING VAN DE STANDAARDEN VAN DE FOD FINANCIËN</b> .....	<b>41</b>
12.1	AFSPRAKEN VOOR NAAMGEVING.....	<b>ERREUR ! SIGNET NON DEFINI.</b>
12.1.1	<i>Bestanden</i> .....	<i>Erreur ! Signet non défini.</i>
12.1.2	<i>Namen van de objecten voor de opslag van gegevens</i> .....	41
12.1.3	<i>Namen van de objecten voor de verwerking van gegevens</i> .....	42
12.1.4	<i>Tool voor business intelligence</i> .....	42

# 1. Presentatie van het document

## 1.1 Voorwerp van het document

Dit document voorziet de noodzakelijke informatie met het oog op een goede kennis van de technische architectuur van het datawarehouse-project voor het beheer van de risico's die zijn gerelateerd met ondernemingen.

Dit project is tevens gekend onder de namen « Data warehouse II » of « Data warehouse release 3 ».

## 1.2 Referentiedocumenten

Letterwoord, gebruikt in het document	Referentie	Titel
SUPDEV_NOMMAGE	Convention_de_nommage-3.doc	Gegevens/Afspraken inzake naamgeving
DCC_ETL_ACCEPTANCE	20100120_DCC ETL Technical Acceptance Criteria_v110.pdf	Technische acceptatiecriteria voor ETL-project
DCC_DS_NAMING	20101129_v200_DataStage_Naming_Conventions.pdf	Afspraken inzake naamgeving voor DataStage
DWH2-<X> <sup>1</sup> -REQ	DWH2_<X>_UserRequirements_vy.z.doc <b>StarTeam path:</b> \DWH2\FUP_02_Requirements	User Requirements
DWH2-<X> <sup>1</sup> -SRC	DWH2_<X>_DescriptionSources_vy.z.doc <b>StarTeam path:</b> \DWH2\FUP_03_Analysis	Analyse van de bronnen
DWH2-<X> <sup>1</sup> -LOGICMOD	DWH2_<X>_ModèlesLogiques_vy.z.doc <b>StarTeam path:</b> \DWH2\FUP_03_Analysis	Logisch gegevensmodel
DWH2-<X> <sup>1</sup> -ETLSPEC	DWH2_<X>_SpecificationsETL_vy.z.doc <b>StarTeam path:</b> \DWH2\FUP_04_Implementation	Gedetailleerde specificaties van de ETL's
DWH2-OPSMANUAL	DWH2_ProceduresOperationnelles_vx.x.doc <b>StarTeam path:</b> \DWH2\FUP_09_Environment\FUP_01_Guidelines	Handleiding
IAM_LOGGING_DESIGN	AIM Logging Design - V1.4.odt	Identity and Access Management Logging Design Specifications

<sup>1</sup>: <X> stemt overeen met de releasecode.

De volledige documentatie met betrekking tot dit project is opgeslagen onder StarTeam, in het project « DWH2 ».

## 2. Inleiding

---

### 2.1 Scope van het document in zijn huidige versie

De FOD Financiën heeft begin jaren 2000 een moderniseringsproject opgestart dat Coperfin werd gedoopt. Het kadert in de Copernicus-hervorming van de federale administratie.

Er werden verschillende programma's uitgewerkt om de nieuwe modaliteiten in de praktijk te brengen. Een van de zestien programma's van het reorganisatieproject is het programma Risicobeheer (op het vlak van assistentie, controle en invordering). Het project *Data Warehouse / Datamining – Analyse van de risico's* maakt deel uit van dit programma.

In dit kader werd een voorstudie uitgevoerd, niet alleen om de exacte behoeften op het vlak van de gebruikers te preciseren en om de verhoopde voordelen te evalueren van een analytische omgeving voor Risicobeheer, assistentie, controle en invordering, maar ook om een eerste inventaris op te maken van de informatiebronnen en een architectuur van topniveau te definiëren naast een algemeen gegevensmodel.

Dit document behandelt de technische aspecten van de uitvoering van een tweede versie van het Data Warehouse voor het beheer van de risico's, de assistentie, de controle en de invordering. Deze versie bestrijkt het onderwerp "Onderneming" dat alle aspecten omvat die te maken hebben met de thema's Persoon, Aangifte-aspect, Specifieke Verwerking en Risico.

### 2.2 Beperkingen

De projecten "Data Warehouse/Business Intelligence" zijn projecten waarvan de uitvoering erg verschilt van de projecten voor transactionele softwareprogramma's (couranter binnen de FOD Financiën). Het grootste gevolg hiervan is het feit dat de FUP-methodologie gedeeltelijk van toepassing is op dit project.

Dit document tracht de FUP-methodologie zo goed mogelijk te respecteren. Er zullen evenwel bepaalde afwijkingen nodig zijn; deze worden verduidelijkt in de betreffende hoofdstukken.

### 2.3 Definities en afkortingen

Begrippen / Afkortingen	Verklaring
ODS	Operational Data Store
DWH	Data Warehouse (Opslagplaats voor gegevens)
DM	Data Mart (Magazijn met gegevens)
FTP	File Transfert Protocol
ETL	Extract-Transform-Load (Integratie van gegevens)
OLAP	On Line Analytical Processing
BI	Business Intelligence
SSRS	Microsoft SQL Server Reporting Services
SSAS	Microsoft SQL Server Analysis Services

# 3. Doelstellingen, vereisten en mechanismen van de architectuur

---

## 3.1 Naleving van de standaarden

<i>Mechanisme voor Design</i>	<i>Mechanismen Implementering</i>	<i>voor</i>	<i>Versie</i>	<i>Verantwoordelijke dienst</i>
Integratie van gegevens	IBM InfoSphere DataStage		8.1 FixPack 3	DCC
Opslag van gegevens	IBM DB2		9.5	DCC
Business Intelligence	Microsoft SQL Server		2008 R2	DCC
Scheduling	Absyss Visual TOM		-	OPS
Identity Authentication Management	IAM		-	IAM
Modellering van gegevens	Embarcadero ER/Studio Data Architect		>8.5	SupDev
Beheer van de code en de documentatie	Borland StarTeam		-	SupDev
Opvolging van problemen	HP Quality Center		-	SupDev

De volgende tools worden niet gebruikt omdat ze niet geschikt zijn voor dit type van project:

- Borland Caliber
  - Software voor inzameling van de functionele "requirements" van softwareprogramma's
  - Gericht op scenario's (gebruikscases), niet aangepast aan de integratie van gegevens of aan de productie van rapporten.
- Borland Together
  - Software voor modellering van softwaretoepassingen
  - Hoofdzakelijk gericht op de modellering van software via UML
  - Er zijn capaciteiten voor modellering van databases beschikbaar maar die zijn beperkt ten opzichte van Embarcadero ER/Studio. Met name:
    - Geen mogelijkheid om gegevenstransformaties in het model te documenteren (data lineage)
    - Logisch model met UML-weergave: minder duidelijk dan weergave met Entiteiten-Relaties, massaal gebruik van stereotypes, enz ...
    - Tal van productiviteitsfunctionaliteiten ontbreken ten opzichte van de oplossing van Embarcadero
- HP LoadRunner
  - Software om tests te automatiseren
  - Vooral geschikt om transactionele software te testen, de resultaten zouden niet voldoende representatief zijn voor dit project.
  - LoadRunner zou geen mogelijkheid bieden om de tests van de BI-tools volledig te automatiseren (geen commando-interface beschikbaar)



## 3.2 Ergonomie

In dit project wordt de ergonomie gedefinieerd door de gebruikte tools (IBM InfoSphere Information Server en Microsoft SQL Server) en kan ze niet worden gewijzigd.

## 3.3 Prestatievermogen

De oplossing moet grote gegevensvolumes kunnen verwerken in beperkte tijdvensters.

De impact op de bronsystemen met gegevens moet zo gering mogelijk zijn. Daartoe moeten de gegevens worden uitgewisseld via vlakke bestanden of via referentiedatabases (kopie van de productiesystemen).

De te behalen prestatiedoelen moeten worden gedefinieerd voor elke gegevensstroom. De aan te geven criteria zijn dus:

- **Voor de stromen voor gegevensintegratie**
  - Tijd voor verwerking en uploading van een bepaald volume van brongegevens. Dit volume moet zodanig worden gekozen dat het representatief is voor het gegevensvolume dat gewoonlijk zal worden verwerkt.
- **Voor de databases data warehouse en data mart**
  - Maximaal toelaatbare tijd om de voaf gedefinieerde query's uit te voeren. Deze query's moeten representatief zijn voor een karakteristiek gebruik van het data warehouse en alle data marts.
- **Voor het BI-gebruik**
  - Maximaal toelaatbare tijd om vooraf gedefinieerde BI-scenario's uit te voeren. Deze scenario's moeten representatief zijn voor een karakteristiek gebruik, bijvoorbeeld weergave van een specifiek rapport.

De prestatietests moeten gebeuren in belastingsomstandigheden die een optimaal verloop van de verwerkingsactiviteiten mogelijk maken (CPU, geheugen en I/O).

De metingen van de prestaties moeten tot stand komen op de meest geschikte wijze voor elk type van test, te weten:

- **Voor de stromen voor gegevensintegratie**
  - Gebruik van de operationele metagegevens (zie hoofdstuk **Erreur ! Source du renvoi introuvable.**: "Oogpunt gegevens" voor meer details)
- **Voor de databases data warehouse en data mart**
  - Gebruik van standaardtools van de DB2-clients of van het commando UNIX "time"
- **Voor het BI-gebruik**
  - Meting met chronometer

## 3.4 Capaciteit

De dimensionering moet naargelang de ontwikkelingen van de stromen kunnen worden aangepast in functie van de gegevensvolumes, het aantal gebruikers en de noodzakelijke prestaties.

De volgende tabel toont de grootteordes die in acht worden genomen voor de uitwerking van de architectuur en de ontwikkelingen.

Maat	Eenheid	Grootteorde
Gegevensvolume in Data Warehouse	Gb	~500

Gegevensvolume per data mart	Gb	~100
Totaal aantal gebruikers	Gebruikers	<50

### 3.5 Beschikbaarheid en betrouwbaarheid

Het data warehouse is ontworpen om een hoge beschikbaarheid te bieden, vooral tijdens de normale kantooruren. Het data warehouse wordt echter niet beschouwd als een kritiek systeem. Het is dus niet ontworpen om een permanente beschikbaarheid te garanderen (type 24/7).

De infrastructuur en de gebruikte softwareprogramma's zijn opgebouwd volgens de standaarden van de FOD Financiën (servers, netwerk, databases,...) en worden beschouwd als voldoende betrouwbaar.

De projectarchitectuur concentreert zich bijgevolg op het solide karakter van de gegevensverwerkingen en de capaciteiten voor heropstarten in geval van problemen.

Het solide karakter van de gegevensverwerkingen wordt met name verzekerd door de integratie in meerdere opeenvolgende stappen. Elk van die stappen vervult elementaire functies die van een toenemende complexiteit zijn (voorbereiding van de gegevens, opslag met historiek, transformatie naar DWH- en DM-gegevensmodellen). De verwerkingen die de minste risico's inhouden, worden prioritair uitgevoerd. Bij ernstige problemen die niet kunnen worden beheerd met de standaardmethodes van de DataStage-tool, kunnen de verwerkingen worden hervat vanaf de voorgaande stap, wat vermijdt dat men operaties moet hernemen die reeds succesvol waren uitgevoerd.

Om de diagnoses te vergemakkelijken, vullen bijkomende operationele metagegevens de metagegevens aan die standaard worden gegenereerd door DataStage. Er is dus eersterangsinformatie over de status van de gegevensstromen beschikbaar.

### 3.6 Veiligheid

Het platform dat in dit project wordt gebruikt, bestaat uit meerdere aparte componenten die alle een sterk verschillend gedrag vertonen ten aanzien van de veiligheid. Verderop in deze paragraaf maken we systematisch een onderscheid tussen de volgende componenten:

- Ontwikkelings- en onderhoudstools voor de integratie van gegevens (voorbeelden: DataStage Designer, DataStage Director, Information Server Console...). Er wordt hieronder naar verwezen als "Clients Information Server".
- Database van het data warehouse (IBM DB2). Er wordt hieronder naar verwezen als "DWH DB2" ci-dessous
- BI-objecten, toegankelijk via web (voorbeelden: Rapports Microsoft SQL Server Reporting Services Web Service - SSRS). Er wordt hieronder naar verwezen als "BI Web reports".
- BI-objecten, toegankelijk via fat clients (voorbeelden: Cubes Microsoft SQL Server Analysis Services - SSAS). Er wordt hieronder naar verwezen als "Other BI objects".

#### 3.6.1 Identificatie

De volgende tabel geeft een samenvatting van de identificatiemethodes ("login") die worden gehanteerd door elke component van de oplossing:

Component	Identificatie
Clients Information Server	Software clients
DWH DB2	Software clients
BI Web reports	IAM Access Manager
Other BI objects	Software clients

### 3.6.2 Authenticatie

De huidige structuur van de authenticatiediensten van de FOD Financiën en de gebruikte softwareprogramma's schrijven belangrijke eisen voor ten aanzien van de authenticatie.

Bijgevolg kan er een onderscheid worden gemaakt tussen twee categorieën van gebruikers:

- De "ETL-ontwikkelaars" en de technische gebruikers
  - Deze gebruikers hebben toegang tot de technische voorzieningen via een login/wachtwoord, specifiek bedoeld voor de UNIX/Solaris-platformen.
- De gebruikers die als consumenten van gegevens staan geboekt
  - Deze gebruikers hebben toegang tot de voorzieningen via de login/wachtwoord van de IAM-dienst of het Microsoft Windows-domein van de FOD Financiën (de gebruikers en de groepen, gedefinieerd in de Active Directory, worden bediend met de informatie van IAM).
  - Indien nodig kan de toegang tot de gegevens van het data warehouse tot stand komen via technische gebruikers, die op transparante wijze worden beheerd door de servers die diensten verstrekken aan de gebruikers.

De onderstaande tabel biedt een overzicht van de authenticatiemethodes die worden gehanteerd door elke component van de oplossing:

Component	Authenticatie
Clients Information Server	Operating system (LDAP/OPS)
DWH DB2	Operating system (LDAP/OPS)
BI Web reports	IAM Policy Manager
Other BI objects	Windows Active directory

### 3.6.3 Autorisatie

Om dezelfde redenen als voor de authenticatie spelen ook hier belangrijke eisen mee. De gebruikte softwareprogramma's bieden niet de mogelijkheid om de autorisaties subtiel te beheren via een aparte directory.

De volgende tabel biedt een overzicht van de autorisatiemethodes die worden gehanteerd door elke component van de oplossing:

Component	Autorisatie
Clients Information Server	Server "IBM Information Server"
DWH DB2	Database DB2
BI Web reports	IAM Access Manager
Other BI objects	Windows Active directory

### 3.6.4 Vertrouwelijkheid van de persoonsgegevens

Uit bezorgdheid om de bescherming van de persoonlijke levenssfeer is het noodzakelijk om de toegang tot gegevens, waarmee personen (natuurlijke of rechtspersonen) kunnen worden geïdentificeerd, te beperken.

Op het niveau van het data warehouse mogen alleen de geavanceerde veiligheidsfuncties van de

database IBM DB2 (LBAC) worden gebruikt om de toegang tot dit type van informatie te beperken.

Voor de data marts kunnen verschillende technieken worden gehanteerd in functie van de behoeften en de specifieke aspecten van de data mart: geavanceerde beveiliging (DB2/LBAC of gelijkaardige functies van SQL Server), anonimisering door samenvoeging van gegevens, afwezigheid van naamgegevens, codering of opsplitsing van naamgegevens,...

### 3.6.5 Audit

De acties van de gebruikers ten aanzien van vertrouwelijke gegevens moeten worden opgeslagen in de loggingservice van IAM. Er is gedetailleerde informatie over deze service beschikbaar in een afzonderlijk document. Zie paragraaf 1.2 voor de referentie naar deze documentatie.

De volgende tabel biedt een overzicht van de opgetekende acties en de techniek voor inzameling van de informatie.

Component	Geauditeerde actie(s)	Inzameling van informatie
Information Server	Toegang tot de gegevens	Via DWH DB2
DWH DB2	Toegang tot de gegevens	Via auditfuncties van DB2 of triggers op belangrijke tabellen (te bepalen) <sup>1</sup>
BI Web reports	Toegang tot de rapporten	Via logging van SSRS <sup>2</sup>
Other BI objects	Toegang tot de rapporten en kubussen	Via logging van SSRS en tracing van SSAS <sup>2</sup>

<sup>1</sup>: In het geval van toegang tot de gegevens via de SQL Server-services, is het niet mogelijk om de oorspronkelijke gebruiker te kennen. Alleen de technische gebruiker, gebruikt door SQL Server, zal kunnen worden geïdentificeerd.

<sup>2</sup>: De oorspronkelijke gebruiker zal in dit geval worden geïdentificeerd.

## 3.7 Beheer

Het beheer van het data warehouse hangt voornamelijk af van de gebruikte tools (IBM InformationServer, IBM DB2, MS SQL Server). De handleidingen voor het beheer van deze tools zijn dus de belangrijkste referenties:

Voor IBM InfoSphere Information Server 8.1 :

<http://publib.boulder.ibm.com/infocenter/iisinfsv/v8r1/index.jsp>

Voor IBM DB2 9.5 :

<http://publib.boulder.ibm.com/infocenter/db2luw/v9r5/index.jsp>

Voor Microsoft SQL Server 2008 :

<http://technet.microsoft.com/en-us/library/bb418439%28SQL.10%29.aspx>

De procedures voor het operationele beheer zullen worden gedetailleerd in een apart document, beschikbaar onder StarTeam. Zie paragraaf 1.2 voor de referentie naar deze documentatie.

## 4. View gebruikscase

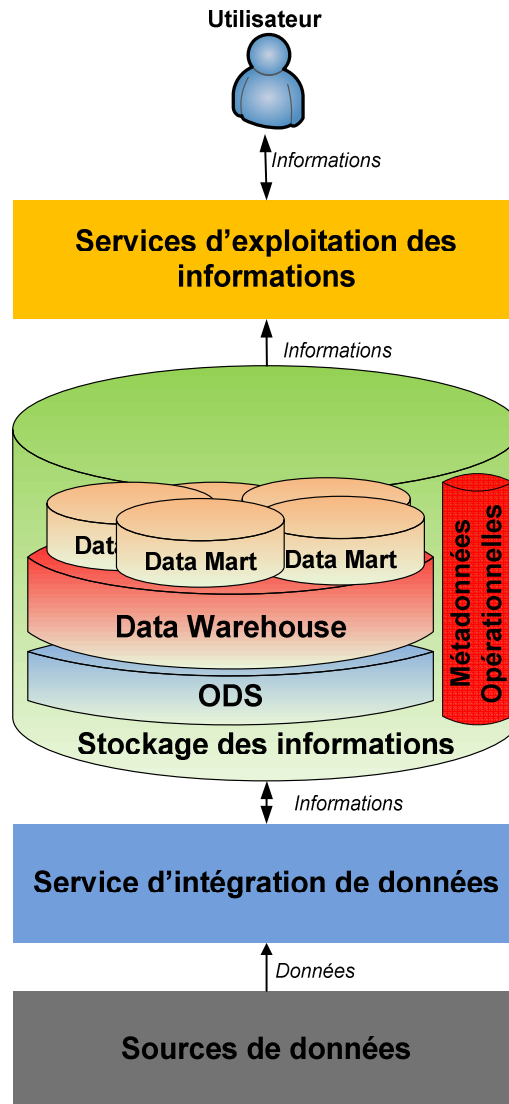
---

De gebruikscases zijn niet van toepassing in het kader van dit project.

# 5. Logische view

## 5.1 Logische architectuur

Het systeem kan op hoog niveau worden weergegeven aan de hand van volgend schema.



### Gegevensbronnen

De gegevensbronnen vertegenwoordigen alle systemen buiten het project die het project voorzien van "ruwe gegevens". Deze bronnen worden beschouwd als "zwarte dozen" en alleen hun interfaces met de service voor gegevensintegratie zijn gekend door het project.

Deze interfaces zullen voor elke bron worden gedetailleerd in het document met de beschrijving van de bronnen, beschikbaar onder StarTeam (zie paragraaf 1.2 voor de volledige referentie). Er zijn verschillende types mogelijk, bijvoorbeeld vlakke bestanden via FTP, databasetabellen, enz ...

De gegevensstromen gaan in een enkele richting, vanuit de gegevensbronnen naar de service voor gegevensintegratie. Alleen de leestoegang tot de bronnen is bijgevolg toegestaan.

### ***Service voor gegevensintegratie***

Deze service heeft als opdracht de gegevens uit de bronnen te distilleren, te transformeren en te uploaden (ETL) in de vorm van geconsolideerde informatie, bestemd voor de gebruikers.

### ***Opslag van de informatie***

De informatie, klaargemaakt door de service voor gegevensintegratie, wordt op dit niveau opgeslagen.

De gegevens worden er georganiseerd in drie opeenvolgende lagen:

De **ODS** (Operational Data Store) vormt een technische laag – niet toegankelijk voor de gebruikers – die de kans biedt om alle gegevens van de bronnen op te slaan en er historische informatie aan toe te voegen.

Het **data warehouse** slaat de geconsolideerde informatie op.

De **data marts** slaan de informatie op die is afgeleid van het data warehouse en werd klaargemaakt voor precieze behoeften.

De gegevensstroom met de service voor gegevensintegratie kan in twee richtingen verlopen omdat de opslag wordt georganiseerd in meerdere interne lagen.

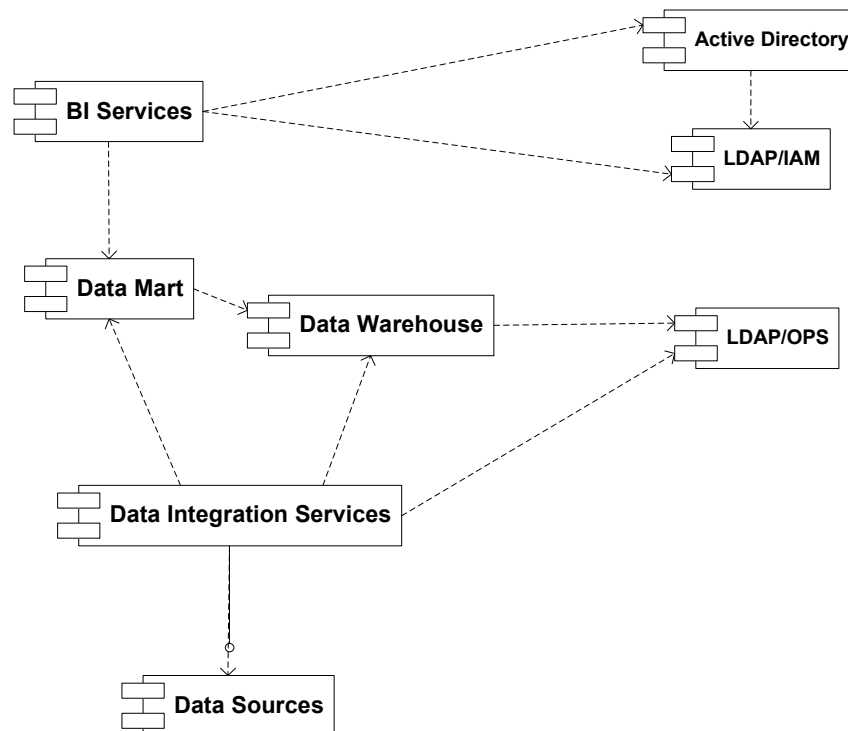
De zone van de **operationele metagegevens** slaat informatie op met betrekking tot de opvolging van de gegevensverwerkingen en de kwaliteit van de gegevens.

### ***Services voor gebruik van de informatie***

Deze services bieden de gebruikers de mogelijkheid om de gegevens op intuïtieve wijze te benutten, zonder noodzakelijk alle onderliggende technische details te kennen. Voorbeelden van services: Rapportering, Analyse ("OLAP"), Data Mining (apart project los van dit project, maar client van de gegevens), enz...

## 6. View implementering

### 6.1 Globaal overzicht – UML-diagram van de Componenten



De service voor gegevensintegratie verwerkt de gegevens die afkomstig zijn van de bronnen via zijn interface. Deze service wordt verzekerd door een specifieke tool (IBM Information Server – DataStage), de interne structuur staat bijgevolg los van dit project.

De verwerkte en geconsolideerde gegevens worden opgeslagen in het data warehouse volgens een relationele en genormaliseerde structuur.

Een deel van de gegevens wordt vervolgens afgeleid van de gegevens van het data warehouse door de service voor gegevensintegratie met het oog op hun opslag in de data marts. De opslagstructuren worden aangepast aan het finale gebruik van de data mart en weerspiegelen over het algemeen specifieke businessaspecten van de consumenten van de gegevens.

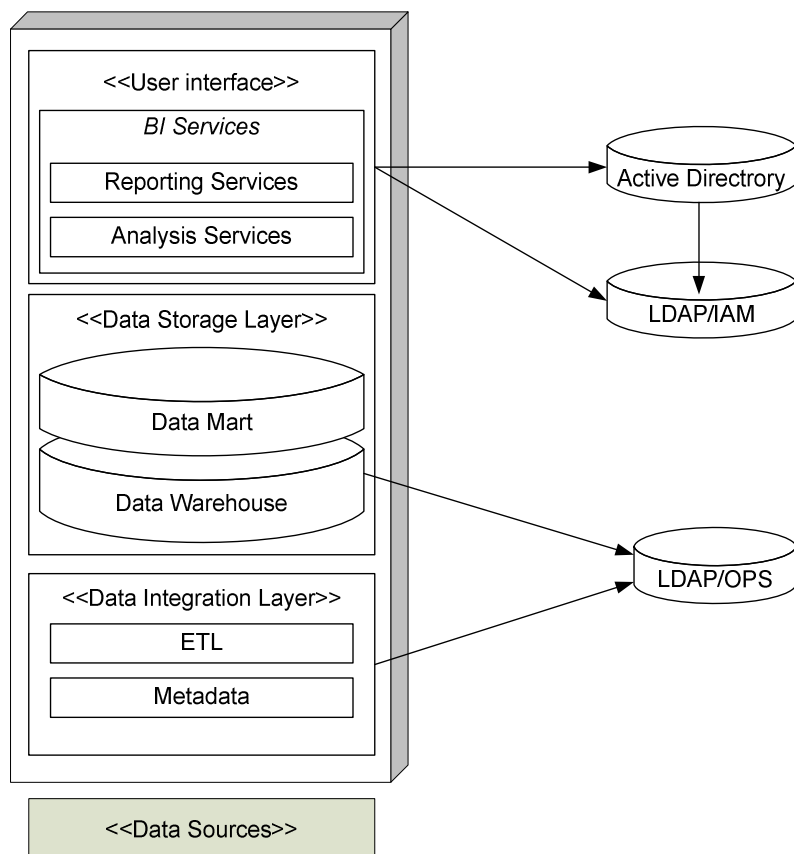
De gebruikers hebben toegang tot de gegevens met behulp van BI-services. Deze services kunnen toegankelijk zijn via fat clients of via het web. Deze services worden verzekerd door een specifieke tool (Microsoft SQL Server), de interne structuur staat bijgevolg los van dit project.

De authenticatie van de gebruikers van de service voor gegevensintegratie (ontwikkelaars van ETL-stromen) en van de gebruikers van het data warehouse wordt verzekerd door het besturingssysteem (Solaris) dat zich baseert op een eigen directory (LDAP/OPS).

De authenticatie van de gebruikers van BI-services gebeurt via een "active directory" van Windows als er een fat client wordt gebruikt en via IAM wanneer de toegang tot stand komt via het web. De "active directory" wordt gesynchroniseerd met de LDAP-directory die is gekoppeld aan de IAM-service.



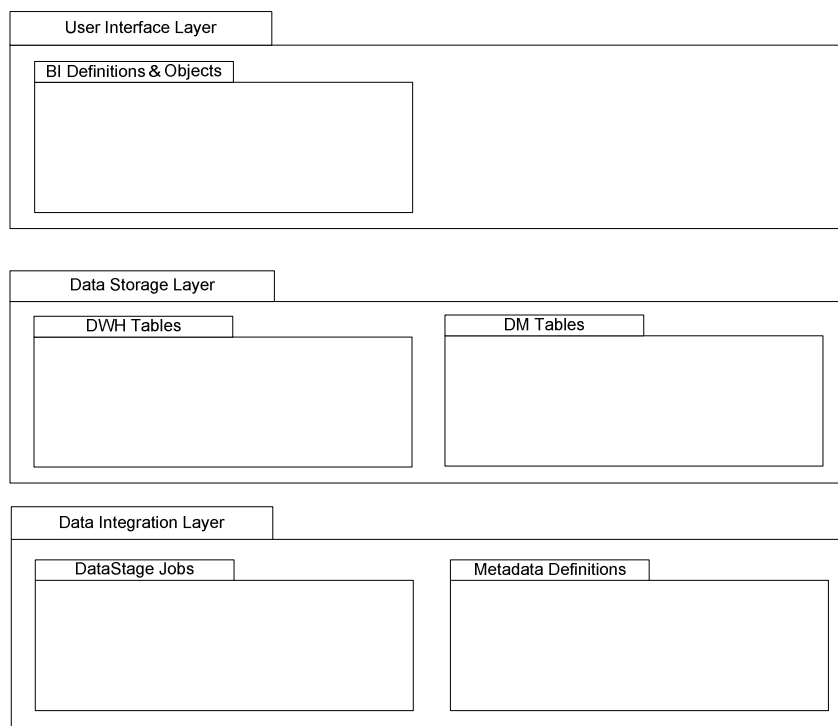
## 6.2 Lagen – UML-diagram van de Componenten



De DWH2-toepassing kan worden opgedeeld in drie hoofdlagen:

- **De laag "Integratie van gegevens"**
  - Functies voor verwerking van gegevens (ETL)
  - Metagegevens, gerelateerd met deze verwerkingen
- **De laag "Opslag van de gegevens"**
  - Data warehouse
  - Data marts
- **De laag "Gebruikersinterface"**
  - Services voor rapportering
  - Service voor gegevensanalyse
  - Andere services voor de benutting van de gegevens

## 6.3 Structurering in packages



### ***Data Integration Layer***

Bevat de objecten die de integratie van de gegevens in het data warehouse en de data marts mogelijk maken.

#### **DataStage Jobs**

Bevat alle DataStage-processen voor de verwerking van de gegevens

#### **Metadata Definitions**

Bevat de technische en businessdefinities van de gegevens van het data warehouse en de data marts

### ***Data Storage Layer***

Bevat de objecten waarmee de gegevens kunnen worden opgeslagen.

#### **DWH Tables**

Bevat de tabellen van het data warehouse

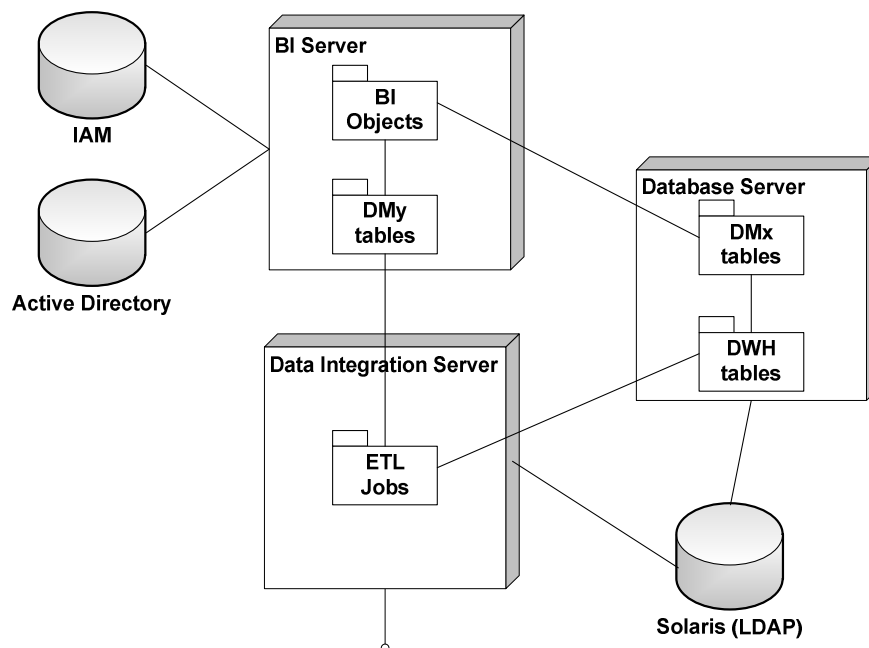
#### **DM Tables**

Bevat de tabellen van de data marts

### ***User interface Layer***

Bevat de objecten waarmee de gegevens kunnen worden benut.

## 6.4 Packaging / Structurering in eenheden voor ingebruikneming



De packaging zal bestaan uit verschillende elementen die in gebruik worden genomen op elke node van het systeem.

### **Database data warehouse**

Het in gebruik te nemen package bestaat uit een tekstbestand dat de beschrijving bevat van de tabellen en de andere objecten van de database (DDL-bestand).

### **Database data mart**

Het in gebruik te nemen package bestaat eveneens uit een tekstbestand dat de beschrijving bevat van de tabellen en de andere objecten van de database (DDL-bestand).

De data marts moeten normaliter worden opgeslagen op het niveau van de databaseserver van het warehouse. Ze mogen echter eventueel ook worden opgeslagen op het niveau van de BI-server om een optimale integratie te garanderen met de services van het BI-platform. In dit geval moet men garanderen dat er geen enkele geavanceerde veiligheidsfunctie noodzakelijk is op het niveau van de BI-server en dat de administratie van de gegevens op een minimaal niveau blijft.

In het geval van de kubussen wordt een gedeelte van de gegevens altijd opgeslagen op het niveau van de BI-server (structuur en gebundelde of vooraf berekende gegevens).

De keuze van de positie van elke data mart zal worden verduidelijkt en verantwoord in de documenten met betrekking tot het gegevensmodel van elke release (zie paragraaf 1.2 voor de referenties).

### **Server voor integratie van gegevens**

Het in gebruik te nemen package is een standaardpackage van de tool DataStage dat de definitie van de tabellen van het data warehouse en de data marts bevat, naast de beschrijving van de processen die de gegevens verwerken. Deze packages worden geproduceerd en gebruikt via de functies voor export en import van de tool DataStage Designer.

### **BI-server**

Het in gebruik te nemen package is een standaardpackage van de tool SQL Server dat de definitie van de tabellen van de data marts bevat naast de beschrijving van de BI-objecten (rapporten, kubussen...). Deze packages worden geproduceerd en gebruikt via de functies voor export en import van de tool.

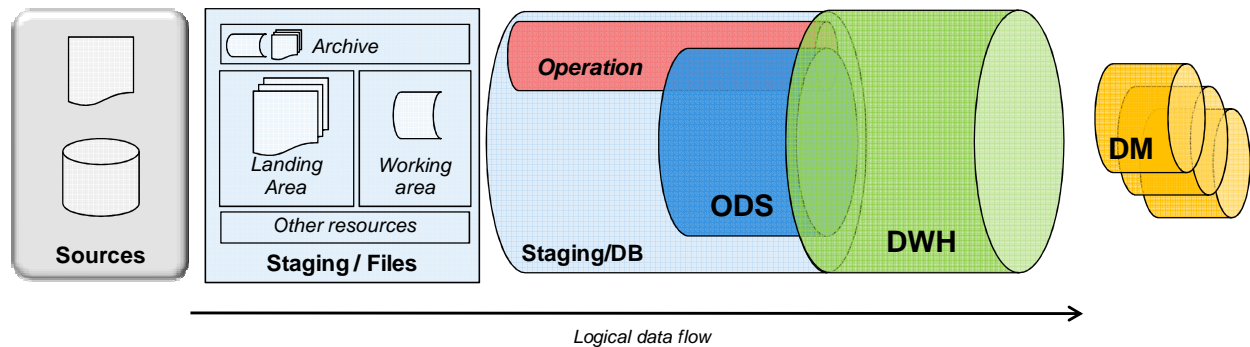
## **6.5 Realisaties in structurerende gebruikscases – UML-diagrammen van Categorieën en Interacties (verplicht)**

Niet van toepassing op dit project.

# 7. View gegevens

## 7.1 Logische gegevensarchitectuur

Het volgende schema toont de organisatie van de gegevens.



Er worden vier logische gegevenszones gedefinieerd:

### **Gegevensbronnen**

De gegevensbronnen zijn alle gegevens buiten het DWH2-project. Dit omvat alle interne en externe gegevensbronnen van de FOD Financiën, met inbegrip van de "FTP-zone".

### **Tussenvliegende gegevens (Staging/Files en Staging/DB)**

De tussenvliegende gegevens zijn de technische (en tijdelijke) gegevens en metagegevens die noodzakelijk zijn voor de verwerkingen van de gegevens.

Deze zone bestaat uit een bestandssysteem en een database, en omvat verschillende gedeelten:

- De landingszone (**landing area**) is de buffer waar de gegevens binnenkomen in de vorm van te verwerken bestanden.
- De zone met de werkbestanden (**working area**) slaat de tijdelijke bestanden op die nodig zijn om de verwerkingen uit te voeren.
- De zone voor **archivering** slaat gegevens op die noodzakelijk kunnen zijn voor onderhouds- en herstelbehoeften.
- De tussenvliegende databasezone (**staging database area**) slaat tijdelijke gegevens op voor de verwerkingen binnenin de database, naast blijvende technische gegevens.
- De ODS (**Operational Data Store**) slaat de brongegevens op door ze historisch in een formaat onder te brengen dat de latere verwerkingen moet vergemakkelijken, met name bij problemen.
- De zone met de operationele metagegevens (**operational metadata area**) wordt gebruikt om metagegevens op te slaan met betrekking tot de verwerking van de gegevens die niet standaard worden voorgesteld door het platform voor gegevensintegratie (IBM Information Server)

### **Data warehouse**

De DWH-zone is de belangrijkste opslagzone. Hier worden de geconsolideerde gegevens opgeslagen.

### **Data marts**

De data marts zijn bestemd voor de opslag van gegevens die werden klaargemaakt voor specifiek

gebruik, zoals rapportering en OLAP-analyses.

## 7.2 Technische organisatie van de gegevens

### 7.2.1 Bronsystemen

De precieze beschrijvingen van elk bronsysteem zullen worden gedetailleerd in specifieke documenten, beschikbaar onder StarTeam. Zie paragraaf 1.2 voor de referentie naar deze documentatie.

De volgende paragrafen verstrekken evenwel enkele aanbevelingen om een betrouwbare integratie van de gegevens te garanderen.

#### 7.2.1.1 Voorkeurformaten

Technisch gezien zijn er uiteenlopende formaten van brongegevens mogelijk. Sommige formaten leveren echter betere resultaten op inzake betrouwbaarheid en prestatieniveau.

Bij wijze van informatie: de volgende tabel toont de hiërarchie van de formaten in volgorde van voorkeur. Het is niet de bedoeling van deze tabel om wijzigingen binnen een bronsysteem op te leggen die tot het voorkeurformaat leiden, maar wel om het best beschikbare formaat te kiezen zonder zware aanpassingen.

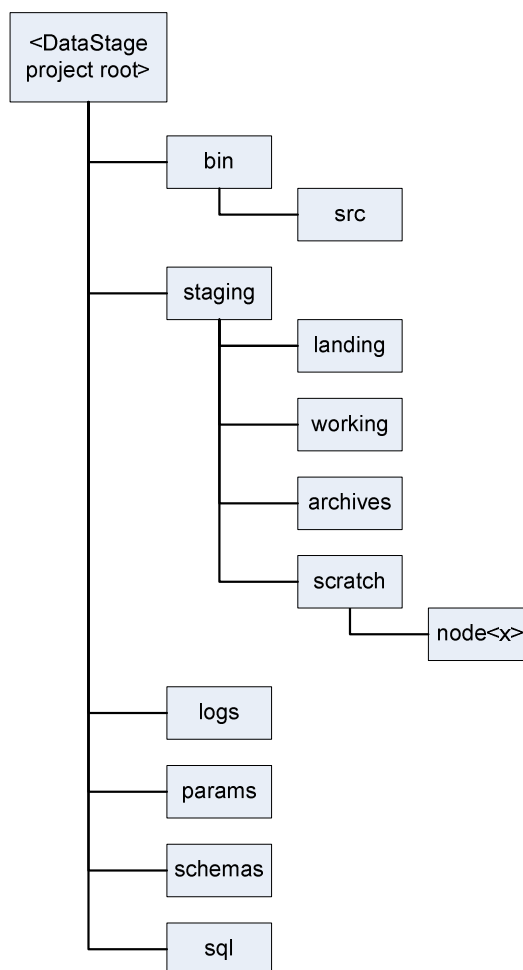
Voorkeur	Type bron	Formaat/Versie	Opmerking
1	MQ-stroom	CDC-stroom (Change Data Capture)	Aanbevolen voor integratie in real time.
2	Database	Views of tabellen DB2 9.5	De views zijn aanbevolen (indien mogelijk)
3		Views of tabellen DB2 9.7	
4		Views of tabellen m.b.t. andere databases, toegankelijk via DataStage	
5	Vlakke bestanden	Vaste velden (fixed width)	
6		Afgebakende velden	Risico om het begrenzingsteken terug te vinden in de inhoud van de velden
7		Andere complexe formaten	

Binaire bestanden zoals Microsoft Excel-bestanden worden niet ondersteund.

In alle gevallen moet het interface-formaat het voorwerp uitmaken van een contract tussen de bron en de ETL, dit om de stabiliteit van de services voor gegevensintegratie te garanderen.

### 7.2.2 Bestandssysteem

Het volgende schema toont de organisatie van het bestandssysteem binnen een DataStage-project.



- *<DS project root>* is de root van het DataStage-project, bijvoorbeeld *"/dsdata/dwmdsadd/DWT\_Dev"* in de ontwikkelingsomgeving.
- *<x>* is de naam van de parallelle verwerkings-node in DataStage. Bijvoorbeeld 1, 2, 3...

De logische zones met gegevens van het type "bestanden" zijn bijgevolg als volgt verdeeld:

Logische zone	Logische beschrijving	Fysieke weg
Staging	Landing area	<i>&lt;DS Project root&gt;/staging/landing</i>
	Working area ( <i>DataStage resources disks</i> ) ( <i>DataStage scratch disks</i> )	<i>&lt;DS Project root&gt;/staging/working</i> <i>&lt;DS Project root&gt;/staging/scratch</i>
	Archive	<i>&lt;DS Project root&gt;/staging/archives</i>

De volgende tabel toont de verdeling van de andere voorzieningen, specifiek voor de gegevensverwerking:

Logische beschrijving	Fysieke weg
DataStage Schema files	<i>&lt;DS Project root&gt;/schemas</i>
DataStage parameters and configuration	<i>&lt;DS Project root&gt;/params</i>

Sql scripts	<DS Project root>/sql
Log files	<DS Project root>/logs
Tools binaries	<DS Project root>/bin
Tools sources	<DS Project root>/bin/src

## 7.2.3 Databases

### 7.2.3.1 Tussenliggende gegevens en data warehouse

De tussenliggende gegevens en de geconsolideerde gegevens van het data warehouse worden opgeslagen in een IBM DB2-database. Deze database wordt DWT\_<x> genoemd, waarbij <x> staat voor het type van omgeving. De mogelijke waarden zijn:

- "D" voor ontwikkelingsomgeving
- "A" voor omgeving voor integratie en acceptatie
- "P" voor productieomgeving

De logische gegevenszones zijn verdeeld in schema's van de database, zoals aangegeven in onderstaande tabel.

Logische zone	Logische beschrijving	Schema
Staging	DB Working area	STGWRK
	Operational Data Store	ODS
Data Warehouse	Données consolidées	DWH

### 7.2.3.2 Data marts

De data marts zijn bedoeld om gegevens te verschaffen in formaten die zijn aangepast aan het uiteindelijke gebruik dat er door de gebruikers van zal worden gemaakt. Een deel van de problemen met de gegevensstructuur, het prestatievermogen en de veiligheid kan op dit niveau worden opgelost, bijvoorbeeld door de historiek te beperken, of door de gegevens samen te voegen en/of te denormaliseren (schema in stervorm).

De exacte organisatie van elke data mart zal worden gedetailleerd in specifieke documenten, beschikbaar onder StarTeam. Zie paragraaf 1.2 voor de referentie naar deze documentatie.

### 7.2.3.3 Operationele metagegevens

De operationele metagegevens worden opgeslagen op twee niveaus.

De "**native**" operationele metagegevens worden standaard beheerd door het IBM Information Server-platform en geëxploiteerd via de tools DataStage Director en Metadata Workbench. Ze verschaffen nauwkeurige inlichtingen op het niveau van de verwerkingen van gegevens (Jobs en DataStage-sequenties).

Deze metagegevens en de tools worden dus niet gedetailleerd in dit document.

De **bijkomende metagegevens** geven informatie over de operationele status van de verwerkingen tussen de bronnen, de ODS, het data warehouse en de data marts op het niveau van de gegevensstromen. De informatie m.b.t. de gegevens van de ODS waarvoor problemen werden vastgesteld, wordt eveneens opgeslagen in deze zone.

De opslagdetails worden vermeld in de volgende tabel:



Logische zone	Logische beschrijving	Schema
Operation	Métadonnées opérationnelles	OPMETA

## 7.3 E/R-diagram van de Blijvende Gegevens

De E/R-diagrammen van de gegevens worden gedetailleerd in afzonderlijke documenten, beschikbaar onder StarTeam. Zie paragraaf 1.2 voor de referentie naar deze documentatie.

De volgende paragrafen belichten de richtinggevende principes die worden toegepast voor de uitwerking van E/R-diagrammen.

### 7.3.1 ODS

#### 7.3.1.1 Regels voor modellering

Het model met de gegevens van de ODS moet praktisch identiek zijn met de modellen met de brongegevens. De verschillen berusten in de historische schikking van de aanpassingen van de gegevens van de bronsystemen. Elke unieke registratie in een bronsysteem kan dus meerdere keren worden opgeslagen in de ODS, in functie van het aantal aanpassingen dat werd vastgelegd of ontdekt door het systeem voor gegevensintegratie.

De originele unieke sleutel moet dus worden uitgebreid tot een veld dat de bijbehorende datum van aanpassing geeft.

Er moet een bijkomende veld worden voorzien om het type van de aanpassing te vermelden.

Het volgende voorbeeld illustreert de volledige levenscyclus van een unieke registratie in een bronsysteem.

Beschrijving van de aanpassing	Code van het aanpassingstype	Datum
Oorspronkelijke invoeging	I	$T$
Eerste aanpassing	U	$t+1$
Tweede aanpassing	U	$t+2$
Verplichte vernieuwing van de gegevens <sup>1</sup>	R	$t+3$
Xde aanpassing	U	$t+n$
Verwijdering	D	$t+n+1$

<sup>1</sup> : De gegevensvernieuwing kan worden gehanteerd in het geval van problemen met synchronisatie met de bron.

Met het oog op een efficiënte opslag en een efficiënt prestatieniveau mag het fysieke gegevensmodel geen eisen bevatten (zoals "NOT NULL"), met uitzondering van eisen met betrekking tot de primaire sleutels. Het is de bedoeling een zo getrouw mogelijk beeld van de gegevensbronnen te kunnen opslaan, met inbegrip van mogelijke inbreuken op eisen.

#### 7.3.1.2 Technische velden

Elke tabel van de ODS moet bovendien naast de velden van de brontabel ook de volgende technische velden bevatten:

- **O\_I\_IDF** (Bigint): interne referentie van de registratie in de ODS.
- **S\_I\_ODS\_INS** (Timestamp): datum en uur van de aanmaak van de registratie in de ODS.

- **C\_I\_MOD\_TYPE** (char(1)) : code van het type van de aanpassing. I = Insert, U = Update, D = Delete, R = Refresh (verplichte vernieuwing van alle gegevens).
- **O\_INS\_FLOW\_IDF** (Bigint): referentie naar de stroom die de registratie heeft geüpload.

## 7.3.2 Data warehouse

### 7.3.2.1 Regels voor modellering

Het gegevensmodel van het data warehouse moet relationeel zijn en moet minstens genormaliseerd zijn tot de derde normale vorm.

#### **Referentiële integriteit**

De referentiële integriteit tussen de gegevenseenheden moet worden verzekerd door technische sleutels (surrogate keys) en niet door businesssleutels (natural keys). Deze regel wordt over het algemeen beschouwd als een goede praktijkregel om de volgende redenen:

- Beperking van het volume van de (primaire en vreemde) sleutelgegevens en de indexen.
- Optimalisering van de prestaties van de verbindingen tussen tabellen.
- Grotere onafhankelijkheid en stabiliteit ten aanzien van de businesssleutels (geen risico's op conflicten en accidentele denormalisering, geringe impact bij wijziging van de businesssleutels ...)

De tabellen die codes beschrijven, kunnen het stellen met businesssleutels en moeten dus geen technische sleutels bevatten.

#### **Beschrijvende gegevens**

DWH 2 moet tijdelijk onafhankelijk zijn van de externe beschrijvende gegevens, zoals DWH 1. Problemen met het uploaden van zijn beschrijvende gegevens mogen dus geen impact hebben op de werking van DWH2.

Om te voldoen aan deze behoefte, bevat het gegevensmodel van het data warehouse een vereenvoudigde vorm van beschrijvende gegevens («minibesrijving»). Deze slaat enkel de beschrijvende gegevens op die noodzakelijk zijn voor de werking van het data warehouse 2.

#### **Beveiliging**

De DWH2-database moet ook in staat zijn om de identificatie van mensen te verbieden om redenen van vertrouwelijkheid. Deze behoefte wordt uitsluitend gedekt door middel van de geavanceerde beveiligingsfuncties van de DB2-database (LBAC-functies). De gegevens waarmee een persoon of een onderneming kan worden geïdentificeerd, kunnen dus onmogelijk worden gebruikt door niet-gerechtigde gebruikers. Voorbeeld: nationaal nummer, ondernemingsnummer, ...

Gevorderde gebruikers die toegang hebben tot het data warehouse kunnen gebruik maken van de "surrogate keys" (statische technische sleutels) als unieke logins.

### 7.3.2.2 Technische velden

Elke tabel van het data warehouse moet minstens de volgende technische velden bevatten:

- **O\_INS\_FLOW\_IDF** (Bigint): referentie naar de stroom die de registratie creëert.
- **S\_I\_INS** (Timestamp): datum en uur van de oorspronkelijke creatie van de registratie.
- **O\_UPD\_FLOW\_IDF** (Bigint): referentie naar de laatste stroom die de registratie heeft gewijzigd.
- **S\_I\_UPD** (Timestamp): datum en uur van de laatste aanpassing van de registratie.

### 7.3.3 Data marts

#### 7.3.3.1 Regels voor modellering

De volgende regels moeten de uitwerking van de data marts in goede banen leiden.

- De gegevens van de data marts moeten normaliter worden opgeslagen in DB2-databases. De gegevens echter die zijn bestemd om te worden gebruikt door de BI Microsoft-services, mogen (gedeeltelijk voor de kubussen) eventueel worden opgeslagen op het niveau van de Microsoft SQL Server-server indien dit de mogelijkheid biedt om een optimale integratie van de functionaliteiten te garanderen. In dit geval moet men ook garanderen dat de SQL Server-database slechts een minieme administratie vereist, ook vanuit het oogpunt van de beveiliging. Geavanceerde beveiligingsfunctionaliteiten mogen enkel door DB2/LBAC worden verzekerd.
- De gegevens van de data marts worden altijd afgeleid van de referentiegegevens van het data warehouse. Nieuwe gegevens worden nooit rechtstreeks ingevoerd in een data mart. Op die manier kan een data mart in het geval van een catastrofe altijd worden gereconstrueerd op basis van het data warehouse.
- De data marts, bedoeld voor de rapportering, zijn normaal gezien gemodelleerd in stervorm om redenen van analytische prestaties (tabellen met feiten en dimensies). Genormaliseerde relationele modellen en modellen in de vorm van "sneeuwvlokken" (snowflake) worden afgeraden.

#### 7.3.3.2 Technische velden

Elke tabel van de data marts zou minstens de volgende technische velden moeten bevatten:

- **O\_INS\_FLOW\_IDF** (Bigint): referentie naar de stroom die de registratie heeft gecreëerd.
- **S\_I\_INS** (Timestamp): datum en uur van de oorspronkelijke creatie van de registratie.
- **O\_UPD\_FLOW\_IDF** (Bigint): referentie naar de laatste stroom die de registratie heeft gewijzigd.
- **S\_I\_UPD** (Timestamp): datum en uur van de laatste aanpassing van de registratie

### 7.3.4 Operationele metagegevens

#### 7.3.4.1 Operationele status van de stromen

De operationele status wordt opgeslagen in de tabel **OPMETA.FLOW\_STATUS**. Deze tabel bevat ook de statistieken van elke gegevensstroom (aantal gedistilleerde, verworpen, getransformeerde, enz. gegevens).

De informatie over de gegevens van de ODS die fouten bevatten, wordt opgeslagen in de tabel **OPMETA.ODS\_DATA\_ERROR**. Deze tabel bevat verwijzingen naar de registraties in de ODS, de gegevensstroom die de registratie heeft verworpen en de reden voor de verwerping. De redenen voor verwerping zijn vooral businessredenen, zoals beschreven in de documentatie van elke gegevensstroom.

De gedetailleerde beschrijving van deze tabellen wordt meegegeven in een afzonderlijk document, beschikbaar onder StarTeam. Zie paragraaf 1.2 voor de referentie naar deze documentatie.

## 7.4 Back-ups

Het gehele systeem moet na een catastrofe kunnen worden hersteld. De volgende paragrafen

beschrijven de elementen waarvan een back-up moet worden gemaakt.

#### 7.4.1 Service voor integratie van gegevens

Er moet een back-up worden gemaakt van het platform IBM InfoSphere Information Server volgens de beheerregels, beschreven in de documentatie van de software ("IBM Information Server 8.1 : Administration Guide").

Een minimale periode van 15 dagen voor de bewaring van de gegevens is aanbevolen om de langste periodes zonder operationele support te kunnen overbruggen (lange weekends, bruggen, enz...).

#### 7.4.2 BI-service

Er moet een back-up worden gemaakt van het platform Microsoft SQL Server volgens de beheerregels, beschreven in de documentatie van de software.

Een minimale periode van 15 dagen voor de bewaring van de gegevens is aanbevolen om de langste periodes zonder operationele support te kunnen overbruggen (lange weekends, bruggen, enz...).

#### 7.4.3 Bestandssysteem

Er moet een back-up worden gemaakt van de volgende directory's:

Directory	Beschrijving	Aanbevolen minimale frequentie	Aanbevolen minimale periode van bewaring
<DS Project root>/staging/archives	Archive	Alle dagen	60 dagen
<DS Project root>/schemas	DataStage Schema files	Bij elke release	1 release
<DS Project root>/params	DataStage parameters and configuration	Bij elke release	1 release
<DS Project root>/sql	Sql scripts	Bij elke release	1 release
<DS Project root>/logs	Log files	Bij elke release	1 release
<DS Project root>/bin	Tools binaries	Bij elke release	1 release

#### 7.4.4 Databases

Er moet een back-up worden gemaakt van de volgende schema's:

Schema	Beschrijving	Aanbevolen minimale frequentie	Aanbevolen minimale periode van bewaring
OPMETA	Operationele metagegevens	Alle dagen	15 dagen
STGWRK	Working area (db)	Alle dagen	15 dagen
ODS	Operational Data Store	Alle dagen	15 dagen
DWH	Data Warehouse	Alle dagen	15 dagen
Data marts	Data Marts	Bij elke release	1 release

De data marts kunnen op eender welk ogenblik worden hersteld op basis van de gegevens van het data warehouse. Alleen hun structuur moet dus worden bewaard.

Het data warehouse zou eventueel kunnen worden hersteld op basis van de gegevens van de ODS. Dit zou echter heel veel verwerkingstijd kunnen vergen en het platform lange tijd onbeschikbaar kunnen maken. Er wordt dus aanbevolen minstens alle dagen een back-up te maken van de gegevens van het

data warehouse.

## 8. View processen

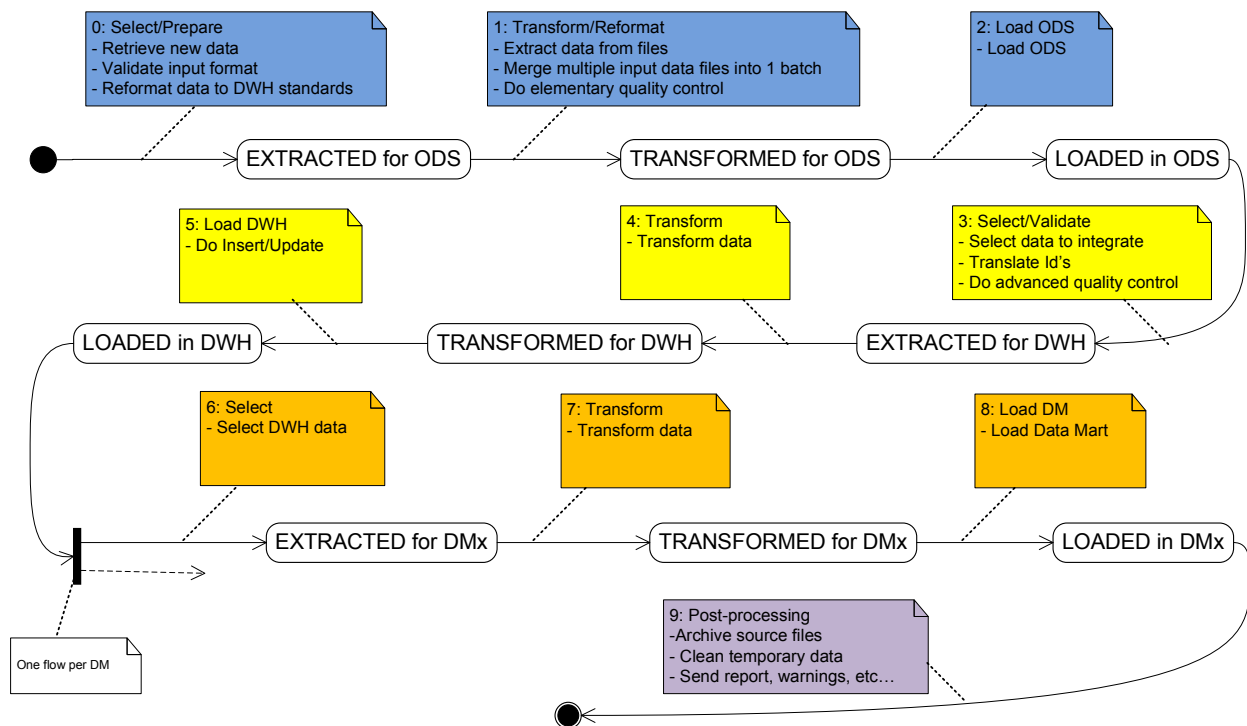
### 8.1 Standaardstroom

Dit gedeelte beschrijft de standaardverwerkingen van de gegevens met het oog op hun integratie in het data warehouse en de data marts. Het is hierbij de bedoeling de coherentie van de verwerkingen voor gegevensintegratie te bevorderen, dit om het onderhoud te vergemakkelijken.

Deze standaard vormt geen absolute regel. Sommige gegevens kunnen verschillende verwerkingen vergen, die in tegenspraak zijn met deze standaard. In dit geval moet men vooral voorrang geven aan de helderheid en het onderhoudsgemak van deze verwerkingen.

Het volgende diagram met statussen-transities geeft een beschrijving op hoog niveau van de standaardstroom voor het uploaden van het data warehouse en de data marts.

De beschreven statussen (EXTRACTED-TRANSFORMED-LOADED) zijn uitsluitend logische statussen. Als de verwerkingen weinig complex zijn, kunnen er meerdere transitie tussen deze statussen worden gebundeld in een enkele DataStage-job om een hoger prestatieniveau te garanderen.



De drie belangrijkste stappen zijn de ODS, het data warehouse en de data marts. Deze stappen vormen "haltes", d.w.z. dat, indien bepaalde verwerkingen mislukken of als gegevens gecorrumpereerd zijn, het steeds mogelijk is om de verwerkingen opnieuw op te starten op basis van de vorige stap.

Hoewel de verwerkingen in deze stappen redelijk verschillend zijn, volgt de stroom altijd de logica Extractie – Transformatie - Uploading.

## **8.1.1 Stap 0: Voorbereiding van de gegevens - Selectie**

### **8.1.1.1 Beschrijving**

Deze stap heeft tot doel de complexiteit van de gegevensbronnen op het vlak van codering en formaat weg te houden van de latere verwerkingen.

### **8.1.1.2 Inputs**

Bronbestanden of database.

### **8.1.1.3 Logische verwerkingen**

1. De stroom in de operationele metagegevens initialiseren met de status "**STARTED**"
2. *Voor bronnen van het type "bestand"*: de nieuwe bestanden opsporen in de FTP-zone (of andere bestandssystemen)  
*Voor bronnen van het type "database"*: de (nieuwe) gegevens distilleren
3. In bestanden de gegevens verwerpen die technisch slecht zijn geformatteerd en/of onleesbaar zijn (bijvoorbeeld : ongeldig datumformaat in een tekstbestand, buitensporige lengte van een tekstveld,...)
4. Indien nodig de verschillen (delta) berekenen ten opzichte van de laatst gekende situatie
5. Bij een complex bestand de gegevens opnieuw formatteren in het normale inputformaat van het data warehouse en deze gegevens in de "landing area" schrijven.
6. Indien mogelijk de informatie van de nieuwe gegevens registreren in de operationele metagegevens met de status "**EXTRACTED**"

### **8.1.1.4 Outputs**

Als de bron een database of een eenvoudig bestand is: gegevensstroom DataStage (interne stroom of datasets)

Als de bron een complex bestand is: normale inputbestanden voor DWH.

### **8.1.1.5 Implementering**

Eender welke programmeringstechniek, in staat om de gegevens op eenvoudige, efficiënte en heldere wijze klaar te maken. De volgende tools worden echter aanbevolen in volgorde van voorkeur:

1. DataStage parallel job
2. DataStage server job
3. Perl-scripts (betere prestaties en native mogelijkheden voor verwerking van erg uitgebreide gegevens, eenvoudige syntaxis)
4. Unix Shell-scripts (sh, Ksh, Csh, Awk...) (zwakke prestaties, uitgebreide mogelijkheden voor verwerking van gegevens, erg eenvoudige syntaxis)
5. Andere tools

## **8.1.2 Stap 1: Formattering (Transformatie)**

### **8.1.2.1 Beschrijving**

Deze stap heeft tot doel de gedistilleerde gegevens klaar te maken voor uploading in de ODS. Het is een stap die systematisch moet zijn en die moet losstaan van de latere verwerkingen, dit om aanpassingen te beperken wanneer de behoeften evolueren.

### **8.1.2.2 Inputs**

Gegevensstroom DataStage (interne stroom of datasets)

Standaardbestanden van de landingszone.

### **8.1.2.3 Logische verwerkingen**

1. Indien bestanden klaargemaakt in de "landing area", de bestanden lezen en indien nodig, ze samenbrengen in één enkel te verwerken lot
2. De verplichte technische velden toevoegen (datum van uploads, id van uploadstroom, enz ...)
3. Indien gedefinieerd, de elementaire markeringen van de kwaliteitscontrole toevoegen op het niveau van elke registratie. Voorbeeld: ontbrekende gegevens, buiten de limieten, enz...
4. De operationele metagegevens updaten met de status "**TRANSFORMED**"

### **8.1.2.4 Outputs**

Gegevensstroom DataStage (interne stroom of datasets).

### **8.1.2.5 Implementering**

Uitsluitend DataStage-processen van het parallelle type.

## ***8.1.3 Stap 2: Upload van de ODS***

### **8.1.3.1 Beschrijving**

Deze stap heeft tot doel de nieuwe gegevens op te slaan in de ODS.

### **8.1.3.2 Inputs**

Gegevensstroom DataStage (interne stroom of datasets).

### **8.1.3.3 Logische verwerkingen**

1. De klaargemaakte gegevens lezen
2. De ODS uploaden
3. De operationele metagegevens updaten met de status "**LOADED**"

### **8.1.3.4 Outputs**

Tabellen vande ODS

### **8.1.3.5 Implementering**

Uitsluitend DataStage-processen van het parallelle type.

## ***8.1.4 Stap 3: Selectie en validering van de DWH-gegevens***

### **8.1.4.1 Beschrijving**

Deze stap selecteert en valideert de gegevens die moeten worden geïntegreerd in het data warehouse.



#### 8.1.4.2 Inputs

ODS

#### 8.1.4.3 Logische verwerkingen

1. De stroom initialiseren in de operationele metagegevens met de status "**STARTED**"
2. De gegevens van de ODS selecteren die moeten worden geïntegreerd in het data warehouse (nieuwe gegevens en gegevens die voordien foutief of als te hergebruiken werden verklaard).
3. De businesssleutels vertalen in technische logins
4. Indien gedefinieerd, de geavanceerde kwaliteitscontroles uitvoeren (referentiële integriteit, validering van de businessregels, enz...). De redenen voor de verwerping invoegen naast de referentie naar de foutieve gegevens van de ODS in de operationele tabel ODS\_DATA\_ERROR.
5. Indien nodig, de gegevens van de minibeschrijving aanvullen om de referentiële integriteit van het data warehouse te garanderen.
6. De informatie van de nieuwe gegevens registreren in de operationele metagegevens met de status "**EXTRACTED**".

#### 8.1.4.4 Outputs

Gegevensstroom DataStage (interne stroom of datasets).

#### 8.1.4.5 Implementering

Uitsluitend DataStage-processen van het parallelle type.

### 8.1.5 *Stap 4: Gegevenstransformatie*

#### 8.1.5.1 Beschrijving

Deze stap transformeert de geselecteerde en gevalideerde gegevens om ze aan te passen aan de structuur van het data warehouse. De essentie van de complexiteit van de verwerkingen zit in deze stap geconcentreerd.

#### 8.1.5.2 Inputs

Gegevensstroom DataStage (interne stroom of datasets)

#### 8.1.5.3 Logische verwerkingen

1. De gegevens transformeren
2. De operationele metagegevens updaten met de status "**TRANSFORMED**"

#### 8.1.5.4 Ouputs

Gegevensstroom DataStage (interne stroom of datasets).

#### 8.1.5.5 Implementering

Uitsluitend DataStage-processen van het parallelle type.

## **8.1.6 Stap 5: Uploaden van het DWH**

### **8.1.6.1 Beschrijving**

Deze stap zorgt voor de uploading van de getransformeerde gegevens in het data warehouse.

### **8.1.6.2 Inputs**

Gegevensstroom DataStage (interne stroom of datasets).

### **8.1.6.3 Logische verwerkingen**

1. De DWH-gegevens uploaden
2. De operationele metagegevens updaten met de status "**LOADED**"

### **8.1.6.4 Outputs**

Tabellen data warehouse

### **8.1.6.5 Implementering**

Uitsluitend DataStage-processen van het parallelle type.

## **8.1.7 Stappen 6,7 en 8: Constructie van de Data Marts**

### **8.1.7.1 Beschrijving**

Deze stappen transformeren de gegevens van het data warehouse om ze aan te passen aan de formaten van de data marts. Deze verwerkingen kunnen erg specifiek zijn voor elke data mart en moeten dus apart worden gedocumenteerd.

### **8.1.7.2 Inputs**

Tabellen data warehouse

### **8.1.7.3 Logische verwerkingen**

1. De stroom in de operationele metagegevens initialiseren met de status "**STARTED**"
2. De gegevens van het DWH selecteren die moeten worden opgenomen in de data mart
3. De informatie van de nieuwe gegevens registreren in de operationele metagegevens met de status "**EXTRACTED**".
4. De gegevens transformeren
5. De operationele metagegevens updaten met de status "**TRANSFORMED**"
6. De gegevens uploaden in de data mart
7. De operationele metagegevens updaten met de status "**LOADED**"

### **8.1.7.4 Outputs**

Tabellen data mart

### 8.1.7.5 Implementering

DataStage-processen van het parallelle type.

## 8.1.8 Stap 9: Post-processing

### 8.1.8.1 Beschrijving

Deze stap bundelt alle verwerkingen die niet rechtstreeks noodzakelijk zijn voor de integratie van de gegevens, maar die de mogelijkheid bieden om het platform in een gezonde toestand te houden.

### 8.1.8.2 Inputs

Operationele parameters van het systeem (historiek van de logs, enz...)

### 8.1.8.3 Logische verwerkingen

- De brongegevens archiveren
- De tijdelijke gegevens verwijderen
- De historiek van de gegevens in de ODS, het data warehouse en de data marts beheren
- De operationele historiek beheren (logs, statussen, ...)
- ...

### 8.1.8.4 Outputs

Log van de activiteiten

### 8.1.8.5 Implementering

Eender welke programmeringstechniek, in staat om de noodzakelijke functies te vervullen op eenvoudige, efficiënte en heldere wijze. Het kan gaan om DataStage-jobs, UNIX Shell-scripts of andere tools.

## 8.2 Strategie voor uitvoering van de verwerkingen

De processen voor de verwerking van gegevens worden in reeksen uitgevoerd via DataStage-sequenties, dit om de correlaties te beperken en het operationele onderhoud te vereenvoudigen bij problemen. Deze sequenties worden geconfigureerd om te kunnen heropstarten op het niveau van de laatste verwerking zonder fouten.

De verwerkingen verlopen volgens drie verplichte « haltes »: de ODS, het Data Warehouse en de data marts. Een stap kan niet worden aangevat zolang de vorige stap niet volledig met succes is afgerond.

De automatische opstart van de sequenties wordt verzekerd door de standaardtool van de FOD Financiën: Absyss Visual TOM. Deze is geconfigureerd volgens de opstartschema's die nodig zijn voor de verwerkingen (dagelijks, wekelijks, maandelijks,...) en bevat geen enkele logica die te maken heeft met de gegevens (deze logica wordt beheerd door de DataStage-sequenties).

Om de impact van eventuele problemen met de uitvoering van een « job » op andere gegevensstromen te beperken, bestaat de mogelijkheid om de verwerkingen van elkaar te scheiden in onafhankelijke domeinen (gegevens en « jobs »). Op die manier zal bijvoorbeeld een fout op het niveau van de integratie van gegevens in de ODS van een domein niet verhinderen dat de data mart van een ander domein op correcte wijze wordt geüpload.

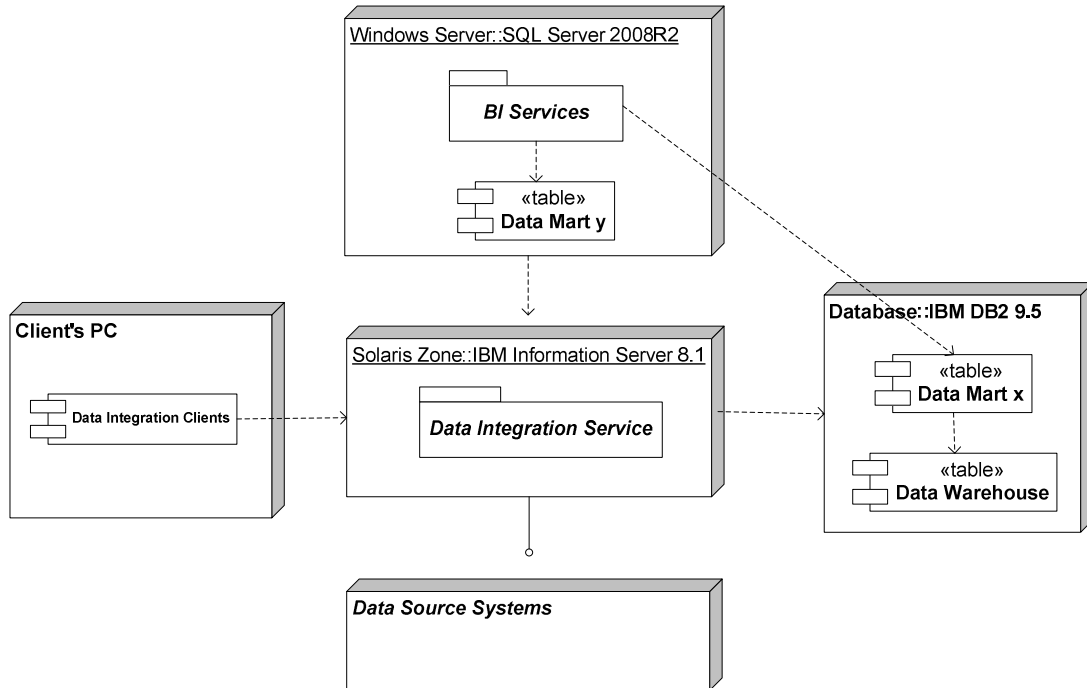
### 8.3 UML-sequentiediagram

Niet van toepassing op dit project.

# 9. View ingebruikneming

## 9.1 Diagram voor ingebruikneming

De verschillende componenten van het platform kunnen op de volgende wijze worden weergegeven.



De componenten van het systeem worden op verschillende servers in gebruik genomen.

### **Server voor integratie van gegevens**

Het gaat om een Solaris-zone die onderdak biedt aan de services voor integratie van gegevens van IBM Information server v8.1/DataStage.

### **Database Data Warehouse**

De gegevens van het data warehouse worden opgeslagen in een IBM DB2 v9.5-database. De server moet worden beschouwd als een server die losstaat van de server voor de integratie van gegevens, omdat hij technisch kan worden geïnstalleerd op dezelfde fysieke server.

Sommige data marts kunnen op deze server worden opgeslagen om prestatieredenen.

### **BI-server**

De gegevens van de data marts kunnen worden opgeslagen op de server die de BI-services verleent, dit om een optimale integratie van alle BI-functionaliteiten te verzekeren.

Deze services worden geleverd door SQL Server 2008R2

### **Client-software**

De ontwikkelaars en beheerders van de processen voor de gegevensintegratie maken gebruik van client-software die communiceert met de server voor gegevensintegratie.

## **10. Omvang en prestaties – FeedBack en suggesties**

---

Dit gedeelte zal gaandeweg worden aangevuld in functie van de operationele feedback die wordt bekomen tijdens de tests.

### **10.1 Stroom voor gegevensintegratie**

#### **10.1.1 Prestatievermogen**

Het prestatievermogen van de stromen voor gegevensintegratie zal worden geraamd door de tijd voor verwerking en uploading van een bepaald volume van brongegevens te meten. Dit volume zal worden gekozen om representatief te zijn voor het gegevensvolume dat gewoonlijk zal worden verwerkt.

Deze prestaties moeten objectief kunnen worden gemeten via de opvolgingstabel van de gegevensstromen in de operationele metagegevens.

Deze prestaties moeten kunnen worden aangepast aan de behoeften, met name door de mate van parallelisme van de DataStage-verwerkingsmotor te regelen.

### **10.2 Databases**

#### **10.2.1 Prestatievermogen**

Het prestatievermogen van de databases zal worden geraamd door de tijd voor de uitvoering van vooraf bepaalde query's te meten. Deze query's zullen representatief zijn voor een typisch gebruik van het data warehouse en elke data mart.

Deze prestaties moeten objectief kunnen worden gemeten via de client-tools van DB2 of via UNIX-scripts met behulp van het commando "time"

Deze prestaties moeten kunnen worden aangepast aan de behoeften, met name door indexen toe te voegen of door de structuur van de gegevens en hun opslagmethode te optimaliseren.

### **10.3 BI-services**

#### **10.3.1 Prestatievermogen**

Het prestatievermogen van de BI-services zal worden geraamd door de tijd voor de uitvoering van vooraf bepaalde BI-scenario's te meten. Deze scenario's zullen representatief zijn voor een typisch gebruik, bijvoorbeeld: weergave van een rapport.

Deze prestaties moeten kunnen worden gemeten met de chronometer.

Deze prestaties moeten kunnen worden aangepast aan de behoeften, met name door de parameters van de BI-server te optimaliseren, en door het gegevensmodel van de data marts te denormaliseren of te optimaliseren.

# 11. Kwaliteit

---

## 11.1 Uitbreidbaarheid / Onderhoud

### 11.1.1 Beschrijving

De uitbreidbaarheid/het onderhoud is de capaciteit om de toepassing te updaten na het einde van de ontwikkelingsfase (om haar aan te passen aan de wijzigende businessbehoeften of om bepaalde bugs te corrigeren).

### 11.1.2 Oplossing

De uitbreidbaarheid/het onderhoud wordt verzekerd door diverse middelen:

- Een technisch design dat een duidelijke scheiding te zien geeft tussen de verschillende functies van de oplossing.
- Het systematische hergebruik van beproefde procedures voor gelijkaardige verwerkingen.
- Een volledige documentatie waarmee niet-ingewijden de technische implementering van de oplossing kunnen begrijpen.
- Het gebruik van tools met grafische interfaces die een duidelijke en eenvoudige kijk bieden op de diepere werking van de oplossing.
- Het gebruik van een parallelle motor voor gegevensintegratie (ETL) die de uitbreidbaarheid van de voorzieningen mogelijk maakt op transparante wijze ten aanzien van de logica van de verwerkingen.

## 11.2 Overdraagbaarheid

### 11.2.1 Beschrijving

De overdraagbaarheid is de capaciteit om de toepassing te benutten in een andere omgeving.

### 11.2.2 Oplossing

Alle componenten van de oplossing kunnen worden gemigreerd naar gelijkaardige platformen vanuit het oogpunt van hardware en besturingssysteem, en dit met een minimum aan aanpassingen.

Voor migraties naar verschillende hardware en besturingssystemen, is elke component verschillend.

#### **Database van het data warehouse**

IBM DB2 9.5 is beschikbaar voor een brede waaier van hardware (x86-64, POWER, System Z, SPARC...) en besturingssystemen (Windows, Linux, AIX, Solaris...). Zie de technische documentatie van het product voor een volledige lijst.

#### **Services voor gegevensintegratie**

IBM InfoSphere Information Server v8.1 is beschikbaar voor een brede waaier van hardware (x86, POWER, SPARC...) en besturingssystemen (Windows, Linux, AIX, Solaris...). Zie de technische documentatie van het product voor een volledige lijst.

IBM InfoSphere Information Server v8.1 ondersteunt een brede waaier van databases, wat belangrijke evoluties van de database van het data warehouse mogelijk maakt.

#### **BI-services en databases van de data marts**

Microsoft SQL Server 2008R2 is uitsluitend beschikbaar voor de platformen met een besturingssysteem van het type Windows. Er kunnen verschillende versies en edities van Windows worden gebruikt. Zie de technische documentatie van het product voor een volledige lijst.

De BI-services van SQL Server 2008R2 kunnen gegevens van verschillende databases gebruiken, maar er zijn in dat geval bepaalde beperkingen mogelijk.

#### **Client-software**

De client-software ondersteunt enkel x86-platformen met een Windows-besturingssysteem.

## **11.3 Monitoring**

### **11.3.1 Beschrijving**

De monitoring is de capaciteit om de verschillende componenten van het systeem te controleren, dit om hun goede werking te verzekeren, fouten vast te stellen en statistieken te produceren.

### **11.3.2 Oplossing**

De gegevensstroom kan worden gecontroleerd door de specifieke tools van het pakket IBM Information Server v8.1 (DataStage Director, Metadata Workbench, ...)

Er wordt eveneens informatie doorgestuurd naar HP OpenView met het oog op een monitoring van hoog niveau. De online commandotools van OpenView zullen door DataStage worden opgeroepen voor de communicatie.

## **11.4 Logging**

### **11.4.1 Beschrijving**

De logging is de capaciteit om de componenten van de toepassing individueel te controleren met het oog op debugging

### **11.4.2 Oplossing**

De logging gebeurt op twee niveaus:

- De logging van de individuele verwerkingen wordt standaard verzekerd door de tools die de services verlenen (IBM Information Server, DB2 en SQL Server)
- De logging van de gegevensstromen maakt een opvolging op hoog niveau mogelijk van de gegevensverwerkingen in hun geheel. Deze logging moet de mogelijkheid bieden om snel de stap te identificeren die voor problemen zorgt in een verwerking. Uit deze logs kunnen ook upload-statistieken worden gedistilleerd.



# 12. Toepassing van de standaarden van de FOD Financiën

---

## 12.1 Afspraken inzake naamgeving

### 12.1.1 Bestanden

#### 12.1.1.1 Bestanden uit de landingszone

De bestanden uit deze zone krijgen de volgende naamgeving

**FWH\_<source name>\_[<schema>\_<table>][<source file name>]\_<timestamp>.TXT**

waarbij

**<source name>** de afgekorte naam is van de gegevensbron

**<schema>** en **<table>** de namen zijn van het schema en de gegevensbrontabel in het geval van bronnen van het type « databases ».

**<source file name>** de afgekorte naam is van het gegevensbronbestand in het geval van bronnen van het type « bestand ».

**<timestamp>** staat voor de datum en het uur waarop het bestand werd aangemaakt ; het formaat moet zijn "%C%y%m%d%H%M%S" (schrijfwijze UNIX/POSIX).

### 12.1.2 Namen van de objecten voor de opslag van gegevens

De afspraken inzake naamgeving voor de gegevensmodellen moeten zoveel mogelijk de standaarden van de FOD Financiën/SupDev volgen. Deze afspraken inzake naamgeving zijn echter hoofdzakelijk voorzien voor de conceptie van objectgerichte software met gebruikmaking van de UML-schrijfwijze. Deze afspraken zijn niet helemaal aangepast aan de conceptie Entiteit-Relatie, die noodzakelijk is binnen een datawarehouse-project. Er zijn dus enkele aanpassingen noodzakelijk.

De volgende paragrafen beschrijven de verschillen tussen de standaardafspraken inzake naamgeving van SupDev en de afspraken die worden gehanteerd in dit project.

Zie deel **Erreur ! Source du renvoi introuvable.** voor de referentie naar het document "SUPDEV\_NOMMAGE" met betrekking tot de afspraken inzake naamgeving van SupDev.

#### 12.1.2.1 Logisch gegevensmodel

De logische gegevensmodellen kunnen worden beschouwd als zijnde gelijkwaardig met de conceptuele schema's binnen een objectgericht ontwerp.

De volgende tabel toont de overeenkomsten tussen objectgerichte ontwerpen en ER-ontwerpen (entiteit-relatie).

OG-concept	ER-concept	Afspraak naamgeving voor ER
Klasse	Entiteit	We beginnen met E ( <i>Entity</i> ) in plaats van K ( <i>Klasse</i> ). Voorbeeld: <i>EPersoon</i> , <i>EjuridischeSituatie</i>
Attribuut	Attribuut	Afkortingen mogen worden gebruikt in de velden <i>Mn...n</i> maar zouden moeten worden vermeden met het oog op een betere leesbaarheid
Associatie	Relatie	Schrijfwijze "Crow's foot". Geen rollen.

### **12.1.2.2 Fysiek gegevensmodel**

De fysieke gegevensmodellen moeten de afspraken van SupDev en RDC volgen.

Voor de data marts bestaat echter de mogelijkheid om de namen van de objecten aan te passen aan de technische of businessseisen van de eindgebruikers.

### **12.1.3 Namen van de objecten voor de verwerking van gegevens**

De gegevensverwerkingen verlopen met IBM DataStage.

De DataStage-objecten moeten hun naam krijgen volgens de afspraken van DCC. Zie het document DCC\_DS\_NAMING in paragraaf 1.2 voor de referentie.

### **12.1.4 Tools voor business intelligence**

De BI-tool, gebruikt voor de exploitatie van de gegevens, is Microsoft SQL Server (Reporting services/SSRS en Analysis services/SSAS).

In overeenstemming met de standaardpraktijken van DCC is er geen enkele strikte norm noodzakelijk voor de naamgeving van de objecten.

Aangezien de BI-objecten zijn bestemd voor de eindgebruikers, wordt niettemin aanbevolen om namen te gebruiken die het object op volledige en nauwkeurige wijze beschrijven voor de gebruikers.