

connect.  
engage.  
succeed.

**BUSINESS INTELLIGENCE  
&  
DYNAMICS AX 2012**

**Auteur:** Daniël Bosmans



Versie: 1.0  
Datum: 06-10-2014

Afstudeerder: Daniël Bosmans  
Studentnummer 11061022  
Opleiding: Informatica  
Instelling: Haagse Hogeschool  
Contact Bleiswijkseweg 37  
2712 PB Zoetermeer  
Nederland  
070 445 7200

Eerste examiner: Dhr. A.A. Nederend  
Tweede examiner: Dhr. R. Ruijsenaars

Bedrijf: Centric Netherlands B.V.  
Afdeling: Supply Chain Solutions  
Contact Antwerpseweg 8  
2803 PB Gouda  
Nederland  
+31 182 34 50 00

Opdrachtgever: Dhr. M.E Helmer  
Bedrijfsmentor: Dhr J. te Molder

Copyright 2014, Centric Netherlands B.V.

Niets uit deze uitgave mag worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen of enige andere manier, zonder schriftelijke toestemming van Centric.

## Referaat

D. Bosmans, 11061022, Afstudeerverslag, Het ontwikkelen van BI dashboards voor de Tradeworld & International Shipments module binnen Microsoft Dynamics AX 2012 R3, oktober 2014.

Dit verslag is geschreven in het kader van het afstuderen aan de opleiding informatica op De Haagse Hogeschool in Zoetermeer. De afstudeeropdracht heeft plaats gevonden bij het bedrijf Centric Nederland te Gouda en binnen de business unit Supply Chain Solutions. Dit afstudeerproject is gestart op 12 mei 2014 en is afgerond op 06 oktober 2014.

De gegevensbron analyse is niet deze versie van het verslag weggelaten, in verband met concurrentie gevoelige productinformatie

### Descriptoren:

- Business Intelligence (BI)
- Dynamics AX 2012
- Microsoft Integration Services 2012 (SSIS)
- Datawarehouse lifecycle Methodiek
- Ralph Kimball
- Dimensioneel modelering
- SQL

## Voorwoord

Mijn naam is Daniël Bosmans en ik volg de opleiding Informatica aan de Haagse Hogeschool te Zoetermeer. Na een lange school carrière, gestart op VMBO, naar het MBO en ten slotte het HBO, ben ik op dit punt aangekomen: afstuderen!.

Om te kunnen afstuderen ben ik op zoek gegaan naar een afstudeerplek. Deze heb ik uiteindelijk gevonden bij Centric op de afdeling Supply Chain Solutions in Gouda. Centric bood een interessante opdracht op het gebied van Business Intelligence aan. Gedurende dit project heeft Centric mij alle hulp geboden en de vrijheid gegeven om deze opdracht uit te kunnen voeren. Ik kijk terug naar een fijne en leerzame periode bij Centric. Ik kan Centric als een goede afstudeerplek aan andere aanbevelen.

Ik wil van deze gelegenheid gebruik maken om iedereen te bedanken voor de steun gedurende deze afstudeerperiode. Vanuit Centric wil ik in het bijzonder Joost te Molder bedanken voor zijn begeleiding en vertrouwen in mij. Van de Haagse Hogeschool wil ik graag mijn studieloopbaan begeleider Arno Nederend bedanken voor al zijn steun, advies en luisterend oor gedurende de hele opleiding. Als laatste wil ik mijn familie bedanken voor hun steun en begrip.

Daniël Bosmans  
Afstudeerder

Gouda, oktober 2014

## Inhoudsopgave

<b>1</b>	<b>PROJECT ACHTERGROND .....</b>	<b>2</b>
1.1	BEDRIJF .....	2
1.2	PROBLEEMSTELLING .....	3
1.3	DOELSTELLING .....	3
1.4	RESULTAAT .....	4
1.5	SCOPE .....	4
<b>2</b>	<b>PROJECT AANPAK EN FASERING.....</b>	<b>5</b>
<b>3</b>	<b>ANALYSE.....</b>	<b>7</b>
3.1	ORIËNTATIE .....	7
3.2	ANALYSE HUIDIGE SITUATIE .....	7
3.3	ANALYSE VERLADINGEN BEDRIJFSPROCES .....	8
3.4	ANALYSE BRONSISTEEM .....	10
<b>4</b>	<b>REQUIREMENTS.....</b>	<b>15</b>
4.1	BUSINESS REQUIREMENTS.....	15
4.2	FUNCTIONELE REQUIREMENTS.....	16
4.3	NON-FUNCTIONELE REQUIREMENTS .....	17
4.4	TECHNISCHE REQUIREMENTS .....	18
<b>5</b>	<b>DATAWAREHOUSE ARCHITECTUUR .....</b>	<b>20</b>
5.1	STAGING OMGEVING.....	21
5.2	DATAWAREHOUSE OMGEVING .....	21
5.3	POWERPIVOT.....	22
5.4	BI APPLICATIONS .....	23
5.5	KIMBALL: CONFORM DIMENSIES .....	23
5.6	REALISEREN ARCHITECTUUR .....	25
<b>6</b>	<b>DIMENSIONEEL MODEL ONTWERP PROCES .....</b>	<b>26</b>
6.1	CONCEPTUEEL ONTWERP .....	26
6.2	LOGISCH ONTWERP .....	27
6.3	FYSIEK ONTWERP.....	37
<b>7</b>	<b>ETL ONTWIKKELINGSPROCES .....</b>	<b>38</b>
7.1	ETL WORKFLOW ONTWERP .....	38
7.2	RECORD GESCHIEDENIS PROCEDURE .....	41
<b>8</b>	<b>BI DASHBOARDS .....</b>	<b>43</b>
<b>9</b>	<b>TESTEN DATAWAREHOUSE.....</b>	<b>44</b>
9.1	VALIDATIE REQUIREMENTS .....	44
9.2	SEMANTISCHE TEST .....	44
9.3	SYNTACTISCHE TEST .....	45
9.4	TEST CASES GESCHIEDENIS.....	46
9.5	REGRESSIE TEST .....	46
9.6	PERFORMANCE TEST.....	47
<b>10</b>	<b>EVALUATIE .....</b>	<b>48</b>
10.1	PROCES .....	48
10.2	(TUSSEN) PRODUCTEN.....	49
10.3	BEROEPSTAKEN.....	49
<b>11</b>	<b>LITERATUUR / BRONNENLIJST .....</b>	<b>53</b>
<b>12</b>	<b>REFERENTIE .....</b>	<b>55</b>

## Inleiding

Dit verslag bevat een verslaglegging van het afstudeerproject voor het ontwikkelen van BI Dashboards voor de module Tradeworld & International Shipments binnen Microsoft Dynamics AX. Deze afstudeeropdracht is uitgevoerd door D. Bosmans in opdracht van Centric Supply Chain Solutions te Gouda. Hier verder de afstudeerder genoemd.

Dit afstudeerverslag dient u inzicht te geven in het proces, de uitgevoerde activiteiten en de gemaakte beslissingen van de afstudeerder ten aanzien van dit afstudeerproject. Tevens dienen de beide examinatoren en de externe gecommiteerde, op basis van dit verslag, te beoordelen of de afstudeerder de competenties bezit, zoals verwacht van een afgestudeerd HBO informatica student.

De afstudeerder heeft de Kimball Lifecycle methodiek toegepast. De hoofdstukken in dit verslag volgen op hoofdlijnen de fasen binnen de methodiek. Het eerste hoofdstuk beschrijft de context van de opdracht evenals de probleemstelling, doelstelling en het verwachte resultaat van dit project. Vervolgens wordt er beschreven in hoofdstuk twee, hoe de afstudeerder invulling heeft gegeven aan de Kimball methodiek.

In het derde hoofdstuk wordt er verslag gedaan van verschillende uitgevoerde analyses, zoals de gegevensbronanalyse. In hoofdstuk vier zijn de requirements gespecificeerd. In alle hoofdstukken kunt u verwijzingen vinden naar deze requirements. Dit om de traceerbaarheid van de requirements ten aanzien van de gemaakte ontwerpbeslissingen te tonen.

De datawarehouse architectuur wordt beschreven in hoofdstuk vijf. In de hoofdstukken zes en zeven worden alle aspecten rond om dimensioneel modeleren en data transformaties beschreven. Hoofdstuk acht beschrijft de analyse en rapportage mogelijkheden van de datawarehouse.

Het testen van de datawarehouse staat beschreven in hoofdstuk negen. De testresultaten kunt u vinden in de bijlagen. De proces evaluatie als mede de evaluatie van de beroepstaken kunt u vinden in hoofdstuk tien.

De laatste twee hoofdstukken, respectievelijk elf en twaalf, bevatten de literatuur/bronnenlijst als de referentie van de bedrijfsmentor ten aanzien van de afstudeerder.

Bij dit verslag zijn een aantal bijlagen toegevoegd. Deze bijlagen dienen als bewijsvoering van wat er in dit verslag wordt beschreven.

## 1 Project achtergrond

Het project betreft het ontwikkelen van Business intelligence in de vorm van dashboards voor klanten van het bedrijf Centric. Centric verkoopt ERP (Enterprise Resource Planning) pakketten aan bedrijven en configureert en ondersteunt hen bij het onderhouden ervan.

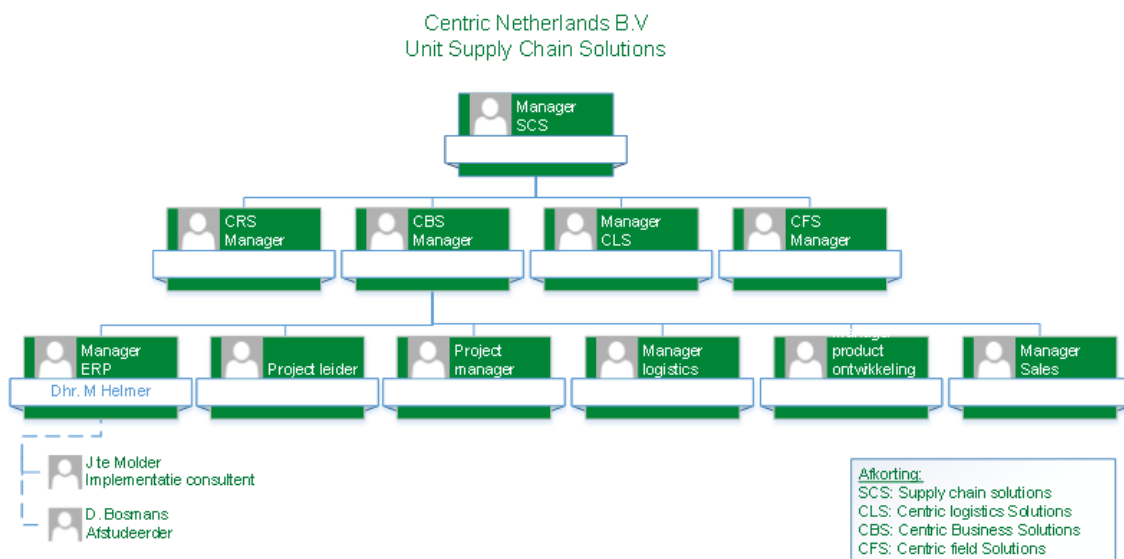
Daarnaast ontwikkelt Centric zogenaamde 'value added' functionaliteit binnen Dynamics AX voor specifieke marktsegmenten. Een voorbeeld van een segment zijn de internationale shipments van goederen door groothandels. Dit proces wordt geadministreerd binnen een module van Dynamics AX genaamd 'Tradeworld & international shipments'. Centric heeft binnen deze module extra functionaliteit gebouwd om verladingen van goederen te managen (bijvoorbeeld het zorgen voor een optimale container vulling). De gebruikers van deze module; de groothandels, ervaren dat door de toegenomen complexiteit van het shipment proces, het overzicht hierop te verliezen.

Centric wenst een Business Intelligence oplossing aan te bieden aan de groothandels om het shipment proces inzichtelijker voor hen te maken. De kennis van Business Intelligence (BI) is ontoereikend binnen de Centric unit Business Solutions. Daarom heeft Centric een afstudeerder van de Haagse Hogeschool ingehuurd, om een BI oplossing voor deze groothandels neer te zetten.

### 1.1 Bedrijf

Deze afstudeeropdracht is uitgevoerd binnen het bedrijf Centric It Solutions B.V. Centric heeft 4.944 (2012) werknemers en vestigingen in Nederland, België, Duitsland, Zwitserland, Noorwegen, Zweden en Roemenië. De organisatie boekte in 2012 een omzet van 525 miljoen euro met een netto bedrijfsresultaat van ruim 12 miljoen euro

Centric is opgedeeld uit een aantal divisies, één daarvan is Supply Chain Solutions. De afstudeerder is werkzaam op de divisie; Supply Chain Solutions op de locatie Gouda. De divisie heeft als doelstelling om software -en hardware oplossingen te bieden voor de volledige Supply Chain van bedrijven. De module Tradeworld & International shipments binnen Microsoft Dynamics AX is één van de software oplossingen die Centric op dit marktsegment aanbiedt. In figuur 1 zit u het organogram van de business unit: Supply Chain Solutions.



Figuur 1 Organogram

Onder deze business units vallen een aantal activiteiten dat vervolgens weer wordt aangestuurd worden door een manager. Deze afstudeeropdracht wordt uitgevoerd binnen deze business unit en onder business solutions. De directe manager en tevens opdrachtgever van dit project is dhr. M Helmer; manager ERP. De bedrijfsmentor (begeleider) tijdens dit project en tevens een directe collega dhr J. te Molder; implementatie consultant.

## 1.2 Probleemstelling

Centric levert en implementeert het Microsoft Dynamics AX ERP pakket bij klanten. Centric ontwikkelt zelf aanvullende modules binnen het standaard ERP pakket. Eén van deze modules is TradeWorld & International Shipments. Met deze module kunnen groothandels hun goederenstromen over de gehele wereld beheren.

Steeds meer handelsactiviteiten vinden in het buitenland plaats. Hierdoor neemt de complexiteit van het proces toe. Hierdoor neemt ook de vraag naar de traceerbaarheid van de in- en uitgaande goederenstroom en transport toe. Met de huidige shipment module ervaart een deel van de klanten dat men op dit moment nog steeds onvoldoende visueel overzicht heeft wat betreft de goederenstromen.

Centric heeft onvoldoende ervaring met het managen en inzetten van grote hoeveelheden aan goederenstromen gerelateerde data. Centric is hierdoor onvoldoende in staat om, voor haar klanten, hierin iets te betekenen. Dit heeft gevolgen voor haar concurrentie positie binnen de markt.

## 1.3 Doelstelling

De doelstelling van de opdracht was om binnen Microsoft Dynamics AX en vervolgens in de module International Shipments een datawarehouse architectuur te ontwerpen, bouwen en te implementeren. Met deze architectuur is het de bedoeling dat men grote hoeveelheden aan data kan verwerken. Door het te verwerken en visueel maken van die datastromen krijgen de groothandels meer inzicht in de goederenstromen en tevens wordt de traceerbaarheid van die stromen verbeterd. De groothandels krijgen zo meer grip op het kritieke shipment proces en kunnen daardoor beter ingrijpen bij calamiteiten.

De architectuur moest leiden tot een systeem, waarin middels een dashboard met een aantal verschillende overzichten en visualisaties, hulp wordt geboden bij het succesvol managen van international shipments. De insteek was om het systeem te ontwerpen met bestaande tooling van Microsoft.

De opdrachtgever ziet deze opdracht als de 'kers op de taart' bij de module International Shipments, waardoor de module beter in de markt te zetten is. De datawarehouse architectuur moest tevens generiek zijn en door verschillende groothandels met minimale aanpassing te gebruiken zijn. Dit zal bereikt worden door een architectuur op te stellen dat flexibel is om met, bijvoorbeeld ontbrekende data, de bedrijfsvragen alsnog te beantwoorden.

Medewerkers van business solutions dienen in staat te zijn om de dashboards en de architectuur te implementeren bij klanten. Om dit te bereiken vindt er aan het einde van dit project een overdracht plaats. In deze overdracht wordt uitgelegd hoe de dashboards en de onderliggende architectuur is opgebouwd. Hierbij is het belangrijkste dat de medewerker inzicht krijgt in de nodige aanpassingen om de oplossing bij klanten te implementeren. Als laatste is het van belang dat hij bij eventuele fouten, dit zelf kan verhelpen. In de architectuur worden middelen zoals het bijhouden van logs of foutafhandeling ingezet om dit te bewerkstelligen.



Het was de wens van de opdrachtgever om binnen vijf maanden (de termijn van de afstudeerstage) de bovenstaande doelstelling te realiseren.

#### 1.4 Resultaat

Het gewenste resultaat van dit project was het ontwerpen en bouwen van een data kubus en een grafische interface om de data in de kubus aan de gebruiker te presenteren. De data in de kubus moet afkomstig zijn uit de Microsoft Dynamics AX2012 module: 'International Shipments'.

Er is toegevoegde waarde aan de module Tradeworld & International Shipments toegevoegd en kan mogelijk zo beter in de ERP markt worden neergezet.

Centric heeft aan het einde van dit project meer kennis op het gebied van Business Intelligence en kan deze inzetten om in de toekomst zelf dergelijke projecten te starten. De ontwikkelde BI oplossingen kan door medewerkers van Centric zelf worden geïmplementeerd bij klanten.

#### 1.5 Scope

De omvang van het Microsoft Dynamics AX ERP pakket is te groot en tijdrovend om deze, in zijn geheel, te analyseren en te bevatten. Daarom beperkt de opdracht zich tot de module International Shipments binnen Dynamics AX. Mocht het voor het correct functioneren van de dashboards noodzakelijk zijn dat er vanuit andere bronnen/module moet worden onttrokken, dan valt dit binnen de scope. Deze andere gegevensbronnen dienen echter wel een relatie te bevatten met het verladingenproces. Daarnaast zijn niet alle gegevens die een relatie hebben met het verladingenproces bruikbaar voor een BI oplossing. Een voorbeeld hiervan zijn gegevens over de (actuele) voorraad. Deze gegevens hebben een relatie met het verladingenproces maar het is niet van belang om bij te houden waar een product fysiek in een magazijn wordt opgeslagen. Het detail niveau is te hoog.

Het is niet de bedoeling om nieuwe modules binnen Dynamics AX te ontwikkelen, noch andere taken uit te voeren die niet bijdragen aan het behalen van de doelstellingen van deze opdracht.

Binnen dit project zal geen BI pakket selectie plaats vinden omdat Centric aangeeft een Business intelligence oplossing binnen een Microsoft omgeving te willen implementeren. Wel vindt er een selectie plaats binnen de tooling/technologieën binnen de BI stack van Microsoft. Deze selectie is afhankelijk van het architectuurontwerp.

Het viel tevens buiten de scope van deze afstudeeropdracht om de gegevensarchitectuur voor Business Intelligence te implementeren bij klanten. Wel kunnen er aanbevelingen worden gegeven op dit gebied. Echter is het wel de doelstelling van dit project om een werkend prototype te realiseren, die uit te bouwen is, tot een volwaardige BI oplossing.

## 2 Project aanpak en fasering

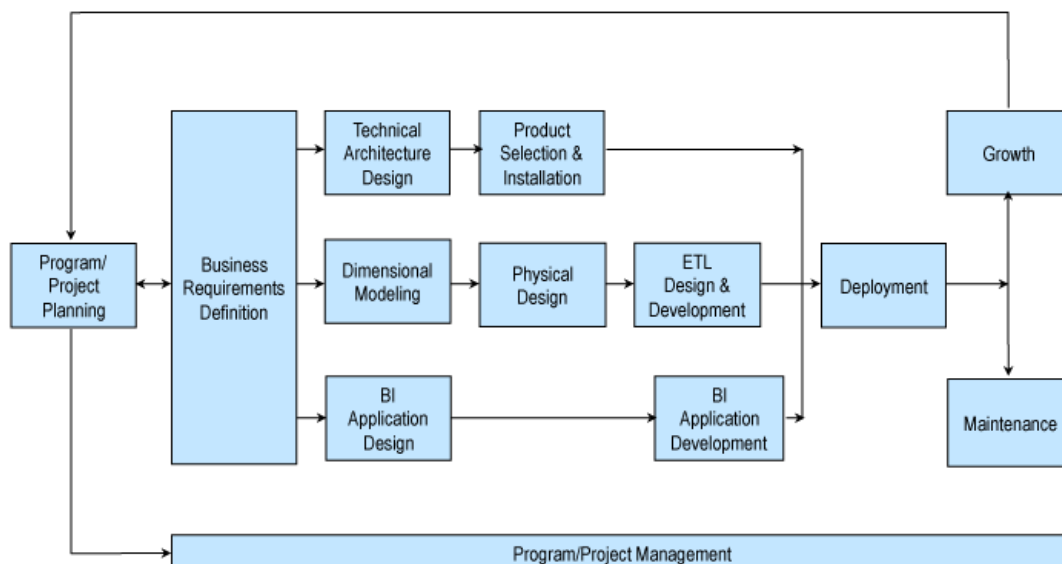
De methodiek die de afstudeerder, voor dit project, heeft gebruikt is de Kimball datawarehouse lifecycle, zoals beschreven in het boek 'The datawarehouse lifecycle toolkit druk 2'. De Kimball lifecycle is een veel gebruikte methodiek voor datawarehouse projecten. Daarnaast worden een aantal opvattingen over datawarehousing en de 'best practises' van Kimball door de afstudeerder gedeeld. Deze betreffen de opvatting van de definitie -en architectuur van een datawarehouse. Van deze architectuur is het gebruik van het dimensioneel datamodel de voornaamste.

Binnen Centric bestond geen vastgestelde project methodiek of ontwikkelstaat voor BI projecten. Centric maakte wel gebruik van SCRUM<sup>1</sup>, maar werd ingezet om ontwikkelingstrajecten in ERP software te managen. Er is dan ook besloten om geen gebruik te maken van SCRUM maar van de Datawarehouse lifecycle methodiek<sup>2</sup>. Deze methodiek bracht meer structuur aan het project en bood meer houvast voor de taken die er uitgevoerd moesten worden. SCRUM biedt op dit gebied, niet de juiste handvatten en is meer gericht op het managen van softwareontwikkelingstrajecten.

Het gebruik van de Kimball architectuur staat in principe los van de lifecycle methodiek. De afstudeerder meent dat de keuze voor de Kimball architectuur in nauw verband staat met de keuze voor de lifecycle methodiek.

Binnen dit project wordt enkel bestaande softwaretools gebruikt om de BI oplossing te realiseren. Er vindt dan ook geen softwareontwikkeling plaats binnen dit project.

Het was niet de bedoeling om de methodiek in detail en in zijn volledigheid uit te voeren. Er is echter wel vast gehouden aan de hoofdlijnen. Een aantal activiteiten binnen de methodiek zijn niet relevant of toepasbaar voor dit project. Mocht u een aantal activiteiten van de methodiek binnen dit project missen, dan was dit met opzet. In figuur 2 ziet u een visuele weergave van de methodiek met de verschillende fase/disciplines.



Figuur 2 Kimball lifecycle diagram

Binnen de Kimball lifecycle worden verschillende project taken (tracks) en disciplines zoveel mogelijk parallel uitgevoerd. De volgende tracks zijn gedefinieerd; architectuur,

<sup>1</sup> SCRUM is een agile softwareontwikkelingsmethodiek dat net als de lifecycle van Kimball ook iteraties of 'sprints' kent

data en BI applicaties. Binnen de track worden taken sequentieel uitgevoerd (bijvoorbeeld; dimensioneel naar een fysiek model).

In de praktijk zal de afstudeerder eerst de architectuur vaststellen en vervolgens het dimensioneel datamodel ontwikkelen. Omdat dit project door één persoon wordt uitgevoerd zullen de disciplines niet parallel worden uitgevoerd. Ook is er in de praktijk veel overlap tussen de verschillende disciplines. Bijvoorbeeld de keuze voor het datamodel is deels afhankelijk van de gekozen architectuur.

De Kimball methodiek is iteratief. De afstudeerder zal tijdens de uitvoering van dit project één volledige iteratie doorlopen. De afstudeerder meent dat het niet haalbaar is om binnen de afstudeerperiode een tweede iteratie te doorlopen, dit vanwege de schaal van de opdracht. Er is dan ook besloten om het ontwikkelingsproces van één dashboard te tonen met, de bijbehorende documentatie.

Een iteratie binnen deze methodiek houdt in dat het gehele ontwikkelingstraject wordt herhaald (van planning tot realisatie). Dit betekent ook dat het ontwikkelde dimensioneel model als mede de ETL<sup>3</sup> processen moeten worden herzien. Door enkel één iteratie te beschrijven heeft u een globaal beeld van een iteratie. De opvolgende iteraties verlopen vergelijkbaar met de eerste, al dan niet op een kleinere schaal.

Er is voorgaande aan dit project een plan van aanpak opgesteld. In dat plan staat in detail beschreven hoe de methodiek is toegepast en welke activiteiten er per fase zijn uitgevoerd. Het plan van aanpak vindt u in bijlage A.

---

<sup>3</sup> ETL is de afkorting voor Extract Transform Load en staat voor het proces dat moet worden doorlopen om data geschikt en beschikbaar te maken voor opslag in een datawarehouse.

### 3 Analyse

Dit hoofdstuk beschrijft de voorbereidende activiteiten en analyses die de afstudeerder heeft uitgevoerd voor de aanvang van dit project. De analyse van de gewenste situatie staat deels beschreven in het plan van aanpak. De elicitering naar de (business) requirements vormen een ander onderdeel van de analyse naar de gewenste situatie.

#### 3.1 Oriëntatie

De afstudeerder heeft als eerste taak contact gelegd met de direct betrokken afdeling en kennis gemaakt met collega's. De afstudeerder heeft daarmee in kaart gebracht hoe de bedrijfsstructuur van het bedrijf Centric in elkaar zit. Het bedrijf Centric kent vele divisies en subdivisies. Deze structuur verandert ook regelmatig. De afstudeerder heeft de bedrijfsanalyse beperkt tot de divisie die verantwoordelijk is voor het implementeren, ontwikkelen en beheren van softwarepakketten, zoals Microsoft Dynamics AX.

Het uitvoeren van deze analyse trof als doel om vast te stellen welke plaats de afstudeerder binnen het bedrijf inneemt. Evenals te bepalen waar de afstudeerder zich moet wenden voor eventuele vragen of hulpmiddelen. Daarnaast is er ook geanalyseerd welke stakeholders<sup>4</sup> bij dit project betrokken zijn. Deze zaken staan beschreven in het plan van aanpak (bijlage A).

Voor de aanvang van dit project was duidelijk dat Centric de rol van opdrachtgever en klant heeft. Centric heeft uit haar klantenbestand vernomen dat er behoefte was aan een business intelligence oplossing in combinatie met Dynamics AX. Centric heeft zelf het project geïnitieerd zonder deze specifiek op een klant te richten. Bij deze opdracht ontbreekt de klant als stakeholder en zijn requirements ten aanzien van de dashboards. De rol van de klant werd binnen dit project vervuld door een pre sales consultant van ERP applicaties binnen Centric. Hij kan vervolgens bepalen of de doelstelling van dit project behaald zijn. Tevens kan hij een uitspraak doen, op basis van zijn contacten met klanten, of de dashboards aan non-functionele requirements<sup>5</sup> voldoen.

De afstudeerder meent dat dit bovenstaande geen ideale situatie is en dat het klant perspectief belangrijk is voor de verzameling van requirements. Klanten (bedrijven) kunnen immers verschillen van mening, over welke bedrijfsgegevens, er in de dashboards getoond moeten worden. Ook kunnen de bedrijven anders tegenover historische gegevens of performance aankijken en er belang aan hechten. Waar deze zaken niet naar boven zijn gekomen, worden deze door de afstudeerder ingevuld en later geverifieerd.

#### 3.2 Analyse huidige situatie

Voorgaande aan dit project is er een analyse uitgevoerd van de huidige situatie. De huidige situatie is in kaart gebracht door informele gesprekken op de afdeling te voeren en door algemene observatie.

Centric heeft onvoldoende ervaring met business intelligence projecten gebaseerd op Microsoft oplossingen. Binnen Centric lopen een aantal projecten met de BI-tools; Cognos en Qlikview. Deze projecten worden uitgevoerd binnen andere business units. Deze business units richten zich op andere marktsegmenten en bieden elk andere ERP software pakketten, zoals SAP waarop deze BI-tools zijn toegepast. De kennis en ervaring die wordt opgedaan wordt gedeeld in een zogeheten 'BI competente center'.

<sup>4</sup> Met stakeholders zijn mensen die betrokken zijn bij dit project, het eind product gaan gebruiken of belang hebben bij het succesvol slagen hiervan.

<sup>5</sup> Non-functioneel requirement zoals; zijn de dashboards overzichtelijk en gemakkelijk te gebruiken?

In de business unit Supply Chain Solutions zijn geen soortgelijke BI projecten gestart in combinatie met Microsoft Dynamics AX. Dit heeft ten delen als oorzaak dat Dynamics al deels standaard BI functionaliteit bevat. Deze standaard functionaliteit biedt de mogelijkheid om via interfaces eenvoudig een kubus en rapporten te genereren. Een aantal klanten van Centric gebruiken deze rapporten, maar ervaren dat deze onvoldoende informatie bevat over verladingen. In de standaard kubus van Microsoft zijn namelijk de gegevens die in de module Tradeworld zijn opgeslagen niet meegenomen. Tevens worden rapporten op basis van de Microsoft kubus traag geladen. Het traag laden van de rapportages komt deels door de opbouw van de kubus en de omvang van Dynamics. De kubus is namelijk geïntegreerd in Dynamics zelf.

Centric heeft een aantal modules toegevoegd naast wat standaard wordt aangeboden binnen Dynamics. De data die hieruit wordt gegenereerd, is niet zichtbaar binnen de huidige gegenereerde kubus noch in de rapportages. Dynamics biedt een ontwikkelingsomgeving aan ontwikkelaars om de kubus en de rapportages aan te passen. De ontwikkelaars ervaren deze omgeving als beperkt en hebben onvoldoende inzicht in het overzetten (mappen) van data uit deze zelf ontwikkelde modules. Dit gegeven verklaart het gebrek van een adequate BI oplossing in combinatie met Dynamics.

De business unit richt zich voornamelijk op het bieden van logistieke oplossingen in combinatie met Dynamics. Hoewel Dynamics een ERP is, ontbrak het van de nodige functionaliteit om het verladingproces effectief te kunnen administreren. Dit was de situatie in de 2009 versie van de ERP. Centric heeft om deze reden de ontbrekende functionaliteit toegevoegd aan het bestaande ERP. De door Centric ontwikkelde functionaliteit heeft een hoge interne koppeling met de andere modules van Dynamics zelf. Met als consequentie dat Centric sterk afhankelijk is van de ontwikkelaar; Microsoft, voor het tijdig melden van software aanpassingen. De functionaliteit kan mogelijk na een update niet meer werken.

Voor de aanvang van dit project was Centric begonnen met het migreren van bestaande code van Dynamics 2009 naar 2013. De module Tradeworld en international shipments was zodoende ook gemigreerd. Tevens is de huidige release (R3) van Tradeworld nog niet getest met als gevolg dat deze nog niet is geïmplementeerd bij klanten. Gedurende dit project werd de Dynamics 2012 R3 release als uitgangspunt gebruikt voor het ontwikkelen van de BI oplossing.

Omdat de nieuwe release nog niet verkocht is, ontbreekt het klant perspectief gedurende dit project.

### 3.3 Analyse verladingen bedrijfsproces

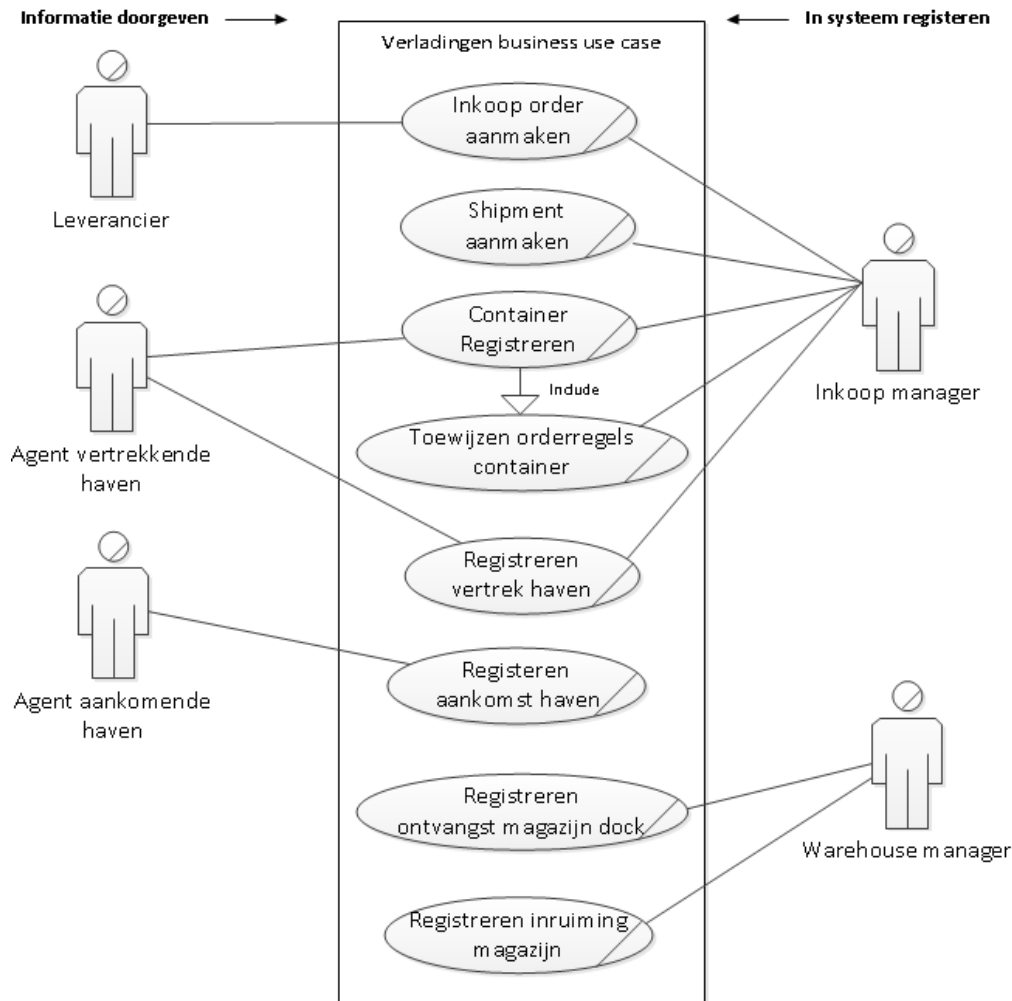
Om een business intelligence oplossing te kunnen ontwikkelen is het belangrijk om inzicht te hebben in het betrokken bedrijfsproces. Door deze analyse binnen dit project uit te voeren kon er worden vastgesteld dat dit bedrijfsproces ondersteund kan worden met business intelligence. Daarnaast droeg het bij aan het verzamelen van de requirements doordat er meer inzicht kwam in de belevingswereld van de stakeholders en eventuele problemen die zij ervaren.

Als laatste droeg het ook bij aan het interpreteren van de bedrijfsdata binnen de Dynamics module en het verladingen proces. Het verzamelen van de meta data<sup>6</sup> werd hierdoor vergemakkelijkt.

---

<sup>6</sup> Meta data is informatie over data; zoals; gebruik, type of betekenis.

Ter ondersteuning van deze analyse is er een business use case diagram<sup>7</sup> opgesteld met bijbehorende beschrijving. In figuur 3 het diagram.



Figuur 3 business use case diagram

In het business use diagram kunt u zien door welke medewerker (actor), data in het systeem invoert. Door het visualiseren van het proces is het duidelijk geworden dat data achteraf wordt geregistreerd en afkomstig is van externen. De correctheid van de data is dus afhankelijk van die derde actoren. Dit is een aandacht punt voor de organisatie voor het naleven van de geldige bedrijfsregels. De externe medewerkers kunnen ook niet worden benaderd.

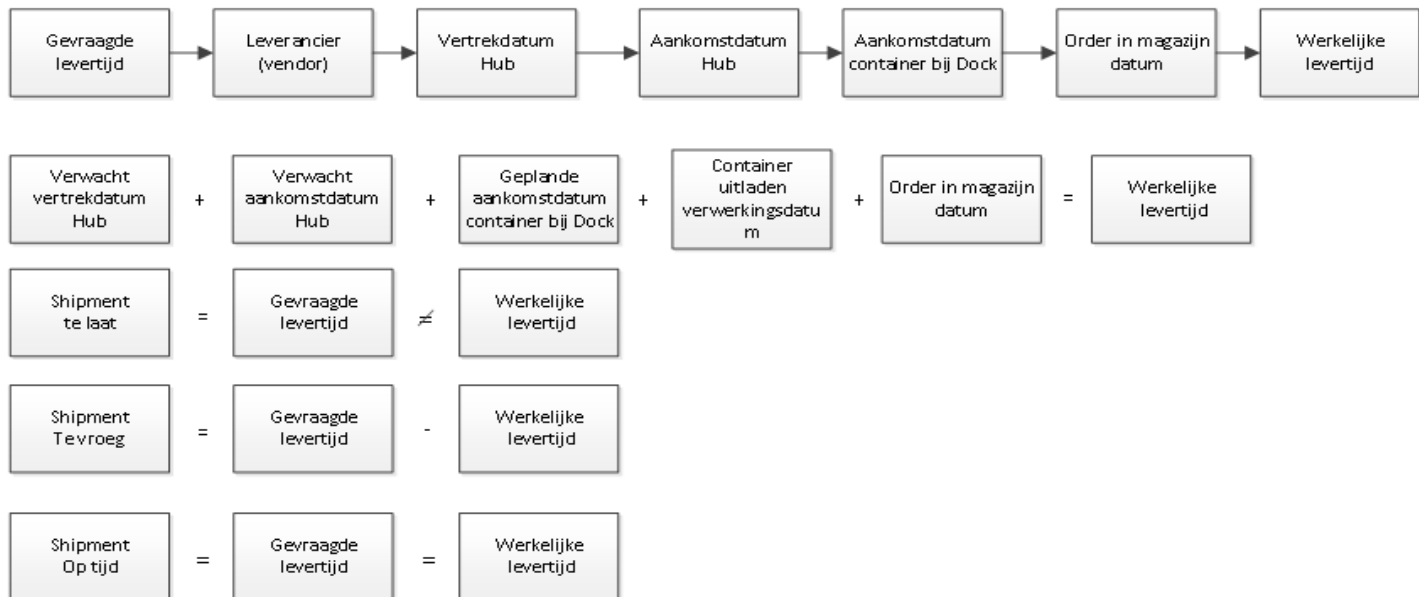
Naast het visualiseren van het proces is deze ook deels beschreven. Deze beschrijvingen zijn reeds aanwezig binnen Centric en het had dan ook geen toegevoegde waarde om deze zelf toe te voegen. Een grafische representatie was echter nog niet aanwezig.

In figuur 4 ziet u een overzicht van de verschillende data in het shipment proces. Door deze analyse uit te voeren is er meer inzicht ontstaan in het proces. In een later stadium kunnen deze berekeningen in een rapport worden opgenomen om bijvoorbeeld een overzicht van vertraagde shipments te tonen. Het verladingen proces is in kaart gebracht tijdens een interview van de pre-consultant Dhr Aldewereld. Dit interview werd gehouden voor de verzameling van de requirements.

<sup>7</sup> Business use case diagram toont grafisch welke bedrijfsactor betrokken is bij het proces. Het betreft hier geen applicatie.



Data verlading proces



Figuur 4 Datum shipment proces

### 3.4 Analyse bronsysteem

Na het analyseren van de huidige situatie binnen de business unit Supply Chain Solutions en van het verlading proces volgde de analyse naar het bronsysteem; Dynamics AX 2012 R3.

Voor de aanvang van deze analyse hebben een aantal gesprekken plaatsgevonden met één van de ontwikkelaars van de betreffende module. De ontwikkelaar heeft de werking van de module uitgelegd en geassisteerd met het opzetten van een aparte werkomgeving met daarop Dynamics geïnstalleerd.

Zowel de werkomgeving als de applicatie moest nog deels worden geconfigureerd voordat deze gebruiksklaar was. Het was niet nodig om een selectie te maken van de benodigde tooling of ontwikkelingsplatform voor dit project. Centric is een gecertificeerde Microsoft partner. Naast de voordelen die dit brengt, legt het ook de beperking op dat Centric voor Dynamics oplossingen voornamelijk Microsoft producten moet gebruiken. In de Kimball lifecycle methodiek is een fase ingericht voor het selecteren en installeren van software. Deze fase verviel. Daarnaast was het gebruik van een DBMS<sup>8</sup> in deze fase al noodzakelijk om toegang te krijgen tot de brondatabase van Dynamics.

De verkregen werkomgeving fungeerde ook als een testomgeving voor de Dynamics release en als zodoende bevatte het enige demo data. Echter was er geen demo data beschikbaar voor de Tradeworld module. Een set aan test data was nog niet beschikbaar omdat de module nog niet getest is. Deze data heeft de afstudeerder zelf in de module toegevoegd. Hierdoor ontbrak representatieve data en dit vertraagde de analyse. Dit droeg wel bij aan de kennis van de module. Dit gegeven is geen wenselijke situatie, maar was geen reden voor de afstudeerder, om het project te staken.

Een bijkomende uitdaging was dat de module nog niet geconfigureerd was. Het configureren is noodzakelijk voor het correct werken van de interface en het kunnen invoeren van data. Door Dynamics deels zelf te configureren en test data toe te voegen

<sup>8</sup> DBMS staat voor DataBase Management System en wordt gebruikt om databases in te kunnen zien en te beheren

ontstond er meer van de data architectuur. Via de 'developer workspace' van Dynamics kon worden achterhaald waar de zojuist ingevoerde data werd opgeslagen. Dynamics AX is een generiek ERP systeem met een flexibel karakter wat het mogelijk maakt om voor elke klant deze anders te configureren en te implementeren. Het was dan ook niet te anticiperen hoe toekomstige gebruikers Dynamics zouden kunnen gaan gebruiken. Voorbeeld hiervan is dat een klant zijn magazijnen heeft onderverdeeld in verschillende geografische locaties (landen). Een andere klant heeft minder aantal magazijnen en heeft zijn organisatie structuur hierop aangepast. De onderliggende database tabellen konden als gevolg wel bij de ene klant geïmplementeerd zijn maar niet bij de ander.

Zowel Dynamics als Tradeworld zijn flexibel opgebouwd om grote verscheidenheid aan organisaties te kunnen ondersteunen. Bepaalde velden in de Dynamics worden wel door de ene organisatie gebruikt maar niet door de ander. Ook is er sprake van grote overlapping in functionaliteit tussen Tradeworld en Dynamics. Een voorbeeld hiervan is het creëren van verladingen (shipments). Deze functionaliteit is toegevoegd door Centric in Dynamics 2009 maar is inmiddels door Microsoft zelf al standaard toegevoegd in 2012. Beide zijn aanwezig in de huidige release.

De database is ook deels dubbel uitgevoerd maar data wordt niet in beide modules opgeslagen. Wel is er geconstateerd dat sommige database tabellen werden gedeeld. Het werkte verwarrend en vertraagde de analyse. Een ontwikkelaar omschreef de bovenstaande situatie als het volgende: "Tradeworld is geen module het is eigenlijk een bouwsteentje dat versmolten is tussen de andere bouwstenen (modules) van Dynamics".

Om toch enige structuur in de gegevensbronanalyse aan te brengen zijn een aantal Entity Relationship Diagrammen<sup>9</sup> (ERD) opgesteld. In deze diagrammen zijn de tabellen enigszins gegroepeerd (verschillende kleuren) per deelgebied of entiteit<sup>10</sup>. De relaties tussen de verschillende entiteiten wordt aangegeven met de 'Crow-foot'<sup>11</sup> notatie. Er is geconstateerd dat enkele attributen van tabellen, dezelfde naam kunnen hebben als attributen van een gerelateerd tabel. Deze attributen zijn dan niet automatisch foreign keys.

In figuur 5 zit u een voorbeeld van het ERD deelgebied: verladingen (shipments). In dit figuur kunt u zien dat vrijwel alle relaties optioneel zijn. De reden hiervan is dat er geen foreign keys of indexen zijn gedefinieerd in de Dynamics database. De referentiële integriteit<sup>12</sup> wordt enkel bewaakt in Dynamics zelf.

In bijlage C gegevensbronanalyse vindt u de overige ERDs van de verschillende deelgebieden.

<sup>9</sup> ERD staat voor entity relation diagram en toont een visuele representatie van de opbouw van een database.

<sup>10</sup> Met een entiteit wordt in deze context; een collectie van database tabellen dat tezamen bijvoorbeeld een product of order vormen.

<sup>11</sup> Crow-foot is een gestandaardiseerde notatie om relaties tussen tabellen in een ERD weer te geven.

<sup>12</sup> Referentiële integriteit betekent de waarborging van de consistentie tussen gerelateerde tabellen.





Figuur 5 ERD betreffende shipments

In de ERDs ziet u enkel de belangrijkste tabellen en attributen. Het was niet praktisch om deze in zijn volledigheid weer te geven. De tabel met inkooporders bevatte bijvoorbeeld al 50± database attributen. Enkel een beperkt aantal van de geanalyseerde tabellen zijn in de ERDs opgenomen. Deze ontbrekende database tabellen bevatte geen relevante gegevens ten aanzien van BI. Deze conclusie werd gebaseerd op het gegeven onderdeel uitmaakte van het bedrijfsproces verladings. In de ERDs zijn ook niet alle attributen weergegeven die wel relevant waren. De reden hiervoor was dat de ERDs anders te groot en daardoor onleesbaar werden. Enkel de belangrijkste attributen zijn opgenomen, om u een idee te geven van de inhoud van het betreffende tabel.

Naast het bronsysteem visueel in kaart te brengen zijn deze ook tekstueel beschreven met behulp van zogenoemde data dictionaries. In tabel 1 ziet u een dergelijke dictionary van het database tabel shipments. Dit database tabel komt ook voor in figuur 5. In de data dictionaries is niet opgenomen of een attribuut lege waardes mag bevatten. In de Dynamics database zijn alle lege waardes vervangen door een lege string. In tabel 1 ziet u de data dictionary van shipments (volgende pagina). Een aantal tabellen en attributen dragen de prefix: CTWINT, dit staat voor Centric Tradeworld International.

#### Dictionary tabel Shipment

Beschrijving: In deze tabel wordt informatie omtrent verladings (shipments) opgeslagen. Dit tabel is onderdeel van de Centric Tradeworld & International Shipment module en vormt de spil van de module. Aan een shipment (verlading) kunnen meerdere containers worden toegevoegd. Binnen deze containers kunnen vervolgens één of meerdere orderlines aan worden toegevoegd. Het Reclid zorgt voor uniciteit<sup>13</sup>. De dictionary:

<sup>13</sup> Primary keys worden gebruikt in tabellen om een regel in een tabel uniek te identificeren (uniciteit)

Kolomnaam	Type	Lengte	Betekenis	Opmerkingen
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED] [REDACTED] [REDACTED]	[REDACTED] [REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED] [REDACTED] [REDACTED]	[REDACTED] [REDACTED] [REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED] [REDACTED] [REDACTED]	[REDACTED] [REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED] [REDACTED] [REDACTED] [REDACTED]	[REDACTED] [REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED] [REDACTED] [REDACTED] [REDACTED]	[REDACTED] [REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED] [REDACTED] [REDACTED] [REDACTED]	[REDACTED] [REDACTED] [REDACTED] [REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED] [REDACTED] [REDACTED]	[REDACTED] [REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED] [REDACTED] [REDACTED]	[REDACTED] [REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED] [REDACTED] [REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED] [REDACTED] [REDACTED]	[REDACTED] [REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED] [REDACTED] [REDACTED]	[REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED] [REDACTED] [REDACTED]	[REDACTED] [REDACTED] [REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED] [REDACTED] [REDACTED]	[REDACTED] [REDACTED] [REDACTED] [REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED] [REDACTED] [REDACTED]	[REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED] [REDACTED] [REDACTED]	[REDACTED] [REDACTED] [REDACTED] [REDACTED]
[REDACTED]	[REDACTED]	[REDACTED]	[REDACTED] [REDACTED] [REDACTED]	[REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED]

Tabel 1 Data Dictionary tabel shipments

De dictionaries boden een goed naslag werk gedurende de opvolgende fasen. Het droeg bij aan het correct mappen van data uit het bronsysteem naar de datawarehouse. Daarnaast diende de dictionaires ook als meta data voor de attributen in de dimensies. In veel gevallen was met het oog niet vast te stellen of een database attribuut een vreemde sleutel bevatte of niet. Dit was met namen het geval in de module Tradeworld. De verantwoordelijke ontwikkelaars van de module hebben deze onvoldoende gedocumenteerd. Dit resulteerde in meer complexere query's om vast te stellen hoe de structuur van het bronsysteem in elkaar zat.

In figuur 6 ziet u een voorbeeld van een SQ<sup>14</sup> query voor het verzamelen van leveringsadressen. In de query ziet u de hoeveelheid joins die nodig waren om het volledige adres te op te bouwen. Dit schetst een goed voorbeeld van de genormaliseerde opbouw van de Dynamics database en de hoeveelheid aan tabellen.

```
select distinct
[is].SITEID as Sitename,
[is].NAME AS SiteNameDiscription,
il.INVENTLOCATIONID as WarehouseLocation,
il.NAME as WarehouseDescription,
addressCountry.COUNTRYREGIONID as WarehouseCountry,
[state].STATEID as WarehouseState,
district.NAME as WarehouseDistrictName,
logloc.DISTRICTNAME,
district.DESRIPTION as WarehouseDistrictDescription,
city.NAME as WarehouseCity,
logloc.CITY,
zip.zipcode WarehouseZipcode,
logloc.ZIPCODE,
zip.STREETNAME WarehouseStreetName,
logloc.ADDRESS as WarehouseCombinedAddress
from dbo.INVENTLOCATIONLOGISTICSLOCATION koppel
join INVENTLOCATION il on koppel.INVENTLOCATION = il.RECID
    join INVENTSITE [is] on il.INVENTSITEID = [is].SITEID
join LOGISTICSLOCATION locloc on koppel.location = locloc.RECID
    join LOGISTICSPOSTALADDRESS logloc on locloc.RECID = logloc.LOCATION
    left join LOGISTICSADDRESSCOUNTRYREGION addressCountry on
logloc.COUNTRYREGIONID = addressCountry.COUNTRYREGIONID
    left join LOGISTICSADDRESSSTATE [state] on logloc.[STATE] = [state].STATEID
    left join LOGISTICSADDRESSCITY city on logloc.CITYRECID= city.RECID or
logloc.CITY = city.NAME
    left join LOGISTICSADDRESSZIPCODE zip on logloc.ZIPCODERECID = zip.RECID or
logloc.ZIPCODE = zip.RECID
    left join LOGISTICSADDRESSDISTRICT district on district.RECID =
logloc.DISTRICT
where il.DATAAREAD = 'usrt' or [is].DATAAREAD = 'usrt'
order by [is].SITEID asc, il.INVENTLOCATIONID asc
```

Figuur 6 Voorbeeld Query

De complexiteit en omvang van de data architectuur binnen Microsoft AX 2012 R3 inclusief Tradeworld kan het beste worden uitgedrukt in cijfers. Namelijk; Dynamics bevat 96064± datavelden verspreidt over 6087± gebruiker definieerde database tabellen. Uiteindelijk zijn er 51 tabellen met 452 attributen verzameld die bruikbaar kunnen zijn voor een BI oplossing. Deze tabellen en attributen bevatten zowel meetwaarden als beschrijvende gegevens die, in een later stadium, dimensies en feittabellen kunnen vormen. Een aantal attributen en tabellen in de module Tradeworld worden enkel gebruikt om de werking van de module te ondersteunen.

Een ontwikkelaar verantwoordelijk voor de Tradeworld module heeft geassisteerd bij de gegevensbronanalyse. De ontwikkelaar heeft een deel van de ERDs gevalideerd.

<sup>14</sup> SQL staat voor structured query language en met deze taal kan een database worden bevraagd

## 4 Requirements

De volgende fase binnen dit project omvat het verzamelen van de requirements voor de dashboards. Voorafgaande aan dit project is door de opdrachtgever een opdrachtbeschrijving geformuleerd. In de opdrachtoomschrijving zijn door de opdrachtgever een aantal globale requirements beschreven. Deze beschrijving kunt u vinden in bijlage A. Deze requirements dienden als scope voor de gegevensbronnalyse, maar voor de requirements analyse het startpunt.

Door de gegevensanalyse vooraf uit te voeren ontstond er inzicht in de mogelijkheden van de data voor een BI oplossing. Het voorkwam namelijk dat er zogenoemde 'gouden bergen' werden beloofd die, met de huidige gegevens in Tradeworld, achteraf niet mogelijk waren. Ook leverde de analyse naar het bedrijfsproces verladingen meer inzichten op naar hoe een BI oplossing dit proces kan ondersteunen. Deze voorgaande analysis droegen tezamen bij aan de verzameling van de requirements.

De verzameling van de requirements<sup>15</sup> verliep op een incrementele wijze en werd gaande dit project gedetailleerder. Voor een initiële richting heeft een formeel interview plaats gevonden met een pre-sales consultant; dhr. Aldewereld. Dit interview vormde de basis voor de business requirements. Naast dit formele interview vonden er ook een aantal informele gesprekken plaats met de heren M. E. Helmer (opdrachtgever) en J..te Molder (implementatie consultant). Deze gesprekken droegen bij aan de verzameling van zowel de functionele als de non-functionele requirements.

Aan elke requirement is een identificatienummer toegewezen. Dit identificatienummer wordt gebruikt om de traceability<sup>16</sup> van een requirement aan te tonen. Als er in het ontwerp van de datawarehouse een bepaald aspect wordt behandeld dan kan dit worden gerelateerd aan een requirement. Dit is om aan te tonen dat er aan de desbetreffende requirement wordt voldaan.

De requirements zijn onderverdeelt in business, functioneel -en non-functionele requirements. Daarnaast zijn de functionele requirements geprioriteerd volgens de Moscow<sup>17</sup> regels.

### 4.1 Business requirements

Voor dit project zijn een aantal business requirements opgesteld. Deze zijn verzameld om na de implementatie van de Business Intelligence oplossing te kunnen valideren of dit project van toegevoegde waarde is voor de groothandels. Daarnaast zijn deze requirements opgenomen om de zaken die aanbod zijn gekomen in het interview met de pre-sales consultant recht aan te doen. De hier opgestelde business requirements kunnen niet worden gevalideerd omdat de implementatie fase geen onderdeel is van dit project als mede dat er geen klant is om deze te valideren. De business requirements:

Nr.	Omschrijving
<b>B001</b>	Centric wenst een generiek business intelligence oplossing te bieden aan haar klanten die gebruik maken van de tradeworld & international shipments module in Dynamics AX 2012 R3. Het betreft hier alleen het verladingen onderdeel binnen de desbetreffende module.
<b>B002</b>	Centric wenst een aantal dashboards aan haar klanten te bieden, die op een strategisch – en tactisch niveau informatie geven over het verladingen bedrijfsproces. De focus van Centric en haar klanten liggen op het tactisch niveau.

<sup>15</sup> Requirements vertaald in het Nederlands zijn eisen en wensen

<sup>16</sup> Traceability dit begrip wordt in deze context het verwijzen of aantonen dat er aan een wens (requirement) is voldaan

<sup>17</sup> MoSCoW is een prioritering techniek en staat voor: Must, Should, Could en won't have.

<b>B003</b>	De dashboards dienen generiek te zijn en moeten te gebruiken zijn voor meerdere klanten. Hierbij worden de ontwikkelingskosten over meerdere klanten verdeeld
<b>B004</b>	Klanten krijgen beter grip en overzicht op het bedrijfsproces verladings en kunnen deze als gevolg beter managen. Indirect ontstaan daardoor minder fouten of vertragingen in de leveringen. Met als resultaat dat de klanttevredenheid wordt verhoogd.
<b>B005</b>	De rapportages en dashboards dienen aan klanten van Centric beschikbaar worden gesteld. Daarnaast dienen deze ook verspreid te worden. Dit als doel om de informatievoorziening en inzicht in de bedrijfsprocessen te verhogen.
<b>B006</b>	Klanten van Centric dienen door doormiddel van analyse tools (bijvoorbeeld Excel of powerpivot) de eigen bedrijfsdata, binnen de Tradeworld & international shipments en van het verladings proces, op ad-hoc basis, te kunnen analyseren.
<b>B007</b>	Centric wenst aan haar klanten de mogelijkheid te bieden om ad-hoc rapportages te kunnen maken. Deze ad-hoc rapportages betreffen zaken over het verladings bedrijfsproces. De ICT afdeling is als gevolg niet meer verantwoordelijk voor het bieden of beschikbaar stellen van informatie. Klanten kunnen deze nu zelf verzamelen en delen.
<b>B008</b>	Het moet mogelijk zijn voor zogenoemde 'power users' om standaard rapportages te ontwikkelen. Hierbij gebruikmakende van tools om direct vanuit de datawarehouse informatie op te vragen. Hierdoor kunnen klanten gemakkelijker en sneller inspelen op nieuwe informatiebehoefte zonder hulp van een externe partij.
<b>B009</b>	Kosten reductie door het bieden van standaard rapportages die binnen de organisatie hergebruikt kunnen worden.
<b>B010</b>	Eenduidig overzicht en inzicht in haar eigen bedrijfsdata. Verschillende afdelingen/gebruikers rapporteren op dezelfde wijze naar het management. De informatievoorziening wordt hierbij betrouwbaarder en is beter te vergelijken.
<b>B011</b>	De meerwaarde van het Business Intelligence dashboard moet aantoonbaar zijn. Dit wordt uiteindelijk bepaald door de gebruiker van de dashboards zelf. Door het gebrek aan het klant perspectief wordt de gebruiker vervangen door de implementatie consultant.

## 4.2 Functionele requirements

De requirements die hier zijn opgesteld beperken zich tot de behoefte om inzicht te krijgen in de performance van het verladings proces. De requirements met de prioriteit 'must' hebben de hoogste prioriteit en dienen binnen dit project bereikt te worden. De prioriteit is in overleg met dhr. J. te Molder (implementatie consultant Dynamics) tot stand gekomen. Deze prioritering is gaande het project gewijzigd om het behalen van de minimale requirements tijdens dit project te garanderen. Daarnaast gaf het richting en kon er een onderscheid worden gemaakt tussen de hoofd –en bijzaken. De functionele requirements beperken zich op de dashboard delivery performance.

De volgende functionele requirements heeft de afstudeerder geformuleerd, gesorteerd op prioriteit:

Nr.	Omschrijving	Prioriteit
<b>F001</b>	De dashboards dienen de mogelijkheid te bieden om gegevens op verschillende detail niveaus aan de klant te tonen. Hierbij denkende aan zogenoemde 'drill-down' functionaliteit	Must
<b>F002</b>	Consultants moeten in staat zijn om het Business Intelligence dashboard in te richten.	Must
<b>F003</b>	Overzicht met shipments onderweg van loshaven naar warehouse dock per week.	Must

<b>F004</b>	De dashboards moeten KPI's <sup>18</sup> bevatten die, in één oogopslag de huidige situatie weergeeft.	Must
<b>F005</b>	Overzicht van het aantal shipments die worden verscheept, per week	Must
<b>F006</b>	Overzicht van shipments die op tijd, te laat of te vroeg aankomen per leverancier; week, Delivery performance KPI, hierbij is week het belangrijkste	Must
<b>F007</b>	Aantal aankomende containers per week	Must
<b>F008</b>	Gemiddelde vullingsgraad van containers per week, maand, jaar.	Could
<b>F009</b>	De groothandels kunnen ad-hoc analyses uitvoeren op een data kubus doormiddel van powerpivot. Daarmee maakt het ook mogelijk om data uit externe bronnen in analyses mee te nemen	Must
<b>F010</b>	Klanten van Centric wensen een data kubus tot hun beschikking te hebben om rapportages/analyses in Excel te maken.	Should
<b>F011</b>	De klant dient, doormiddel van filters of parameters de wijze waarop het dashboard of rapport gegevens toont, aan te passen.	Should
<b>F012</b>	Overzicht van het aantal shipments die worden verscheept, per jaar en maand	Should
<b>F013</b>	Overzicht van totale waarde verscheepte orderregels per maand en jaar.	Should
<b>F014</b>	Aantal aankomende containers per maand, week, dag per magazijn, en dock	Should
<b>F015</b>	De dashboards moeten worden getoond in Dynamics via sharpoint webparts. In de webparts moet het mogelijk zijn om een powerpivot kubus te bevragen.	Should
<b>F016</b>	Leveringsverschillen (aantallen) tussen de ontvangende partij of de verzendedepartij. Delivery performance.	Could
<b>F017</b>	De dashboards dienen meertaligheid te ondersteunen. Hierbij zijn de talen Engels en Nederlands de belangrijkste.	Could
<b>F018</b>	De dashboards dienen meerdere valuta's te ondersteunen. Hierbij is het wenselijk dat de gebruiker de mogelijkheid heeft om van valuta te wisselen. De belangrijkste valuta is dollar en euro.	Could
<b>F019</b>	Overzicht van herkomst leveranciers per land.	Could
<b>F020</b>	Aantal aankomende containers per aankomst haven en vertrekkende haven over week, maand, jaar	Could
<b>F021</b>	Overzicht van intercompany shipments	Won't

### 4.3 Non-functionele requirements

Voor dit project zijn een aantal non-functionele requirements opgesteld. De belangrijkste zijn hier genoteerd. De classificatie van de requirements zijn afkomstig uit de ISO/IEC 9126 normen.

Nr.	Omschrijving	Classificatie
N001	De datawarehouse dient uitbreidbaar en flexibel genoeg te zijn om nieuwe rapportage mogelijkheden te kunnen ondersteunen.	Onderhoudbaarheid
N002	Fouten in de gegevens in de datawarehouse moeten tijdig kunnen worden opgelost. Door fouten in het ETL proces te monitoren.	Onderhoudbaarheid
N004	De gegevens in de datawarehouse dienen correct te worden opgeslagen. Er zijn vrijwel geen discrepanties tussen de datawarehouse en het bronsysteem.	Betrouwbaarheid
N005	Rapportages dienen een volledig, eenduidig en correct overzicht te geven van de stand van zaken binnen de organisatie.	Betrouwbaarheid
N006	Gegevens in de database mogen niet onvolledig zijn dit betekent ook voor eventuele gekoppelde gegevens. Het mag de betrouwbaarheid	Betrouwbaarheid

<sup>18</sup> Kpi staat voor Key Performance Indicator



	van de rapportages niet beïnvloeden.	
N007	De rapportages dienen overzichtelijk te zijn voor gebruikers. De gebruikers kunnen doormiddel van parameters het rapport aanpassen of filteren zodat deze overzichtelijker wordt.	Bruikbaarheid
N008	De dashboards dienen een aantrekkelijke uitstraling te hebben. Dit verhoogd de kans dat de tradeworld module met de dashboards worden verkocht. Dit dient bereikt te worden door het gebruik van Powerview en PowerMap. Deze tools geven leveren extra grafische opties voor de analyse. Bijvoorbeeld met behulp van een scatter chart.	Aantrekkelijkheid
N009	Gebruikers dienen ad-hoc analyses met relatief gemak te kunnen maken. Zonder daarvoor complexe query's te hoeven bouwen.	Bruikbaarheid
N010	De beschrijving van de gegevens is begrijpelijk voor gebruikers. De gebruikers kunnen gemakkelijk gegevens selecteren voor analysis.	Bruikbaarheid
N011	De rapportages kunnen eenvoudig door de organisatie worden verspreidt.	Bruikbaarheid
N012	Het laden van een rapport mag hoogstens een paar seconden duren.	Efficiëntie

#### 4.4 Technische requirements

Naast de business, functioneel –en non-functionele requirements heeft de afstudeerder een aantal andere requirements opgesteld. Deze requirements zijn niet afkomstig van het bedrijf en hebben een hoger technisch aspect. Daarnaast zijn deze globaal toepasbaar op elk BI project. Deze overige requirements zijn samen met een ontwikkelaar van Dynamics tot stand gekomen. De ontwikkelaar simuleert, bij gebrek aan een klant, de ICT afdeling waar de BI oplossing wordt geïmplementeerd. Het valideren van deze requirements is deels door de ontwikkelaar uitgevoerd. Er is besloten om deze requirements apart te behandelen. Door het apart behandelen van deze requirements doet men recht aan deze belangrijke aspecten van de BI oplossing.

#### Bedrijfsdata

De requirements betreffende de bedrijfsdata:

Nr.	Omschrijving
<b>BG001</b>	Klanten van Centric dienen door doormiddel van analyse tools, data uit andere bronnen anders dan de module tradeworld & international shipments analyseren.
<b>BG002</b>	Binnen Dynamics zijn released products leidend. Producten die toegevoegd zijn maar nog niet zijn goedgekeurd, hoeven niet te worden meegenomen.
<b>BG003</b>	Gegevens die geen onderdeel uit maken van het inkoopproces dienen niet te worden verzameld.
<b>BG004</b>	De hoofdzakelijk entiteiten die moeten worden verzameld zijn: shipments, inkooporders, containers, losplaatsen, docks, agents, magazijn locaties, logistieke adressen en de groothandels. Ook attributen binnen deze entiteiten die geen relevantie hebben in het verladingen proces dienen niet te worden opgeslagen in de datawarehouse.
<b>BG005</b>	Gebruikers mogen geen hinder ervaren tijdens het verzamelen van gegevens.
<b>BG006</b>	Waar dit mogelijk is moeten de gegevens incrementeel worden ingeladen om de datavolumes tijdens het verzamelen te reduceren.
<b>BG007</b>	Het datavolume in de datawarehouse dient, waar dit mogelijk is, gereduceerd te worden.
<b>BG008</b>	Het verzamelen van transactionele gegevens zoals; voorraad of product verkopen, dienen niet te worden verzameld indien deze niet geaggregeerd kunnen worden.

<b>BG009</b>	De datawarehouse moet één keer per dag, via een geautomatiseerd proces, de data uit de bronsysteem laden in de datawarehouse.
<b>BG010</b>	Het totale ETL proces dient binnen een nacht voltooit te zijn

### Historie

De requirements voor het bijhouden van geschiedenis in de datawarehouse. Het verklaart hoe de datawarehouse moet omgaan met wijzigingen in het bronsysteem.

Nr.	Omschrijving
<b>H001</b>	Wijzigingen in adresgegevens, product prijzen en berekening van leveringstijden dienen bewaard te blijven.
<b>H002</b>	Updates die plaats vinden op ondersteunde data zoals container typen, product categorieën, proces fasen of classificatie van docks en havens moeten overschreven kunnen worden in de datawarehouse. Deze gegevens bevatten dezelfde waardes als die in het bronsysteem.
<b>H003</b>	De datawarehouse moet minstens vijf jaren aan geschiedenis kunnen bevatten.
<b>H004</b>	Gegevens die in het bronsysteem worden verwijderd dienen gedetecteerd te worden en in de datawarehouse als zodanig gemarkeerd te worden.

### Design constraints

De design constraints waar rekening moet worden gehouden gedurende dit project.

Nr.	Omschrijving
<b>DC001</b>	Het primaire bronsysteem is Dynamics AX 2012 R3. Het moet dus mogelijk zijn, met behulp van de juiste drivers, gegevens eruit te kunnen onttrekken.
<b>DC002</b>	De afdeling supply chain solutions ontwikkelt voornamelijk met Microsoft producten. Dit wordt deels door Microsoft afgedwongen. Het gebruik van de business intelligence stack van Microsoft staat voor dit project vast.
<b>DC003</b>	Voor het gebruik van powerpivot moet sql server 2012 r2 zijn geïnstalleerd.
<b>DC004</b>	Om gebruik te kunnen maken van DAX binnen Powerpivot dient de gegevensbron een tabular OLAP kubus te zijn. Anders moet men MDX gebruiken voor het browsen in de kubus.



## 5 Datawarehouse architectuur

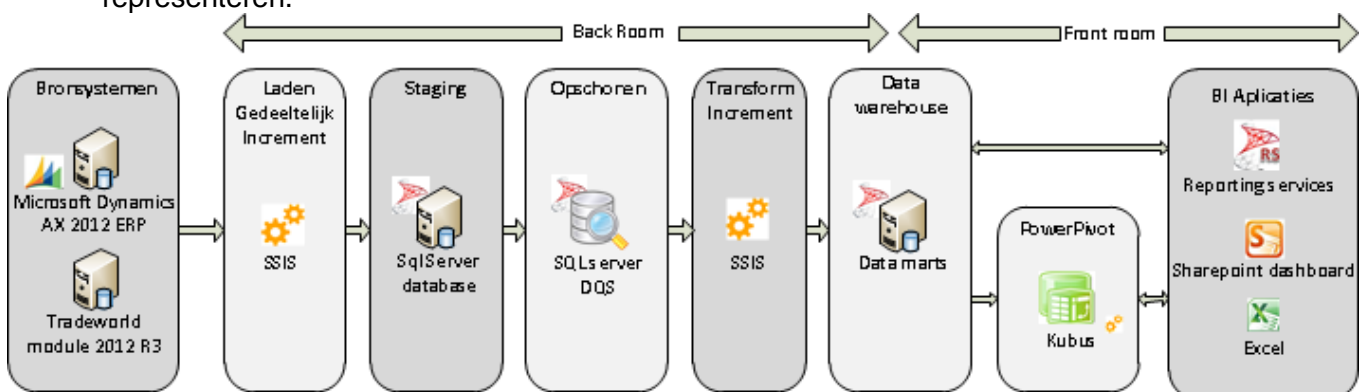
Op basis van de requirements heeft de afstudeerder een datawarehouse architectuur geselecteerd, aangepast en gerealiseerd. Er is besloten om de bus architectuur van Kimball als basis te nemen voor de datawarehouse architectuur. De gekozen project aanpak, de Kimball lifecycle, sluit goed aan bij het voorgenomen datawarehouse ontwerp. De architecturale beslissingen worden onderbouwd en er wordt gerefereerd naar de corresponderende requirement (d.m.v. het identificatienummer).

De bus architectuur van Kimball is onderverdeeld in een front –en backroom. De backroom is het technische en infrastructuur gedeelte van de architectuur. De frontroom zijn de verschillende interfaces of applicaties die de data uit de datawarehouse aan de gebruiker grafisch kan tonen. In het architectuur ontwerp zijn deze onderdelen los van elkaar behandeld. De backroom architectuur wordt in dit hoofdstuk behandeld. De data architectuur wordt beschreven in het hoofdstuk dimensioneel modeleren. De frontroom is echter nog niet beschreven in dit document. De dashboards zijn, ten tijden van het schrijven van dit verslag, nog niet af.

Deze scheiding in het architectuurontwerp is aangebracht om de grafische rapportage omgeving van de technische infrastructuur te scheiden. De datawarehouse met de bedrijfsdata kan meerdere zogenoemde frontends, of terwijl grafische interfaces bevatten waar gebruikers door de data kunnen navigeren. De backroom is echter statisch en verborgen voor de gebruikers en dient enkel om de data geschikt te maken voor rapportage en analyse doeleinden.

In dit project worden enkel Microsoft producten zoals; reporting services of Powerpivot gebruikt voor het bouwen van de database en de frontends (DC002). Het blijft echter mogelijk om ook andere software aanbieders, zoals Endeca of Qlikview te gebruiken om frontends te ontwikkelen. Dit toont de flexibiliteit en uitbreidbaarheid van de architectuur aan doordat deze een lagere koppeling heeft met de rapportage omgeving (N001).

In figuur 7 ziet u een overzicht van de datawarehouse architectuur met de ondersteunende producten en tools en waar de scheiding tussen de front –en backroom zich bevindt. Daarnaast kunt u zien welke rapportage tools er worden gebruikt om door de data in datawarehouse te navigeren en deze visueel te representeren.



Figuur 7 Datawarehouse architectuur

Binnen dit project is Microsoft Dynamics AX 2012 R3 (ERP systeem) met de tradeworld module als primair bronsysteem geïdentificeerd. De bedrijfsdata in Dynamics is opgeslagen in één Microsoft Sql 2012 database. De module tradeworld maakt gebruik van dezelfde database. De brondatabase heeft gedurende dit project

zijn huidige architectuur behouden. Er is enkel toegang verleend om door middel van Microsoft integration services de benodigde data eruit te onttrekken. In de backroom van het architectuur ontwerp zijn ook een aantal lagen, of terwijl omgevingen aangebracht. Data uit het bronsysteem wordt door een aantal omgevingen getransporteerd voordat deze getoond kan worden aan de gebruiker. Elke omgeving transformeert de data en draagt zorg voor de correctheid van de data. Het transformatie proces staat beschreven in hoofdstuk 7 ETL ontwikkelingsproces. In de volgende paragrafen de omschrijving van elke omgeving.

### 5.1 Staging omgeving

In de staging omgeving wordt de bedrijfsdata, die uit het bronsysteem is onttrokken, tijdelijk opgeslagen. De bedrijfsdata wordt zonder grote transformaties in een staging database opgeslagen. Het behoudt hierbij grotendeels dezelfde structuur als in het bronsysteem. De transformaties die worden uitgevoerd hebben te maken met het verkleinen van de ruimte die een attribuut in neemt. Door een extra fysieke laag in de architectuur aan te brengen wordt de koppeling tussen de datawarehouse en bronsysteem verlaagd. Zo wordt de datawarehouse beter afgeschermd van wijzigingen in het bronsysteem. Het wordt tevens eenvoudiger om nieuwe gegevensbronnen toe te voegen en op te schonen, zonder dat de bestaande ETL processen moeten worden aangepast (BG001). Een voorbeeld hiervan zijn twee externe Excel sheets met gegevens over havens (hubs) en adressen.

Daarnaast wordt het bronsysteem ontlast doordat zware transformaties op de staging database worden uitgevoerd (BG005). Met transformaties wordt het omzetten van het relationeel model van het bronsysteem naar het dimensioneel model bedoeld.

Het merendeel van de brontabellen worden in zijn volledigheid onttrokken. Er is echter besloten om een deel van de data uit het bronsysteem incrementeel in te laden. Het incrementeel laden van data vereist het gebruik van een aantal technieken die, bij het extraheren van data, vertragend kunnen werken. Een voorbeeld hiervan is het gebruik van hash<sup>19</sup> techniek om te controleren of een record is gewijzigd. Daarom worden enkel de tabellen shipments en inkooporders incrementeel in geladen. Deze twee tabellen bevatten grote volumes aan data maar kunnen bijvoorbeeld op basis van een order status onttrokken worden. Op deze manier blijft de primaire functie van de staging database behouden, maar wordt het datavolume beperkt (BG006). De staging database is op een andere SQL instantie geplaatst zodat deze niet alsnog het bronsysteem belast.

Door het hoge genormaliseerde karakter van het bronsysteem is de data over relatief veel tabellen verspreid. Om het ETL proces beter te kunnen managen is er besloten om gebruik te maken van Microsoft Integration Services (SSIS). Met SSIS zijn een aantal ETL workflows aangemaakt. Deze workflows tezamen zorgen voor een inzichtelijk en ordelijk ETL proces (N002). SSIS biedt ook de mogelijkheid tot monitoring en foutafhandeling van het gehele ETL proces. Voor de beschrijving het ETL ontwikkelingsproces verwijst ik u naar hoofdstuk 7.

### 5.2 Datawarehouse omgeving

De volgende laag in de architectuur is de daadwerkelijke datawarehouse. Kimball omschrijft dat een datawarehouse uit één of meerdere data marts kan bestaan. De data marts bevatten informatie deelgebieden van bijvoorbeeld een bedrijfsproces. Binnen dit project wordt voortsnog één data mart gebouwd met alle gegevens over het verladingen proces.

<sup>19</sup> Hash techniek is het omvormen van een record in een hexadecimaal format zodat deze later kan worden vergeleken met de hash van een binnenkomend record.

De data marts zijn, in tegendeel tot het bronsysteem, volgens het dimensioneel model opgebouwd. De transformatie van het relationele bronsysteem naar de dimensionele datawarehouse vindt plaats op de staging database. Daarnaast wordt de data opgeschoond en wordt er gecontroleerd of deze aan de geldende bedrijfsregels voldoen. Het opschonen van data is beperkt mogelijk omdat de datawarehouse generiek moest worden opgebouwd (B003). De geldende bedrijfsregels zijn per bedrijf verschillend.

De datawarehouse omgeving kan mogelijk gebruikt worden om standaard analyses via Microsoft Reporting Services (SSRS)<sup>20</sup> en SQL te bouwen. Het blijft daardoor mogelijk om zonder een kubus of powerpivot analyses te maken voor de gebruikers die de DAX<sup>21</sup> of MDX talen niet machtig zijn. Als er geen wensen zijn op het gebied van ad-hoc rapportering dan is het mogelijk dat de kubus in zijn geheel niet wordt geïmplementeerd. Dit is afhankelijk per klant. De kubus en de datawarehouse zijn beiden aparte omgevingen of lagen in de architectuur. De koppeling tussen de datawarehouse en de kubus is echter hoog. De dimensies en feiten in de warehouse vormen de basis voor de kubus. Zodra er structurele wijzigingen in datawarehouse plaats vinden, dan moet de kubus opnieuw worden opgebouwd.

### 5.3 Powerpivot

Naast de staging -en datawarehouse omgevingen is er een data kubus ingericht. De kubus vormt de laatste laag binnen de backroom architectuur. Een data kubus is ingericht omdat het op langere termijn onvermijdelijk is, dat zodra gebruikers na enige tijd met standaard rapportages en dashboards werken, deze uiteindelijk nieuwe wensen gaan ontwikkelen. De kans bestaat dat gebruikers de behoefte hebben om ook data uit andere bronnen te analyseren en hiervoor zelf een omgeving in te richten (Excel). Op de deze manier ontstaat er een wildgroei aan data die niet meer beheersbaar is. Om deze behoefte toch te kunnen faciliteren wordt gebruik gemaakt van Microsoft PowerPivot (F008, B005).

Het is mogelijk om met powerpivot het datamodel uit te breiden. Een voorbeeld hierbij is het toevoegen van een aggregatie of berekening die niet in het standaard datamodel zijn opgenomen (B006). Het zelfstandig verzamelen van gegevens en het flexibel buiten de standaard rapportages analyses maken wordt ook wel 'self service BI' genoemd. Een kanttekening hierbij is echter dat Centric of haar klanten enige kennis moeten bezitten van zowel Dax en Excel om deze uitbreidingen te realiseren. Microsoft beweert dat zogenoemde 'power users' in staat kunnen zijn om deze taken op zich te nemen, mits er enige training is aangeboden. Er is besloten om deze aanname van Microsoft over te nemen. Mede omdat de klanten van Centric enige kennis bezitten van Excel (N008). Daarnaast is er ook voldoende kennis binnen Centric op het gebied van Excel en van SQL databases. Een implementatie consultant van Dynamics moet in staat zijn een klant te assisteren met het opzetten en uitbreiden van een Powerpivot datamodel.(F002).

Er bestaan ook een aantal nadelen van het gebruik van Powerpivot. De voornaamste is de beperking van de opslag capaciteit. In powerpivot word de data in zogenoemde met behulp van VertiPaq in tabular formaat omgezet en ingeladen in het geheugen. Hoewel Powerpivot, afhankelijk van een aantal factoren, de data tot ~10% van de oorspronkelijke grote comprimeert blijft deze substantieel. Het is mogelijk dat dit een onevenredige hoge belasting op het geheugen van de server legt. Echter is de verwachting dat de datavolumes beperkt zullen blijven. Volgens een pre-salesconsultant bestaan de klanten van Centric uit voornamelijk middelgrote bedrijven.

<sup>20</sup> SSRS staat voor sql server reporting services en kan gebruikt worden om rapporten mee te bouwen.

<sup>21</sup> Dax is een script taal binnen powerpivot om bijvoorbeeld ad-hoc aggregaties te kunnen maken.

Powerpivot kan de datavolume die deze bedrijven genereren aan. Dit inclusief historische gegevens (N003).

Door het gebruik van verschillende datamarts per bedrijfsproces, bijvoorbeeld verladingen is de data gepartitioneerd in kleinere delen. Dit maakt het mogelijk om alleen data in het geheugen te laden waar men analyses op wil uitvoeren. Dit gegeven verhoogd de performance van de dashboards en rapportages (N011). Tevens wordt het datavolume gereduceerd.

In figuur ziet u dat de Powerpivot kubus zich in de frontroom bevindt. De reden hier voor is dat Powerpivot, naast dataopslag ook een grafische interface bezit. De interface kan worden benaderd vanuit Excel of Sharepoint. De data in de Kubus wordt op de Sharepoint server of binnen een Excel workbook geladen. De Powerpivot kubus heeft een sterke koppeling met de datawarehouse omgeving. Het datamodel van de datawarehouse vormt de basis voor het model van de PowerPivot kubus. Veranderingen in het datamodel van de datawarehouse betekend dat de kubus opnieuw moet worden gedeployed. Omdat het datamodel van de datawarehouse dimensionaal is opgebouwd is de Powerpivot kubus dat ook. De afstudeerder meent dat men hierdoor kan spreken van een HOLAP<sup>22</sup> kubus in tabular formaat.

## 5.4 Bi applicaties

De laatste omgeving of laag in de architectuur zijn de tools (grafische interfaces) waarmee gebruikers de data kunnen analyseren en vervolgens rapporteren. Deze tools behoren tot de frontroom van de architectuur. De tools die hier beschreven worden zijn door Microsoft ontwikkeld.

Het is de wens van Centric om de dashboards in Sharepoint te tonen (DC004). Om deze dashboards overzichtelijk en aantrekkelijk er uit te laten zien en tevens interactief, wenst Centric gebruikt te maken van Powerview en PowerMap (NF008). Deze tools bevinden zich in Powerpivot. Na onderzoek is gebleken dat het enkel mogelijk is om met een powerpivot kubus in tabular mode<sup>23</sup>, in combinatie met Sharepoint te gebruiken (MSDN). Het bouwen van een SQL Server Analysis Services (SSAS)<sup>24</sup> kubus is, door het gebruik van Powerpivot, binnen dit project te komen vervallen.

Met reporting services blijft het mogelijk om zonder een kubus rapportages te maken. De rapportages kunnen vervolgens in Sharepoint worden ingeladen. Doormiddel van stored procedures kunnen de rapportages deels door de gebruiker zo worden aangepast dat deze voor hem overzichtelijker wordt. Hoewel er geen gebruik gemaakt wordt van reporting services is het belangrijk dat deze mogelijkheid blijft bestaan. Het kan namelijk voorkomen dat een klant niet de juiste Sql Server versie heeft of wenst geen gebruik te maken van Sharepoint. De BI oplossing moest immers generiek zijn en dient door meerdere klanten te gebruiken zijn. (B001, B006, B008).

## 5.5 Kimball: Conform dimensies

Kimball beschrijft in zijn methodiek de practice van geconformeerde dimensies aan. Dit zijn dimensies die over de gehele datawarehouse dezelfde structuur of gegevens bevatten. Een dimensie is niet geconformeerd als er binnen de datawarehouse een dimensie bestaat, die dezelfde attributen heeft, maar bijvoorbeeld enkel een ander primaire sleutel bevat. Een dimensie is daarnaast conform als deze dezelfde betekenis heeft voor alle feit tabellen waaraan deze gekoppeld wordt. Voor informatie over het

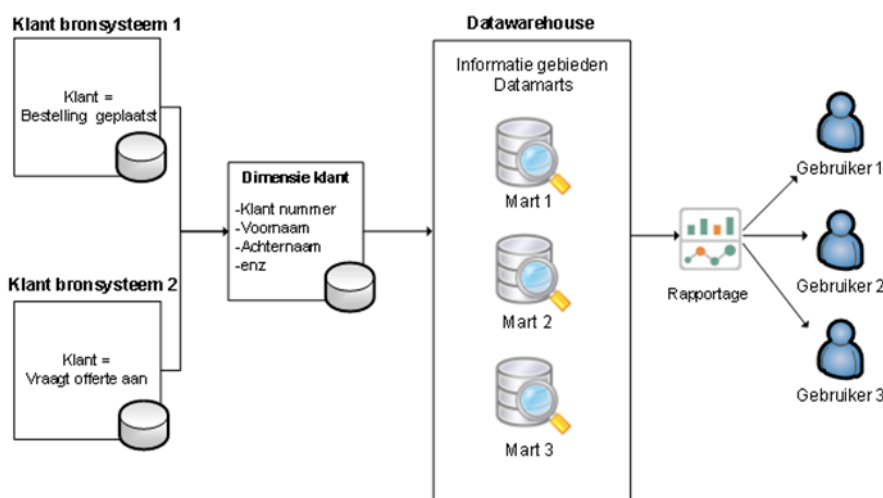
<sup>22</sup> HOLAP staat voor Hybrid Online Analytical Processing en betekend dat de kubus zowel relationele als dimensionele kenmerken heeft. Het datamodel van de datawarehouse is gebouwd als een snowflake.

<sup>23</sup> Tabular mode is een in memory database

<sup>24</sup> SSAS is een tool waarmee een data kubus kan worden gebouwd

dimensioneel ontwerp, verwijst ik u naar hoofdstuk 7: dimensioneel model ontwerp proces.

Geconformeerde dimensies kunnen alleen gebouwd worden als de organisatie dezelfde definitie van bijvoorbeeld een klant hanteert. Geconformeerde dimensies bevorderen consistentie rapportering, want de gehele organisatie kan de rapportages eenduidig interpreteren (N005). In Microsoft Dynamics zijn de klant gegevens al geconsolideerd in het ERP pakket. Hierdoor ontstaan al conforme definities van wat bijvoorbeeld een klant is omdat deze al wordt opgeslagen in dezelfde database van Dynamics. Maar in Dynamics is er veel ruimte voor het invullen van eigenschappen van klanten. De tabel waar klanten in worden opgeslagen bevat een groot aantal attributen. Het is dan ook afhankelijk hoe Dynamics is ingericht, hoe het klant gegeven bij een betreffend bedrijf is ingevuld. Een implementatie consultant heeft een standaard vastgesteld die in veel gevallen generiek zal kunnen zijn voor meerdere bedrijven (B003). In figuur 8 ziet u een voorbeeld van een architectuur met meerdere data marts.



*Figuur 8 Voorbeeld proces conformed data marts*

De data marts in figuur maken allen gebruik van één en dezelfde klant dimensie, of een kopie daarvan. Deze architectuur maakt het mogelijk om over meerdere data marts, consistente feit tabellen te bouwen. Binnen dit project is enkel één data mart gebouwd met dimensies die relevant zijn voor het verladingen proces. De dimensies die zijn gedefinieerd voor het verladingen proces kunnen ook worden ingezet in een ander bedrijfsproces. Een voorbeeld hiervan is het afhandelen van crediteuren, waar bij aan het inkoopfactuur feit de dimensie vendor (groothandel) is gekoppeld. Het betalen van binnengekomen goederen moet aan dezelfde vendor worden gekoppeld als waarbij deze is ingekocht. De feit tabellen shipments en inkooporder moeten aan dezelfde geconformeerde vendor dimensie worden gekoppeld.

Volgens kimball::

Eén definitie en betekenis entiteit	= Conformed dimensie
Datawarehouse	= Informatiegebieden (datamarts) tezamen.
Consistente rapportering	= Conformed dimensies + informatie gebieden.

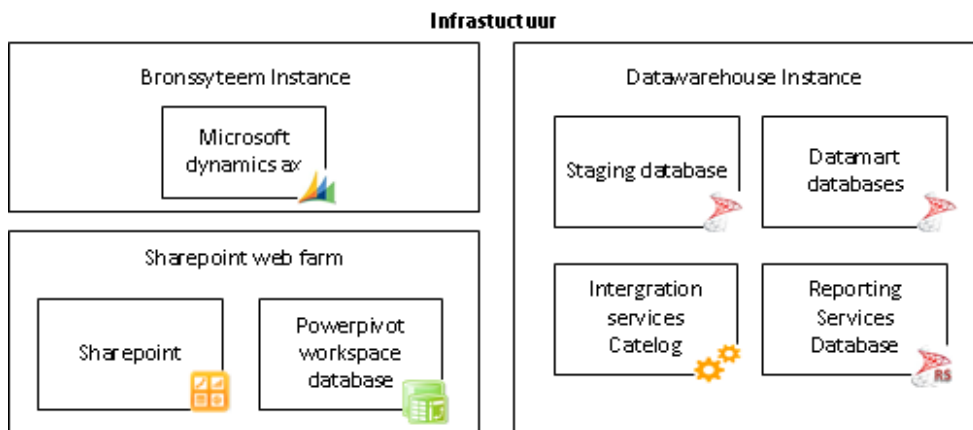
Om de conformiteit van dimensies te garanderen is in de ETL workflow één plek ingericht waar alle dimensies worden geladen. Het laden van de feit tabellen gebeurt apart van de dimensies en kan op meerdere plekken plaats vinden. Het laden van dimensies op één plek voorkomt het dupliceren van dimensies.



## 5.6 Realiseren architectuur

De afstudeerder heeft de beschreven architectuur geïmplementeerd binnen de eigen virtuele testomgeving. Het is echter de afstudeerder nog niet gelukt om Sharepoint te installeren. Binnen Sharepoint was het de bedoeling om de dashboards, doormiddel van PowerPivot webparts te tonen. Als alternatief is de PowerPivot workspace database (de kubus) tijdelijk ondergebracht in een aparte SQL Server instantie. Het is de intentie van de afstudeerder om Sharepoint met de PowerPivot dashboards bij de afstudeerzitting, geïmplementeerd te hebben.

Ter illustratie heeft de afstudeerder een grafische weergave gemaakt van de database instantie infrastructuur, zoals deze is geïnstalleerd en geconfigureerd in de eigen virtuele machine. In figuur 9 ziet u de instantie infrastructuur.



*Figuur 9 Fysieke infrastructuur databases*

Zoals u kunt zien is de datawarehouse in een aparte SQL Server instantie geplaatst. Als de datawarehouse geïmplementeerd wordt bij een klant, dient de instantie wel op een andere (virtuele) server geplaatst te worden.

## 6 Dimensioneel model ontwerp proces

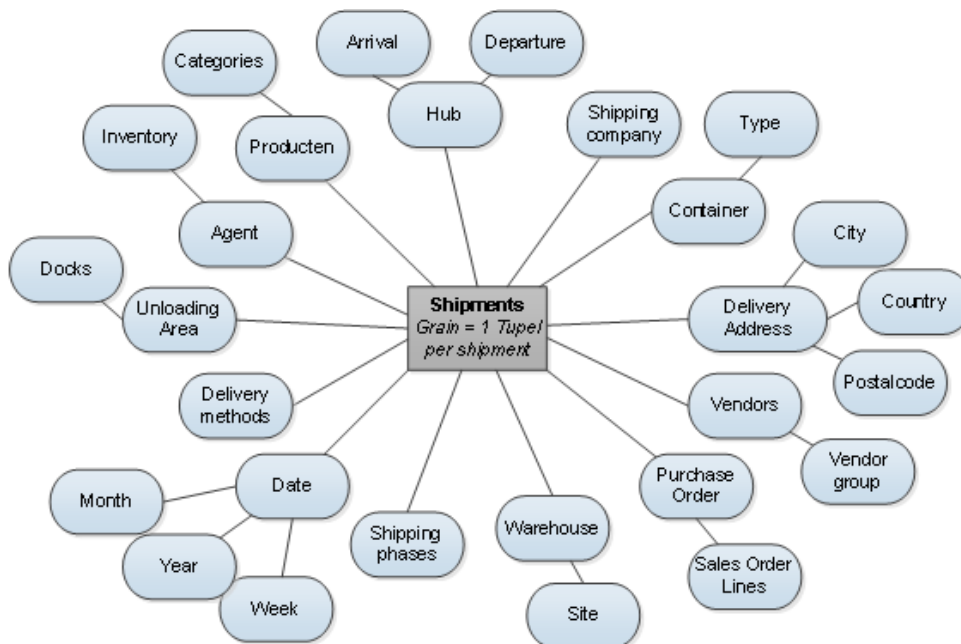
Na het selecteren, beargumenteren en realiseren van Kimball's datawarehouse architectuur, heeft de afstudeerder een datamodel opgesteld. Ook het ontwerpproces van een datamodel volgt de aanbevelingen van Kimball. Het dimensioneel model ontwerp proces verloopt hoofdzakelijk volgens het 4-stappen model. Het stappenmodel bevat de volgende stappen: selecteer het bedrijfsproces, stel de granulariteit vast, identificeer de dimensies en identificeer de feiten. In dit hoofdstuk wordt enkel het dimensioneel model voor de wekelijkse leveringsperformance beschreven. Met deze keuze worden de eerste twee stappen genomen, namelijk het bedrijfsproces is: verladingen en de granulariteit is: wekelijks.

Om meer structuur in het ontwerp proces aan te brengen is er voor gekozen om deze in drie delen op te splitsen. De delen zijn het conceptueel, logisch en fysiek model.

### 6.1 Conceptueel ontwerp

Na het vaststellen en beargumenteren van een geschikt database model is een conceptueel database ontwerp opgesteld. Voorafgaande aan het ontwerp proces zijn de mogelijk database entiteiten geïnventariseerd. De inventarisatie heeft plaats gevonden op basis van de verzamelde tabellen in de gegevensbronanalyse. In bijlage C vindt u de data dictionaries. Centraal bij de inventarisatie stond het bedrijfsproces verladingen met bijbehorende gegevens en de requirements.

Bij het in kaart brengen van mogelijke dimensies is gebruik gemaakt van een bubble schema. In figuur 10 ziet u het bubble schema.



Figuur 10 Bubble schema shipments

Het gebruik van het bubble schema wordt door Kimball, tijdens het ontwikkelen van het dimensioneel model voorgeschreven. Het werd in deze situatie ingezet om de begeleider op een hoog conceptueel niveau een stermodel voor te leggen. De begeleider kon zonder technische details verifiëren of alle entiteiten horende bij het shipment proces in kaart waren gebracht.

Dit schema droeg daarnaast ook bij aan het structureren van de inventarisatie van mogelijke dimensies. Het fungeerde deels als mindmap en vormde de basis van het conceptueel model.

## 6.2 Logisch ontwerp

Het logisch database ontwerp bouwt voort op het conceptueel ontwerp. Na de identificatie van de entiteiten zijn de verschillende entiteiten in dimensies ondergebracht. Tijdens het verzamelen van de data heeft er al een selectie plaats gevonden van de attributen van entiteiten die moeten worden opgenomen in de datawarehouse. Dit op basis van het gekozen bedrijfsproces verladingsen.

### 6.2.1 Ontwerp dimensies

Door de hoge mate van normalisatie binnen Dynamics zijn de entiteiten verspreid over meerdere tabellen. Een voorbeeld is het entiteit product, dat uit ~10 tabellen bestaat. Om van deze collectie van genormaliseerde entiteiten een aantal dimensies te creëren is er in plaats van normalisatie<sup>25</sup> juist een denormalisatie techniek gebruikt. Na de denormalisatie bestond bijvoorbeeld de dimensie warehouse uit enkel één tabel (dimensie). Door het verminderen van het aantal tabellen worden tegelijkertijd ook het aantal SQL joins verminderd. Dit resulteert in performance winst (N012). Daarnaast verlaagt het de complexiteit van query's (N009). In de bronanalyse zijn 51 tabellen verzameld. Na het denormalisatie proces zijn er 13 basis dimensies (tabellen) overgebleven.

De dimensies zijn ontworpen om in de breedte veel data op te slaan maar niet in de diepte. Dit wil zeggen dat het aantal tupels in een dimensie beperkt zal blijven (diepte) en dat een dimensie veel omschrijvende gegevensattributen moet bevatten (breedte). Entiteiten die niet aan deze voorwaarde kunnen voldoen zijn hoofdzakelijk transactionele gegevens. Een voorbeeld hiervan zijn shipments. Deze gegevens zijn dan ook in niet een dimensie opgenomen. In de datawarehouse zijn transactionele gegevens omgevormd tot feit tabellen, door deze al dan niet te aggregeren (BG008). Een voorbeeld hiervan is het opslaan van shipments geaggregeerd per week in plaats van dagelijks (F005).

Daarnaast is er besloten om meetwaarden, zoals gewicht of prijs niet in de product dimensie op te nemen. Er is tot dit besluit gekomen omdat deze meetwaarden relatief snel kunnen veranderen. Deze meetwaarden zijn van SCD type 2<sup>26</sup>. De betreffende dimensie groeit hiermee exponentieel (in de diepte) en daarmee ook het datavolume. Dit is onwenselijk. Deze meetwaarden worden dan ook in de feittabellen opgenomen in plaats van de oorspronkelijke dimensie.

Door het verplaatsen van meetwaarden naar de feit tabellen kan er mogelijk gegevensverlies optreden. Dit verlies kan ontstaan doordat bijvoorbeeld een product nog niet verscheept is. De meetwaarden van het betreffende product worden dan niet opgeslagen doormiddel van een feit. Daarmee gaat de geschiedenis ook verloren. De basis attributen van de product dimensie worden daarentegen altijd opgeslagen.

Binnen dit project is er besloten om voor deze snel wijzigende meetwaarde attributen geen SCD type 2 geschiedenis bij te houden. Het is geen requirement om analyses uit te voeren op bijvoorbeeld prijs of gewicht van individuele producten. Het opslaan van deze meetwaarden in dimensies is dus overbodig. Voor een ander bedrijfsproces kan deze requirement wel bestaan. Mocht dit het geval zijn dan kan er een SCD type 4<sup>27</sup> dimensie worden toegevoegd waarin deze meetwaarden apart worden ondergebracht.

<sup>25</sup> Normalisatie / denormalisatie is een techniek voor het samenvoegen of juist uit elkaar trekken van database attributen in een tabel

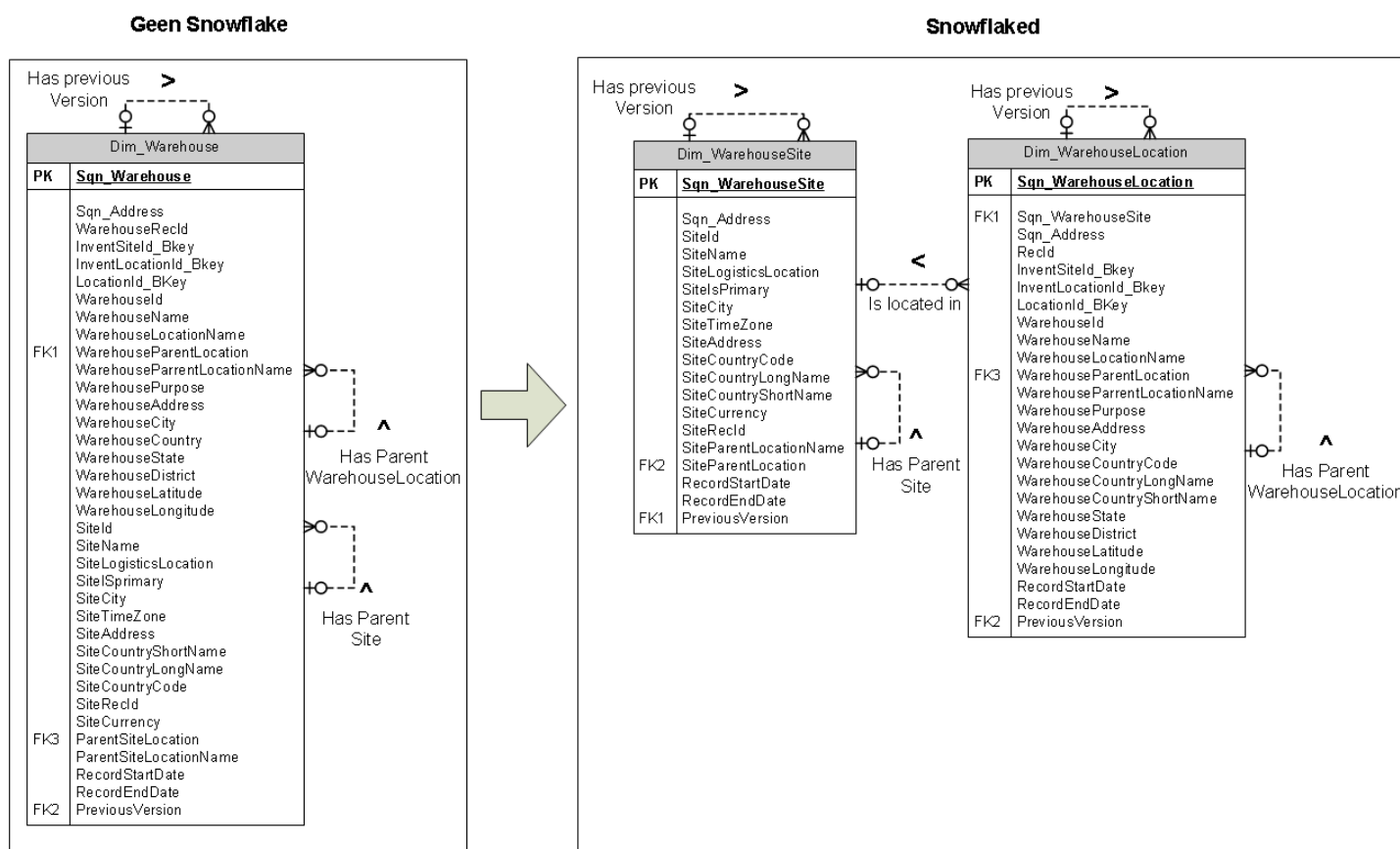
<sup>26</sup> SCD type 2 het bijhouden van geschiedenis, door zowel het oude als de gewijzigde record op te slaan.

<sup>27</sup> SCD type 4 is een ander variant voor het bijhouden van geschiedenis. De geschiedenis wordt bijgehouden in een apart tabel



Er wordt wel geschiedenis bewaard van adres wijzigingen en statussen van verladingen (Shipment phase). De geschiedenis van adres wijzigingen wordt met SCD type 2 bijgehouden. De overige attributen worden afgehandeld met SCD type 1.

Een uitbreiding op het traditionele dimensionale stermodel is het Snowflake model. Door het gebruik van het snowflake model moeten enkele dimensies weer deels naar de tweede normaalvorm (Codd) worden genormaliseerd. Voorbeeld hiervan is de warehouse dimensie met de redundante sites<sup>28</sup>. In figuur 11 ziet u een voorbeeld van de warehouse dimensie voor-en-na de normalisatie.



Figuur 11 Normalisatie warehouse dimensie

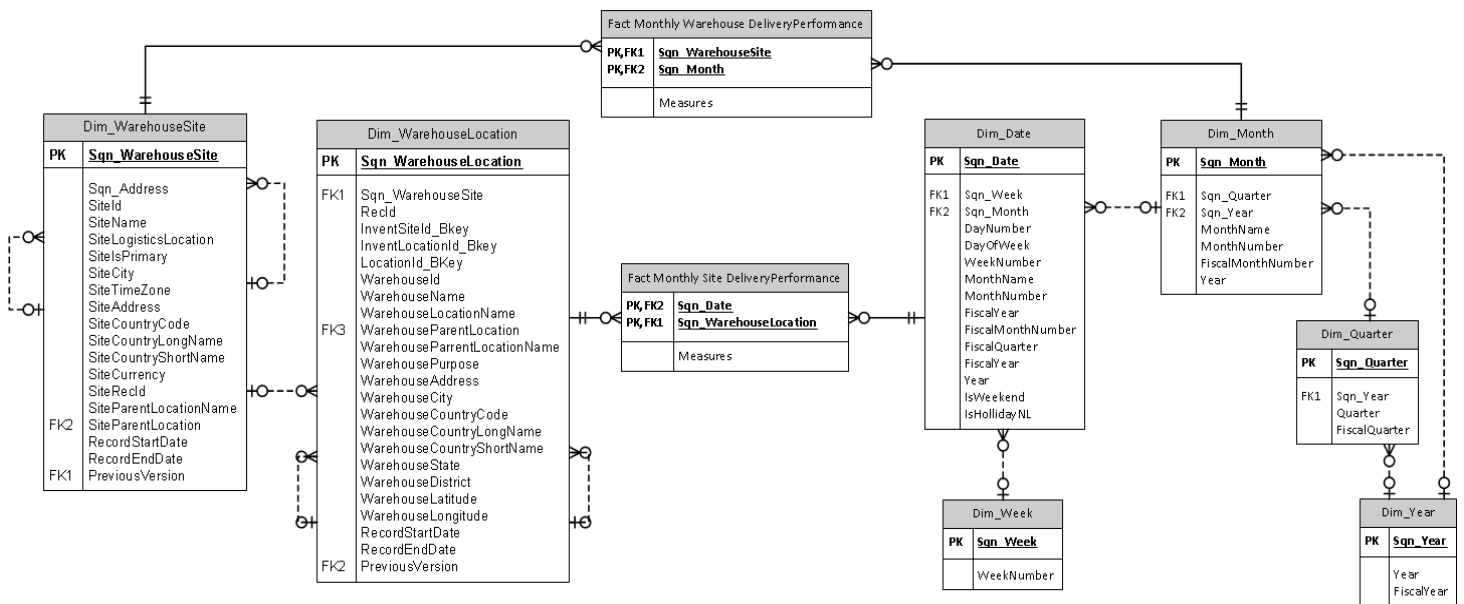
Het Snowflake<sup>29</sup> model wordt in dit project toegepast om meer flexibiliteit in het koppelen van dimensies met hogere granulariteit<sup>30</sup>. (N001). Een voorbeeld hiervan is een pre-geaggregeerd feit tabel met wekelijkse verlading rapportages. Aan dit feit wordt, in plaats van de hogere granulaire datum dimensie, de lagere week dimensie gekoppeld. De genormaliseerde (snowflaked) dimensies kunnen op deze wijze in andere datamarts aan feiten worden gekoppeld zonder de andere gerelateerde dimensies met non-relevante data daarbij betrokken worden.

<sup>28</sup> Een site is een geografische onderverdeling van meerdere magazijn locaties. Zie data dictionary

<sup>29</sup> Snowflake is een uitbreiding van het stermodel en ziet er letterlijk in het Nederlands uit als een sneeuwvlok

<sup>30</sup> Granulariteit is de mate van detail in gegevens. Letterlijk betekend het korreligheid

In figuur 12 ziet u een voorbeeld van twee feit tabellen van verschillende granulariteit.



Figuur 12 Meerdere granulariteit feiten

In figuur 12 zijn beide feit tabellen gerelateerd aan dimensies met een andere granulariteit. Het feit tabel maandelijkse levering performance is gekoppeld aan zowel de dimensie maand als de warehouse site. Het feit tabel dagelijkse levering performance is daar in tegen gekoppeld aan dag (date) en warehouse location. Op deze wijze ontstaan er feiten (aggregaties) op verschillende niveaus (F006).

Daarnaast voorkomt het gebruik van snowflake dimensies er voor dat er bij de analyse onnodige attributen en tupels met hoge granulariteit bij de berekening van de aggregatie worden betrokken. Een voorbeeld hiervan is de dimensie warehouselocation<sup>31</sup> dat een hogere granulariteit heeft dan de dimensie warehousesite. Het koppelen van de dimensie warehouseSite in plaats van warehouseLocation voorkomt dat er extra attributen worden meegenomen, die geen relevantie hebben bij het feit. Dit voorkomt verwarring bij gebruikers (N010).

Door meerdere hiërarchieën (snowflakes) in het model op te nemen wordt het eenvoudiger om zogenoemde 'drill-down' functionaliteit mogelijk te maken (F001). De snowflake dimensies zijn omgevormd tot hiërarchieën in de powerpivot kubus.

Als laatste wordt het datavolume beperkt door redundantie van data in de dimensies te voorkomen. Het gebruik van het snowflake model kent ook een aantal nadelen. Zo kan het de complexiteit van SQL query's vergroten. Dit omdat er additionele Joins moet worden gelegd om data te verzamelen. Echter door het gebruik van Powerpivot wordt de gebruiker van deze complexiteit afgeschermd. Powerpivot is namelijk in staat deze relaties automatisch te detecteren en deze te presenteren aan de gebruiker als één tabel (N009). De afstudeer meent dat het gebruik van het Snowflake model voor dit project, gezien de beschreven voordelen, gepast is.

De afstudeerder heeft 13 basis dimensies ontworpen. Naast deze basis dimensies zijn er ook additioneel 15 genormaliseerde snowflake dimensies toegevoegd. Een voorbeeld van een snowflake dimensie is warehouse locatie, die genormaliseerd is uit de basis dimensie warehouse. Het totaal aantal ontworpen dimensies komt daarop op 28 te staan. Dit aantal loopt gestaag op vanwege de gefineerde dimensie hiërarchieën.

<sup>31</sup> Zie data dictionaries

In het dimensioneel ontwerp heeft de afstudeerder een adres dimensie hiërarchie opgenomen. Deze dimensie wordt niet bij de feit tabellen gebruikt voor het bijhouden van het afzender en leveringsadres van de shipment. De dimensie warehouselocatie is zelf onderdeel van een hiërarchie. Een warehouselocatie kan worden onderverdeeld in meerdere locaties. Het adres van een warehouse (magazijn) heeft daarnaast andere attributen dan als een adres van een pand of persoon. De afstudeerder ziet de dimensie adres en warehouselocatie als aparte entiteiten. Het opnemen van warehouselocaties in de adres dimensie zal deze niet meer conform maken,.

In bijlage E vindt u twee diagrammen met daarin alle ontworpen dimensies en de gekoppelde snowflakes.

### 6.2.2 Mapping

De afstudeerder heeft een aantal dimensies ontworpen uit de verzamelde entiteiten en bijbehorende attributen uit de gegevensbronanalyse. Gegevens van deze entiteiten zijn van de staging database naar de datawarehouse overgebracht. Deze gegevens (attributen) konden niet één op één worden overgebracht zonder kwaliteitscontroles en transformaties uit te voeren. Dit proces wordt ook wel conversie genoemd.

Voorafgaande aan de gegevensconversie heeft de afstudeerder een aantal zogenoemde mappingschema's gemaakt. Deze schema's zijn gemaakt om aan te geven waar de attributen uit de staging database moeten worden opgeslagen in de datawarehouse. In figuur 13 ziet u een mappingschema van het entiteit warehouse locatie.

Source tabel	Source Attribuut	Bron type	Target Attribuut	Type	Bijzonderheid
Toegevoegd	Toegevoegd	/	Sqn_WarehouseLocation	Int	Primary / surrogate key SCD 0
Dim_WarehouseSite	Sqn_WarehouseSite	int	Sqn_WarehouseSite	Int	Foreign key SCD 2
Dim_WarehouseSite	SiteRecId	Bigint	WarehouseSiteRecId	Bigint	Alternative Foreign key SCD 2
InventSite	InventSiteId	nvarchar(30)	warehouseSiteId	nvarchar(30)	SCD 2
InventLocation	InventLocationId	Nvarchar(30)	WarehouseId	Nvarchar(30)	Alternative business key SCD 1
InventLocation	InventLocationDescription	Nvarchar(255)	WarehouseName	Nvarchar(60)	SCD 1
LogisticsLocation	LocationId	Nvarchar(30)	WarehouseLocationId	Nvarchar(30)	SCD 1
LogisticsLocation	Description	Nvarchar(255)	WarehouseLocationDescription	Nvarchar(60)	SCD 1
LogisticsLocationRole	Name	Nvarchar(30)	WarehousePurposeName	Nvarchar(30)	SCD 1
LogisticsLocationRole	Type	int	WarehousePurposeType	int	SCD 1
LogisticsPostalAddress	Address	Nvarchar(255)	WarehouseAddress	Nvarchar(150)	SCD 1
LogisticsAddressCountryRegion	CountryRegion	Nvarchar(255)	WarehouseCountryName	Nvarchar(50)	SCD 1
LogisticsAddressCountryRegion	CountryRegionId	nvarchar(30)	WarehouseCountryCode	nchar(3)	SCD 1
LogisticsAddressCity	Name	Nvarchar(60)	WarehouseCity	Nvarchar(60)	SCD 1
LogisticsAddressState	StateId	Nvarchar(10)	WarehouseState	Nvarchar(10)	SCD 1
LogisticsAddressDistrict	Name	Nvarchar(60)	WarehouseDistrict	Nvarchar(60)	SCD 1
LogisticsAddressZipcode	ZipCode	nvarchar(10)	WarehouseZipcode	nvarchar(10)	SCD 2
LogisticsPostalAddress	StreetName	nvarchar(60)	WarehouseStreet	nvarchar(60)	SCD 1
InventLocationLogisticsLocation	IsPrimary	int	WarehousesPrimary	Bit	SCD 1
InventLocationLogisticsLocation	IsPostalAddress	int	WarehousesPostalAddress	Bit	SCD 1
LogisticsPostalAddress	Longitude	Numeric(32,16)	WarehouseLongitude	Numeric(9,6)	SCD 1
LogisticsPostalAddress	Latitude	Numeric(32,16)	WarehouseLatitude	Numeric(9,6)	SCD 1
LogisticsLocation	ParentLocation	Bigint	WarehouseParentLocationId	Bigint	Unair relatie key SCD 1
Toegevoegd	Toegevoegd	/	WarehouseParentLocationName	nvarchar(30)	Unair Relatie attribuut SCD 1
InventLocation	RecId	Bigint	WarehouseRecId	bigint	Business key
InventLocation	DataAreald	Nvarchar(4)	DataAreald	Nvarchar(4)	SCD 0
InventLocation	Partition	Bigint	Partition	bigint	SCD 0
Toegevoegd	Toegevoegd	/	RecordStartDate	DateTime	SCD 1
Toegevoegd	Toegevoegd	/	RecordEndDate	DateTime	SCD 1
Toegevoegd	Toegevoegd	/	RecordHash	Binary	SCD 1

Figuur 13 Mapping schema WarehouseLocation

In dat mapping schema ziet u de herkomst van een attribuut afkomstig uit de staging database met de uiteindelijke bestemming in een datawarehouse dimensie. Bij bijzonderheden staat vermeldt van welke type geschiedenis (SCD) het betreffende attribuut is. U kunt zien dat het attribuut 'warehouseZipcode' van type 2 (nieuwe regel) is. De overige adresgegevens zijn van type 1 (overschrijven). De reden hiervoor is dat het aanpassen van een postcode een bewuste keuze is. Het veranderen van bijvoorbeeld de straatnaam kan bijvoorbeeld het gevolg zijn van een spellingsfout. Hoewel het mogelijk blijft dat postcodes ook foutief worden ingevuld is de kans echter kleiner. Het aantal nieuwe tupels als gevolg van het bijhouden van geschiedenis, wordt daarmee kleiner.

Naast het mappen van bestaande attributen uit het bronsysteem zijn er ook een aantal toegevoegd. Deze attributen zijn surrogaat sleutels, business sleutels en alternatieve sleutels. Attributen met de prefix; sqn zijn surrogaat sleutels en dienen als primaire sleutels voor de dimensies.

De business sleutels (natural keys) dienen als identificatie voor een bepaald gegeven in een dimensie. Ook hebben deze betekenis voor het bedrijf, zoals het attribuut inventLocationId (warehouseId). Dit attribuut is onderdeel van de gecombineerde primaire sleutel. Het identificeert, samen met de sleutels dataAreald en partition een warehouse record. Zie de datadictionary van inventlocation in bijlage D.. De sleutels dataAreald en partition hebben geen betekenis voor een gebruiker en zijn daarom geen goede kandidaten om als business keys te dienen. Hoewel deze onderdeel zijn van de primary key. De business key warehouseId is door dit gegeven niet uniek en

kon niet gebruikt worden om te controleren of het betreffende warehouse was gewijzigd in het bronsysteem. In de tabel inventlocation was een andere kandidaat primaire sleutel aanwezig, het attribuut ReclD. De beweegredenen om in plaats van één attribuut een gecombineerde sleutel te kiezen, is niet duidelijk. Om een warehouse locatie uniek vanuit het bronsysteem te kunnen identificeren is er voor gekozen om het attribuut ReclD in de dimensie op te nemen. (N006) Dit attribuut bevat enkel een groot getal van het type Bigint en heeft geen betekenis. De attributen partition en dataAreald zijn ook opgenomen in de dimensie. Deze attributen hebben alleen waarde als men een overzicht wenst te maken van alle aanwezige warehouse locaties in het bedrijf als geheel.

Er is niet voor gekozen om van het attribuut ReclD een primaire sleutel voor de dimensie te maken. Omdat dit attribuut in het geval van een SCD type 2 dimensie niet de uniciteit van het entiteit blijft garanderen. In elke dimensie is dan ook een surrogaat sleutel toegevoegd om te blijven garanderen dat een gegeven uniek geïdentificeerd kan blijven worden. Een duplicate business key vormt door het toevoegen van een surrogaat sleutel geen probleem meer.

Alle tabellen in het bronsysteem bevatten het attribuut ReclD. Helaas is er geconstateerd dat het ReclD attribuut niet consistent op dezelfde wijze wordt gebruikt. In de module Tradeworld zijn in alle gevallen het attribuut ReclD de primaire sleutel. Ook in dit geval blijft de uniciteit van een record niet gewaarborgd en is de toevoeging van een surrogaat sleutel noodzakelijk.

### 6.2.3 Conversie

Tijdens de gegevensbronanalyse heeft de afstudeer een indruk gekregen van de aanwezige bedrijfsdata. Door het ontbreken van 'echte' bedrijfsdata is de kwaliteit van de test data bekeken. De test data bevatte de standaard bedrijfsdata van het fictief bedrijf Contoso<sup>32</sup>. Het controleren van de gegevenskwaliteit had geen toegevoegde waarde voor dit project. De data wordt immers vervangen door 'echte' bedrijfsdata zodra deze wordt geïmplementeerd bij een klant. Hierdoor komen de domein –en bedrijfsregelcontroles te vervallen. De data in de module Tradeworld is door de afstudeerder, volgens een handleiding en generieke instellingen ingevoerd. Het ligt tevens buiten de scope van deze opdracht om te controleren of in de module alle bedrijfsregels worden gehandhaafd.

De afstudeerder meent echter dat het mogelijk is om eventuele kwaliteitsproblemen in de bedrijfsdata te anticiperen. Er is in de gegevensbronanalyse niet alleen gekeken naar de data maar ook naar de gebruikersinterface van Dynamics. Hieruit viel op te maken dat Dynamics bijvoorbeeld controleert of er duplicaten worden ingevoerd of dat vereiste attributen niet worden ingevuld. De bedrijfsregels worden door de implementatie consultants per bedrijf ingesteld. De interpretatie of een gegeven aan de bedrijfsregels is afhankelijk van implementatie.

Het is echter nog steeds mogelijk om bijvoorbeeld het attribuut InventlocationId<sup>33</sup> met daarachter een lege string (spatie) in te voeren. Met als gevolg dat er in feite duplicaten ontstaan. Of dat een niet bestaand adres wordt ingevoerd. Tijdens de integriteitscontroles is geconstateerd dat enkele tupels ontbraken in het koppelen van warehouses aan adressen/locaties. Het attribuutstype van locationId is een string en tegelijkertijd diende deze als foreign key in de tabel met adressen. De sleutel verwees naar een ander adres. Tijdens de gegevensbron analyse heeft de afstudeerder een warehouse toegevoegd met een zogenoemde 'trailing space'. Deze fout is dus tijdens de integriteitscontroles naar boven gekomen. Deze 'kleine' controles en correcties zijn

<sup>32</sup> Contoso is een fictief bedrijf bedacht door Microsoft

<sup>33</sup> Zie data dictionary

in het ETL proces uitgevoerd. In figuur 14 ziet u een voorbeeld van een eenvoudige conversie binnen integration services.

Derived Column Name	Derived Column	Expression	Data Type	Length
AddStreetNumber	<add as new column>	TRIM(Street + " " + StreetNumber)	Unicode string [DT_WSTR]	276
ConvertIsPrimary	<add as new column>	(DT_BOOL)IsPrimary	Boolean [DT_BOOL]	
ConvertIsPostalAddress	<add as new column>	(DT_BOOL)IsPostalAddress	Boolean [DT_BOOL]	

Figuur 14 SSIS Gegevens Conversie

De afstudeerder heeft aantal attribuut types heeft veranderd. De rede hiervoor was dat er meer ruimte gereserveerd wordt voor attributen dan dat er in de interface van Dynamics mogelijk was. Om het data volume te beperken is er besloten om de ruimte reservering te verminderen. Helaas heeft dit slecht beperkt effect want alle string attributen binnen Dynamics zijn Nvarchar<sup>34</sup>. Dit neemt extra ruimte in, maar biedt ook meer ondersteuning voor andere karaktersets. De attributen 'isPrimary' en 'isPostalcode' zijn in het bronsysteem integers maar zijn, na inspectie, boolean<sup>35</sup> waarden. Deze zijn geconverteerd naar het type 'bit', het equivalent van een boolean binnen SQL Server. Dit maakt het in de rapportages gemakkelijker om te filteren op bijvoorbeeld primaire warehouse locaties (F011).

De afstudeerder heeft ook twee additionele Excel sheets in de staging locatie toegevoegd. Deze Excel sheets worden gebruikt om onberekende gegevens in het bronsysteem aan te vullen. In het geval van hubs (havens) ontbreekt locatiegegevens. Deze ontbrekende gegevens worden uit het Excelsheet gelezen op basis van de naam of code van de haven. In het ETL proces wordt dit een 'lookup' genoemd. Door deze gegevens toe te voegen het mogelijk om de herkomst van een shipment te achterhalen (F003). De gegevens in beide Excel sheets moesten naar unicode string worden geconverteerd.

Dynamics ondersteunt meertaligheid. Tijdens het koppelen van alle gegevens kwamen een aantal duplicate tupels in het resultaat set naar voren. Een voorbeeld hiervan is het entiteit product, waarvan de naam in meerdere talen was opgeslagen. Het opnemen van het meer waardig attribuut productnaam resulteerde in duplicate producten in de dimensie. Er is besloten om het productnaam attribuut op te splitsen in meerdere attributen. Elk attribuut bevat de naam van het product in een andere taal. Op basis van het attribuut 'languageid', wordt bepaald in welke taal de productnaam is opgeslagen. Vervolgens wordt er doormiddel van een conversie regel in het ETL proces de productnaam in de juiste vertaling attribuut opgeslagen (F017).

De primaire sleutels van de dimensies is een zogenoemde surrogaatsleutel. Deze sleutel vervangt de originele primaire sleutel van het entiteit uit het bronsysteem. In Dynamics bestaat de primaire sleutel in veel gevallen uit meerdere attributen, of terwijl het is een geclusterde sleutel. De uniciteit<sup>36</sup> van een entiteit kan daarnaast ook nog worden afgedwongen door het attribuut: Recld. In een aantal gevallen waar het niet mogelijk was om een business sleutel te definiëren is de recld van het overkoepelende entiteit gebruikt. De keuze voor het niet gebruiken van het attribuut recid voor de primaire sleutel was het gegeven dat het datatype een bigint<sup>37</sup> was. Het gebruik van een bigint zal het datavolume in gekoppelde feit tabellen onverenigbaar laten toenemen. Omdat de indexen geplaatst op het feittabel ook in volume zullen toenemen (BG007).

<sup>34</sup> Nvarchar sql server data type dat niet- Latijnse alfabet ondersteund

<sup>35</sup> Boolean sql server data type dat alleen de waarde 'true' of 'false' bevatten

<sup>36</sup> Unicité betekend of een record uniek is binnen het zelfde tabel

<sup>37</sup> Bigint sql server data type dat een groter datatype ondersteund dat het type integer.



#### 6.2.4 Onderwerp feit tabellen

De volgende stap in het ontwerpproces is het ontwerpen van feit –en aggregaat tabellen. De afstudeer heeft een aantal feittabellen ontworpen voor het beantwoorden van standaard bedrijfsvragen binnen het verladingen proces op tactisch niveau (B002). In deze paragraaf worden de feit tabellen beschreven. In de bijlage (hier verwijzing) kunt u de feit tabellen vinden. Het feit tabel 'Warehouse product arrival' is opgenomen in dit verslag. Deze bedrijfsvragen zijn geïnventariseerd en beschreven in hoofdstuk 4 requirements.

De bedrijfsvragen zijn niet in één feittabel te verenigen omdat deze om een ander detail niveau vragen. Een voorbeeld hiervan zijn gegevens over shipments. Een shipment kan meerdere containers bevatten. Containers bevatten één of meerdere inkooporderregels van producten. Voor het beantwoorden van de vraag; aantal geleverde containers per magazijn (warehouse), kan men niet alleen het aantal shipments summeren. De individuele containers in de shipments worden bij deze aggregatie niet meegenomen. Door dit gegeven is besloten om meerdere feit tabellen te ontwerpen. Elk feit tabel ondersteunt een set aan bedrijfsvragen, dat getoond kan worden in een standaard rapport of dashboard. Elk feit is zo ontworpen dat alle meetwaarden dezelfde granulariteit hebben. In powerpivot kan de gebruiker een meetwaarde van een feit naar een hoger granulariteit niveau aggregeren. Een voorbeeld hiervan is het feit delivery performance op maand niveau in plaats van wekelijks (Should F012). De snowflake dimensies in de datawarehouse zijn in de PowerPivot kubus tot hiërarchieën getransformeerd. Hierdoor kan de gebruiker navigeren tussen verschillende detailniveaus. Dit worden ook wel 'drill-down/up' genoemd (F001, F009).

De afstudeerder heeft alleen aggregaties in de datawarehouse opgenomen die het meest frequent worden gebruikt. Volgens de pre-sales consultant dhr. Aldewereld worden veel bedrijfsvragen wekelijks gesteld. Aggregaties op week niveau ligt daarom het meest voor de hand. Het blijft mogelijk voor gebruikers om zonder pre-geaggregeerde feittabellen, via Powerpivot aggregaties te maken. De pre-geaggregeerde feiten dienen in dit geval enkel voor performance redenen. De feit tabellen zijn allen gekoppeld aan één of meerdere dimensies. Een dimensie kan gekoppeld worden aan meerdere feit tabellen. Om een overzicht te houden van het gebruik van de dimensies heeft de afstudeer gebruik gemaakt van de Kimball bus matrix. De matrix toont het gebruik van de conform dimensies bij verschillende bedrijfsprocessen en/of feiten. De matrix kan worden uitgebreid als de datawarehouse meerdere bedrijfsprocessen van analyses en rapportages moet worden voorzien. De voordelen van het gebruik van geconformeerde (consistente) dimensies werd door het gebruik van deze matrix zichtbaar. Het assisteerde bij het ontwerp van de feit tabellen.

De afstudeerder heeft drie feit tabellen ontworpen en gebouwd. In de matrix kunt u zien vervolgens zien welke dimensies er aan het feit zijn gekoppeld. De verschillende dimensie hiërarchieën zijn hierbij weggelaten. In figuur 15 ziet u de Kimball Bus Matrix.

Proces / feiten	Conformed Dimensions												
	Date	Warehouse	Vendor	Container type	Hubs	Shipping Agents	Shipping Type	Delivery method	Product	Unloading area	Shipment company	Address	Shipment phase
Warehouse Product Arrival	x	x	x	x	x	x	x	x	x	x	x	x	x
Warehouse Container Arrival	x	x	x	x	x	x	x		x	x	x	x	
Shipments	x	x	x		x	x	x		x	x	x		
Delivery performance	x	x				x				x	x		

Figuur 15 Kimball Bus Matrix

Naast het gebruik van de Bus matrix heeft de afstudeerder ook een aantal ERDs gemaakt van de datawarehouse om deze grafisch te weer te geven. Het grafisch modeleren van het totale database schema was onpraktisch vanwege de hoeveelheid relaties die ontstonden. De afstudeerder heeft het datamodel opgesplitst in zogenoemde analyse modellen waarin elk één feit tabel centraal staat. In dit verslag wordt enkel het feit tabel 'Warehouse Container Arrival' getoond. In bijlage **nummer** kunt u de overige analyse ERD modellen vinden.

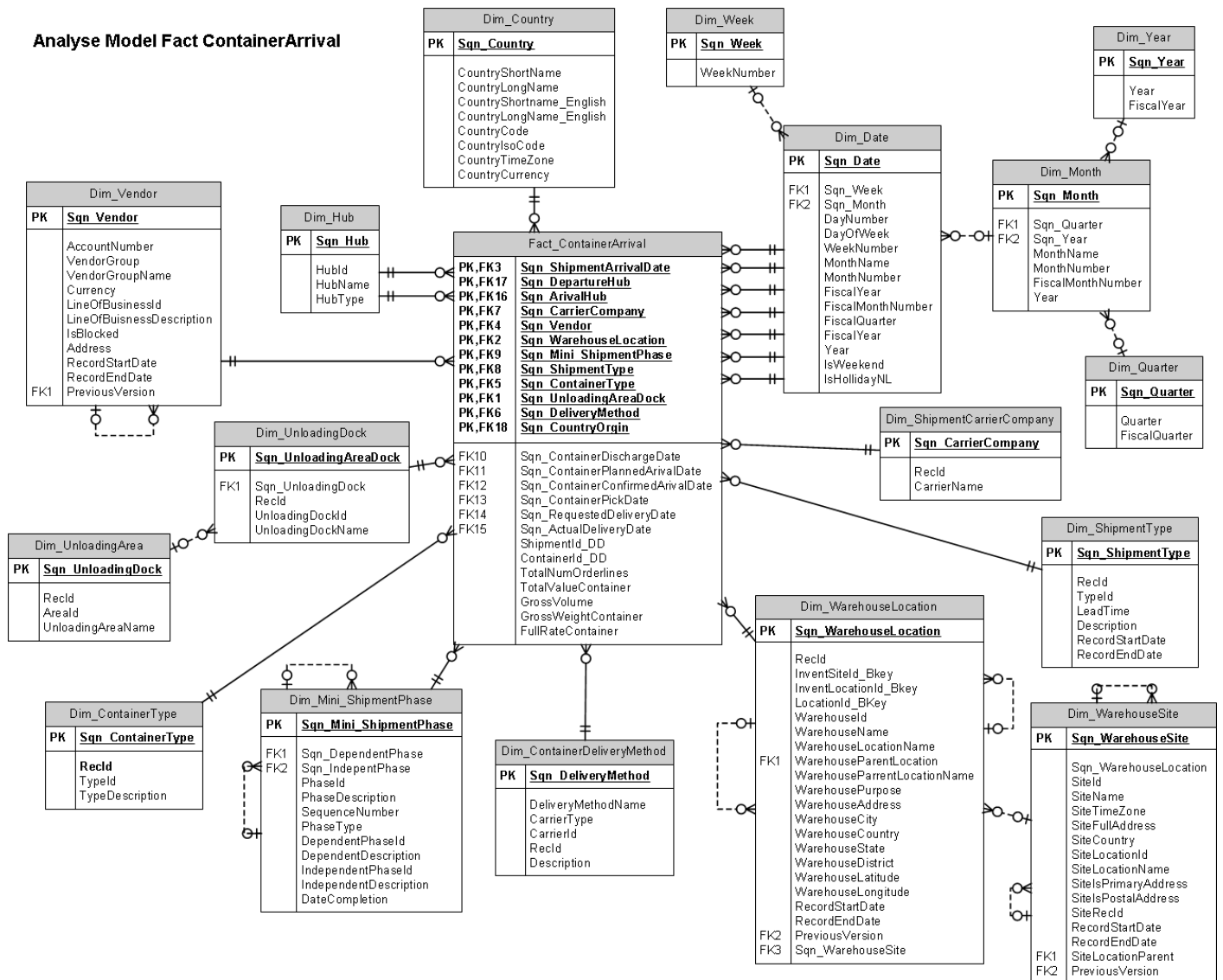
Om verlies van detailgegevens te beperken is er besloten om een feit tabel te ontwerpen van dagelijks aankomende goederen bij een warehouse (Product arrival). Het granulariteit niveau van het feit tabel is dermate hoog dat alle basis dimensies inclusief alle tupels aan dat feit gekoppeld kunnen worden (N004 N006). Dagelijkse feiten (handelingen), zoals het registreren van ontvangst goederen en daarbij betrokken gegevens worden in de datawarehouse opgeslagen. Het feit tabel 'Warehouse Product Arrival' is wel ontworpen maar niet gebouwd. Het heeft een te hoge granulariteit niveau voor de focus van de datawarehouse. De bedrijfsvragen liggen voornamelijk op het tactisch niveau. De operationele gegevens<sup>38</sup>, bijvoorbeeld over individuele orderregels, zijn hierbij minder relevant Dit valt dan ook buiten de scope van de opdracht. De afstudeerder heeft dit feit tabel ontworpen omdat het 'best practice' is om ten minste één feit met de hoogste granulariteit te bouwen (Kimball). In dit geval is dat order regel (producten). Gebruikers behouden hier bij de mogelijkheid om te aggregeren van container tot shipment (F001, F005, F007, N009). In een later stadium kan er alsnog besloten worden om dit feit tabel te bouwen, mocht hiervoor een gebruiker requirement ten grondslag liggen.

In figuur 13 ziet u het analyse model van 'container arrival. Voor dit model is een design pattern gebruikt. Dit pattern heet; accumulating snapshot (Kimball). In een datawarehouse worden feit tabellen niet geupdate. Een wijziging van een feit zou betekenen dat deze fout is geregistreerd. In een accumulating snapshot table worden feiten wel geüpdatet. Het feit 'container arrival' is een bedrijfsproces met vast gedefinieerde fases met een begin en einddatum zie Figuur . Het feit wordt telkens geüpdatet zodra één van de data wijzigt. De geschiedenis blijft bewaard omdat men het tijdsverschil tussen bijvoorbeeld de verwachte en de uiteindelijke leverdatum kan berekenen. Dit laatste is vervolgens als meetwaarde opgenomen in datzelfde feit. De overige attributen van het feit worden niet gewijzigd.

<sup>38</sup> Gegevens die gebruikers dagelijks muteren. Daarnaast zijn deze gegevens niet geaggregeerd en geven enkel inzicht op het operationele nieuwe van een proces.



## Analyse Model Fact ContainerArrival



Figuur 16 Snapshot feit container arrival

In het analyse model ziet u dimensies met unair<sup>39</sup> relatie verbanden. In de vendor dimensie verwijst de unaire relatie naar het vorige valide record uit de totale geschiedenis van het vendor record. De dimensie 'ShippingPhase' bevat een 'flattened' hiërarchie<sup>40</sup> van de verschillende fasen; zoals onderweg of uitgeladen, binnen het shipment proces. Deze fasen kunnen door elke klant worden anders worden ingevuld. Het is geen vaste generieke definieerbare hiërarchie. In de dimensie zijn alle mogelijke paden in de hiërarchie opgenomen, zodat deze door elke klant gebruikt kan worden (B001, B003). In bijlage F vindt u de Sql query die de betreffende hiërarchie omzet.

In het analyse model; feit container arrival (figuur 16), ziet u dat het feit tabel meerdere relaties heeft met de datum dimensie. Kimball noemt een dergelijk dimensie een 'role-playing dimension'. Een join uit te voeren op alle relaties zal verwarrend zijn voor gebruikers, omdat deze zogenoemde 'Alias'<sup>41</sup> dienen te gebruiken. Om dit te voorkomen, heeft de afstudeerder van elke relatie naar de datum dimensie, een view gebouwd. De gebruiker kan een join leggen op een view in plaats van de fysieke datum dimensie. Voor de gebruiker bestaat de datum dimensie uit meerdere tabellen. Bij het

<sup>39</sup> Unair of self referencing relatie betekent dat een dimensie tabel een foreign key bevat die verwijst naar een key uit datzelfde tabel.

<sup>40</sup> Flattend hiërarchie betekend letterlijk platgeslagen en wordt in deze context het denormaliseren van de verschillende tabellen waaruit de hiërarchie uit was omgebouwd.

<sup>41</sup> Alias betekend een synoniem

maken van powerpivot rapporten kan de gebruiker de view gebruiken in plaats van het niets zeggende generieke datum dimensie (N009). Voor het ontwerp van de views verwijs ik u naar bijlage F implementatie query's

### 6.3 Fysiek ontwerp

De laatste stap in het dimensioneel ontwerp proces is het beschrijven van de fysieke (opslag) structuur van de database. De afstudeerder heeft doormiddel van het gebruik van Sql Server Data Tools (SSDT) het datamodel omgezet naar een fysiek datamodel.

De afstudeerder heeft met behulp van SSDT ook de staging database gebouwd. De SQL scripts voor elk tabel, zijn ondergebracht in aparte scriptbestanden. De staging scripts zijn daarnaast ook weer ondergebracht in een ander zogenoemde SSDT packages. De scripts voor de indexen en constraints zijn eveneens apart opgeslagen. Op deze wijze heeft de afstudeerder structuur gebracht aan het database ontwikkelingsproces. Daarnaast wordt het managen en bijhouden van het fysiek implementatiemodel van zowel de staging als de datawarehouse database eenvoudiger. SSDT heeft ook de mogelijkheid om, via een grafische interface zelf tabellen te ontwikkelen. Op deze wijze kunnen de consultants eventueel zelf de database eenvoudiger uitbreiden (F002, N001). Binnen SSDT is gebruik gemaakt van SVN<sup>42</sup> om eventuele calamiteiten op te vangen. De database is uiteindelijk gedeployed naar een SQL Server 2012 instantie, hiervoor hoeft de gebruiker enkel op een knop te drukken.

Tijdens het ontwikkelen van de datawarehouse heeft de afstudeerder onderzocht welke opslag structuren gebruikt dienen te worden. Er is besloten om de tabellen in de staging database op te slaan volgens een heap. De relatie hiervoor is dat er snel en efficiënt data tijdelijk kan worden opgeslagen.

De feit tabellen in de datawarehouse zijn gepartitioneerd opgeslagen op basis van de datum key uit de datum dimensie. Bijvoorbeeld het feit tabel 'container arrival' groeit dagelijks relatief snel in volume. Selecties uitvoeren op grote database tabellen kan resulteren in performance problemen. Door het tabel te partitioneren per maand worden selecties op het feit tabel efficiënter (N012). Daarnaast wordt het onderhouden van de database data bestanden eenvoudiger. Verouderde data kan op deze manier per partitie gemakkelijker worden gearchiveerd (H003). In figuur 14 ziet u het SQL script voor het creëren van een partitie functie en schema. In dit voorbeeld wordt er per maand naar een andere filegroep geschreven.

```
--Partitie functie t/m september
CREATE PARTITION FUNCTION [PartitionFunc_FactShip](datetime)
AS RANGE LEFT
FOR VALUES (
N'2014-04-01T00:00:00',
N'2014-05-01T00:00:00',
N'2014-06-01T00:00:00',
N'2014-07-01T00:00:00',
N'2014-08-01T00:00:00',
N'2014-09-01T00:00:00')

-- Partitie schema naar filegroepen
CREATE PARTITION SCHEME [PartitionScheme_FactShip] AS PARTITION
[PartitionFunc_FactShip]
TO (ShipApril2014, ShipMay2014, ShipJune2014, ShipJuly2014, ShipAugust2014, ShipSeptember2014, [PRIMARY])
```

Figuur 17 Partitioneren van feit\_shipment

In bijlage F implementatie query's kunt u een aantal voorbeelden van indexen, secundaire indexen en constraints.

<sup>42</sup> SVN staat voor subversion en wordt gebruikt voor het back-uppen van bestanden

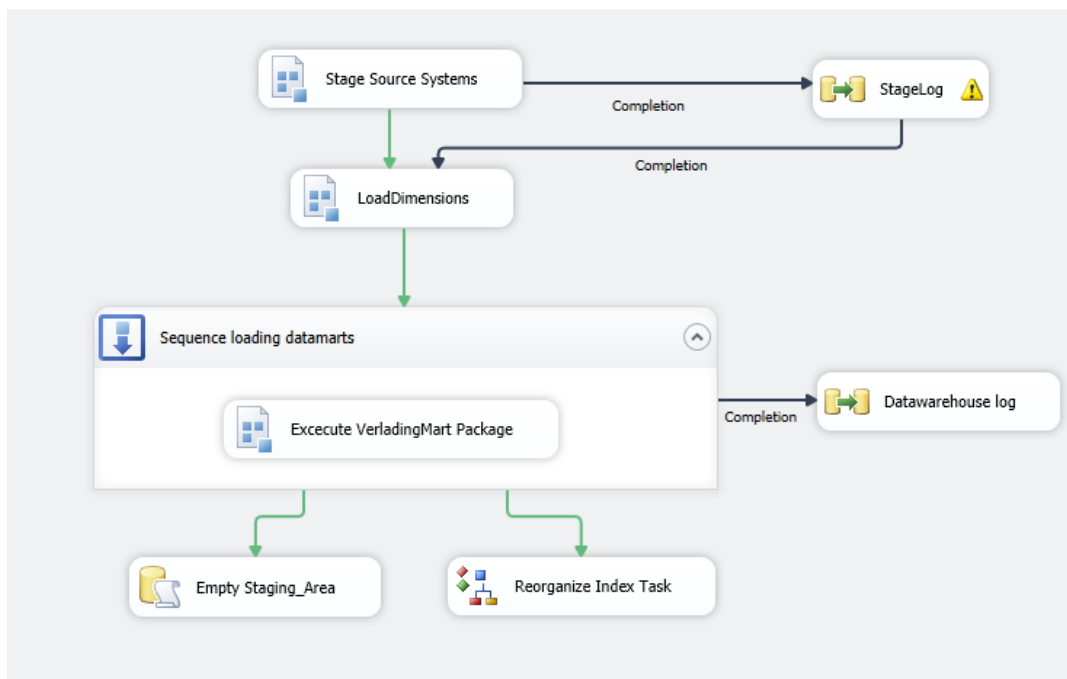
## 7 ETL ontwikkelingsproces

Volgende op het dimensioneel ontwerp proces is het ontwikkelen van de daadwerkelijke ETL dataflows. Voor het ontwikkelen van de ETL dataflows is SQL server integration services 2012 (SSIS) van Microsoft gebruikt. Dit hoofdstuk beschrijft in het kort, het ETL ontwikkelingsproces in SSIS en welke technieken hierbij gebruikt zijn.

### 7.1 ETL workflow ontwerp

De afstudeerder heeft de SSIS dataflows hiërarchisch en modulair opgebouwd. De ETL dataflows van de staging en de datawarehouse databases zijn beide in aparte packages (containers) geplaatst. Op deze manier kunnen de dataflows onafhankelijk van elkaar worden onderhouden en uitgevoerd.

In het ETL onderwerp is een zogenoemde 'master' package toegevoegd. Deze package staat aan de top van de dataflow hiërarchie en vormt de 'main' flow van het ETL proces, zoals beschreven in het architectuur ontwerp. Op dit niveau worden alle packages met de dataflows in volgorde uitgevoerd. In Figuur 18 ziet u de SSIS master package.



Figuur 18 Master package

Hierin kunt u zien dat er meerdere packages worden aangeroepen en dat de verschillende packages in verbinding met elkaar staan. Zodra een package is afgerond, wordt de volgende uitgevoerd. De staging package wordt als eerste uitgevoerd, vervolgens worden de dimensies geladen en als laatste de, datamarts packages met feit tabellen. Als de datawarehouse succesvol is geladen, wordt de staging database geleegd.

Om conformiteit van dimensies te garanderen, is er slechts één package ingericht voor het laden van de dimensies. Het centraal bijhouden van de dimensies verhoogd de overzichtelijkheid en de onder houdbaarheid van de dimensies (N001). Het reduceert dat er een wildgroei van dimensies kan ontstaat (B010).

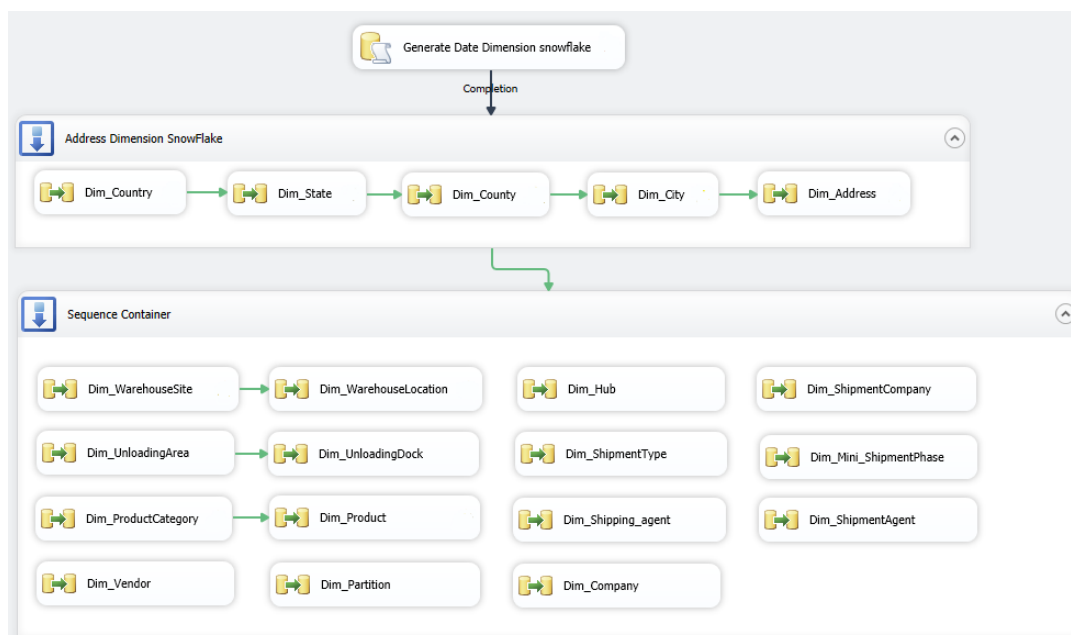
Het laden van de feit tabellen vindt apart van de dimensies plaats. Omdat de feit tabellen te kunnen laden moet er logischerwijs eerst de dimensies worden geladen.

Een ander doel van deze scheiding is echter dat de koppeling tussen dimensies en feiten, op deze wijze, kleiner wordt. Feit tabellen worden gekoppeld aan dimensies, dimensies worden niet gekoppeld aan feiten. De afstudeerder bedoelt hiermee dat er in feite oneindig aantal feit tabellen (analyse modellen) kunnen worden gemaakt door deze te koppelen aan verschillende constellaties van dimensies. Door het gebruik van de Kimball bus matrix kan bijgehouden worden welke combinaties van feiten tabellen zijn gebouwd.

Door dit gegeven kan men in feite spreken van een één op veel relatie tussen de SSIS packages verantwoordelijk voor het laden van feiten en dimensies. De SSIS package is om deze reden in een zogenoemde 'sequence' container geplaatst. De package voor het laden van dimensies hoeft slechts één keer worden gekoppeld aan de container. Binnen de container kunnen er meerdere packages worden gebouwd voor het laden van feit tabellen. Deze worden allen uitgevoerd zodat de dimensies zijn geladen.

Het volgende niveau binnen de SSIS dataflow hiërarchie zijn de packages waar alle dataflows bij elkaar gegroepeerd zijn. De afstudeerder heeft op dit niveau twee packages gebouwd, de staging en de datawarehouse package. In de staging package worden de tabellen shipment en orderlines op basis van een 'modified' datum incrementeel ingeladen (BG008).

In figuur 19 ziet u de package met de verschillende dimensie dataflows. Als een dimensie onderdeel is van een hiërarchie dan wordt de betreffende dimensie in volgorde geladen.

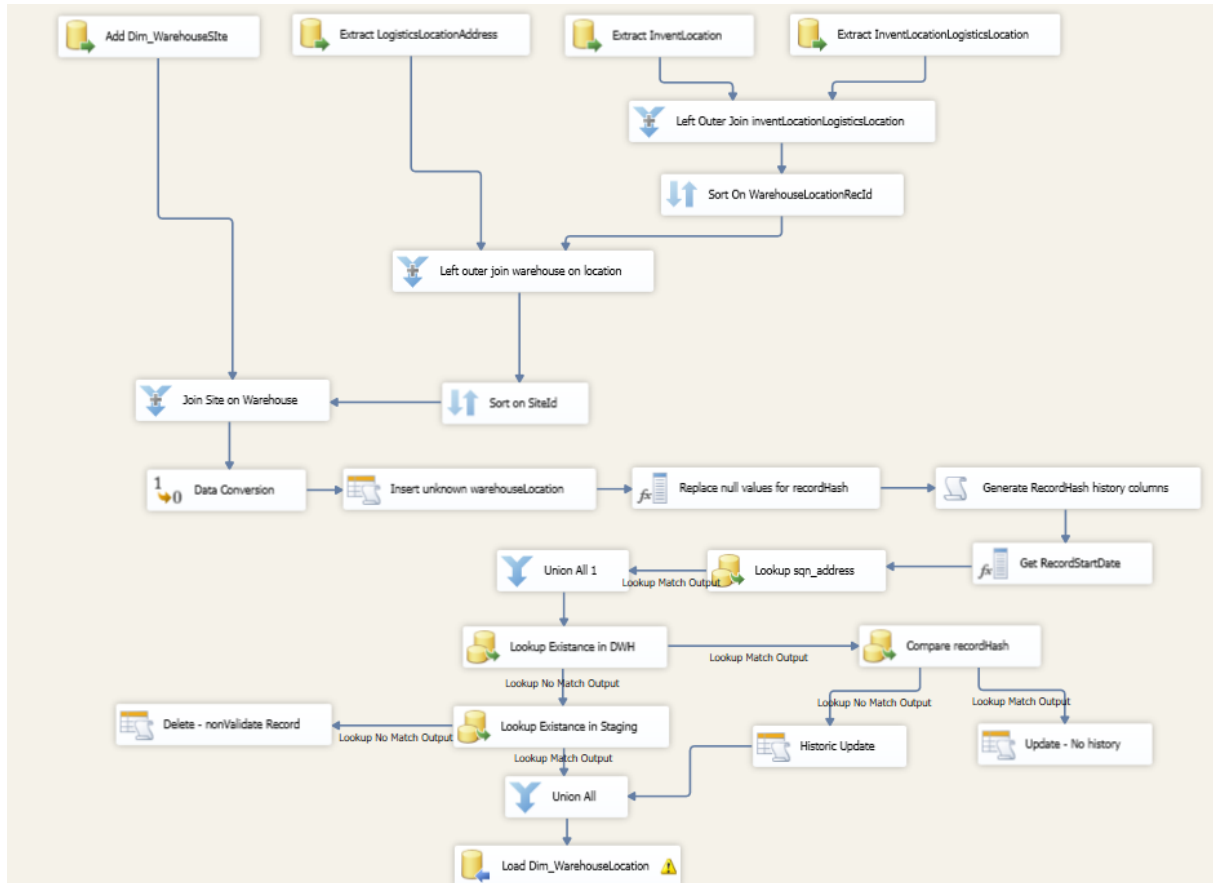


Figuur 19 Dataflow laden dimensies

Om de dimensies generiek te maken voor meerdere klanten, is niet afgedwongen als een foreign key lookup geen resultaten oplevert (B001). Een voorbeeld hiervan is dat bepaalde klanten geen gebruik maken van warehouseSites. Het ontbreken van een site levert geen semantische fout op (Zie hoofdstuk 10 testen datawarehouse). De betreffende dataflow wordt als gevolg, succesvol afgerond, zonder dat er data is overgebracht.

Voor het laden van de datum dimensie wordt niet data onttrokken vanuit het bronsysteem. De afstudeerder heeft een SQL stored procedure<sup>43</sup> geschreven voor het genereren van de datum dimensie hiërarchie. Ook de datum dimensie is generiek opgebouwd. Aan de stored procedure worden een aantal parameters meegegeven. Met deze parameters kan worden ingesteld wat de eerste dag van de week is en wanneer het fiscaal jaar begint (B001). Deze stored procedure kunt u vinden in Bijlage F.

Het laagste niveau binnen de ETL flows zijn de uiteindelijk dataflows met de logica voor het converteren en transformeren van data. In figuur 20 wordt de dataflow voor het laden van de warehouselocatie dimensie getoond.



Figuur 20 SSIS Flow dimensie warehouseLocation

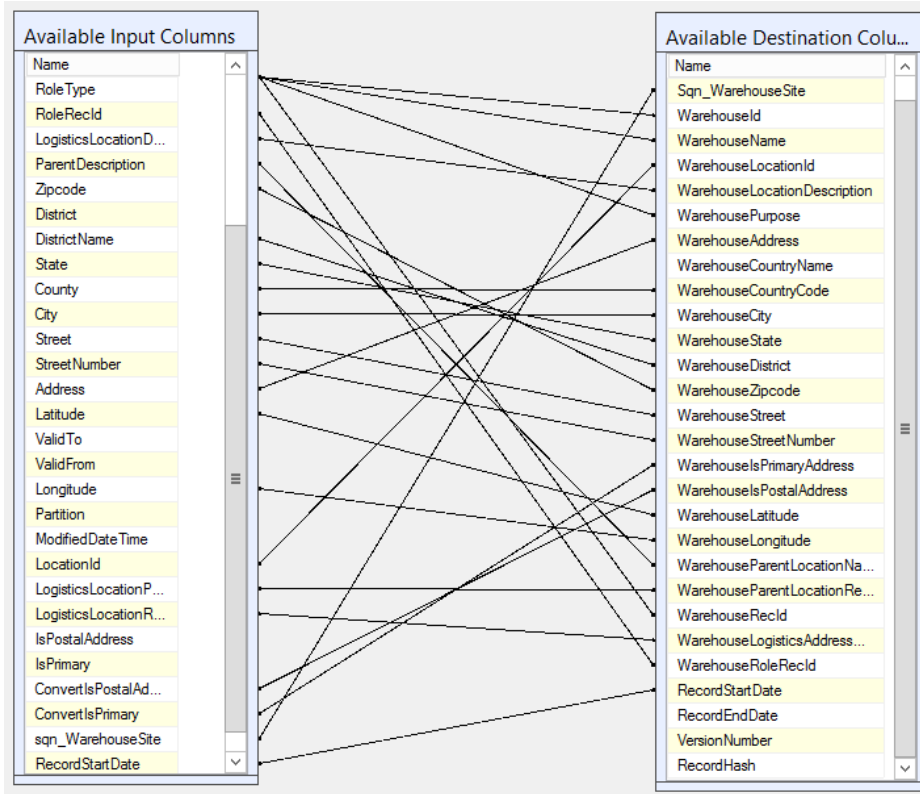
Tijdens het ontwerpen van de dataflows is rekening gehouden met performance aspecten. Zo wordt er bijvoorbeeld enkel attributen geselecteerd van een tabel die ook daadwerkelijk nodig zijn. Als het mogelijk was om het aantal tupels te verminderen is de meest discrimineerde 'where clause' gebruikt (BG007). Voor het joinen van twee dataflows (gegevens set) is, waar dit mogelijk was, eerste beide dataflows op dezelfde wijze gesorteerd op de foreign of de primary key. Door het joinen van twee gesorteerde dataflows kan SSIS, in plaats van een hash een merge join<sup>44</sup> gebruiken. Dit resulteert in een efficiëntere join query. Bij het laden van data wordt een 'bulk insert'<sup>45</sup> gebruikt zodat er minimaal, binnen SQL server, gelogd wordt. Als laatste heeft de afstudeerder ervoor gezorgd dat SSIS zoveel mogelijk dataflows parallel en in 'memory' uitvoert. Dit is bereikt door alle dataflows gescheiden van elkaar te laden en onderlinge afhankelijkheden te vermijden. Door deze maatregelen wordt het uitvoeren van het ETL proces efficiënter (BG010).

<sup>43</sup> Stored procedure

<sup>44</sup> Hash join is binnen SSIS minder efficiënt dan een merge join

<sup>45</sup> Bij een bulk-insert wordt er minimaal gelogd

De afstudeerder heeft in hoofdstuk 6.2 het gegevens mapping proces beschreven. In SSIS zijn deze mappings geïmplementeerd. Aan het einde van elke dataflow wordt de gegevens set gecombineerd en in memory opgeslagen. In SSIS heeft de afstudeerder aangegeven waar de attributen in de gegevens set naar de datawarehouse moeten worden weggeschreven. In figuur 21 ziet u een grafische weergave van de mapping, zoals deze plaats vindt in SSIS.



Figuur 21 SSIS Mapping

In het figuur lopen de mapping lijnen door elkaar heen, wat niet overzichtelijk is. Onder dit figuur is het echter mogelijk om handmatig de attributen te mappen. Daarnaast kan SSIS een aantal mappings automatisch detecteren en verbinden.

## 7.2 Record geschiedenis procedure

De datawarehouse bevat historische gegevens. Voor het bijhouden van geschiedenis heeft de afstudeerder een procedure opgesteld. Attributen met SCD type 1 worden overschreven. Attributen van het type SCD 2 worden echter niet overschreven. Om geschiedenis van deze attributen bij te houden is er gebruik gemaakt van een zogenoemde 'hash' techniek. Deze techniek is efficiënter dan het vergelijken van meerdere individuele attributen van een record. Enkel van de attributen van het SCD type 2 wordt een hash gedefinieerd. Deze hash wordt vervolgens gecontroleerd met de hash uit de datawarehouse. Als deze 'matchen' is het record niet gewijzigd. Is dit niet het geval dan wordt het oude record afgesloten, doormiddel van een record validatie datum. Vervolgens wordt er een nieuwe record in de database aangemaakt, die alle waarden van de vorige record versie bevat, inclusief de wijzigingen.



De afstudeerder heeft zelf, binnen SSIS, een C# script geschreven voor het genereren van een record hash. De C# code:

```
Public override void Input0_ProcessInputRow(Input0Buffer Row){
    var shaIProvider = HashAlgorithm.Create("SHA1");
    StringBuilder columnValues = new StringBuilder();
    PropertyInfo[] pi = Row.GetType().GetProperties();

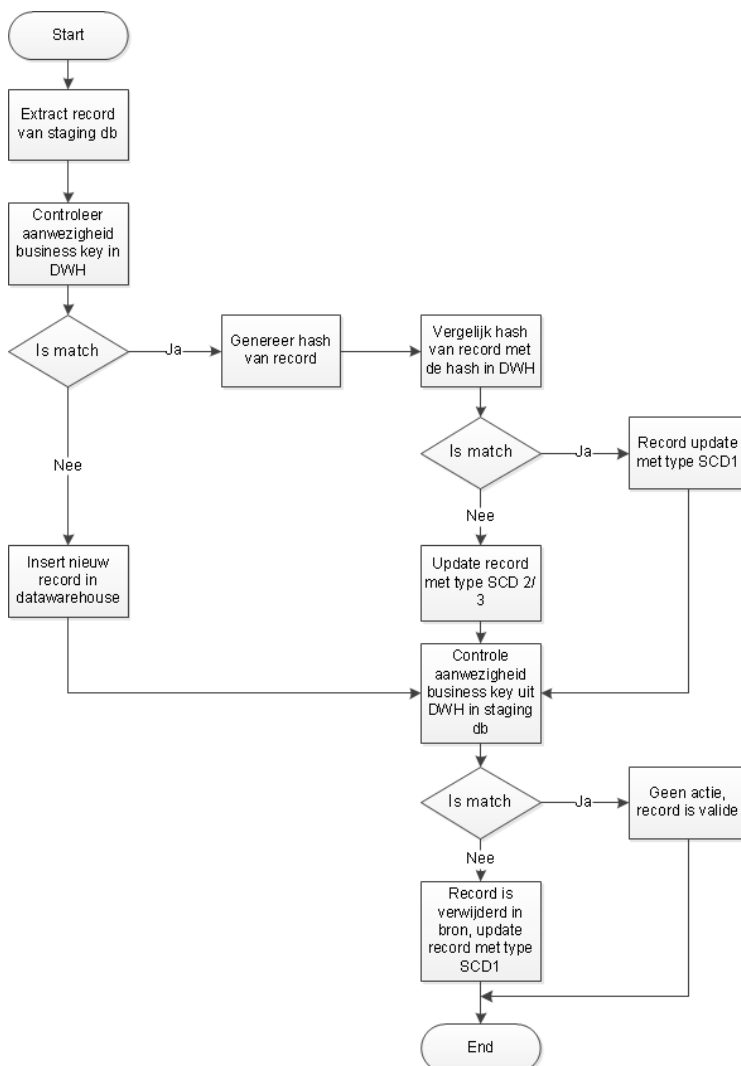
    foreach (PropertyInfo p in Row.GetType().GetProperties()){
        try{
            //Skip columns of type nullable
            if (!object.ReferenceEquals(p.PropertyType, typeof(bool))){
                //Add column values to the string array
                columnValues.Append(p.GetValue(Row, null).ToString());
            }
        }
        catch (Exception ex){
            ex.ToString();
            continue;
        }
    }

    //Generate Hash byte array from column values in the string array and convert to unicode
    byte[] binHash = shaIProvider.ComputeHash(enc.GetBytes(columnValues.ToString()));

    //Make the hash sql friendly
    Row.RecordHash = Row.RecordHash + "0x" + BitConverter.ToString(binHash).Replace("-", "");
}
```

Figuur 22 SSIS Scripttask C# hash functie

Om de geschiedenis procedure verder inzichtelijk te maken, heeft de afstudeerder een flowchart gebouwd. In figuur 23 de betreffende procedure flowchart.



Figuur 23 Flowchart geschiedenis

## 8 BI dashboards

Hier komen voorbeelden van de dashboards. Deze zijn nog niet af en worden na het inleveren van het verslag verder ontwikkeld. Om u toch een voorbeeld geven van de mogelijkheden van analyse modellen, worden hier een aantal voorbeelden gegeven. Dit hoofdstuk dient daarnaast om de traceability van de functionele requirements aan te tonen.

De volgende analyse modellen met bijbehorende referentie naar de requirement:

- Overzicht van shipments (F003)
- Overzicht aankomende containers bij magazijn (F007)
- Delivery performance kpi (F006)
- Laagste granulariteit feit: warehouse product arrival (F016)

## 9 Testen datawarehouse

De afstudeerder heeft geen formele testmethodiek gebruikt voor het testen van de datawarehouse. Echter zijn er wel testtechnieken toegepast. In de analyse van de gegevensbronnen en de gegevens conversie is vastgesteld wat de huidige data kwaliteit is. Dit is de eerste stap in het testen van de datawarehouse en de data. Namelijk het vaststellen van de huidige gegevenskwaliteit en daarvoor conversie regels op te stellen. In de requirements staat beschreven welke eisen er aan de data wordt gesteld, qua analyses of het bijhouden geschiedenis. Dit is de baseline waaraan de correctheid en kwaliteit van de data aan moet voldoen. Door de huidige bedrijfsdata met de baseline te vergelijken kan worden vastgesteld of de gewenste correctheid en kwaliteit van de data is behaald. In ten samen omvat het plan voor het testen van de datawarehouse.

### 9.1 Validatie requirements

Binnen dit project is telkens geverifieerd of de gemaakte ontwerpbeslissing in lijn is met het behalen van de eisen en wensen van de klanten. Het verifiëren en valideren van de requirements is deels door traceability bereikt en door deze voor te leggen aan een ontwikkelaar of implementatie consultant. Elke belangrijke ontwerpbeslissing werd getoetst aan een requirement. Door het niet tijdig, binnen de afgesproken termijn, de dashboards op te leveren zijn veel non-functionele requirements niet getoetst. Met Centric is afgesproken om het project met een maand te verlengen om de dashboards alsnog te realiseren. De validatie van de functionele requirements is hierdoor uitgesteld en wordt niet meer meegenomen in dit verslag.

### 9.2 Semantische test

Een belangrijk aspect van bedrijfsanalyses en rapportages is de correctheid hiervan. De gebruikers moeten immers beslissingen nemen op grond van de analyses en rapportages. Foutieve en ontbrekende gegevens in de datawarehouse kunnen resulteren in foutieve en onvolledige rapportages. Kwaliteit van gegevens kan alleen worden getoetst of deze voldoen aan de geldige bedrijfsregels. Deze toets was tijdens dit project slechts deels mogelijk. De geldige bedrijfsregels worden door Dynamics zelf gehandhaafd. De controle hierop heeft tijdens de gegevensbronanalyse plaatsgevonden. Bedrijfsspecifieke regels zijn niet opgesteld omdat het klant perspectief bij dit project ontbreekt. De toetsing hierop is dan ook niet mogelijk.

De afstudeerder verwachtte althans ook niet dat er in de huidige gegevens set, data kwaliteit problemen zouden bestaan. De beschikbare gegevens set, bestaat uit demo data van het fictieve bedrijf; Contoso. Deze gegevens set is samengesteld om gebruikers een indruk te geven van de functionaliteit van Dynamics AX. De afstudeerder meent de aanname te kunnen maken dat er zorg is gedragen voor het leveren van representatieve data. Voor de module Tradeworld was echter geen demo data beschikbaar, deze is door de afstudeerder tijdens de gegevensbronanalyse toegevoegd. Ook de module handhaaft de bedrijfsregels die door de implementatie consultants worden ingesteld. De afstudeerder heeft echter onvoldoende kennis van de interne werking van Dynamics AX en Tradeworld om deze beweringen te bekrachtigen.

De controles in Dynamics kunnen niet altijd waterdicht zijn. Zo kan een medewerker een procedure fout maken door bijvoorbeeld een verkeerd product te scannen bij aankomst in het magazijn. Het product wordt vervolgens verkeerd gekoppeld aan een shipment. De rapportages zijn gebaseerd op plaatsgevonden feiten en zijn op dat moment de waarheid. Het scannen van een verkeert product is daarmee semantisch gezien correct. Het ligt buiten de scope om deze semantische fouten te corrigeren. Dit is de taak van het bedrijf zelf (Kimball).

Tijdens de bronanalyse en de gegevensconversie is echter wel de referentiele integriteit van het bronsysteem gecontroleerd. De foreign keys werden in de brondatabase niet expliciet gehandhaafd. De afstudeerder heeft bij het verzamelen van de brontabellen alle foreign keys gecontroleerd. Deze controle is uitgevoerd door voor elke key een SQL join te schrijven. Er werd telkens vier type joins geschreven. De eerste join stelde vast of de key daadwerkelijk naar een andere key verwees. Daarmee kon ook de kardinaliteit van de relatie worden vastgesteld.

Vervolgens werd er doormiddel van een left of right join gecontroleerd of de relatie verplicht was. Of terwijl, zijn er tupels aan zowel de linker als de rechter tabel die geen relatie tot elkaar hebben.

Daarna werd er gecontroleerd, in het geval dat een tabel meerdere foreign keys bevatte, of combinatie tussen deze keys semantisch correct is. Bijvoorbeeld het tabel shipment bevat een verwijzing naar een vertrek en aankomst hub (haven). Het is semantisch niet correct dat alleen een shipment aankomt bij een haven, deze is ook vertrokken vanaf een haven. Voor deze controle is een SQL sub query met een 'where in clause' geschreven die controleert of beide foreign keys naar verschillende havens (hub) verwijzen.

Met de laatste test query wordt gecontroleerd of een database relatie over meerdere tabellen correct is. Bijvoorbeeld voor het koppelen van producten aan een shipment moet er door verschillende tabellen worden genavigeerd: *Shipment* → *Container* → *Orderline* → *InventTable* → *EcoResProduct*.

In bijlage H vindt u een paar voorbeelden van dergelijke test query's. De afstudeerder heeft gedurende deze tests geen fouten ontdekt. Echter zijn de relaties die zijn gemoduleerd in de ERDs gedurende de bronanalyse nu wel geverifieerd.

### 9.3 Syntactische test

De afstudeerder meent dat er binnen dit project gefaseerd getest is. In elke ontwerp fase is gevalideerd of deze voldoen aan de requirements. Tijdens het bouwen van de database is gecontroleerd of deze is opgebouwd volgens het analyse model. De werking van de database en de bijbehorende constraints en indexen zijn getest (zie performance test).

Tijdens het bouwen van de ETL proces is ook gefaseerd getest. Zo is er bij het uitvoeren van elke dataflow gecontroleerd of de data fysiek is overgebracht. Intergration services toont na het uitvoeren van de ETL flow of er geen technische fouten zijn ontstaan. Als er, na het uitvoeren van de SSIS dataflow geen fouten zijn opgetreden, dan is de dataflow syntactisch correct. Het kan voorkomen dat niet alle paden in de dataflow worden doorlopen. Eventuele fouten zoals truncatie<sup>46</sup> of foutieve conversies worden dan niet 'at-run-time' gedetecteerd. Er is zorg gedragen dat er bij het testen van de dataflow alle paden worden doorlopen, al dan niet geforceerd.

Tijdens het ontwikkelen van de ETL dataflows heeft de afstudeerder een aantal functies in SSIS gebruikt. De volgende functies zijn ingezet voor het (syntactisch of semantisch) testen van de dataflows:

- Het gebruik van een event handler voor het afvangen van kritieke fouten in de dataflow die door andere wijze niet zijn af te vangen.
- Door condities in te stellen waaraan de data aan moet voldoen. Afwijkende waarden worden automatisch afgevangen. Zoals bijvoorbeeld; een string in

<sup>46</sup> Truncatie betekent het wegvallen van bijvoorbeeld letters van een woord. Dit kan voorkomen als er niet voldoende ruimte voor de string is gereserveerd binnen de database.

plaats van een integer of de verwachte gegevenslengte overstijgt de bestemmingskolom.

- Het bijhouden van een error en audit log. Op deze wijze kan de gebruiker achterhalen welke fouten er tijdens het ETL proces zijn ontstaan (N002).
- Het herleiden van foutieve records uit de dataflow en de fout apart af te handelen door een conversie regel (hoofdstuk 7.23 conversie). Zo ontstaan er meerdere dataflow paden.
- No-match lookup; ontbrekende record toevoegen of verwijzen naar een default record. Bijvoorbeeld een warehouselocatie is geen onderdeel van een warehousesite. Om nul waarden in foreign keys te voorkomen wordt een 'unknown' warehousesite aan de locatie toegewezen

In SSIS wordt elke dataflow in een transactie<sup>47</sup> geplaatst. Alle database bewerkingen in de dataflows worden automatisch terug gedraaid in het geval een fout (rollback). Er is besloten als er tijdens het ETL één dataflow faalt, dan worden alle database bewerkingen, in de gehele SSIS package terug, gedraaid (N002). Zo wordt er voorkomen dat er inconsistenties ontstaan tussen de datawarehouse en het bronsysteem (N004). De rapportages mogen geen onvolledige gegevens bevatten (N006).

#### 9.4 Test cases geschiedenis

De datawarehouse bevat gegevens waarvan geschiedenis moet worden bijgehouden. Om geschiedenis te kunnen bijhouden zijn een aantal technieken gebruikt (zie hoofdstuk 7.2) De afstudeerder heeft een aantal testcases opgesteld om te controleren of de technieken voor het bijhouden van geschiedenis ook daadwerkelijk werken. In Tabel 2 ziet u een voorbeeld van een testcase. De overige testcases kunt u vinden in bijlage.H: test rapport.

Testcase 3	
<b>Actie:</b>	Wijzigen van een bestaand record waarvan één of meerdere attributen van het type SCD 2 zijn. Vervolgens controleren of het betreffende record is geüpdatet in de database
<b>Verwacht resultaat:</b>	<p>In de database is een nieuwe record toegevoegd. Het attribuut 'startDate' van het nieuwe record heeft als waarde de wijzigingsdatum van het record. Het attribuut 'endDate' is leeg. Het attribuut 'previous_version' verwijst naar de primary key van de vorige record versie. De overige attributen van het SCD type 1 of 2 bevatten de gewijzigde waarden.</p> <p>De waarden van de attributen van de vorige record versie zijn onveranderd. Het attribuut 'endDate' heeft als waarde de datum waarop het record is geüpdatet. Het attribuut 'previous_version' blijft leeg.</p>
<b>Geslaagd (Ja / Nee)</b>	Ja

Tabel 2: Voorbeeld testcase

#### 9.5 Regressie test

De afstudeerder heeft een regressie test opgesteld. In deze context wordt bedoeld, dat een test herhaalbaar is. De intentie van deze test is om te controleren of voorafgaande aan de deployment, de datawarehouse architectuur functioneert. Hiervoor heeft de afstudeerder een standaard gegevens set gebouwd, waarin opzettelijk een aantal fouten in zijn aangebracht. Deze gegevens set is een eenvoudige SQL insert statement die in een aparte testpackage kan worden aangeroepen. In zowel de staging als de

<sup>47</sup> Database transactie

datawarehouse package wordt bijgehouden hoeveel tupels de dataflows verwerken. Het aantal tupels wordt geregistreerd in een log, samen met andere audit gegevens over de dataflow. Vervolgens wordt er gecontroleerd of het aantal tupels overeenkomt met het, vooraf verwachte aantal.

Het tellen van het aantal tupels als het bijhouden van audit informatie in de datawarehouse package is, tijdens het inleveren van dit document nog niet aanwezig. Voor de zitting is deze regressie test geïmplementeerd.

## 9.6 Performance test

De afstudeerder heeft, naast de controle van de correctheid van de data, ook een performance test uitgevoerd. De performance test betrof het schrijven van een query op de Dynamics database voor het tonen van alle shipments met bijbehorende data van fase overgangen gesorteerd per week. Vervolgens heeft de afstudeerder een query geschreven op het feit tabel shipments in de datawarehouse. Deze query bestond uit een eenvoudige selectie query met joins naar de datum, leverancier en de warehouselocatie dimensies. Voor beide query's is het query plan van SQL server onderzocht en met elkaar vergeleken. In het query plan kunt u zien dat er voor de query op de staging database meer joins nodig was dan dezelfde query op de datawarehouse. Dit leverde een significante performance wist op (N012). In bijlage H, vindt u de beide query plannen met de bijbehorende statistieken.

Sql server toonde, naast het query plan en de statistieken, ook een aanbeveling voor een index. Volgens SQL server kan de performance van de query verhoogd worden met 43,70 percent. De afstudeerder heeft deze index dan ook geïmplementeerd en heeft de query nogmaals uitgevoerd. Na het analyseren van het plan bleek dat de, gestelde performance winst, minder significant was dat door SQL server werd beweerd. De uitvoering tijd van de query werd slechts gereduceerd van 131 naar 102 milliseconden. De oorzaak van deze discrepantie is niet duidelijk. In figuur 24 ziet u de aanbeveling van SQL server.

```
/*
Missing Index Details from SQLQuery32.sql - AX2012R3.CTWINT2013R3 (CONTOSO\Administrator (60))
The Query Processor estimates that implementing the following index could improve the query cost by 43.7073%.
*/

/*
USE [CTWINT2013R3]
GO
CREATE NONCLUSTERED INDEX [<Name of Missing Index, sysname,>]
ON [dbo].[Fact_Shipment] ([SHIPMENTID])
INCLUDE ([TotalNumOrderlines],[TotalValueShipped])
GO
*/
```

Figuur 24 Aanbeveling index sql server

De uitgevoerde query op de datawarehouse was niet optimaal. De afstudeer had een index geplaatst op het feit tabel, maar deze werd niet door de SQL server gebruikt. De significante performance winst die sql server toont is niet geheel representatief. De reden hiervoor was dat er in de staging database meer data was opgeslagen, dan in de datawarehouse. Door het gebrek van een gegevens set met volume, kan er niet worden getest of de datawarehouse met significante datavolume nog efficiënt werkt.



## 10 Evaluatie

In dit hoofdstuk wordt de persoonlijke evaluatie of reflectie van de afstudeerder besproken. De evaluatie/reflectie is geschreven om kritisch en objectief naar mijn afstuderen te kijken. Het tweede deel van de evaluatie wordt er per beroepstaak geëvalueerd of deze, op het juiste niveau is uitgevoerd. Daarnaast dient de evaluatie als een extra verklarende verdediging van de gemaakte keuzes. In hoofdstuk 13 vindt u de referentie van de bedrijfsmentor. Hij heeft mij gedurende deze afstudeerperiode begeleid.

### 10.1 Proces

Voor de aanvang van het project, had ik een aantal verwachtingen. Ik verwachtte dit project, zonder al te veel obstakels, met gemak zou kunnen afronden. Ik was van mening dat ik alle kennis en ervaring bezat die ik hiervoor nodig zou moeten hebben. Echter na het inlezen van de opdracht en de eerste analyse van het bronsysteem, bleek dat het project mogelijk complexer van aard was, dan dat ik van te voren had voorzien. Dit voedde mijn onzekerheid.

Deze onzekerheid uitte zich, in het continu verzamelen van informatie, zodat ik grip kon krijgen van de complexiteit en omvang van het bronsysteem. Hoe meer ik ging analyseren, hoe meer ik het overzicht verloor. Ik had achteraf meer structuur in de gegevensanalyse moeten aanbrengen en mijn bevindingen meteen moeten documenteren. Ik had meer hulp moeten vragen en aanvaarden, van mijn begeleiders en hen meer bij de analyse moeten betrekken.

Echter, de drempel in eerste instantie was vrij hoog. Dit omdat ik kundig wilde overkomen. Ik wilde aantonen dat ik voldoende kennis van zaken had, om dit project zoveel mogelijk zelfstandig te uit te voeren. Mijn onzekerheid is gaande dit project steeds minder geworden. Helaas had ik al veel vertraging opgelopen, gedurende de gegevensanalyse zodat ik de gemaakte planning, in het plan van aanpak, niet meer kon halen.

De planning van de eerste twee fasen, was niet correct. Ik had meer tijd moeten in plannen voor het opstarten van het project en het verzamelen van de requirements. Het verzamelen van de requirements heb ik uitgesteld omdat ik eerst Dynamics wilde analyseren. De rede hiervoor was dat ik op deze manier beter met de klant, dhr. Aldewereld, kon communiceren over de requirements. Daarnaast voorkwam het dat ik zaken ging beloven, die ik achteraf niet waar zouden kunnen maken met Dynamics. Wel had ik achteraf, in een vroeger stadium, met dhr. Aldewereld contact moeten leggen betreffende de requirements. Zo had ik eerder de scope van de gegevensbronanalyse kunnen bepalen.

Daarnaast vind ik het ook moeilijk om prioriteiten te stellen binnen het project. Het ligt in mijn karakter, om te reageren op impulsen en hier meteen iets aan te doen. Ik verlegde daardoor steeds mijn aandacht op een ander onderdeel van het project of op het verslag. Hierdoor raakte ik deels het overzicht kwijt, terwijl ik in mijn plan van aanpak, duidelijk een chronologische en strikte project structuur en aanpak had bedacht. Ik heb dan ook geleerd, dat het van te voren alles willen vastleggen, deels geen zin heeft. Een flexibelere projectaanpak was voor dit project meer op zijn plaats geweest. Daarnaast wilde ik de Kimball methodiek in detail volgen. Deze opvatting heb ik gaande dit project losgelaten. Ik heb nu de principes van Kimball overgenomen en niet zozeer zijn projectaanpak. Ik sta echter wel achter mijn keuze voor het gebruik van deze methodiek. Het architectuur ontwerp en de principes van het dimensioneel modeleren, waren zeer nuttig voor dit project.

Om met een positief punt te eindigen; ik vind dat ik veel geleerd heb van deze afstudeerstage. Niet alleen inhoudelijk maar ook persoonlijk. Ik moet meer vertrouwen hebben in mijzelf en moet niet invullen wat andere mensen wellicht over mij denken. De bijgevoegde referentie van de bedrijfsmentor (zie hoofdstuk 12) is de bevestiging dat ik wellicht minder negatief over mijzelf moet denken.

## 10.2 (Tussen) Producten

Tijdens dit project heb ik een aantal tussen producten opgeleverd zoals; de requirements, de datawarehouse architectuur en de analyse modellen. Ik ben relatief tevreden over de kwaliteit van de tussen producten. De opgedane kennis, tijdens een business intelligence minor, heb ik goed weten toe te passen. Wel ben ik van mening, dat ik op sommige punten te ver ben doorgeschoten, in het perfectioneren van de producten. Telkens had ik in mijn achterhoofd, als ik deze methodiek toepas of deze functie toevoeg, dan krijg ik wellicht een hoger cijfer. Het laten leiden door deze gedachten heeft geen toegevoegde waarde. Het is belangrijker dat alle producten op tijd af komen.

Helaas is het mij dan ook niet gelukt om de dashboards voor het inleveren van dit document, af te krijgen. Dit terwijl het ontwikkelen van dashboards juist het doel van dit project was. Gelukkig heeft Centric voldoende vertrouwen, dat de dashboards worden gerealiseerd, voor de aanvang van de afstudeer zitting. Ik ben Centric dan ook zeer dankbaar voor deze kans en vertrouwen.

Samengevat, ben ik van mening dat de opgeleverde tussenproducten van voldoende niveau zijn. Over het eindproduct kan ik nog geen oordeel vellen omdat deze nog niet af is.

## 10.3 Beroepstaken

In bijlage B vindt u de competentielijst die voor de aanvang van het afstuderen zijn opgesteld. De competenties zijn afkomstig uit het afstudeerplan en zijn sindsdien niet meer gewijzigd. Voor de evaluatie van de beroepstaken heb ik mij laten leiden door de beschrijvingen en beoordelingscriteria van de beroepstaken zelf.

### **Uitvoeren analyse door definitie van requirements niveau 3**

Voor dit project zijn een aantal requirements verzameld. Het verzamelen hiervan verliep op incrementele wijze. Er heeft één formeel interview met een stakeholder plaats gevonden. De rest van de requirements zijn op informele wijze tot stand gekomen.

De requirements zijn onderverdeelt in business, functioneel –en non-functioneel. Daarnaast zijn deze, volgens de Moscow regels, geprioriteerd. De wijze waarop de requirements tot stand zijn gekomen, beschouw ik als een zwak punt; bij de uitvoering van deze beroepstaak. Daarnaast is het aantal functionele requirements, in verhouding tot de andere typen requirements, te laag. Als laatste heb ik de requirements niet kunnen valideren doordat de dashboards nog niet af zijn.

Ik ben wel van mening, dat ik de traceerbaarheid van de requirements, goed heb aangetoond. Ik refereer na elke belangrijke ontwerpbeslissing naar de requirements. Door het aantonen van traceability, wat één van de complexe en lastige beoordelingscriteria zijn, vind ik dat deze beroepstaak op niveau 3 heb uitgevoerd.

### **Opstellen gegevensmodel voor database niveau 4**

Het opstellen van een gegevensmodel voor een database, was één van de belangrijkste taken binnen dit project. Tijdens het opstellen van het gegevensmodel

van de datawarehouse is gebruik gemaakt van normalisatie als denormalisatie technieken.

Ik heb een aantal snowflake modellen (analyse modellen) gebouwd waarin een aantal design principes van Kimball zijn gebruikt. Een voorbeeld hiervan is het accumulated shapshot feit: container arrival. Ook is er een strategie bedacht voor het bijhouden van historische gegevens. Een voorbeeld hierbij zijn attributen van het SCD type 1 of 2, die doormiddel van een hash op wijzigingen worden gecontroleerd.

Het ontwerpproces is uitgebreid beschreven in hoofdstuk 7: Dimensioneel model ontwerp proces. Door de bovengenoemde voorbeelden kan ik, met redelijke zekerheid, de aannamen wagen, dat ik deze beroepstaak, op niveau 4 heb uitgevoerd.

#### **Ontwerpen, bouwen en bevragen van een database niveau 4**

Tijdens dit project ben ik continu met databases bezig geweest. Hierbij heb ik bijna alle taken uitgevoerd, zoals beschreven in de beoordelingscriteria van de beroepstaak. Ik heb complexe query logica geschreven. Een voorbeeld hiervan is het genereren van de datum dimensie.

Bijna alle ontwerpbeslissingen van de datawarehouse wordt bekeken welke impact deze heeft op performance. Zo is er na gedacht over het gebruik van secondaire indexen en fysieke opslag structuren. Een voorbeeld hiervan is het partitioneren van een feit tabel en het plaatsten van een column store index. Bij het ontwikkelen van de ETL dataflows met SSIS is rekening gehouden met performance aspecten van bepaalde SQL query's. De referentiele integriteit van de datawarehouse is gewaarborgd door op elke foreign key, een constraint te plaatsen.

In bijlage F kunt u een aantal voorbeelden vinden van query's van het implementatie domein.

Echter, heb ik niet na gedacht over locking strategieën, dit omdat er alleen leesacties op de datawarehouse plaats vinden. Het laden van nieuwe records gebeurt enkel door SSIS. Er wordt een bulk insert gebruikt voor het laden van nieuwe records in de datawarehouse. Bovendien staat het recovery model van de database op 'simple'. Dit betekend dat er minimaal wordt gelogd.

Ik heb ook nog niet na gedacht over data beveiliging. Deze zaken zouden aanbod zijn gekomen bij het bouwen van de dashboards. Dhr Aldewereld heeft twee gebruikersrollen gedefinieerd, de inkoop –en magazijnmanager. Hij vond het, in dit stadium van het project, nog niet belangrijk om zaken als beveiliging te bespreken.

Het bouwen van effectieve indexen is nog niet gelukt. Tijdens de performance test heb ik een 'non clustered' index op de feit tabel shipment geplaatst. Deze werd door SQL Server niet gebruikt

Ondanks dat ik zaken als beveiliging niet heb geïmplementeerd en dat de plaatsing van effectieve indexen niet is gelukt. Durf ik met enige twijfel, toch de uitspraak te doen, dat ik deze beroepstaak op niveau 4 heb uitgevoerd. Mede omdat ik toch het concept van indexen begrijp, als dat ik in staat ben om deze te schrijven.

#### **Uitvoeren gegevensconversie niveau 3**

De gegevens conversie is binnen dit project, beperkt gebleven. Er zijn geen integriteit of bedrijfscontroles uitgevoerd. De gegevenskwaliteit was op een voldoende niveau. Daarnaast was het ontbreken van goede testdata een belemmering, om een zinnig uitspraak te doen, van toekomstige gebruiker gegevens. Het definiëren van

bedrijfsregels kon ook beperkt worden uitgevoerd omdat deze kunnen wisselen per klant. Wel heeft de afstudeerder waar dat mogelijk was conversies uitgevoerd.

In de beoordelingscriteria van de beroepstaak staan een aantal taken, dat tijdens een conversie moet worden uitgevoerd. De taken:

- **Er is een juiste analyse van de verschillende databronnen.**  
Zie gegevensbronanalyse hoofdstuk 3.4
- **Er is een juiste analyse van de structuur van de data uit de databronnen.**  
Ja, een deel van de data dictionaries zijn gevalideerd door de bedrijfsmentor.
- **Er is een juiste analyse van de verbeterpunten in de kwaliteit van de data uit de databronnen.**  
Ja, een voorbeeld hiervan is het converteren van een integer naar een boolean.  
Zie Figuur 14 SSIS Gegevens Conversie
- **Er zijn schema's waarin de transformatie van databronnen naar datadoelen zijn vastgelegd.**  
Ja, zie figuur 13 Mapping schema WarehouseLocation
- **De gevolgen van de conversie voor systemen en bedrijfsprocessen zijn juist beschreven.**  
Ja, zie hoofdstuk 9 testen datawarehouse.
- **De conversie wordt ondersteund door scripts.**  
Ja, er wordt gebruik gemaakt van SSIS, zie figuur 20 SSIS Flow dimensie warehouseLocation
- **Er is vastgesteld dat de data uit de databronnen op de juiste wijze in de doeldatabases terecht is gekomen.**  
Deels, zie hoofdstuk 9 testen datawarehouse.
- **Er is vastgesteld dat de beoogde kwaliteitsverbetering is bereikt.**  
Nee

De gegevensconversie, is naar mijn mening, niet op niveau 3 maar op niveau 2 uitgevoerd. Dit niveau wordt ook verwacht van een HBO informatica student. De conversie was weliswaar simpel maar deze is wel zelfstandig uitgevoerd en geïmplementeerd met de juiste tooling.

### Ontwerpen (software)architectuur niveau 3

In hoofdstuk 5 heb ik de datawarehouse architectuur beschreven. Ik heb beschreven waarom ik voor de Kimball architectuur heb gekozen. Vervolgens is deze architectuur geïmplementeerd. De beroepstaak gaat over het ontwerpen van software architectuur. Er is binnen dit project geen softwarearchitectuur ontworpen.

Echter heb ik de definitie van een architectuurontwerp breed geïnterpreteerd Ik ben van mening dat het ontwerpen van ster schema's, de keuze voor de instantie infrastructuur, de implementatie van design patterns en lagen in SSIS en ten slotte het gebruiken van Kimball ontwerp principes, gesproken kan worden van een alomvattend business intelligence architectuurontwerp.

Omdat de beroepstaak het specifiek over een software, in plaats van een business intelligence architectuur heeft. Kan ik niet beoordelen of de beroepstaak haar behoren is uitgevoerd.

### Uitvoeren van en rapporteren over het testproces niveau 3

Voor het testen van de datawarehouse heb ik een performance test uitgevoerd. Deze test, was naar mijn mening, niet objectief. In de query plannen en de statistieken zag je

wel een performance winst, maar dit was slechts het resultaat van het leggen van minder joins.

Ik heb geen gebruik gemaakt van een vooropgezet (master) testplan of andere test methodieken. Wel heb ik een paar testcases opgesteld voor het controleren van gegevensgeschiedenis. In hoofdstuk 9 testen datawarehouse heb ik beschreven welke activiteiten ik heb uitgevoerd om de datawarehouse te testen. Doordat de dashboards nog niet af zijn. Heb ik ook geen acceptatietest kunnen uitvoeren. Ik ben van mening dat ik niet, volgens de beroepstaak, op een juiste wijze heb gerapporteerd over het testproces.

Hiermee trek ik de conclusie, dat ik deze beroepstaak niet op niveau 3, maar op niveau 2 heb uitgevoerd.

## 11 Litratuur / bronnenlijst

### **Ralph Kimball & Margy Ross -The Data Warehouse Toolkit Third Edition**

#### **Chapter 5 – role playing dimensions**

(Acuma) <http://www.acuma.co.uk/index.php/training/data-warehousing>

Info support, ontwikkelstraat endeavor, project sjablonen

Info support, Jesse Gorter consultant en trainer – Best practices data warehousing

Centric intranet: Implementatie document verladingen module (TW2012\_R3 10 verladingen.docx)

Centric intranet, bedrijfsgegevens.

Centric, Joost te mulder implementatie consultant / begeleider – Functionele requiremens & prioritering

Centric, Martin Aldewereld pre-sales dynamics consultant – business requirements

Centric, René van Leeuwen Dynamics AX developer – technische requirements

Dynamics AX home page

<http://www.microsoft.com/dynamics/ax/>

<http://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/>

<http://sqlmag.com/business-intelligence/delivering-bi-through-sql-server-and-sharepoint>

ISO extended 9126

#### **MSDN – Microsoft Dynamics AX 2012 Resource Library**

<http://www.microsoft.com/en-us/dynamics/resource-library.aspx>

#### **Presentatie keuze immon of Kimball**

[http://ioug.itconvergence.com/pls/html/db/DWBISIG.download\\_my\\_file?p\\_file=2346](http://ioug.itconvergence.com/pls/html/db/DWBISIG.download_my_file?p_file=2346)

#### **Definitie HOLAP kubus**

<http://en.wikipedia.org/wiki/HOLAP>

#### **Kimball datawarehouse architectuur**

<http://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/technical-dw-bi-system-architecture/>

#### **4-step beslissing dimensioneel modeleren**

<http://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/dimensional-modeling-techniques/four-4-step-design-process/>

#### **Kimball Dimensional modeling Techniques**

<http://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/dimensional-modeling-techniques/>

#### **Kimball design tip 152 Slowly Changing Dimension Types 0, 4, 5, 6 and 7**

<http://www.kimballgroup.com/2013/02/design-tip-152-slowly-changing-dimension-types-0-4-5-6-7/>

#### **Kimball design tip 85 Using Smart Date Keys to Partition Large Fact Tables**

<http://www.kimballgroup.com/2006/11/design-tip-85-using-smart-date-keys-to-partition-large-fact-tables/>

#### **Kimball design tip 130 Accumulating snapshots for complex workflows**

powerpivot for sharepoint <http://msdn.microsoft.com/en-us/library/ee210692.aspx>



**Stacia Minser SQL Server Pro - Designing SSIS packages for high performance**

<http://sqlmag.com/sql-server-integration-services/designing-ssis-packages-high-performance>

**LeapFrogBi - Youtube channel**

[http://www.youtube.com/channel/UCuxkuyY\\_U1P\\_67f6XFbqygw](http://www.youtube.com/channel/UCuxkuyY_U1P_67f6XFbqygw)

## 12 Referentie

Daniël heeft blijk gegeven van voldoende beheersing van kennis en vaardigheden om de opdracht uit te voeren. Hij is leergierig, ijverig, sociaal, scherpzinnig, ingenieus en gaat voornamelijk zelfstandig te werk. Daniël heeft zich in korte tijd verheven tot een behoorlijke Business Intelligence ontwikkelaar met oog voor business requirements. Je kan bouwen op zijn impliciete kennis. Verder is hij collegiaal en praat op alle fronten gewoon mee. Kortom, Daniel is een fijne en kundige collega.

Referent: Dhr. J. te Molder  
Functie: Implantatie Consultant Dynamics  
Bedrijf: Centric Netherlands B.V.  
Contact: Joost.te.Molder@centric.eu

## Externe bijlagen

Dit afstudeerverslag bevat de volgende bijlagen:

- Bijlage A Plan van aanpak
- Bijlage B competentie lijst
- Bijlage C Bedrijfsproces analyse
- Bijlage D Gegevensbron analyse
- Bijlage E Analyse modellen (dimensioneel database modellen)
- Bijlage F Implementatie query's
- Bijlage G SSIS ETL Workflows
- Bijlage H Test rapport

connect.  
engage.  
succeed.

**BIJLAGE A**

**PLAN VAN AANPAK**

## Versie beheer

### Historie

Versie	Status	Datum	Verandering
1.0	Concept	11-05-2014	Creatie
1.1	Concept	20-05-2014	Toepassing huisstijl van Centric
1.2	Verzoek tot goedkeuring	23-05-2014	Document gefinaliseerd en ter goedkeuring aangeboden aan de bedrijfsmentor
1.3	Op hoofdlijnen goedgekeurd	27-05-2014	Aanwijzingen van de bedrijfsmentor verwerkt. Toegevoegd: hoofdstuk 1.9 en risico impact factor.

### Distributie

Versie	Datum	Aan
1.2	23-05-2014	Dhr. J te Molder (bedrijfsmentor)
1.3	27-05-2014	Dhr. J Cruysen (Expert ERP)
1.3	27-05-2014	Digitale portfolio Haagse hogeschool
1.3	03-06-2014	A.A. Nederend Examinator

## Inhoudsopgave

<b>1</b>	<b>PROJECT ACHTERGROND .....</b>	<b>1</b>
1.1	BEDRIJF .....	1
1.2	PROBLEEMSTELLING .....	1
1.3	DOELSTELLING .....	2
1.4	RESULTAAT .....	2
1.5	SCOPE .....	2
1.6	RANDVOORWAARDEN .....	3
1.7	BENODIGDE HARD –EN SOFTWARE .....	3
1.8	HUMAN RESOURCES .....	4
1.9	KOSTEN EN BATEN .....	4
1.10	RELATIES MET ANDERE PROJECTEN .....	4
<b>2</b>	<b>AANPAK EN FASERING .....</b>	<b>5</b>
2.1	PROJECT INITIATIE & MANAGEMENT .....	5
2.2	BUSINESS REQUIREMENTS DEFINITION .....	5
2.3	TECHNOLOGIE TRACK .....	6
2.4	DATA TRACK .....	7
2.5	BUSINESS INTELLIGENCE TRACK .....	9
2.6	TESTEN .....	9
2.7	DEPLOYMENT .....	9
2.8	GROWTH .....	9
2.9	MAINTENANCE .....	10
<b>3</b>	<b>PROJECT TECHNIEKEN &amp; METHODIEKEN .....</b>	<b>11</b>
<b>4</b>	<b>DELIVERABLES .....</b>	<b>12</b>
<b>5</b>	<b>COMMUNICATIE MANAGEMENT .....</b>	<b>13</b>
5.1	ORGANOGRAM BUSINESS UNIT .....	13
5.2	STAKEHOLDERS .....	13
5.3	COMMUNICATIE PROCEDURE .....	14
5.4	RAPPORTAGE .....	14
<b>6</b>	<b>RISICO MANAGEMENT .....</b>	<b>15</b>
<b>7</b>	<b>PLANNING .....</b>	<b>17</b>
7.1	RANDVOORWAARDEN .....	17
7.2	PROJECTPLANNING .....	17
<b>8</b>	<b>BIJLAGE: A: OPDRACHTFORMULERING .....</b>	<b>18</b>
<b>9</b>	<b>BIJLAGE B GRANTT CHART .....</b>	<b>19</b>



## 1 Project Achtergrond

Het project betreft het ontwikkelen van Business intelligence in de vorm van dashboards voor klanten van het bedrijf Centric. Centric verkoopt ERP (Enterprise Resource Planner) pakketten aan bedrijven en configureert deze vervolgens.

Daarnaast ontwikkelt Centric zogenoemde 'value added' functionaliteit binnen Dynamics AX voor specifieke marktsegmenten. Een voorbeeld van een segment zijn de internationale verschepingen van goederen door groothandels. Dit proces wordt geadministreerd binnen een module van Dynamics AX genaamd 'Tradeworld & international shipments'. Centric heeft binnen deze module extra functionaliteit gebouwd om verladingen van goederen te managen (bijvoorbeeld zorgen voor een optimale container vulling). De gebruikers van deze module; de groothandels, ervaren dat door de toegenomen complexiteit van het shipment proces, het overzicht hierop te verliezen.

Centric wil graag een Business Intelligence oplossing aanbieden aan de groothandels om het shipment proces inzichtelijker te maken. De kennis van Business Intelligence (BI) is ontoereikend binnen Centric. Daarom heeft Centric een afstudeerder van de Haagse Hogeschool ingehuurd, om een BI oplossing voor deze groothandels neer te zetten.

In dit plan van aanpak staat beschreven hoe de afstudeerder een Business Intelligence oplossing voor de groothandels zal gaan realiseren.

### 1.1 Bedrijf

In 1992 start Gerard Sanderink, de Sanderink Groep, de voorloper van het huidige Centric. Dankzij autonome groei en een strategisch acquisitiebeleid is Centric uitgegroeid tot een Europese dienstverlener van formaat.

Het bedrijf Centric levert de volgende producten en diensten: Software Solutions, IT Outsourcing, Business Process Outsourcing en Staffing Services. Met gekwalificeerde professionals, technologische oplossingen en IT- en administratieve dienstverlening, zorgt Centric ervoor dat klanten zich op hun kerntaken kunnen richten.

Centric heeft 4.944 (2012) werknemers en vestigingen in Nederland, België, Duitsland, Zwitserland, Noorwegen, Zweden en Roemenië. De organisatie boekte in 2012 een omzet van 525 miljoen euro met een netto bedrijfsresultaat van ruim 12 miljoen euro.

Centric is opgedeeld uit een aantal divisies, één daarvan is Supply Chain Solutions. De divisie heeft als doelstelling om software en hardware oplossingen te bieden voor de volledige de Supply Chain van bedrijven.

Het project wordt uitgevoerd binnen de divisie: Supply Chain Solutions en de unit Business Solutions. De betreffende divisie houdt zich globaal met de volgende activiteiten bezig:

- Verkoop, implementatie, uitbreiding en ondersteuning van Microsoft Dynamics AX software in combinatie met het TradeWorld-concept
- Verkoop, implementatie, uitbreiding en ondersteuning van ERP-pakketten als Bestmate, Triathlon en WES.

### 1.2 Probleemstelling

Centric levert en implementeert het Microsoft Dynamics AX ERP pakket bij klanten. Centric ontwikkelt zelf aanvullende modules binnen het standaard ERP pakket. Eén

van deze modules is TradeWorld & International Shipments. Met deze module kunnen groothandels hun goederenstromen over de gehele wereld beheren.

Steeds meer handelsactiviteiten vinden in het buitenland plaats. Hierdoor neemt de complexiteit het proces toe. Daardoor neemt ook de vraag naar traceerbaarheid van de in- en uitgaande goederenstroom en transport toe. Met de huidige shipment module ervaart een deel van de klanten dat men op dit moment nog steeds onvoldoende visueel overzicht heeft wat betreft de goederenstromen.

Daarnaast heeft Centric onvoldoende ervaring met het managen en inzetten van grote hoeveelheden aan goederenstromen gerelateerde data. Centric is hierdoor onvoldoende in staat om deze 'big data' inzichtelijk te maken voor haar klanten. Dit heeft gevolgen voor haar concurrentie positie binnen de markt.

### 1.3 Doelstelling

De doelstelling van de opdracht is om binnen Microsoft Dynamics AX en vervolgens in de module International Shipments een 'big data' architectuur te ontwerpen, bouwen en te implementeren. Met deze architectuur is het de bedoeling dat men grote hoeveelheden aan data kan verwerken. Door het te verwerken en visueel maken van die datastromen krijgen de groothandels meer inzicht in de goederenstromen en tevens wordt de traceerbaarheid van die stromen verbeterd. De groothandels krijgen zo meer grip op het kritieke shipment proces en kunnen daardoor beter ingrijpen bij calamiteiten.

De architectuur moet leiden tot een systeem waarin middels een dashboard met een aantal verschillende overzichten en visualisaties hulp wordt geboden bij het succesvol managen van international shipments. De insteek is om het systeem te ontwerpen met bestaande tooling van Microsoft.

De opdrachtgever ziet deze opdracht als de 'kers op de taart' bij de module International Shipments, waardoor de module beter in de markt te zetten is. De architectuur moet tevens generiek zijn en kunnen worden gebruikt door verschillende groothandels. Dit is te bereiken door de architectuur te ontwerpen met in acht neming van de bestaande bedrijfsprocessen omtrent Shipments.

Het is de wens van de opdrachtgever om binnen vijf maanden (de termijn van de afstudeerstage) de bovenstaande doelstelling te realiseren.

### 1.4 Resultaat

Het uiteindelijk op te leveren eindproduct is een cube en een dashboard te ontwikkelen voor de Microsoft Dynamics AX2012 module: 'International Shipments'. Deze dashboards moeten binnen deze module generiek en toepasbaar zijn voor meerdere klanten van Centric.

In de eind scriptie van de afstudeerder staat beschreven hoe een Business Intelligence oplossing in combinatie met Microsoft Dynamics AX kan worden opgezet en ontwikkeld. Centric heeft aan het einde van dit project meer kennis op het gebied van Business Intelligence en kan deze inzetten om in de toekomst zelf dergelijke projecten te starten.

### 1.5 Scope

De omvang van Microsoft Dynamics AX ERP pakket is groot en tijdrovend om deze in zijn geheel te analyseren en te bevatten. Daarom beperkt de opdracht zich tot de module International Shipments binnen Dynamics AX. Mocht het voor het correct

functioneren van de dashboards noodzakelijk zijn dat er vanuit andere bronnen/module moet worden onttrokken, dan valt dit binnen de scope.

Het is niet de bedoeling om nieuwe modules binnen Dynamics AX te ontwikkelen, noch andere taken uit te voeren die niet bijdragen aan het behalen van de doelstellingen van deze opdracht.

Binnen dit project zal geen BI pakket selectie plaats vinden omdat Centric aangeeft een Business intelligence oplossing binnen een Microsoft omgeving te willen implementeren. Wel vindt er een selectie plaats binnen de tooling/technologieën binnen de BI stack van Microsoft. Deze selectie is afhankelijk van het architectuurontwerp.

Het valt tevens buiten de scope van deze opdracht om de gegevensarchitectuur voor Business Intelligence te implementeren bij klanten. Wel kunnen er aanbevelingen worden gegeven op dit gebied. Echter is het wel de doelstelling van dit project om een werkend prototype te realiseren, die uit te bouwen is, tot een volwaardige BI oplossing.

In het geval van meevallers of onvolkomenheden in de planning kan, in overleg met de bedrijfsmentor of de school, de opdracht in scope vergroot of verkleint worden. Vooralsnog worden de dashboards/rapporten en (tussen) producten opgeleverd zoals beschreven in hoofdstuk 4: Deliverables. Eveneens in de requirements zal specifiekere worden beschreven welke eisen en wensen van de opdrachtgever ten aanzien van de dashboard aan voldaan moet worden.

## 1.6 Randvoorwaarden

Het project moet aan een aantal randvoorwaarden voldoen om uitgevoerd te kunnen worden. Hieronder volgt een lijst met de randvoorwaarden die voor het project gelden:

- Toegang tot de relevante bedrijfssystemen en databronnen om deze te kunnen analyseren en data extraheren voor de datawarehouse.
- Beschikbaarheid van een ontwikkelomgeving waar de datawarehouse en ETL processen kunnen worden ontwikkeld. Deze omgeving moet krachtig genoeg zijn om ETL processen uit te voeren.
- Benodigde hard –en software om de Business Intelligence oplossing te kunnen ontwikkelen.
- Voldoende ruimte om de benodigde verslaglegging te doen van de afstudeerstage en het documenteren van de producten.
- Beschikbaarheid van de bedrijfsmentor voor tussentijdse beoordelingen van de begeleidende examiner van de Haagse Hogeschool. Toegang tot de benodigde experts voor technische vragen.
- De scope van het project wordt uitsluitend aangepast in overleg met Centric en de Haagse Hogeschool.
- De documentatie van de ontwikkelde producten als mede de verslaglegging moeten ter beschikking worden gesteld aan de Haagse Hogeschool ter beoordeling van de afstudeerstage. Hierbij rekening houdende met eventuele geheimhouding.
- De documentatie van de ontwikkelde producten gedurende dit project zal uitsluitend met toestemming van Centric aan derden worden vertrekt.

## 1.7 Benodigde hard –en software

Voor het uitvoeren van dit project, vraagt de afstudeerder aan Centric de volgende hard –en software:

- Microsoft SQL Server om de database te beheren en query's te schrijven.
- Microsoft Analysis Services (SSAS) om data cubes te realiseren.
- Voor het bouwen en implementeren van ETL processen: Microsoft Intergration Services (SSIS).
- Het bouwen van rapporten en/of onderdelen wordt Microsoft Reporting Services (SSRS) gebruikt.
- Het realiseren van de dashboards wordt gebruik gemaakt van Sharepoint om data verschillende uit verschillende rapporten te combineren (webparts). In deze dashboards kan ook gebruikt gemaakt worden van Power BI. Met deze tool kan de data in de dashboards aantrekkelijker worden gevisualiseerd. Power BI is alleen beschikbaar vanaf Microsoft office 2013. Het gebruik van Power BI in de dashboards is geen vereiste.
- Testomgeving met daarin toegang tot het Microsoft Dynamics AX R3 ERP pakket, inclusief de module Tradeworld & International Shipments.
- Microsoft office pakket voor documentatie van de producten en verslaglegging van de afstudeerstage.
- Computer/laptop met voldoende capaciteit om complexe en zware ETL processen uit te voeren. Daarnaast is het wenselijk om additionele monitoren aan te sluiten ter bevordering van multitasking.

## 1.8 Human resources

Voor het uitvoeren van dit project is het noodzakelijk dat Centric naast hard –en software ook mensen beschikbaar stelt. In hoofdstuk 5.2: stakeholders, kunt u lezen welke mensen (stakeholders) gedurende dit project betrokken moeten worden.

## 1.9 Kosten en baten

Dit project wordt uitgevoerd door een afstudeerder aan de Haagse Hogeschool met een maandelijkse vergoeding van: 348,00,- euro. De looptijd van dit project is 20 weken. Daarnaast maakt de afstudeerder gebruik van een laptop met de nodige hard – en software (zie hoofdstuk 1.7). De afstudeerder declareert zijn uren (32 per week) via het urenregistratiesysteem van Centric (omaWeb). Alle uitgaven moeten door de manager (Dhr M. Helmer) worden geaccordeerd. Als laatste heeft de afstudeerder wekelijks begeleiding van de bedrijfsmentor (Dhr J. te Molder). Deze begeleidingsuren vormen ook een kostenpost.

Centric heeft op dit moment onvoldoende ervaring op het gebied van Business Intelligence. Eén van de doelstelling van dit project is dat Centric kennis en ervaring op doet op het gebied van Business Intelligence. Hierdoor kan Centric mogelijk nieuwe oplossingen en producten aan haar klanten aanbieden op dit gebied. Naast kennis van Business Intelligence heeft dit project als einddoel om een aantal dashboards te ontwikkelen voor de module international shipments binnen Microsoft Dynamics AX. Deze dashboards kunnen bijdragen om een klant te bewegen bij Centric de module te kopen.

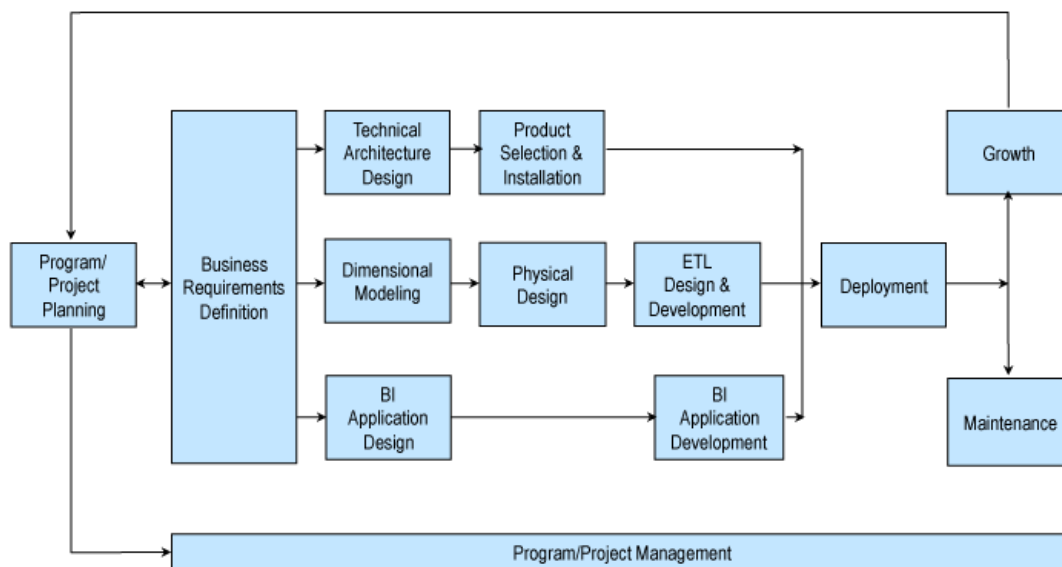
Naast de kosten van de stagevergoeding en de werkplek zijn er vrijwel geen additionele investeringskosten nodig voor dit project.

## 1.10 Relaties met andere projecten

Op de afdeling Business Solutions is nog een afstudeerder werkzaam (Dhr. L. Yu). Zijn opdracht is om te onderzoeken hoe Centric op korte termijn, doormiddel van een externe partij, Business Intelligence oplossingen kan bieden aan klanten. Aantal zaken betreffende zijn onderzoek kan worden meegenomen in dit project.

## 2 Aanpak en fasering

Het project wordt gemanaged en uitgevoerd volgens de Kimball datawarehouse lifecycle methodiek. In figuur 1 ziet u een visuele weergave van de methodiek met de verschillende fase/disciplines:



Figuur1: Kimball lifecycle diagram (bron opzoeken)

Binnen de Kimball lifecycle worden verschillende project taken (tracks) en disciplines zoveel mogelijk parallel uitgevoerd. De volgende tracks zijn gedefinieerd: architectuur technologie, data en BI applicaties. Binnen de track worden taken sequentieel uitgevoerd (bijvoorbeeld; dimensioneel naar een fysiek model). Mochten de requirements van de business of de scope van het project veranderen dan is het mogelijk om binnen deze methodiek een nieuwe iteratie van het project te starten. Om een nieuwe iteratie te starten moeten significante veranderingen plaats vinden. Een iteratie wordt ook gestart als het project in scope wordt uitgebreid. In de volgende hoofdstukken worden de verschillende taken binnen het project beschreven. Deze project aanpak en fasering kunt u zien als een begin van een ontwikkelstaat voor datawarehousing.

### 2.1 Project initiatie & management

Project planning/initiatie & management is de eerste fase binnen deze methodiek. In deze fase worden voornamelijk voorbereidingen voor het project getroffen en vervolgens worden zaken zoals; planning, randvoorwaarden en de scope bepaald.

Daarnaast wordt er in deze fase de bedrijfsprocessen geanalyseerd. De scope van de bedrijf/proces analyse ligt enkel bij shipments/verladingen. Verdiepen in de organisatie, cultuur en de structuur van Centric om eventuele stakeholders en hulpmiddelen in kaart te brengen.

### 2.2 Business requirements definition

Onderzoek en definiëren van de requirements naar de informatiebehoefte van de groothandels op het gebied van verladingen. Het doel van dit onderzoek is het achterhalen of de dashboards die zijn geformuleerd in de opdracht daadwerkelijk bestaan of dat deze gewijzigd moeten worden. Daarnaast zijn de geformuleerde requirements, in de bijgevoegde opdrachtschrijving bij de aanvang van het project, te globaal en zullen voor het ontwikkelen van de dashboards gedetailleerder moeten worden beschreven. Dit is nodig om de kans te vergroten dat de dashboards de juiste

gegevens bevatten als mede dat de informatiebehoefte van de groothandels op het gebied van verladingen voldaan wordt.

Opstellen van de non-functionele requirements met het oog op de aspecten als; gebruikersgemak, performance en onder houdbaarheid en evenals bestaande technische beperkingen. De medewerkers van de afdeling Business Solutions kunnen hierbij adviseren en assisteren. Het is al bekend dat voor deze opdracht de Microsoft tooling moet worden gebruikt. Centric heeft contracten met Microsoft die deze beperking opleggen.

Opstellen van de geldende bedrijfsregels. Hier moet onderzocht worden welke standaarden er binnen een Microsoft Dynamics project gelden.

Een deel van de requirements zijn voor de aanvang van dit project, geformuleerd in de opdrachtschrijving maar zullen gaande het project onderhevig zijn aan veranderingen. Deze veranderingen zullen worden gedocumenteerd.

## **2.3 Technologie track**

In deze track wordt de discipline (gegevens) architectuur ontwerp uitgevoerd. Deze track is onderverdeeld in twee delen; het technisch ontwerp als mede de beschrijving en verantwoording van de architecturale beslissingen. Het tweede deel omvat het installeren en configureren van zowel de ontwikkelomgeving als de testomgeving voor de datawarehouse.

### **2.3.1 Technische architectuur design**

Onderzoeken en selecteren van een geschikte datawarehouse architectuur. Deze architectuur kan bijvoorbeeld gebaseerd zijn op die van Kimball, Immon of een combinatie van beide. De rationale voor de besloten architectuur zal gebaseerd worden op de verzamelde requirements (bijvoorbeeld performance). De significante architectuur beslissingen worden gedocumenteerd en gevisualiseerd d.m.v. UML (Unified Modeling Language).

De opdrachtgever geeft aan dat er op dit moment enige performance problemen bestaan binnen het Microsoft Dynamics pakket. De oorzaak hiervan is niet bekend maar het kan enig effect hebben op de dashboards. Dit zal moeten worden onderzocht. In het ontwerp van de architectuur zal, aan de hand van het onderzoek blijken, of er maatregelen moeten worden genomen.

### **2.3.2 Tooling selectie & ontwikkelomgeving inrichting**

Inrichten van de ontwikkeling omgeving. Centric zal deze omgeving beschikbaar stellen. De benodigde tooling zal in deze omgeving waar nodig worden geïnstalleerd en geconfigureerd. De gebruikte software en tooling staan aan het begin van de opdracht vast namelijk; het gebruik van de Microsoft BI stack/tooling.

De technologie track wordt parallel uitgevoerd met de andere tracks. Het is wel de bedoeling dat het installeren en configureren van de ontwikkelomgeving is afgerond, voordat men start met het realiseren van de fysieke gegevensarchitectuur en de ETL processen.



## 2.4 Data track

In de data track komen alle disciplines rondom data analyse en architectuur aanbod. Deze track bevat drie hoofdtaken namelijk; dimensioneel modeleren (datamarts), fysiek ontwerp (implementatie model) en het ontwerp van de ETL processen.

Inventarisatie van de huidige gegevens architectuur binnen een module van Microsoft Dynamics genaamd; TradeWorld International Shipments. Aan de hand van de requirements en overleg met de beheerders en ontwikkelaars van Microsoft Dynamics wordt er een inschatting gemaakt welke gegevens relevant kunnen zijn voor de dashboards. Als er wijzigingen in de requirements plaats vinden of de rapportage of analysebehoefte veranderen, dan kan er alsnog besloten worden om eventueel ook gegevens uit andere modules bij de analyse te betrekken.

Analyseren van de geïnventariseerde gegevensbronnen en het documenteren van de structuur, type, betekenis en kwaliteit hiervan. Daarnaast moet er worden vastgesteld wie verantwoordelijk is voor de gegevens en indien van toepassing, de beschikbaarheid van deze systemen voor ETL processen. Voor al deze zaken moet er overleg plaatsvinden met de beheerders van het Microsoft Dynamics AX ERP pakket. Voor het documenteren van de beschikbare en bruikbare gegevensbronnen wordt een data dictionary gemaakt. Indien deze reeds aanwezig is dan zal deze worden gebruikt en waar nodig aangepast.

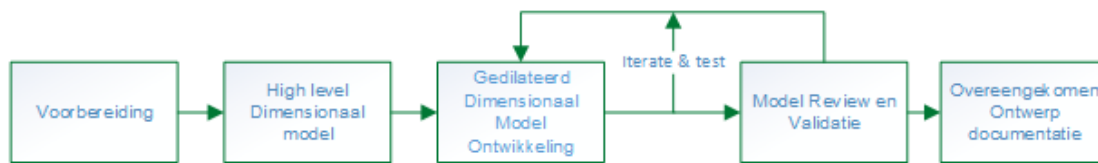
### 2.4.1 Dimensional modeling

Ontwerpen van de gegevensarchitectuur (data model) voor de datawarehouse. Er zal gebruik gemaakt worden van dimensioneel moduleren. Aan de hand van de gegevensanalyse worden de dimensies vastgesteld, waar dit mogelijk is. Het datamodel wordt gedocumenteerd en gevisualiseerd waar mogelijk d.m.v. een ERD.

Het dimensioneel model ontwerp proces verloopt in een aantal stappen (zie figuur.2) Deze stappen zijn gedefinieerd door R. Kimball in zijn boek: the datawarehouse toolkit en omvat het volgende:

- Voorbereidingen treffen aan het begin van het ontwerp proces. In deze eerste stap worden de stakeholders geconsulteerd om consensus te bereiken over zaken als; wat is een klant of product. Zie hoofdstuk: 5 communicatie management voor de beschrijving van de stakeholders Dit moet later in het proces resulteren in geconformeerde dimensies.
- Op een hoog abstract niveau het model ontwerpen en deze voor overleg tonen aan de ontwikkelaars van Dynamics AX. Het model zal in de vorm van een 'Bubble Chart' worden gevisualiseerd. Het doel van dit model is zonder om in details te treden en relatief snel in het ontwerp proces bijsturing van de stakeholders te ontvangen.
- Detail ontwerp realiseren waarin alle tabel attributen, aggregaties en opmerkingen van de ontwikkelaars in zijn opgenomen.
- In de één na laatste stap wordt het gedetailleerde model gecontroleerd of deze de gewenste rapportages en analyses kan ondersteunen. Parallel worden ook de dashboards ontwikkeld. De dashboards zijn gebaseerd op het dimensioneel model. Het model kan aangepast worden als dit nodig is vanwege technische redenen of dat de requirements in de tussentijd zijn veranderd. Dit is een iteratief proces en wordt herhaald indien nodig.
- Als laatste stap in het dimensioneel ontwerp proces is het verantwoorden en documenteren van het gefinaliseerde ontwerp.





**Figuur 2: Dimensioneel ontwerp proces**

Als het dimensioneel datamodel ontwerp is gefinaliseerd volgt de volgende stap namelijk; het overzetten van de bedrijfsgegevens. De bedrijfsgegevens in de Microsoft Dynamics ERD worden overgezet naar de nieuwe gerealiseerde database die volgens het fysieke datamodel is gebouwd. Hiervoor zullen zogenoemde mapping schema's worden opgesteld. Deze schema's beschrijven in detail hoe de gegevens geconverteerd moeten worden en waar deze worden opgeslagen in de nieuwe database. Deze stap geeft garanties dat de data juist wordt overgezet.

Het uiteindelijk converteren en transporteren van de gegevens zal via een geautomatiseerd proces verlopen. Dit proces wordt ook wel ETL of: Extract, Transform en Load genoemd en zal ontworpen worden in de ETL tool van Microsoft Integration services (SSIS).

### 2.4.2 Fysiek datamodel

Ontwerpen van het Fysieke datamodel voor de warehouse als de cubes. In het fysieke datamodel is vastgelegd hoe de data wordt opgeslagen. Het fysiek ontwerp is de uitwerking van bijvoorbeeld een ster schema. In het ster schema kunnen eventueel aggregaties worden opgenomen voor de cubes. Hiervoor moet wel worden vastgesteld welke bevestigingen er op de database in kwestie plaats gaan vinden. De informatiebehoefte is standaard van aard en zal doormiddel van dashboards worden gefaciliteerd.

In de opdracht staat vermeld dat gebruikers de mogelijkheid moeten hebben om via vooraf gedefinieerde parameters het dashboard visueel te kunnen wijzigen. Op deze parameters (tabel velden) zullen indexen worden ontworpen en geplaatst.

### 2.4.3 ETL ontwerp & development

De laatste discipline binnen de data track is het realiseren van het fysieke datamodel.. Het datamodel wordt omgezet in tabellen (dimensies) met daartussen relaties. Een voorbeeld van een dimensie tabel zijn klanten of producten. Alle dimensies tezamen vormen de uiteindelijke database.

Tijdens het ontwerpen van de ETL processen wordt rekening gehouden met eventuele fouten en deze af te vangen zonder dat het proces als geheel vastloopt. Binnen de ETL tool van Microsoft Intergration Services (SSIS) kunnen fouten automatisch worden afgehandeld.

De volgende functies binnen SSIS kunnen bijvoorbeeld worden ingezet bij fouten:

- Event handler; als een specifieke error binnen het proces plaats vindt dan wordt er een event getriggerd waar een vooraf gedefinieerde actie plaats vindt.
- Door condities in te stellen waaraan de data aan moet voldoen. Afwijkende waarden worden automatisch afgevangen.
- Het bijhouden van een error log.

## 2.5 Business intelligence track

Het ontwerpen en bouwen van de rapportages en de dashboards vindt plaats binnen de Business Intelligence track. Deze track wordt niet als laatste uitgevoerd maar loopt parallel aan de andere tracks maar dan wel later in het algehele ontwikkelproces.

De rapportages en dashboards zijn vrijwel het eerste resultaat dat klanten te zien krijgen. Het is daarom van belang om zo snel mogelijk deze aan de klant te tonen, zodat deze opmerkingen en/of nieuwe requirements kan inbrengen. De ontwerpen en producten in de andere tracks moeten, als gevolg, ook worden geüpdatet.

Voor het ontwerpen van de dashboards wordt het volgende gedocumenteerd; doelstelling, autorisatie, parameters, design, berekeningen, filters/sortering en als laatste; welke gegevens de dashboards op gebaseerd zijn.

De dashboards worden ontwikkeld volgens ontwerp. Het aantal ontwikkelde dashboards kan gedurende dit project wijzigen. Als het datamodel flexibel is ontworpen dan moet het mogelijk zijn om, als het nodig is, nieuwe rapportages toe te voegen.

## 2.6 Testen

In dit project wordt een aparte testfase ingericht. Hoewel het testen gebeurt na elke stap/fase binnen dit project. Is het invoeren van een testfase toch noodzakelijk om formeel, volgens een plan, te controleren of de data in de dashboards correct zijn. Foutieve gegevens kan ook resulteren op verkeerde beslissingen van bijvoorbeeld een manager, die deze heeft gebaseerd op deze gegevens. De tests worden ontworpen en gedocumenteerd in een test plan. De testfase is officieel geen onderdeel van de lifecycle methodiek.

Er wordt deels gefaseerd getest door aan het einde van elk ETL proces te controleren of alle data correct is overgebracht. Omdat het om veel data gaat, is het niet mogelijk om deze met de hand te controleren. Het controleren van de data gebeurt dan ook deels geautomatiseerd, al dan niet met een aparte testtool. Daarnaast wordt ook gecontroleerd of het aantal tupels (regels) aan het einde van het ETL proces overeenkomen met de verwachtingen.

Binnen dit project wordt ook telkens geverifieerd of de gemaakte ontwerpbeslissing in lijn is met het behalen van de eisen en wensen van de klanten. Er wordt telkens gevalideerd; zijn de gegevens in de dashboard juist en werkt alles zoals beschreven en gewenst.

## 2.7 Deployment

Het is de wens van Centric om aan het einde van deze afstudeeropdracht de dashboards ontwikkeld te hebben. De 'big data' architectuur zal in overleg met de beheerders van Microsoft Dynamics als eventueel de ICT afdeling worden geïmplementeerd

De dashboards moeten de bestaande module International Shipments binnen Microsoft Dynamics qua functionaliteit aanvullen en niet in de weg zitten. Na de implementatie van de architectuur wordt deze getest om te controleren of deze naar behoren functioneert.

## 2.8 Growth

Informatiebehoeften van klanten zijn niet statisch en kunnen veranderen en zelfs groeien. Binnen de Kimball lifecycle methodiek is er plaats om het project incrementeel mee te laten groeien met deze ontwikkelen. Omdat dit project niet wordt ontwikkeld

voor een klant en deze generiek moet zijn, is de kans kleiner dat dit plaatsvindt. Echter kan er wel een iteratie van het project worden gestart, mocht de scope van het project worden uitgebreid. Gezien de ingeplande duur en heldere scope is de kans aanwezig.

## 2.9 Maintenance

In deze fase wordt beschreven hoe het beheer van de datawarehouse wordt geregeld. Omdat het systeem niet wordt geïmplementeerd bij een klant zijn er geen beheertaken aanwezig. Aan het einde van dit project en de stage wordt de datawarehouse architectuur gedocumenteerd. Naast de documentatie vindt er een overdracht plaats. De overdracht bestaat uit instructies hoe de architectuur geïmplementeerd -en onderhouden moet worden.

### 3 Project technieken & methodieken

Binnen dit project worden een aantal methoden en technieken gebruikt. Deze lijst hieronder zijn een aantal beschreven die voor de aanvang van dit project al te bepalen zijn:

- Dimensioneel modelering is een techniek om de gegevensarchitectuur van een database te visualiseren. Met dit model is het eenvoudiger om zware rapportages uit te voeren. Tegenhanger van dimensioneel is relationeel en word toegepast in transactionele systemen, waaronder het ERP pakket; Microsoft Dynamics AX
- Gebruik van het principe ETL of terwijl; Extract, Transform en Load. Data moet eerst worden geladen uit Dynamics. Vervolgens getransformeerd van een transactioneel naar dimensioneel model. Als laatste het laden van de bedrijfsdata in een datawarehouse.
- Tijdens dit project zal architecturale onderdelen gevisualiseerd worden door UML (Unified Modelling Language) dit om zoveel mogelijk duidelijkheid te garanderen richting de stakeholders. Een selectie uit de mogelijke diagrammen:
  - ERD staat voor Entity Relation Diagram en wordt toegepast om het dimensioneel model te visualiseren.
  - Bubble chart is diagram om een high level overzicht te geven van de dimensies binnen Dynamics.
  - Flowchart voor het visualiseren van de ETL processen.
- De gebruikte projectaanpak methodiek is globaal gebaseerd op Kimballs: Datawarehouse/business intelligence project lifecycle. In het hoofdstuk: aanpak en fasering staat beschreven hoe de methodiek globaal wordt toegepast.
- Het verzamelen van de requirements wordt d.m.v. interviews of eventueel workshops verzameld. Het documenteren en vastleggen van de requirements wordt deels gebruikt gemaakt van de methode van Nicole de Swart, zoals beschreven in haar handboek(2010). Daarnaast worden de requirements ook opgesteld volgens de standaarden voor BI-projecten van het bedrijf Info Support. Als laatste worden de requirements geprioriteerd volgens de MosCow methode.

Gebruik van bepaalde methodieken, technieken, standaarden en best practises, op het gebied van datawarehousing zijn afkomstig van het bedrijf Info Support. Zij hebben een ontwikkelstraat in gericht voor datawarehouse projecten. De beschreven methodieken worden toegepast indien het toepasselijk is binnen de opdracht. Het is niet bedoeling dat het gebruik van een methodiek de overhead binnen dit project te groot wordt.

## 4 Deliverables

Binnen dit project worden de volgende (tussen) producten opgeleverd:

- Plan van aanpak
- Specificatie requirements
- Data dictionary
- Big data ontwerp architectuur
  - Analyse modellen (dimensies, feiten, meetwaarden, hiërarchieën en mapping-schema's)
  - ETL schema en ontwerp
  - Datamodel (Gegevens architectuur van datawarehouse en de data marts)
  - Fysiek datamodel
  - Ontwerp dashboards
- Datawarehouse (in de vorm van data marts) plus ontwerp.
- Verladingen cube en ontwerp
- Dashboards zoals omschreven in de opdrachtomschrijving:
  - Overzicht met shipments die worden verscheept
  - Overzicht met shipments die aankomen,
  - Overzicht met shipments onderweg van loshaven naar warehouse (DC)
  - Overzicht van de piekmomenten in het warehouse (DC)
  - Een visuele map weergave waarop duidelijk is waar de shipments vandaan komen en naar toe moeten
- Testontwerp en testrapportage
- Evaluatie van het project met daarin eventueel lessen voor vervolg projecten.
- Overdracht documentatie

Naast de opdracht voor Centric zal er ook een afstudeerverslag worden opgeleverd. Dit verslag is bestemd voor de Haagse Hogeschool om de uitvoering van de afstudeeropdracht en de stage als geheel te kunnen beoordelen. Het ontwerp van de architectuur dient tevens als documentatie voor de overdracht.

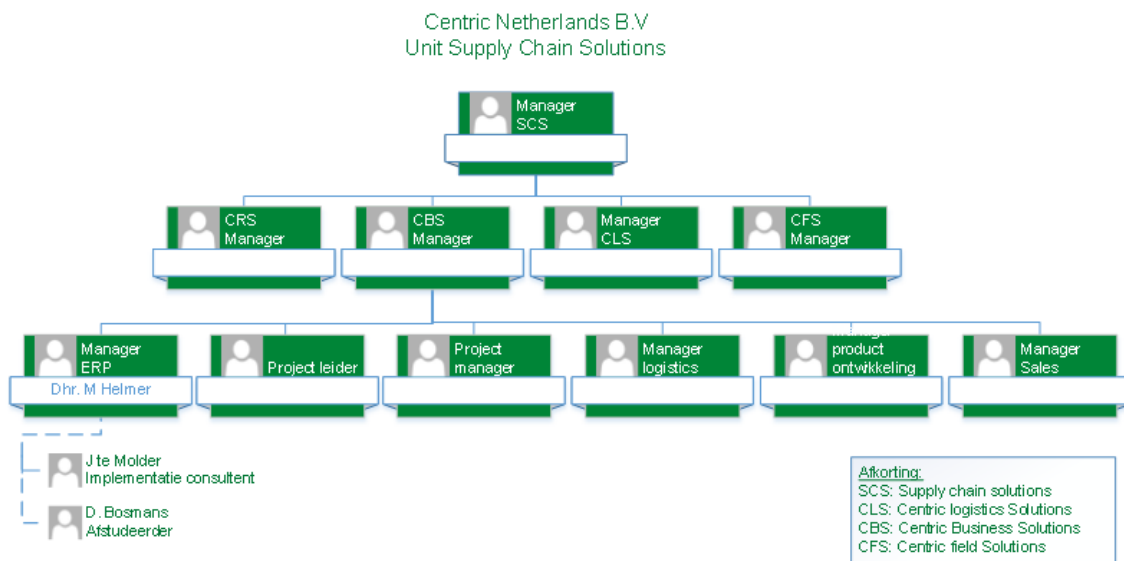
## 5 Communicatie Management

In dit hoofdstuk staat beschreven hoe de communicatie tussen de verschillende belanghebbende binnen dit project communiceren.

### 5.1 Organogram business unit

Deze opdracht wordt intern ontwikkeld binnen Centric en in de business unit Supply Chain Solutions. De directie collega's kunnen assisteren met het analyseren van het ERP pakket; Microsoft Dynamics AX. Daarnaast hebben consultants binnen de unit contact met de klanten (groothandels) als gevolg zijn zij op de hoogte van de requirements van de klanten.

Figuur 2 toont een organogram van de Business unit Supply Chain Solutions en daaronder de direct managers en collega's:



Figuur 3: Organogram

### 5.2 Stakeholders

Binnen dit project zijn er een aantal stakeholders. Met stakeholders wordt bedoeld betrokkenen of belanghebbende in het project. In dit project vindt er geen directe contact plaats met de klanten maar gelden wel als stakeholder, de dashboards worden immers gebruikt door hen. De volgende stakeholders zijn geïdentificeerd binnen het project:

Belanghebbende	Beschrijving
<b>Opdrachtgever</b>	Marcel Helmer is de direct betrokken manager bij het project. Hij neemt uiteindelijk de beslissing over de koers van dit project. Daarnaast kan hij geïnterviewd worden voor de requirements.
<b>Bedrijfsmentor</b>	Joost te Molder is de directe begeleider binnen Centric voor dit project. Hij houdt zich bezig met kwaliteit –en proces bewaking
<b>Afstudeerbegeleider</b>	De toegewezen afstudeerbegeleider van de afstudeerder die vanuit de Haagse Hogeschool (onderwijsinstelling van de afstudeerder) de voortgang bewaakt en begeleid.
<b>Consultants</b>	De consultants kunnen worden geïnterviewd om de requirements van de klanten te achterhalen.
<b>Experts ERP</b>	Voor technische vragen betreffende het Microsoft Dynamics AX ERP pakket kan worden geïnformeerd bij de beheerders

	en ontwikkelaars hiervan.
<b>Klanten</b>	De klanten of groothandels zijn eindgebruikers van de dashboards.
<b>Afstudeerder - ontwikkelaar</b>	Dit project wordt uitgevoerd door een afstudeerder informatica aan de Haagse Hogeschool. Binnen de afdeling werkt een tweede afstudeerder (L. Yu). Zijn opdracht is om te onderzoeken welke kennis er binnen Centric bestaat op het gebied van BI en hoe zij hier invullen aan kunnen geven.

### 5.3 Communicatie procedure

Er zal contact worden opgenomen met de verantwoordelijke stakeholders mochten er vragen zijn betreffende dit project. Het eerste aanspreekpunt binnen Centric is de bedrijfsmentor. Mochten er problemen ontstaan die de bedrijfsmentor niet afdoende kan oplossen dan wordt de opdrachtgever (directie manager) benaderd.

Afspraken inplannen gebeurt via de outlook agenda van Centric. Andere collega's, waaronder de bedrijfsmentor kunnen deze inzien. Afspraken worden op deze wijze formeel vastgelegd. In het geval van ziekte dient de afdeling HR geïnformeerd te worden.

### 5.4 Rapportage

Wekelijks wordt de voortgang van het project, met de bedrijfsmentor, besproken. Bij deze gelegenheid kunnen (tussen) producten ter controle worden overlegd. Naast deze vaste rapportage mogelijkheid kan er tussentijds ook vragen worden gesteld. De afstudeerder neemt eigen initiatief op dit gebied.

De Haagse Hogeschool verlangt van de afstudeer dat alle (tussen) producten via de digitale portfolio worden beschikbaar gesteld aan de afstudeerbegeleider. De afstudeerbegeleider heeft hierbij ook de geheimplicht.



## 6 Risico management

Tijdens het project is het van groot belang dat het duidelijk is welke risico's er voor kunnen komen tijdens het project. Een risico kan gezien worden als bedreiging voor het project en is een onzeker feit of onzekere omstandigheid die, als deze zich voordoet, gevolgen heeft voor het realiseren van de projectdoelstellingen oftewel het resultaat van het project.

De kans dat een risico optreedt en de mate van negatief impact op het project, wordt aangegeven op een schaal van 1 tot 5. De laatste kolom geeft het risico aan (kans \* effect). Op deze wijze kan goed worden afgewogen welke bedreigingen, met de hoogste punten aantal, de meeste aandacht verdienen.

De volgende risico's heeft de afstudeerder geïdentificeerd binnen dit project met de volgende tegenmaatregelen:

Nr.	Risico	Tegenmaatregel	K	I	Factor
1	Het project raakt uit scope omdat de afstudeerder de requirements onvoldoende beheert, waardoor er meer additionele wensen aan het project worden toegevoegd.	Wijzigingen overleggen met de opdrachtgever. De nieuwe requirements toetsen aan zaken zoals; de scope, toegevoegde waarde en technische haalbaarheid.	4	4	16
2	De planning van het project loopt niet op schema door het onjuist stellen van prioriteiten.	De requirements prioriteren volgens de MoScow methode. De hoofd van bij zaken beter te scheiden.	5	4	20
3	Tunnel visie tijdens de ontwerp fase vanwege inflexibiliteit bij het accepteren van technische beperkingen of vasthouden aan het eigen gelijk.	Hulp vragen aan de relevante experts om te verifiëren of de beperking daadwerkelijk bestaat. Rationale opschrijven van zowel eigenstandpunten als die van de tegen partij en deze te vergelijken. Op deze manier ontstaat er inzicht in de argumenten van de andere partij.	3	2	6
4	Het niet kunnen loslaten van onvolkomenheden binnen het project als gevolg van het perfectionisme van de afstudeerder.	Vaststellen of het project zonder een bepaalde feature nog steeds kan worden voldaan aan de doelstellingen van het project. Daarnaast moeten niet kritieke functionaliteit niet de hoogste prioriteit krijgen. Als laatste hulp vragen aan de bedrijfsmentor zaken in perspectief te plaatsen.	4	3	12
5	Gebrek aan toegang tot de stakeholders binnen dit project.	De betreffende stakeholders op de hoogte stellen van de opdracht en de mogelijke resultaten. Dit om de stakeholders te enthousiasmeren en te overtuigen van de toegevoegde waarde van dit project. Daarnaast via de interne agenda/kalender van Centric afspraken inplannen en deze geconfirmeerd te krijgen. De	1	3	3

		manager kan op dit punt ook van betekenis zijn.			
6	Haagse Hogeschool is gesloten in de zomervakantie met als gevolg; onbereikbaarheid van de studiebegeleider.	Kritieke en belangrijke vragen voor de aanvang van de zomervakantie stellen aan de studiebegeleider. Daarnaast afspraken maken voor een langere periode.	5	1	5
7	Microsoft beslist een patch voor Dynamics uit te brengen waardoor het datamodel verandert	De architectuur in logische lagen te ontwerpen om de koppeling van het datamodel van Dynamics AX te verkleinen.	1	3	3
8	Complexiteit van het datamodel van Microsoft Dynamics AX en gebrek aan kennis binnen Centric van dit model.	Het raadplegen van externe bronnen of experts om beter inzicht te krijgen in het datamodel. Inzichtelijk maken van de werking van het ERP pakket en de bedrijfsprocessen. Dit om de betekenis en de relaties tussen deze verschillende gegevens beter te begrijpen.	2	2	4

De bovenstaande tabel bevat een aantal project risico's, maar deze lijst is niet onuitputtelijk. De afstudeerder zal op tijd maatregelen nemen of hulp inschakelen indien nodig.

## 7 Planning

In dit hoofdstuk staat de planning beschreven. Deze planning is gebaseerd op de werkwijze zoals die beschreven is in hoofdstuk 3.3. Aanpak & Fasering.

### 7.1 Randvoorwaarden

Om de beschreven planning op schema te houden zijn er voor dit project de volgende randvoorwaarden gesteld:

- Geen andere taken naast de opdracht die niet bijdragen aan het behalen van haar doelstellingen.
- Beschikbaarheid van hard –en software, zoals beschreven in hoofdstuk 1.7.
- Beschikbaarheid van de verschillende stakeholders.

### 7.2 Projectplanning

De officiële duur van het project is; vijf maanden of van 12-05-2014 t/m 03-10-2014. De zomervakantie behoort niet tot de officiële afstudeerperiode. Dit betekent echter niet dat het project dan ook stilstaat. Het is wel de voornemen van de afstudeerder om, waar dit kan, een aantal dagen vrij te nemen. De zomervakantie wordt als buffer gebruikt om onvolkomenheden binnen dit project op te vangen. Het project heeft totaal een duur van 100 werkdagen of 800 uur.

De globale planning van de activiteiten en oplevering van (tussen)producten tijdens het uitvoeren van deze afstudeeropdracht:

Activiteit / product	Begin datum	Eind datum	Duur in werkdagen
Plan van aanpak	12-5-2014	16-5-2014	5 Dagen
Specificatie requirements	19-5-2014	23-5-2014	5 Dagen
Analyse module International Shipments & gegevensbronnen	26-5-2014	6-6-2014	10 Dagen
Ontwerp architectuur	9-6-2014	18-7-2014	30 Dagen
Realiseren & testen*	21-7-2014	19-9-2014	37 Dagen
Implementatie	18-9-2014	19-9-2014	3 Dagen
Afstudeerverslag	22-9-2014	3-10-2014	10 Dagen

In bijlage B kunt u een Gantt Chart vinden met de detail planning. In deze planning kunt u zien dat de tracks enige overlap met elkaar hebben. Dit betekent dat deze taken dan parallel worden uitgevoerd. De zwarte lijnen representeren de tracks.

## 8 Bijlage: A: opdrachtformulering

### Opdrachtformulering

Ontwerp en ontwikkel een systeem waarin middels een dashboard met een aantal verschillende overzichten en visualisaties hulp wordt geboden bij het succesvol kunnen managen van international shipments. De insteek is om het systeem te ontwerpen met bestaande tooling (Microsoft stack).

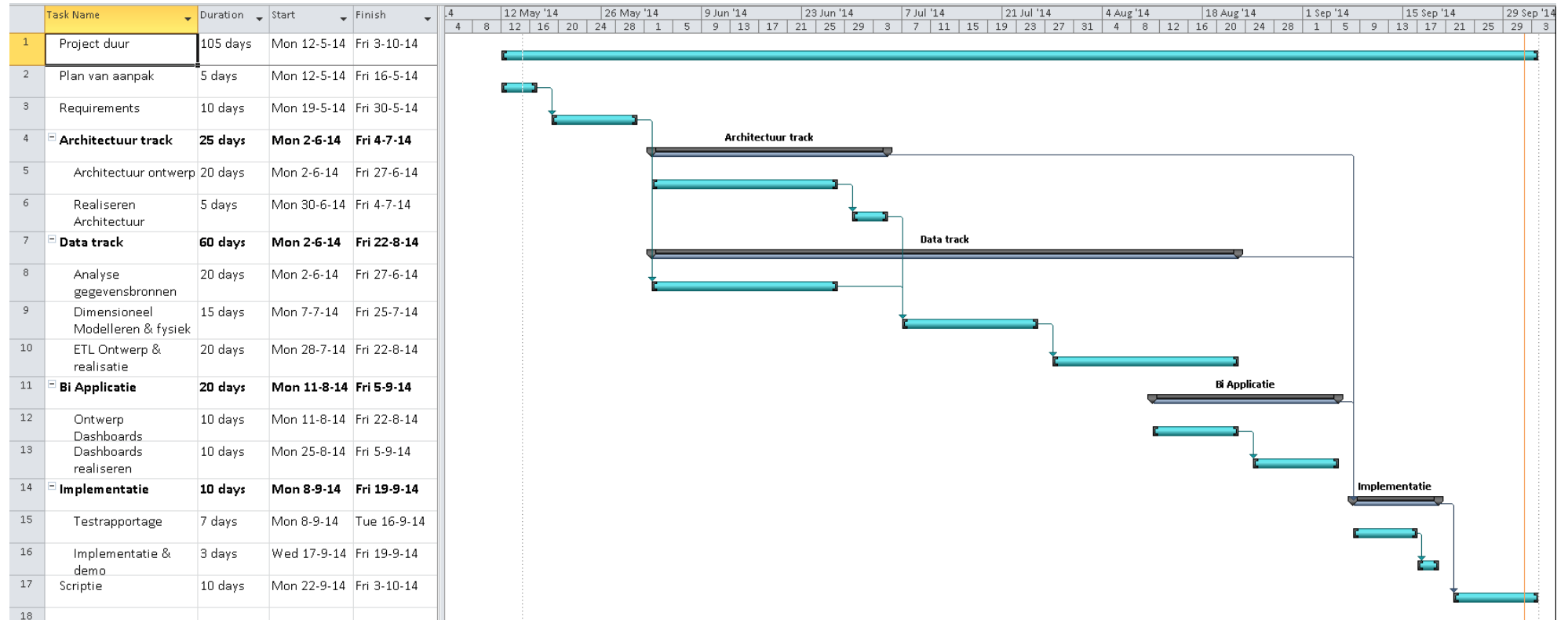
Het uiteindelijk op te leveren eindproduct is een cube en een dashboard met betrekking tot de Microsoft Dynamics AX2012 module 'International shipments'. Hier zijn de volgende elementen in verwerkt:

- Overzichten op basis van transactionele gegevens en de te ontwerpen cube
- Visuele weergaves op basis van de te ontwerpen cube en andere externe bronnen (bv. Bing maps)
- Bepaalde delen van het dashboard zijn voor de gebruiker in te stellen met behulp van parameters
- Bepaalde delen van het dashboard bieden de mogelijkheid om in te zoomen op de details

Bijvoorbeeld:

- Overzicht met shipments die worden verscheept, per jaar, maand, week, dag.
- Overzicht met shipments die aankomen, per jaar, maand, week, dag.
- Overzicht met shipments onderweg van loshaven naar warehouse (DC)
- Overzicht van de piekmomenten in het warehouse (DC) , bv. met behulp van een scatter chart
- Een visuele map weergave (bv. Power Map) waarop duidelijk is waar de shipments vandaan komen en naar toe moeten

## 9 Bijlage B grantt chart



connect.  
engage.  
succeed.

**BIJLAGE B:**

**COMPETENTIELIJST**

## **Te demonstreren competenties en wijze waarop**

De volgende beschreven beroepstaken zal ik demonstreren tijdens het uitvoeren van deze afstudeeropdracht om aan te tonen deze ik deze competenties zoals beschreven in de beroepstaken, beheers:

### **1.4 Uitvoeren analyse door definitie van requirements niveau 3**

Het analyseren en documenteren is ook een onderdeel van deze opdracht. De requirements worden verzameld door zowel de opdrachtgever als de consultants binnen Supply Chain Solutions te interviewen/te bevragen. De opdrachtgever(de manager van de afdeling Supply Chain Solutions) is via zijn contacten met de groothandels (gebruikers van de dashboards) op de hoogte van hun informatiebehoefte. De requirements worden verzameld om te analyseren welke gegevens en parameters de dashboards moeten bevatten. De uitkomsten van deze analyse bepaald welke gegevens verzameld moeten worden voor de dashboards.

De requirements worden onderverdeeld in business, functionele en non-functionele requirements. Voor het beschrijven van de non-functionele requirements wordt van een deel van de ISO-9126 standaard (kwaliteitskenmerken van software) gebruikt.

De requirements kunnen tijdens dit project ook wijzigen als er bij de stakeholders (groothandels, opdrachtgever en de consultants) nieuwe wensen naar boven komen en/of de scope van de opdracht wordt verkleint/vergroot. De veranderingen in de requirements worden genoteerd. De requirements worden alleen gewijzigd in overleg met de opdrachtgever en/of de bedrijfsmentor.

### **2.1 Opstellen gegevensmodel voor database niveau 4**

Een van de activiteiten voor het architectuur ontwerp is het opstellen van een gegevensarchitectuur (datamodel) voor de datawarehouse. Voor dit model moet er bepaald worden hoe de gegevens genormaliseerd zullen gaan worden. Aantal gegevensbronnen zal worden gedenormaliseerd en andere juist weer genormaliseerd. De mate van normalisatie is afhankelijk van de non-functionele eisen zoals performance van rapportages of efficiëntie van data opslag.

Het datamodel zal een groot aantal aan tabellen en relaties bevatten, dit is echter wel afhankelijk van de gekozen dimensioneel model (Ster, Snowflake of een combinatie van beide). Daarnaast kan het aggregaat of andere type tabellen bevatten. Het model kan complex van aard zijn en kan al dan niet worden gevisualiseerd door middel van een ERD.

Het datamodel moet historische gegevens kunnen opslaan. Daarnaast moet het mogelijk zijn om verbanden tussen historische gegevens te kunnen leggen (bv huidige versie tegen over de vorige versie van het gegeven). Naast een datamodel voor een database moet er ook bepaald worden welke aggregaties, hiërarchieën of analyse paden er in de data cubes worden opgenomen.

Door de complexiteit en benodigde kennis van databases voor deze opdracht, kan ik deze beroepstaak op niveau 4 demonstreren.

### **2.2 Ontwerpen, bouwen en bevragen van een database niveau 4**

Aan de hand van het ontworpen datamodel wordt de database gebouwd. Voor het bouwen van de database wordt gebruikgemaakt van de daarvoor bestemde tools. Naast de tabellen worden ook enkele stored procedures gemaakt voor de rapportages.



In deze opdracht moet zowel een multinational database model als een cube worden bevestigd. Het gebruik van MDX behoort niet tot deze opdracht. Het bevestigen van de cubes wordt doormiddel van stored procedures binnen Microsoft Reporting Services (SSRS) geregeld.

Het bevestigen van de database door middel van een data manipulation language (DML) zal tijdens het ETL en test proces uitgebreid aanbod komen (bv. tabel joins of aggregaat functies). Het SQL dialect van Microsoft wordt hiervoor gebruikt.

Ook is het wenselijk dat er een keuze wordt gemaakt voor een fysieke opslag structuur voor de database. Met de fysieke structuur wordt bedoeld het leggen van Indexen of het toepassen van partitionering.

Door de bovengenoemde activiteiten en de kennis die ik tijdens de minor van Info Support heb opgedaan, ben ik er zeker van deze ik deze beroepstaak op niveau 4 kan uitvoeren en demonstreren.

### **2.3 Uitvoeren gegevensconversie niveau 3**

Met behulp van een ETL tool worden de gegevens uit de bronsystemen geconverteerd en opgeslagen in de datawarehouse. Voor de gegevensconversie zal ik mapping schema's opstellen. In deze mapping- schema's staat beschreven welke conversies er zijn uitgevoerd en naar welke tabellen deze worden getransporteerd. Voor de conversie worden de kwaliteit en consistentie van de gegevens gecontroleerd, deze activiteit vindt plaats binnen de analyse fase. Het opstellen van conversie regels is een belangrijk onderdeel van de gegevensconversie. Voorbeelden van conversie zijn het veranderen van data types, oplossen van fouten en consistent maken aan de geldige bedrijfsregels. Dit kan automatisch worden afgehandeld binnen het ETL proces.

Het gebruik van SSIS maakt de gegevens conversie nog steeds niet gemakkelijk en vereist nog veel kennis van databases en SQL.

### **3.1 Ontwerpen (software)architectuur niveau 3**

Voor het bouwen van een datawarehouse moet je ook beslissingen nemen op het gebied van architectuur, net als bij het ontwikkelen van software. Een voorbeeld van een beslissing kan zijn of het nodig is om een datavault (Linstedt) op te zetten. De architectuur beslissingen worden gedocumenteerd. Daarnaast moet er gebruik gemaakt worden van UML om de infrastructuur op een abstract niveau te visualiseren.

Ook worden de dashboards van tevoren functioneel beschreven en gedocumenteerd op welke data deze gebaseerd is met eventueel de gebruikte formules/aggregaties.

### **3.5 Uitvoeren van en rapporteren over het testproces niveau 3**

Het testen gebeurt na elke stap/fase binnen deze afstudeeropdracht. Telkens wordt er geverifieerd of de gemaakte ontwerpbeslissing in lijn is met het behalen van de eisen en wensen van de gebruikers. Er wordt telkens gevalideerd; zijn de gegevens in het dashboard juist en werkt alles zoals beschreven en gewenst.

Tijdens het ontwerpen van de ETL processen wordt rekening gehouden met eventuele fouten en deze af te vangen zonder dat het proces als geheel vastloopt. Binnen de ETL tool van Microsoft Intergration Services (SSIS) kunnen fouten automatisch worden afgehandeld. De volgende functies binnen SSIS kunnen worden ingezet bij fouten:

- Event handler; als een specifieke error binnen het proces plaats vindt dan wordt er een event getriggerd waar een vooraf gedefinieerde actie plaats vindt.

- Door condities in te stellen waaraan de data aan moet voldoen. Afwijkende waarden worden automatisch afgevangen.
- Het bijhouden van een error log.

Aan het einde van een ETL proces wordt gecontroleerd of alle data correct is overgebracht. Omdat het om veel data gaat, is het niet mogelijk om deze met de hand te controleren. Het controleren van de data gebeurt deels geautomatiseerd. Daarnaast wordt ook gecontroleerd of het aantal tupels aan het einde van het ETL proces overeenkomstig zijn met de verwachtingen.

In deze afstudeeropdracht wordt dus getest zoals hierboven is beschreven. Voor deze opdracht ontwerp ik tests, voert deze uit en rapporteer als gevolg het resultaat van de tests in een testrapportage. De beschreven testtechnieken heb ik geleerd tijdens de minor van Info support.










connect.  
engage.  
succeed.

**BIJLAGE C:**

**ANALYSE  
VERLADINGPROCES**

## Omschrijving verladingenproces

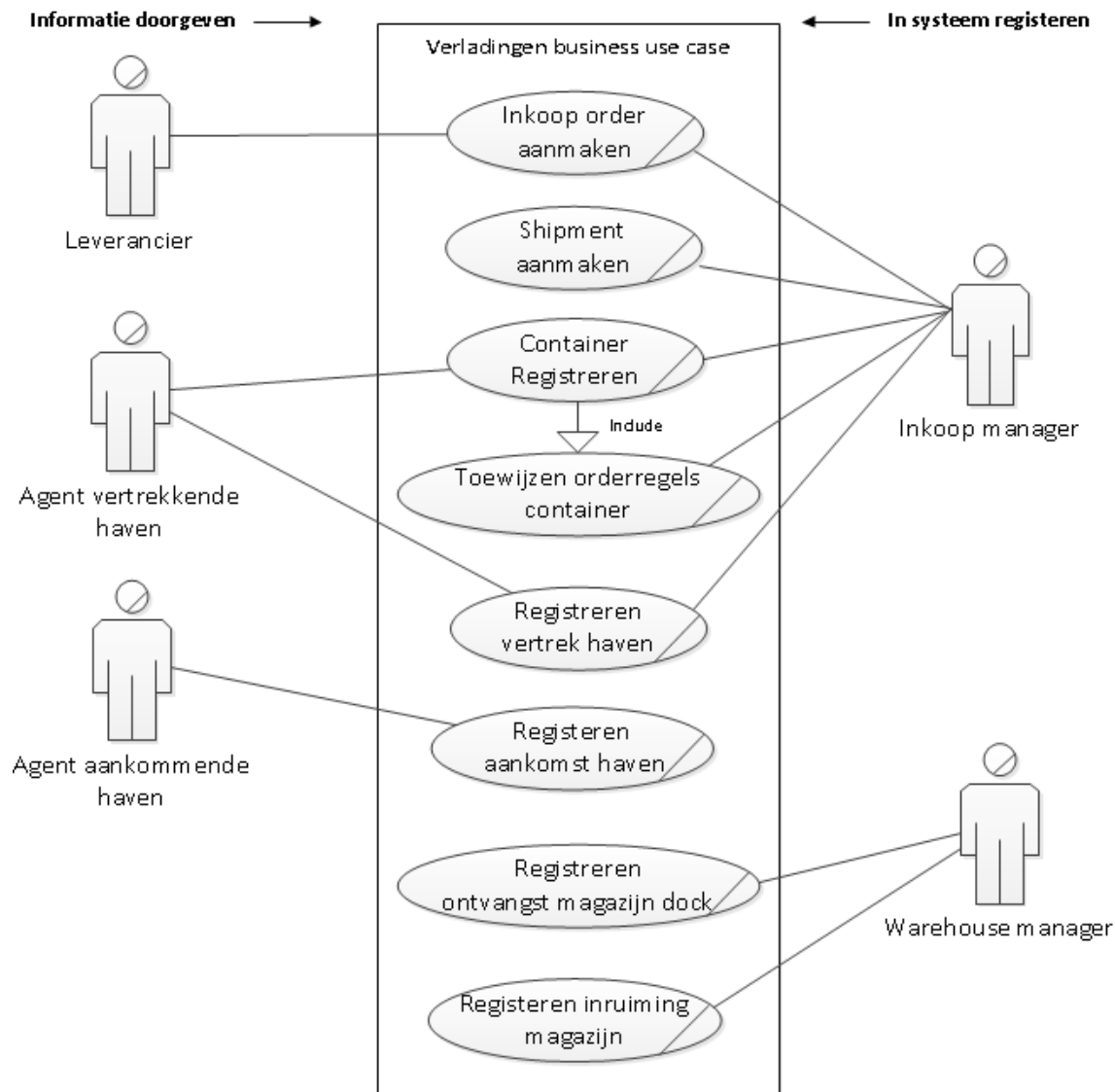
In Figuur 2 ziet u een grafische weergave van het verladingproces. In de linker kolom 'proces' ziet u de vast gedefinieerde verladingen fasen die in de tradeworld module zijn vastgelegd. In de rechter kolom 'verlading fase' ziet u een voorbeeld van fases die per klant, door een implementatie consultant kan worden ingesteld. Deze fasen kunnen per klant verschillen.

	Proces	Verlading fase
         	Signalering behoefte	
	Inkooporder	
	Verlading aangemaakt	Verlading aangemaakt
	Geschatte afvaartdatum	
	Bevestigde afvaartdatum	Varend
	Geschatte aankomstdatum	
	Bevestigde aankomstdatum	Ontvangstdatum bevestigd
	Aankomstdatum haven	
	Geschatte losdatum	
	Bevestigde aankomstdatum	
	Aankomstdatum magazijn	
	Aankomst goederen	Ontvangen
	Pallettransport	
	Verlading afsluiten	

Figuur 2 Grafisch overzicht verlading proces

## Business use case

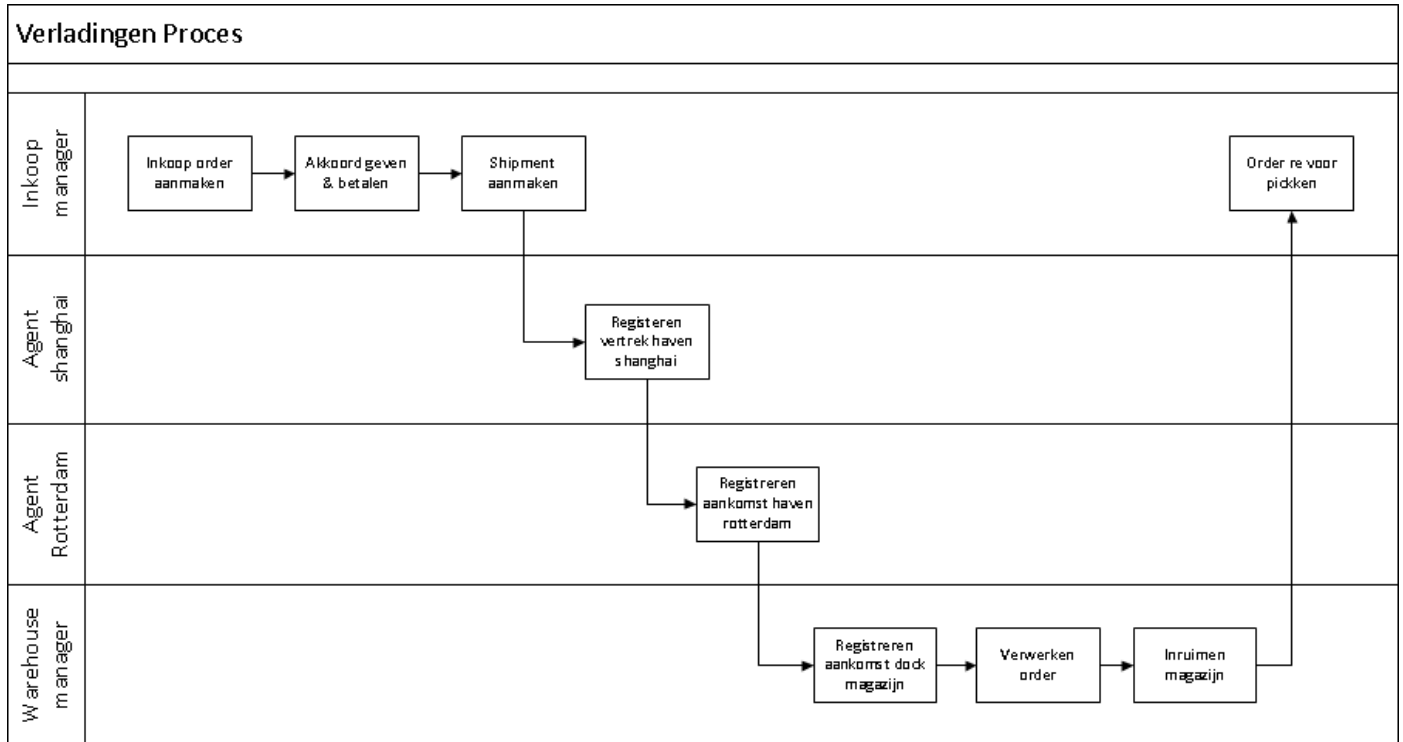
In Figuur 3 ziet u een business use case diagram. Het toont op een hoog niveau de verschillende betrokken stakeholders bij het verladingen proces. De stakeholders aan de linkerkant geven informatie door aan het bedrijf. Deze zijn niet werkzaam bij het bedrijf zelf en zijn extern. De stakeholders aan de rechterkant registreren de ontvangen gegevens van de externe stakeholders in Dynamics.



Figuur 3 Business use case diagram

## Business proces diagram

De bedrijfsmentor, vond voor het modeleren van het verladingen proces, een business proces diagram, gewenst. In Figuur 4 het diagram.



Figuur 4 Business proces diagram verladingen

Analyse van condities en de berekening van data in het verladingenproces. Op deze manier kan er berekend worden of een shipment te vroeg, te laat of op tijd is afgerond.

**Verwachte afrondingsdatum shipment** = doorlooptijd van fases bij elkaar opgeteld

**Verwachte aankomstdatum hub** = Vertrekdatum haven AND Verwachte aankomstdatum hub

### **Geplande aankomstdatum(shipping)**

De geplande aankomst van de container bij het magazijn dock wordt gebaseerd op de verwachte vertrekdatum van shipment bij hub met opgeteld de doorlooptijd(leadtime) van het gekozen vervoerstype. Als ConfirmedArrivalDate in shipment niet het zelfde is als EstimatedArrivalDate in shipment dan word de doorlooptijd (leadTime) van vervoerstype bij ConfirmArrivalDate opgeteld.

**PlannedArrivalDate** = {

ConfirmedArrivalDate = 0 → EstimatedArrivalDate + LeadTime,

ConfirmedArrivalDate = ¬ 0 → ConfirmedArrivalDate + LeadTime

### Container Pickdatum

PickDate = {

ConfirmedArrivalDate = 0 → PlannedArrivalDate + LeadTime

ConfirmedArrivalDate = ¬ 0 → ConfirmedArrivalDate + LeadTime

## Componenten International Shipments

Hieronder de beschrijving van de individuele componenten in de module Tradeworld betreffende verladingen:

Module	Omschrijving
Verladingtypen	Het definiëren van verschillende verladingtypen, ook wel transporttypen genoemd. Voorbeelden zijn een verlading door middel van scheepvaart, wegvervoer, spoorwegen of luchtvaart. Elk type vervoer kent hiervoor zijn eigen specifieke vervoersmiddel.
Verladingfases	Het definiëren van een reeks van fasen. Een verlading kan zich in een bepaalde fase van de reeks begeven. Bijvoorbeeld: in behandeling, varende of aankomst in haven.
Containertypen	Het definiëren van de verschillende containers en bijbehorende eigenschappen zoals het maximum gewicht en volume. Containers worden gebouwd met gestandaardiseerde afmetingen. Een containertype wordt gebruikt om te bepalen hoeveel inkooporder(regels) in 1 container passen.
Hubs	Het definiëren van de verschillende hubs. Een hub is een knooppunt voor het goederenverkeer bestaande uit onder andere rangeerstations, werven, zeehavens, vrachtwagen terminals, of combinaties van deze. Op een hub wordt de lading uitgewisseld tussen voertuigen onderling of tussen vervoerswijzen. Bijvoorbeeld van zeevervoer naar vrachtwagenvervoer, of van vrachtwagen naar luchtvervoer. Een hub kan van het type loshub of laadhub zijn. Een laadhub is het punt waar de goederen voor transport worden aangebracht. De loshub is de bestemmingsplaats waar de goederen gelost moeten worden.
Losplaatsen	Het definiëren van de verschillende losplaatsen. Losplaatsen zijn plekken waar gelost kan worden en dus ook ingericht zijn voor het lossen. Zo is er bijvoorbeeld een laad- en losdok aanwezig. Dit is een plaats tegen of in een gebouw waar een vrachtwagen kan laden en lossen. In een dergelijk dok is de laadruimte van de vrachtwagen vaak op dezelfde hoogte als de vloer van het gebouw, zodat het laden en lossen van de vrachtwagen met een heftruck of steekwagen direct mogelijk is
Vervoerder	Het definiëren van de verschillende vervoerders, ook wel een logistieke dienstverlener of vervoersmaatschappij genoemd. Een vervoerder is een partij die het vervoer van de goederen verzorgt, zoals een transportbedrijf of rederij.



connect.  
engage.  
succeed.

**Bijlage D**

**Gegevensbron analyse**

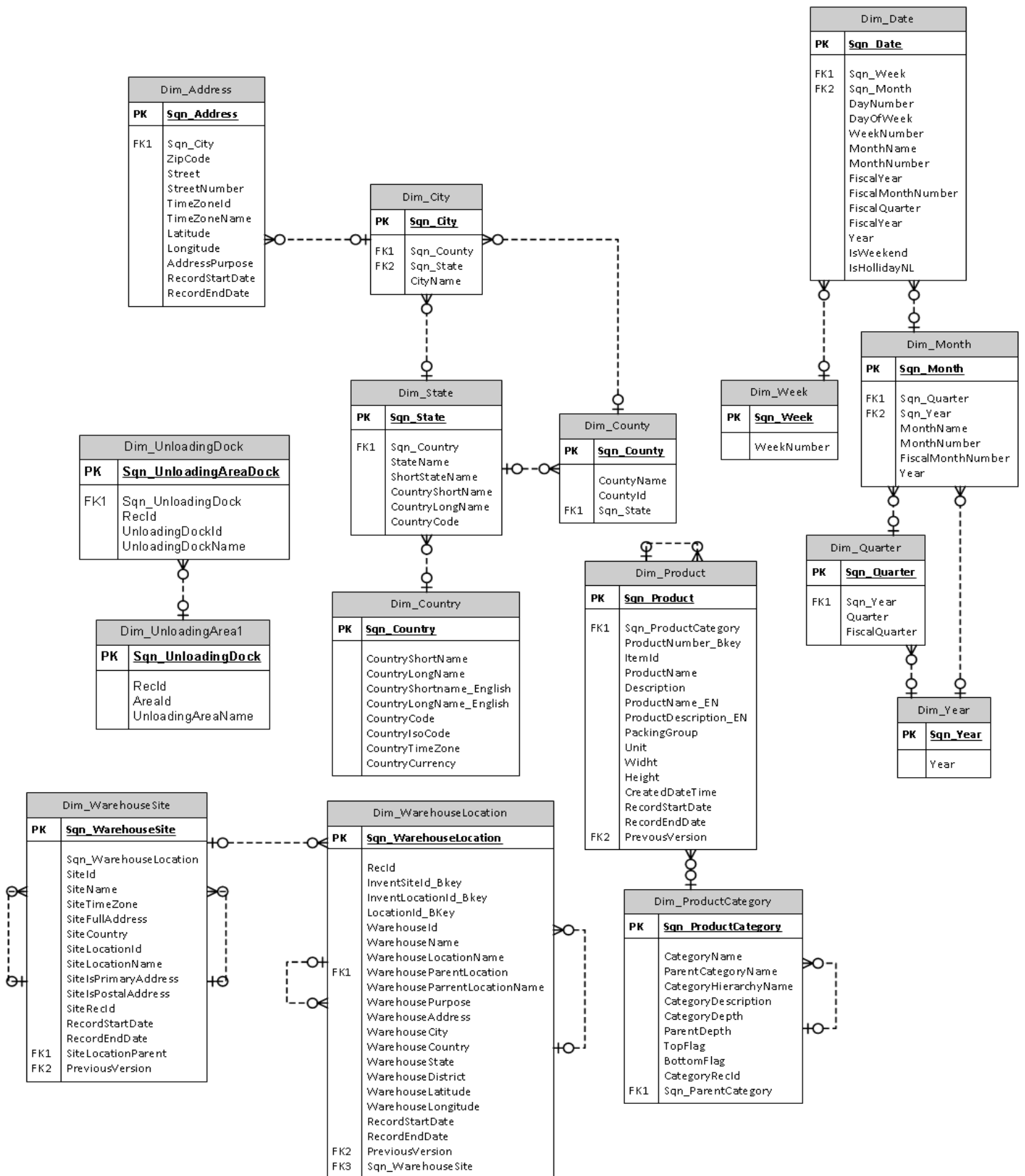
De gegevensbron analyse is niet deze versie van het verslag weggelaten, in verband met concurrentie gevoelige productinformatie

connect.  
engage.  
succeed.

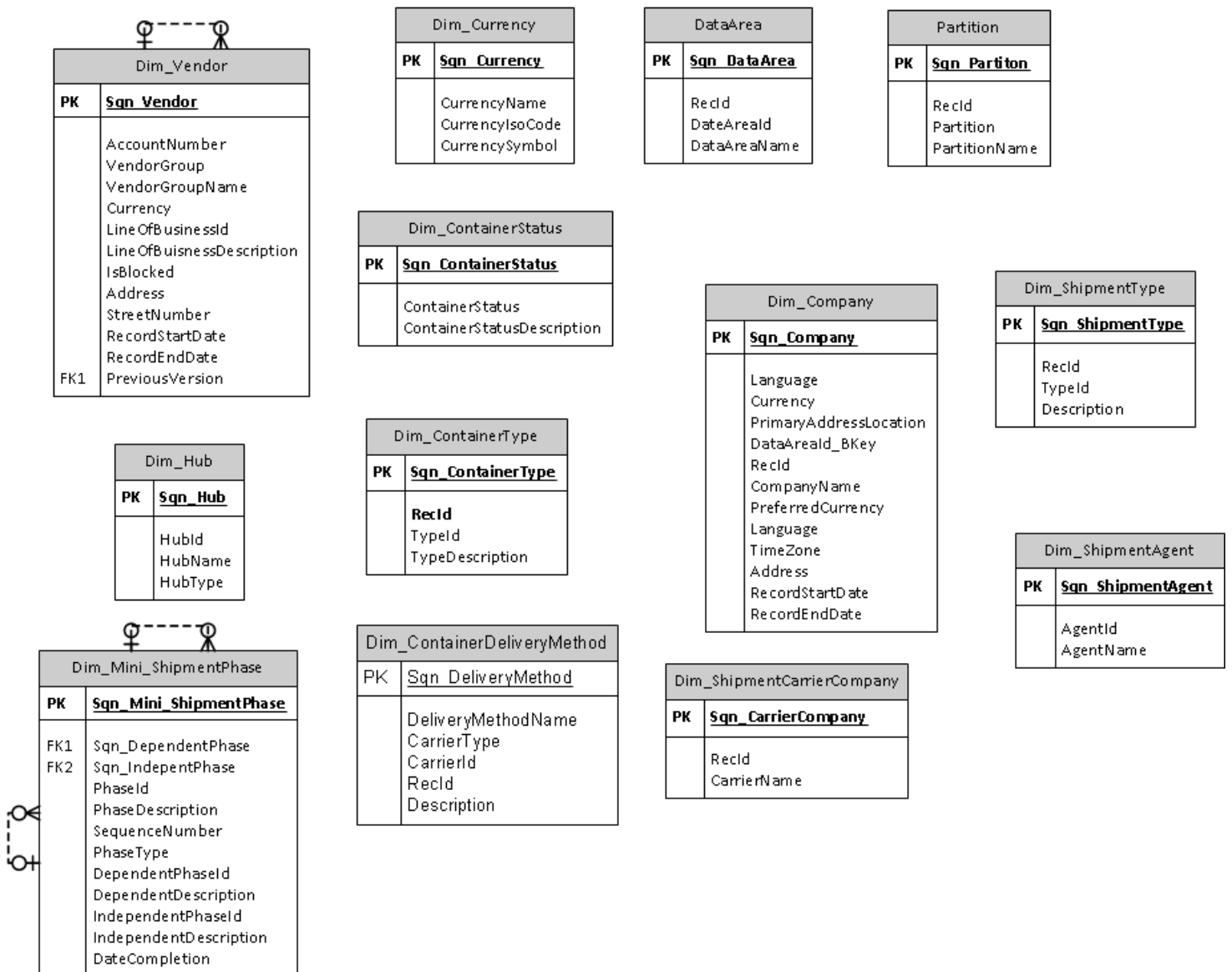
## Bijlage E

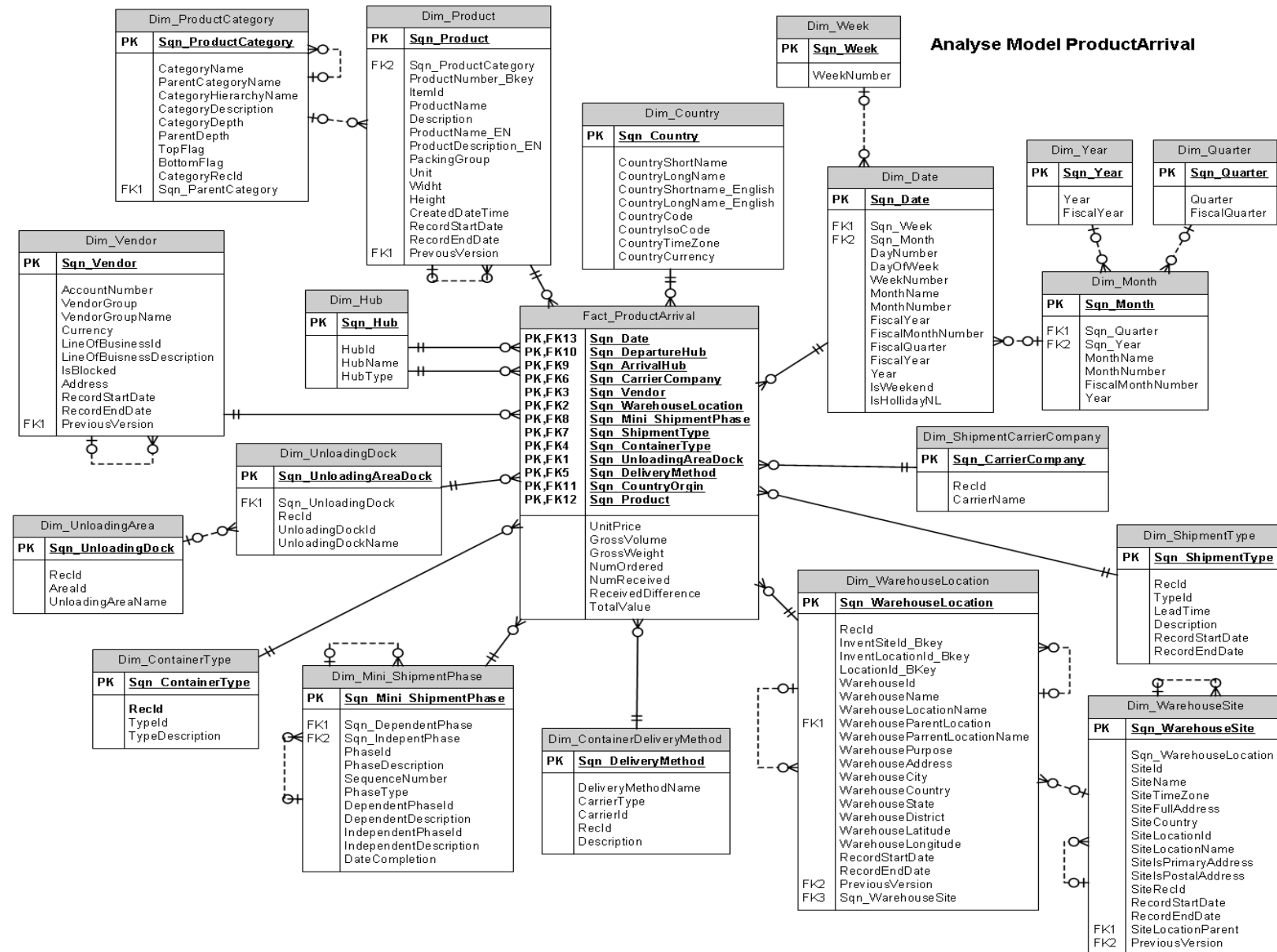
Dimensie ontwerp &  
Analyse modellen

## Snowflake (hiërarchieën) dimensies1

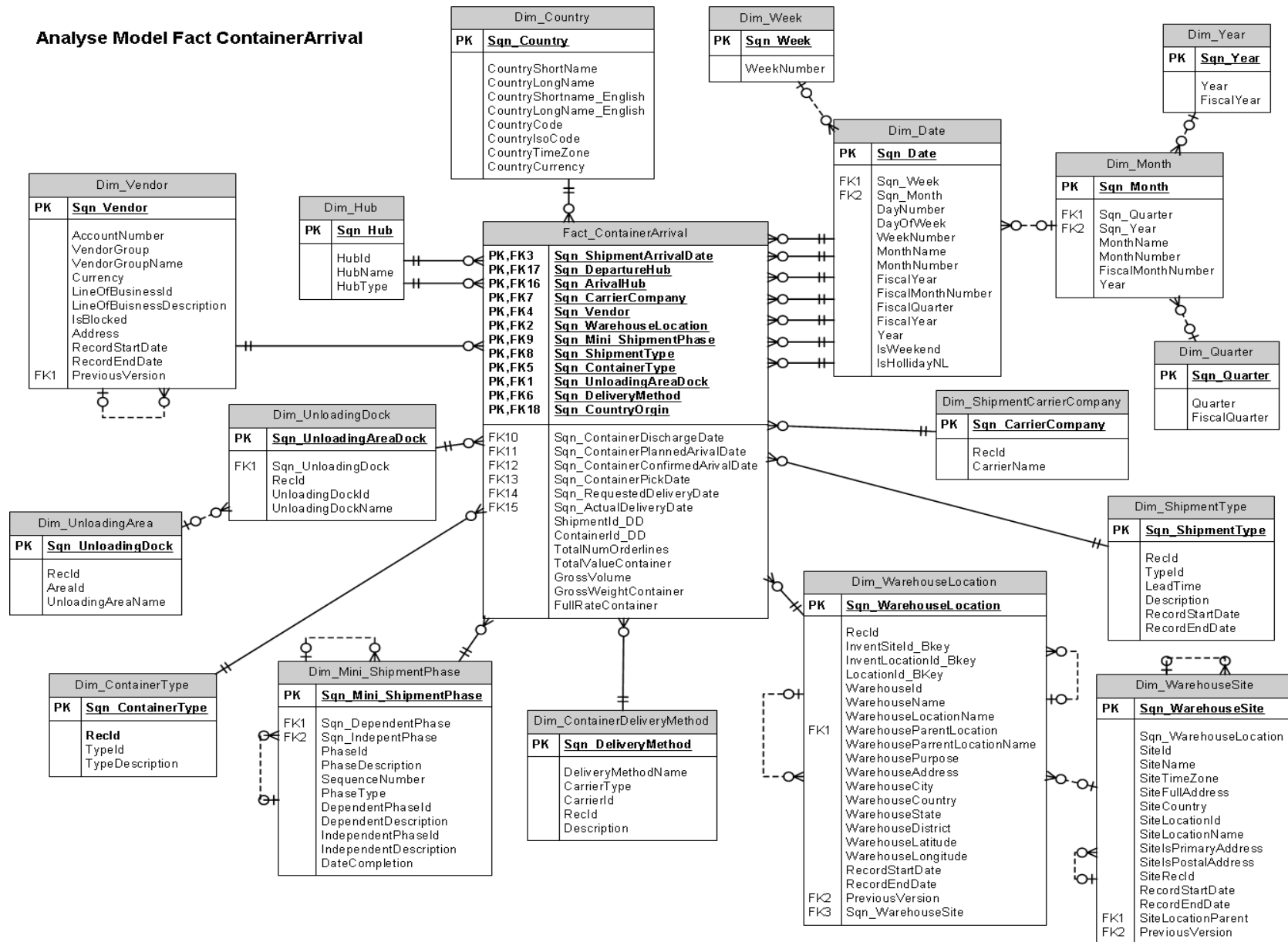


## Dimensies 2



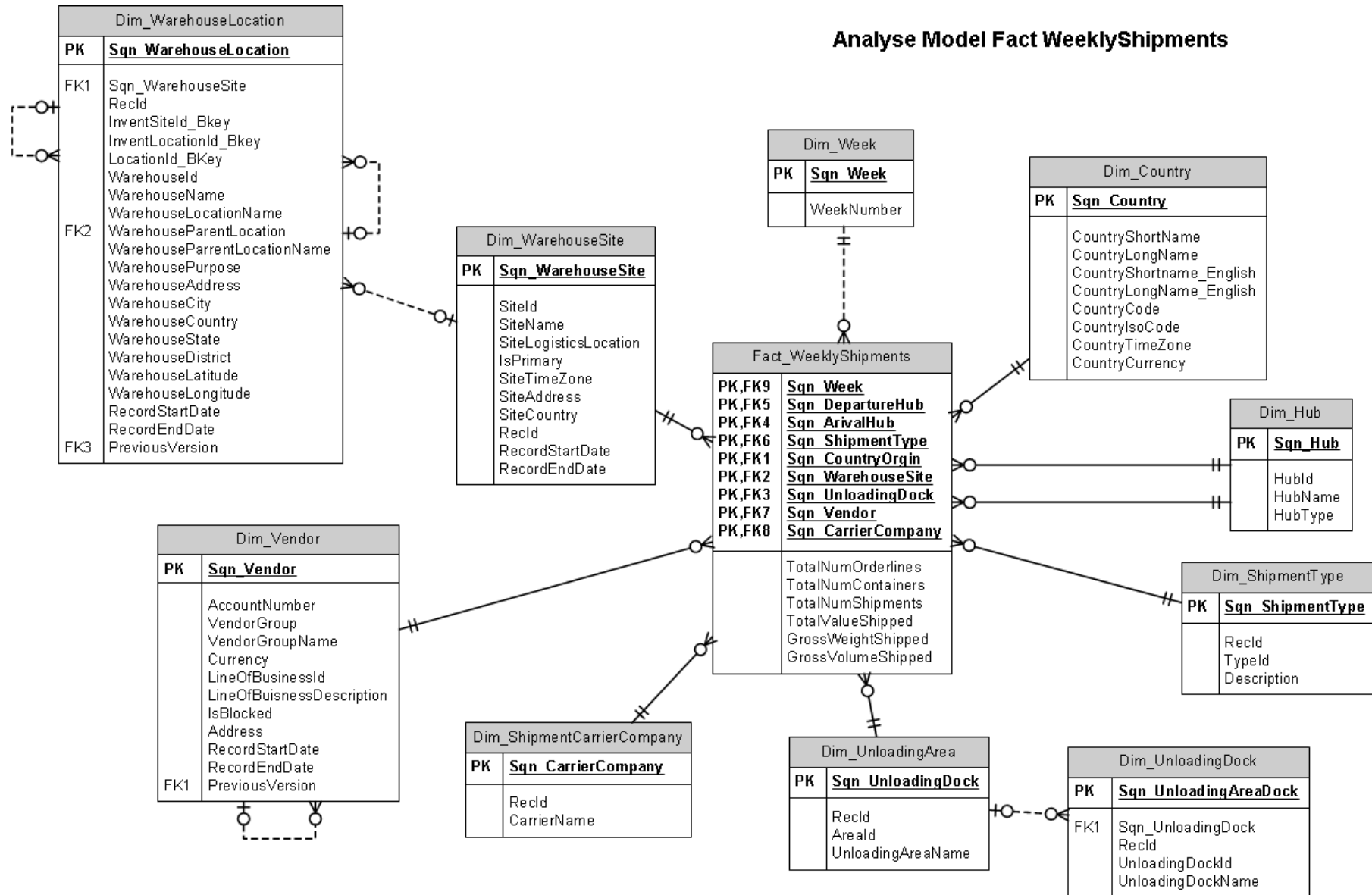


## Analyse Model Fact ContainerArrival

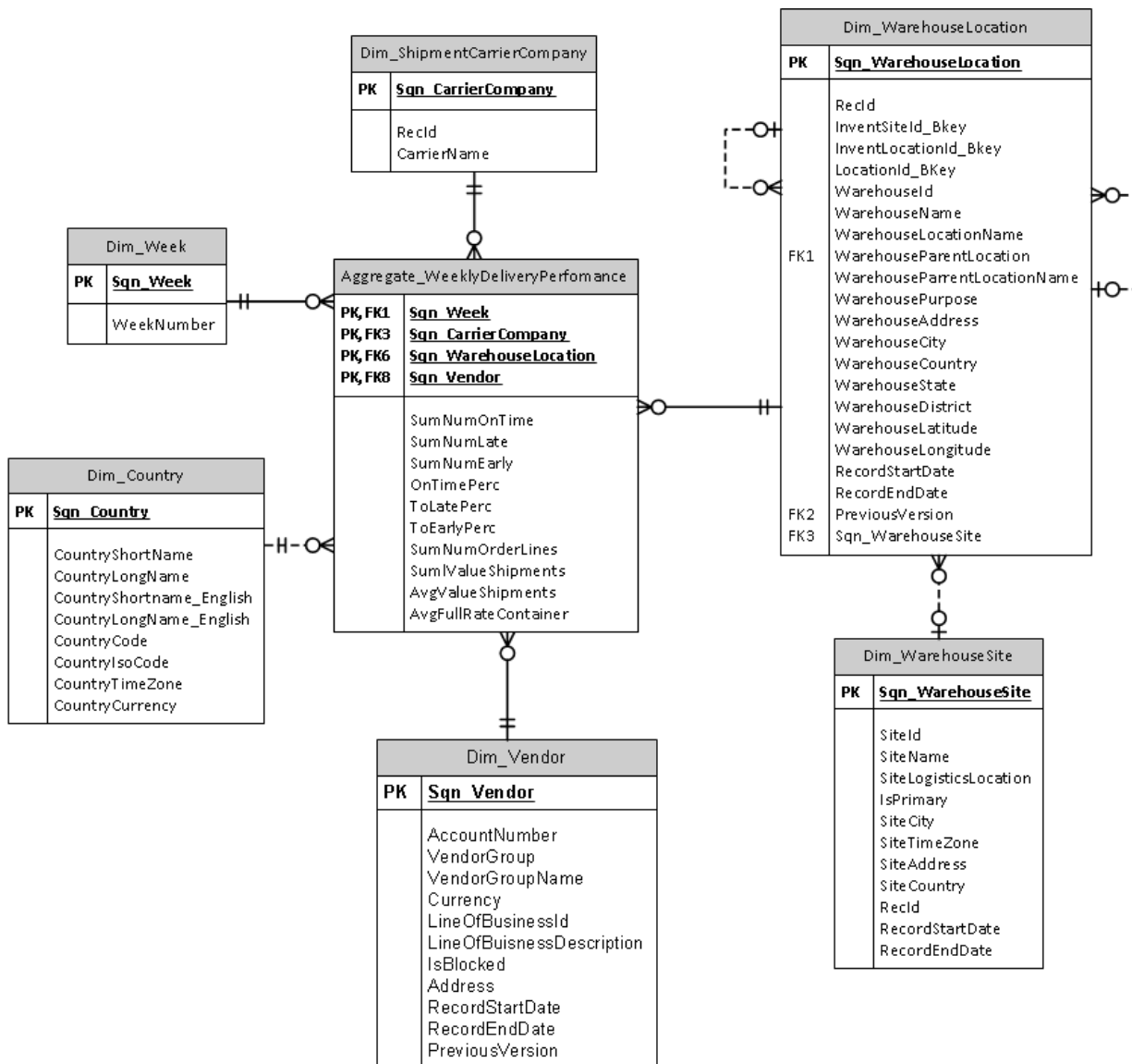




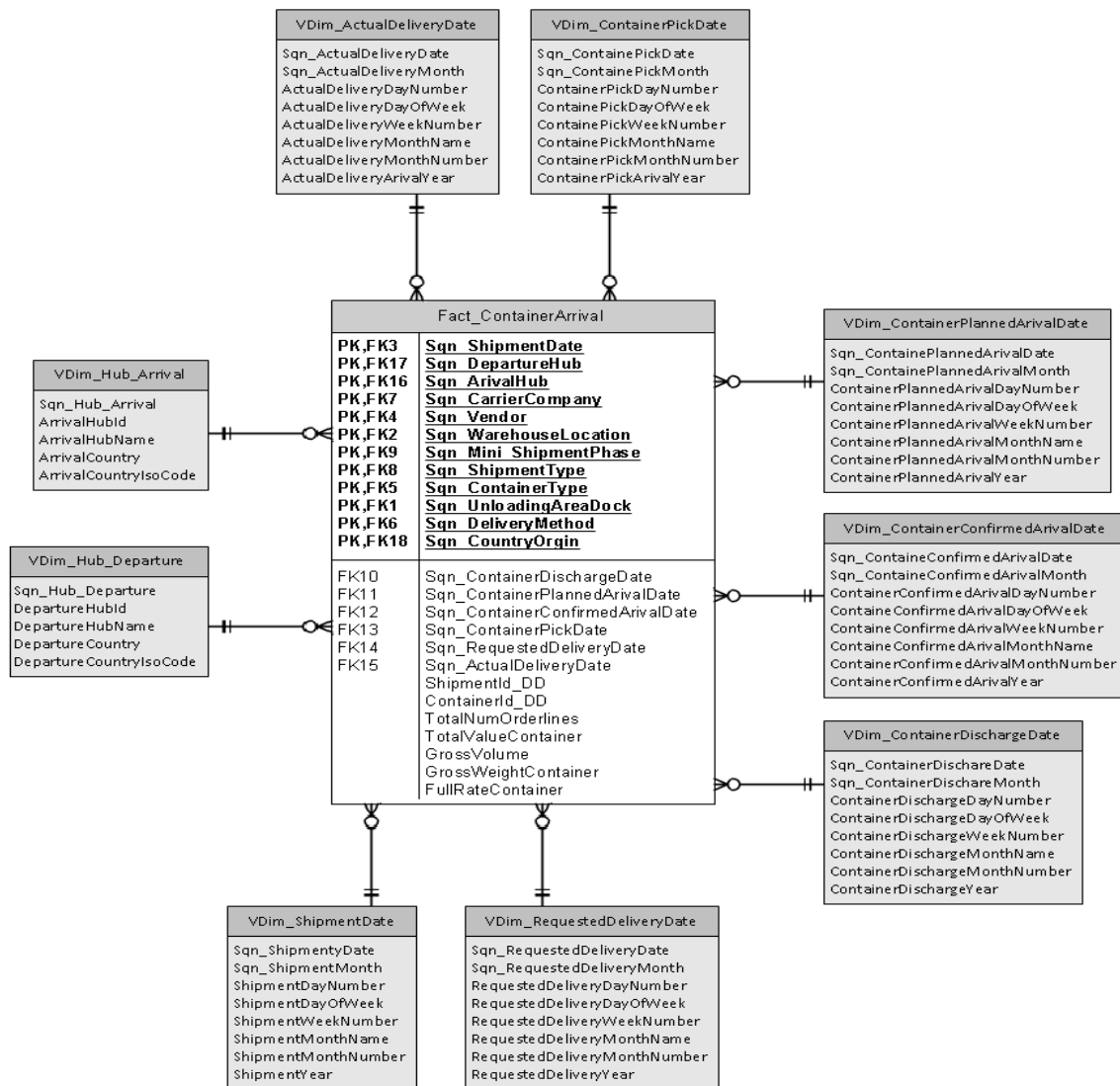
## Analyse Model Fact WeeklyShipments



## Analyse Model Aggregate Weekly delivery performance



## Roleplaying dimensie



## Mapping schema's

## Warehouse location

Source tabel	Source Attribuut	Bron type	Target Attribuut	Type	Bijzonderheid
Toegevoegd	Toegevoegd	/	Sqn_WarehouseLocation	Int	Primary / surrogate key SCD 0
Dim_WarehouseSite	Sqn_WarehouseSite	int	Sqn_WarehouseSite	Int	Foreign key SCD 2
Dim_WarehouseSite	SiteRecId	Bigint	WarehouseSiteRecId	Bigint	Alternative Foreign key SCD 2
InventSite	InventSiteId	nvarchar(30)	warehouseSiteId	nvarchar(30)	SCD 2
InventLocation	InventLocationId	Nvarchar(30)	WarehouseID	Nvarchar(30)	Alternative business key SCD 1
InventLocation	InventLocationDescription	Nvarchar(255)	WarehouseName	Nvarchar(60)	SCD 1
LogisticsLocation	LocationId	Nvarchar(30)	WarehouseLocationId	Nvarchar(30)	SCD 1
LogisticsLocation	Description	Nvarchar(255)	WarehouseLocationDescription	Nvarchar(60)	SCD 1
LogisticsLocationRole	Name	Nvarchar(30)	WarehousePurposeName	Nvarchar(30)	SCD 1
LogisticsLocationRole	Type	int	WarehousePurposeType	int	SCD 1
LogisticsPostalAddress	Address	Nvarchar(255)	WarehouseAddress	Nvarchar(150)	SCD 1
LogisticsAddressCountryRegion	CountryRegion	Nvarchar(255)	WarehouseCountryName	Nvarchar(50)	SCD 1
LogisticsAddressCountryRegion	CountryRegionId	nvarchar(30)	WarehouseCountryCode	nchar(3)	SCD 1
LogisticsAddressCity	Name	Nvarchar(60)	WarehouseCity	Nvarchar(60)	SCD 1
LogisticsAddressState	StateId	Nvarchar(10)	WarehouseState	Nvarchar(10)	SCD 1
LogisticsAddressDistrict	Name	Nvarchar(60)	WarehouseDistrict	Nvarchar(60)	SCD 1
LogisticsAddressZipcode	ZipCode	nvarchar(10)	WarehouseZipcode	nvarchar(10)	SCD 2
LogisticsPostalAddress	StreetName	nvarchar(60)	WarehouseStreet	nvarchar(60)	SCD 1
InventLocationLogisticsLocation	IsPrimary	int	WarehousesPrimary	Bit	SCD 1
InventLocationLogisticsLocation	IsPostalAddress	int	WarehousesPostalAddress	Bit	SCD 1
LogisticsPostalAddress	Longitude	Numeric(32,16)	WarehouseLongitude	Numeric(9,6)	SCD 1
LogisticsPostalAddress	Latitude	Numeric(32,16)	WarehouseLatitude	Numeric(9,6)	SCD 1
LogisticsLocation	ParentLocation	Bigint	WarehouseParentLocationId	Bigint	Unair relatie key SCD 1
Toegevoegd	Toegevoegd	/	WarehouseParentLocationName	nvarchar(30)	Unair Relatie attribuut SCD 1
InventLocation	RecId	Bigint	WarehouseRecId	bigint	Business key
InventLocation	DataAreaid	Nvarchar(4)	DataAreaid	Nvarchar(4)	SCD 0
InventLocation	Partition	Bigint	Partition	bigint	SCD 0
Toegevoegd	Toegevoegd	/	RecordStartDate	DateTime	SCD 1
Toegevoegd	Toegevoegd	/	RecordEndDate	DateTime	SCD 1
Toegevoegd	Toegevoegd	/	RecordHash	Binary	SCD 1

Tabel 3 mapping warehouselocation

## Product en Categorie

Source tabel	Source Attribuut	Bron type	Target Attribuut	Type	Bijzonderheid
Toegevoegd	Toegevoegd	Toegevoegd	Sqn_Product	Bigint	Primary / surrogate key SCD 0
Toegevoegd	Toegevoegd	Toegevoegd	Sqn_ProductCategory	Int	Foreign key SCD 2
EcoResProductIdentifier	ProductNumber	NVarchar(255)	ProductNumber	Varchar(50)	Alternative business key SCD2
EcoResProductTranslation	Name	NVarchar(255)	ProductNumber	Nvarchar(50)	SCD 1
EcoResProductTranslation	Description	Nvarchar(255)	ProductDescription	Nvarchar(150)	SCD 1
EcoResProduct	RecId	bigint	ProductRecId	Bigint	Alternative business key SCD 1
EcoResProduct	Type	NVarchar(255)	ProductType	nvarchar(50)	SCD 1
EcoResProduct	Partition	Bigint	ProductPartition	bigint	SCD 1
EcoResProduct	DataAreald	NVarchar(255)	CompanyId	varchar(4)	SCD 1
InventTable	ItemId	Int	ProductId	Int	Business key SCD 1
InventTable	Height	Numeric(32,16)	Height	Numeric(8,2)	SCD 1
InventTable	Width	Numeric(32,16)	Width	Numeric(8,2)	SCD 1
DataAreald	Name	NVarchar(255)	CompanyName	Nvarchar(50)	SCD 1
InventTableModule	UnitId	Nvarchar(10)	UnitId	Nvarchar(10)	SCD 1
InventTable	CreatedDateTime	timestamp	RecordStartDate	DateTime	SCD 0
Toegevoegd	Toegevoegd	Toegevoegd	RecordEndDate	DateTime	SCD 1
Toegevoegd	Toegevoegd	Toegevoegd	PreviousVersion	Bigint	Unair relatie type SCD 1
EcoResCategory	RecId	bigint	CategoryRecId	Bigint	Business key SCD 1
EcoResCategory	Name	Nvarchar(255)	ParentCategoryName	Nvarchar(50)	Unair relatie attribuut SCD 1
EcoResCategory	Name	Nvarchar(255)	CategoryName	nvarchar(50)	SCD 1
EcoResCategory	Level_	int	CategoryDepth	int	SCD 1
EcoResProductCategory	Description	NVarchar(255)	CategoryDescription	Nvarchar(150)	SCD 1
EcoResCategoryHierarchy	Name	nvarchar(128)	CategoryHierarchyName	Nvarchar(128)	SCD 1
Toegevoegd	Toegevoegd	Toegevoegd	BottomFlag	bit	SCD 1
Toegevoegd	Toegevoegd	Toegevoegd	TopFlag	bit	SCD 1

Tabel 4 Mapping product &amp; categorie

connect.  
engage.  
succeed.

## Bijlage F

### Implementatie Query's



## Query's

De afstudeerder heeft gedurende dit project een groot aantal SQL query's geschreven. In deze bijlage worden een aantal van deze query's eruit gelicht.

In het figuur hieronder ziet u het script voor het creëren van de datawarehouse database. Hierbij wordt vastgelegd waar de data en de log bestanden worden opgeslagen en hoe groot deze mogen worden. De log en data bestanden zijn in de test opstelling op dezelfde schijf geplaatst. Dit script wordt in een later stadium aangepast zodat deze op twee aparte schijven komen te staan. Als laatste wordt het recovery model van de database op 'simple' gezet.

```
If NOT EXISTS(select * from sys.databases where name = 'DWH_CTWINT_2012_R3')
Begin
    CREATE DATABASE DWH_CTWINT_2012_R3
    ON
        ( NAME = DWH_CTWINT_2012_R3_Data,
          FILENAME =
'E:\Datawarehouse_DB\MSAS11.DWH_CTWINT_2012_R3\Datawarehouse\Data\DWH_CTWINT_2012_R3_Data
.mdf',
          SIZE = 10094MB,
          FILEGROWTH = 5094MB
        )
    LOG ON
        (
          NAME = DWH_CTWINT_2012_R3_log,
          FILENAME =
'E:\Datawarehouse_DB\MSAS11.DWH_CTWINT_2012_R3\Datawarehouse\Log\DWH_CTWINT_2012_R3_Log.1
df',
          SIZE = 50MB,
          FILEGROWTH = 25MB
        );
END

ALTER DATABASE DWH_CTWINT_2012_R3 SET RECOVERY SIMPLE WITH NO_WAIT
```

Figuur 5 Script datawarehouse creatie



In het figuur hieronder ziet u een deel van de foreign constraints die in de datawarehouse zijn aangebracht. Bijgevoegd de clustered primary key van het feit tabel 'container arrival'

```
-Dim_WarehouseSite self reference key
ALTER TABLE dbo.Dim_WarehouseSite
ADD CONSTRAINT Dim_WarehouseSite_Dim_Address_FK
FOREIGN KEY (Sqn_Address) REFERENCES dbo.Dim_Address (Sqn_Address)
GO

ALTER TABLE dbo.Dim_WarehouseSite
ADD CONSTRAINT Dim_WarehouseSite_ParentSite_Unair
FOREIGN KEY (SiteParentLocationRecId) REFERENCES dbo.Dim_WarehouseSite (SiteRecId)
GO

--Dim_WarehouseLocation version self reference
ALTER TABLE dbo.Dim_WarehouseLocation
ADD CONSTRAINT Dim_WarehouseLocation_Dim_WarehouseSite_FK
FOREIGN KEY (Sqn_WarehouseSite) REFERENCES dbo.Dim_WarehouseSite (Sqn_WarehouseSite)
GO

--Fact_ContainerArrival WarehouseLocation foreign key
ALTER TABLE dbo.Fact_ContainerArrival
ADD CONSTRAINT Fact_ContainerArrival_Dim_WarehouseLocation_FK
FOREIGN KEY (Sqn_WarehouseLocation) REFERENCES dbo.Dim_WarehouseLocation
(Sqn_WarehouseLocation)

-- Primary key constrains
Constraint pk_Dim_WarehouseLocation_key primary key (Sqn_WarehouseLocation)

--Clustered Primary Constraint op feit tabel container arrival
ALTER TABLE [dbo].[Fact_ContainerArrival] ADD CONSTRAINT [pk_Fact_ContainerArrival_key]
PRIMARY KEY CLUSTERED (
    [Sqn_ShipmentArrivalDate], [Sqn_DepartureHub], [Sqn_ArivalHub],
    [Sqn_CarrierCompany], [Sqn_Vendor], [Sqn_WarehouseLocation],
    [Sqn_Mini_ShipmentPhase], [Sqn_ShipmentType], [Sqn_DeliveryMethod],
    [Sqn_ContainerType], [Sqn_UnloadingAreaDock], [Sqn_CountryOrigin]
)
```

Figuur 6 Script Foreign key constraints

In het figuur hieronder ziet u een Sql script van het creëren van een column store index op het feit tabel shipment.

```
ALTER TABLE ADD NONCLUSTERED COLUMNSTORE INDEX CSI_Fact_Shipment
ON dbo.Fact_Shipment
(
    Sqn_Date,
    Sqn_WarehouseLocation,
    ShipmentId,
    TotalNumOrderLines,
    TotalNumContainers,
    TotalValueShipment,
    GrossWeightShippment,
    GrossVolumeShipped
)
on PartitionScheme_FactShip;
```

Figuur 7 Script Columnstore Index

De volgende query is geschreven om de hiërarchie van shipment fasen te flateren (plat te slaan). Hierbij wordt gebruik gemaakt van een recursieve common table expression (CTE) die door de gehele hiërarchie navigeert. De CTE wordt daaronder aangeroepen via een select query. In de select query wordt meegegeven of de top van de hiërarchie is bereikt.

```

; WITH ProductCategory_cte AS (
SELECT p.[RecId]
      ,p.[ParentCategory],
      p.[CategoryHierarchy]
      ,p.[Name]
      ,p.[IsActive]
      ,p.[Level_]
      ,p.Partition
FROM [STAGING_AREA].[dbo].[ECORESCATEGORY] c
join [STAGING_AREA].[dbo].[ECORESCATEGORY] p on p.RecId = c.ParentCategory

UNION ALL
--Recursief aanroep cte stop conditie: geen child
SELECT c.[RecId]
      ,c.[ParentCategory]
      ,c.[CategoryHierarchy]
      ,c.[Name]
      ,c.[IsActive]
      ,c.[Level_]
      ,c.Partition
FROM ProductCategory_cte p
join [STAGING_AREA].[dbo].[ECORESCATEGORY] c on c.ParentCategory = p.RecId
)
--Select van common table expression de product categorieën
SELECT p.RecId as CategoryRecId,
      p.ParentCategory,
      p.CategoryHierarchy as CategoryHierarchyRecId,
      p.Partition,
      h.Name as CategoryHierarchyName,
      coalesce(c.Name, 'N/A') as ParentName,
      p.Name as ChildName,
      coalesce(c.Level_, 0) as Parent_Depht,
      p.Level_ as Child_Depth
      , 'Top Flag' = CASE WHEN c.level_ > 1 THEN 'N' ELSE 'Y' END
FROM ProductCategory_cte p
left join [STAGING_AREA].[dbo].[ECORESCATEGORY] c on p.ParentCategory = c.RecId
left join [STAGING_AREA].[dbo].[ECORESCATEGORYHIERARCHY] h on p.CategoryHierarchy =
h.RecId
where p.ParentCategory <> 0
GROUP BY p.CategoryHierarchy, h.Name, p.Level_, p.RecId,
p.ParentCategory, p.Name, c.name, c.Level_, p.Partition

```

Figuur 8 Common table expression

Door de afstudeerder geschreven script voor het genereren van de datum dimensie hiërarchie. In de parameter van de stored procedure kan worden meegegeven wat de eerste dag van de week is. Ook wanneer een fiscaal jaar voor een bepaald bedrijf begint. Deze worden automatisch overgenomen vanuit Dynamics. Dit maakt deze query generiek voor meerdere gebruikers.

```
CREATE PROCEDURE GenerateDateDimSnowflake @FirstDayOfWeek int, @FiscalYearStart
varchar(20), @StartDate varchar(20), @EndDate varchar(20)
AS
--sets the first day of week to Dynamics settings
SET DATEFIRST @FirstDayOfWeek
SET NOCOUNT ON

-- Declare and set variables for loop
Declare
@StartDate datetime,
@EndDate datetime,

-- Declare iterator variables for the sqn keys
@Sqn_Date int,
@Sqn_Month int,
@Sqn_Week int,
@Sqn_Year int,
@Week int,

--Set variables for the loop, the Parameters startDate and endDate are the generated data
range
@Date datetime
Set @StartDate = @StartDate
Set @EndDate = @EndDate
Set @Date = @StartDate

--Set Iterator variables for the sqn_keys in the loop
Set @Sqn_Date = 1
set @Sqn_Week = 1
set @Sqn_Month = 1
set @sqn_Year = 1

-- Loop through dates
WHILE @Date <= @EndDate
BEGIN
    IF (DATEPART(wk, @Date) <> 53)
    BEGIN
        set @Week = DATEPART(wk, @Date)
    END
    ELSE
    BEGIN
        set @Week = 1
    END

    -- Insert record in dimension table
    If (@date = DATEADD(MONTH, DATEDIFF(MONTH, 0, @Date), 0))
    BEGIN
        Insert into Dim_Month (
            [Sqn_Month], [Sqn_Year], [MonthName], [MonthNumber],
            [FiscalMonthNumber], [Quarter], [Year]
        )
        values (
            @Sqn_Month, @Sqn_Year, DATENAME(mm, @Date),
            DATEPART(mm, @Date),
            DATENAME(qq, @Date), DATEPART(MM, @FiscalYearStart), Year(@Date)
        )
    END
END
```

Continue →

```

If(@Date = DATEADD(yy, DATEDIFF(yy,0,@Date), 0))
BEGIN
    Insert into Dim_Year([Sqn_Year], [Year])
    values (@Sqn_Year,Year(@Date) )

END

If(@date =DATEADD(ww, DATEDIFF(ww,0,@Date), 0))
BEGIN
    Insert into Dim_Week ([Sqn_Week],[Sqn_Year],[WeekNumber],[Year])
    values (@Sqn_Week,@Sqn_Year, @Week, Year(@Date) )
END

INSERT Into Dim_Date (
    Sqn_Date, [Date], [DayNumber], [DayofYear], [DayofWeek],
    [Dayofweekname], [WeekNumber], [MonthNumber], [MonthName],
    [Quarter], [Year], FiscalYear, Sqn_Week, Sqn_Month
)
Values (
    @Sqn_Date, @Date, Day(@Date), DATEPART(dy, @Date),
    DATEPART(dw, @Date), DATENAME(dw, @Date),
    @Week, DATEPART(wk, @Date), DATEPART(mm, @Date),
    DATENAME(mm, @Date), DATENAME(qq, @Date),
    Year(@Date),DAY(@FiscalYearStart),@Sqn_Week, @Sqn_Month
)

--Checks when loop iterator needs to increase
IF(@Date = DATEADD(MONTH, DATEDIFF(MONTH, 0, @Date + 1), -1))
BEGIN
    Set @Sqn_Month = @Sqn_Month + 1
END

If(@Date = DATEADD(wk, DATEDIFF(wk,0,@Date), 0))
BEGIN
    Set @Sqn_Week = @Sqn_Week + 1
END

If(@Date = DATEADD(year, DATEDIFF(year, -1, @Date), -1))
BEGIN
    Set @Sqn_Year = @Sqn_Year + 1
END

--Next day
Set @Date = @Date + 1
Set @Sqn_Date = @Sqn_Date + 1
END

Insert into Dim_Year( [Sqn_Year], [Year])
values (@Sqn_Year, Year(@Date) )
Insert into Dim_Week ( [Sqn_Week],[Sqn_Year],[WeekNumber],[Year])
values (@Sqn_Week, @Sqn_Year, @Week, Year(@Date) )
GO

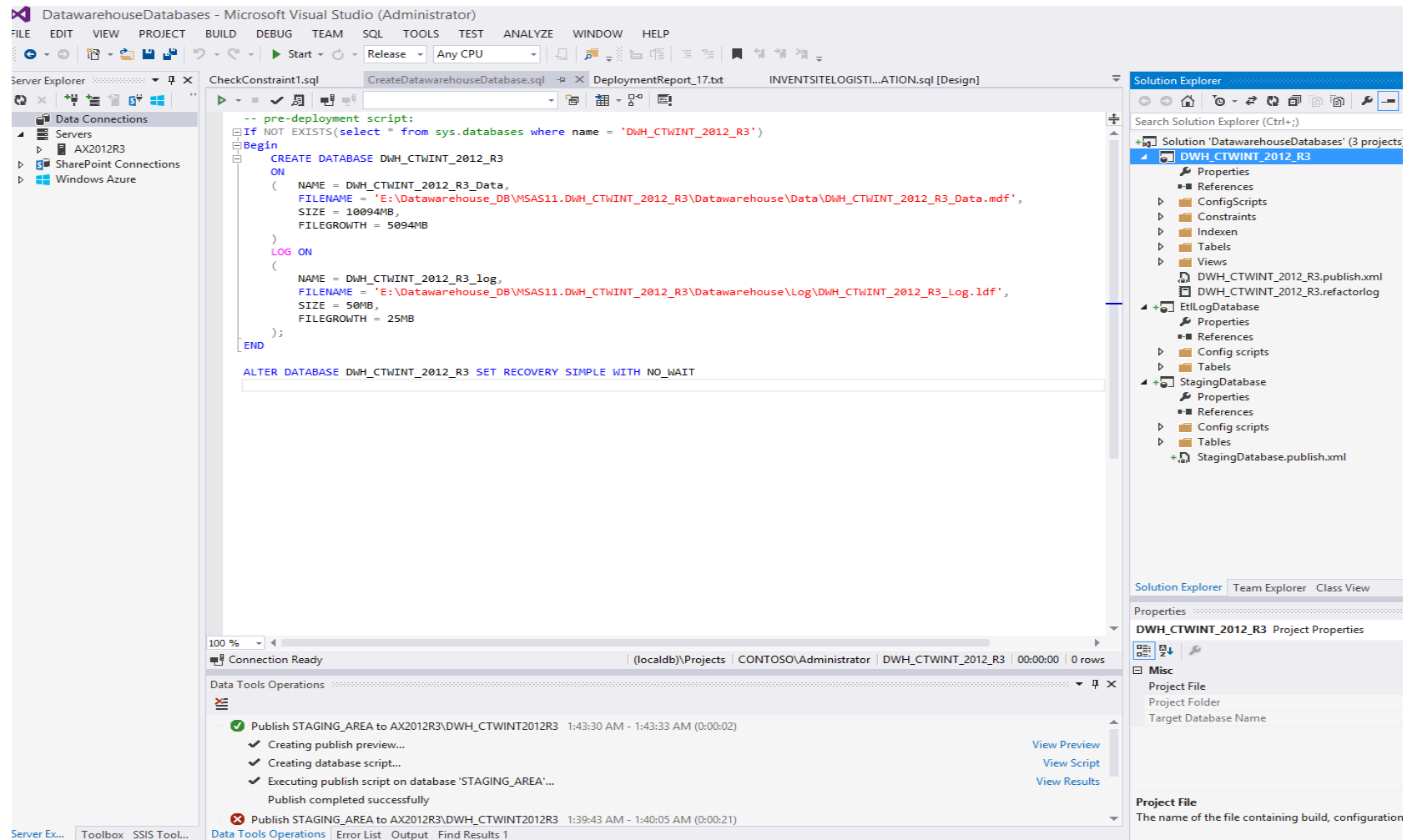
SET NOCOUNT OFF;

```

*Figuur 9 Script generatie date dimensie snowflake*

## Screenshot database ontwikkeling omgeving

In Figuur 10 ziet u de SQL Server Data Tools (SSDT) ontwikkeling omgeving. Hierin kunt u zien hoe de afstudeerder de verschillende Sql scripts binnen SSDT heeft georganiseerd. De afstudeerder wenst hier mee aan te tonen dat het bouwen van de datawarehouse op een gestructureerde wijze is verlopen. Tevens dient dit screenshot als bewijs.



Figuur 10 Sql server datatools IDE

connect.  
engage.  
succeed.

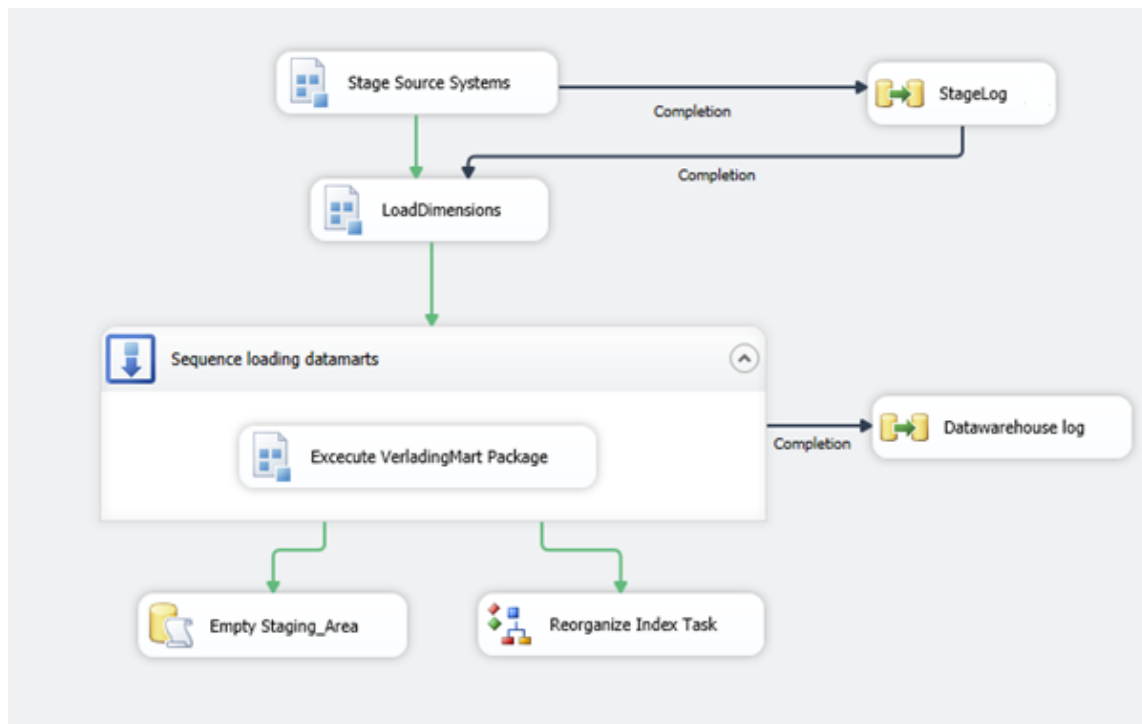
## Bijlage G

### SSIS ETL Workflows

## ETL Workflows

De afstudeerder heeft een aantal SSIS dataflows in deze bijlage toegevoegd. Deze geven u een indruk van de wijze waarop de afstudeerder het ETL proces heeft ontwikkeld. Er wordt van elke package niveau één afbeelding getoond. De overige flows worden hier niet getoond.

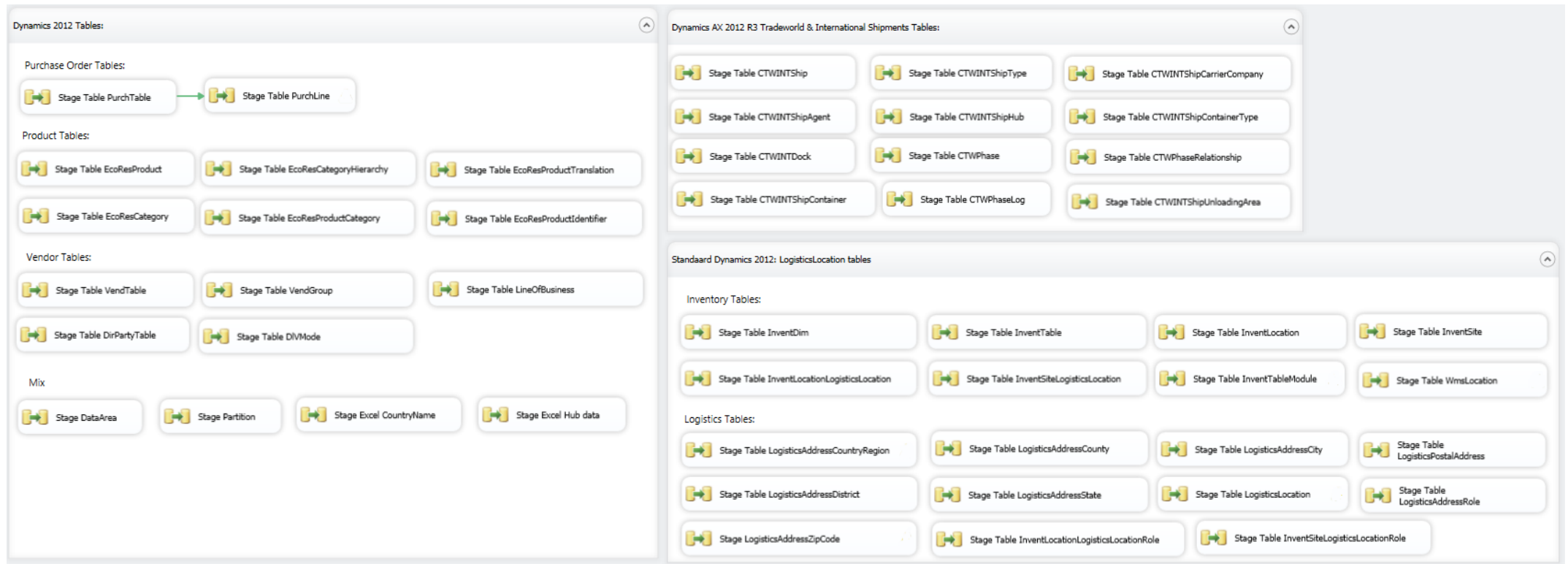
In Figuur 11 ziet u de 'Master package'. Deze package roept de andere SSIS dataflow packages aan. Het loggen van het ETL proces gebeurt op dit niveau.



Figuur 11 SSIS Master package flow

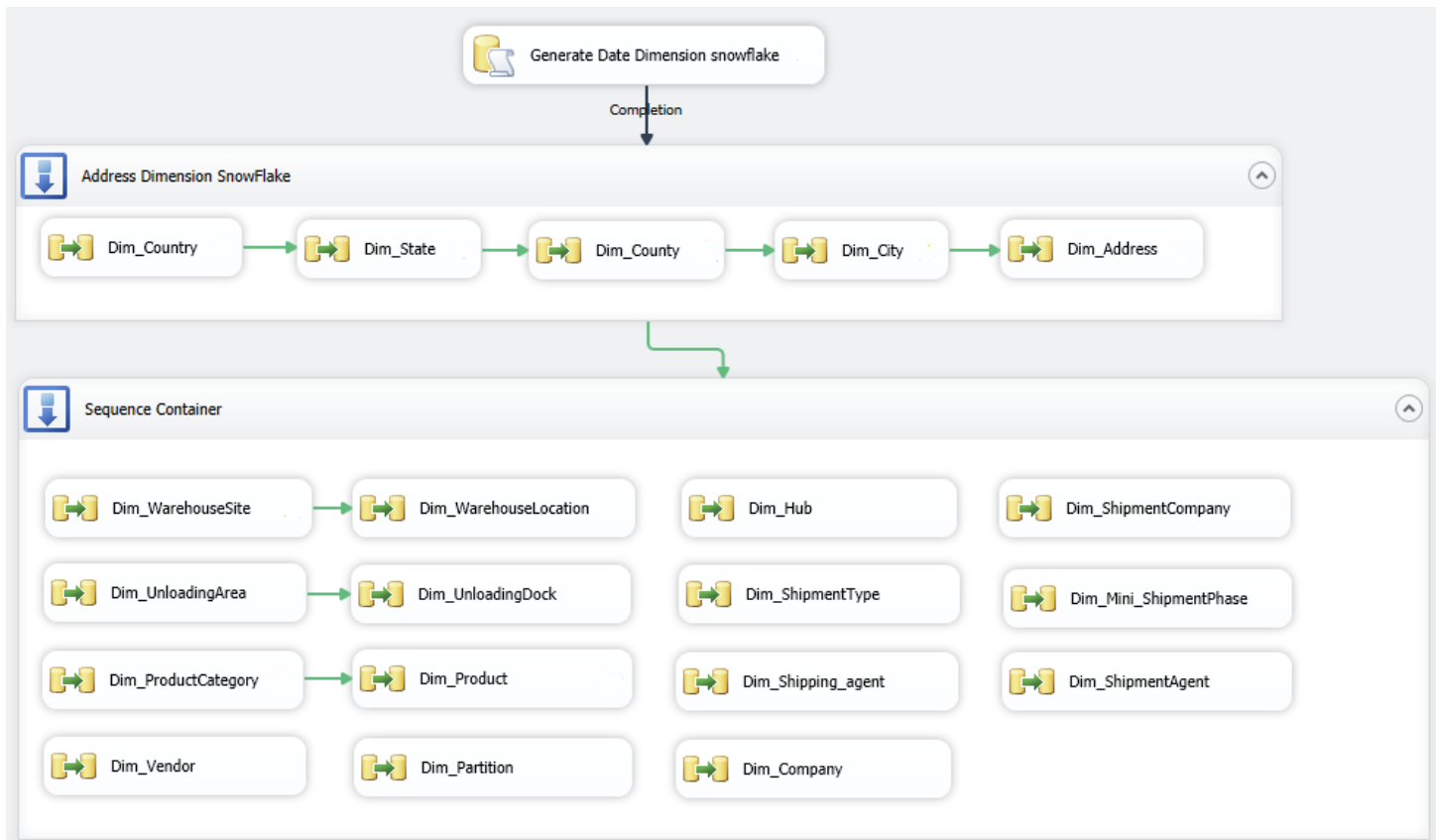


In Figuur 12 ziet u de eerste SSIS package dat wordt aangeroepen vanuit de master dataflow (Figuur 11). In deze package worden alle dataflows aangeroepen die verantwoordelijk zijn voor het 'extracten' van alle geïnventariseerde gegevens uit de brondatabase naar de staging database.



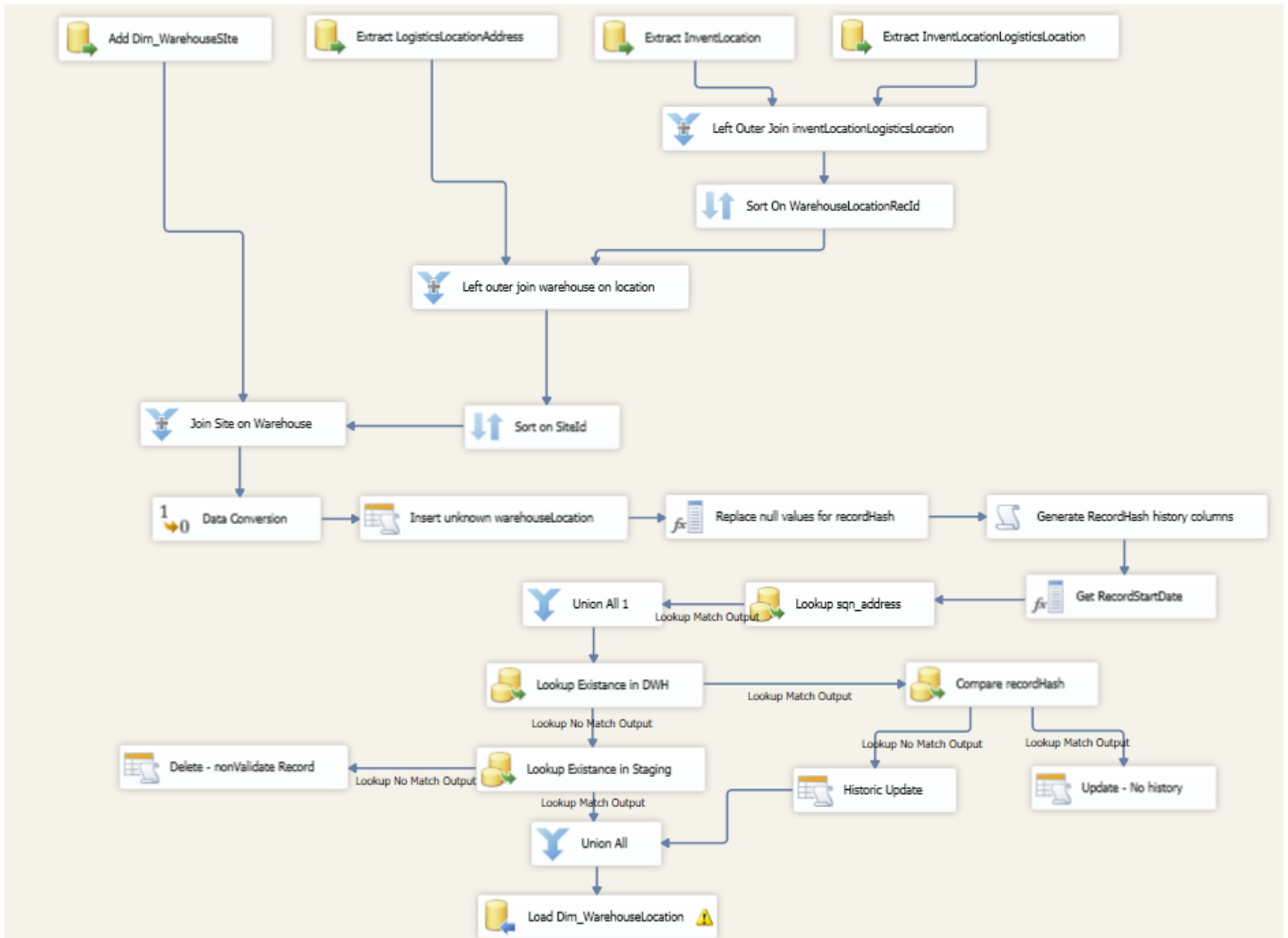
Figuur 12 SSIS Staging package

In Figuur 13 ziet u de ssis dimensie package. In deze package worden alle ETL dataflows aangeroepen om de data uit de staging database te transformeren naar het dimensioneel model. In de staging package (Figuur 12) dataflows worden de data juist onttrokken uit het bronsysteem. Deze package wordt in de master flow (Figuur 11) na de staging package aangeroepen.



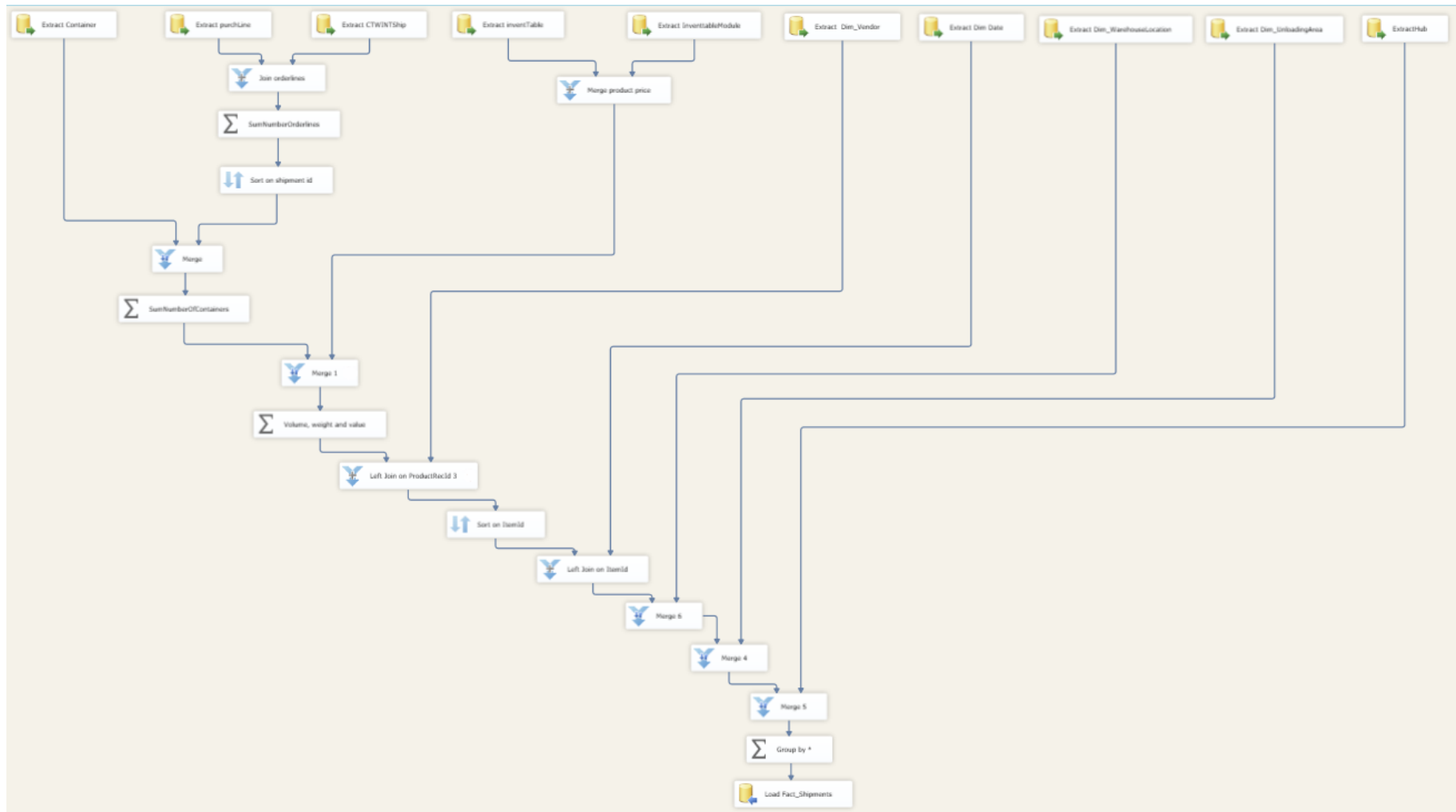
Figuur 13 SSIS dimension package

In Figuur 14 ziet u een ETL dataflow van de dimensie warehouseLocation. Deze dataflow packages vormen het laagste niveau binnen de ETL workflow. Hier wordt de data daadwerkelijk geconverteerd en getransformeerd.



*Figuur 14 SSIS flow laden warehouselocation dimensie*

In Figuur 15 vindt u de ETL flow van de feittabel shipments. Door de omvang van het figuur is deze verkleind. U kunt de ETL flow alleen op hoofdlijnen zien.



Figuur 15 ETL Flow Fact\_Shipment

connect.  
engage.  
succeed.

**Bijlage H**

**Test rapport**

## Controle semantiek database relaties

De afstudeerder heeft gedurende de gegevensbron analyse een aantal test query's geschreven om de referentiele integriteit te testen. Deze controles noemt de afstudeerder in deze context semantische controles.

In Script 1 ziet u de SQL query voor de controle of alle shipments in de database een vertrek –en aankomst hub hebben. Het attribuut 'hubtype' bepaald het type hub.

Hubtype 1 = loadinghub  
Hubtype 0 = dischargehub.

Voor meer informatie over de tabel 'hubs' verwijst ik u naar de data dictionaires.

```
-- Sub query controle combinatie vertrek -en aankomsthubs
SELECT *
FROM dbo.CTWINTSHIP
WHERE DEPARTUREHUBID in
(
    SELECT RECID
    FROM dbo.CTWINTSHIPHUB
    WHERE HUBTYPE = '0'
)
AND ARRIVALHUBID IN
(
    SELECT RECID
    FROM dbo.CTWINTSHIPHUB
    WHERE HUBTYPE = '1'
)

-- Sub query controle ontbrekende combinatie vertrek -en aankomsthubs
SELECT *
FROM dbo.CTWINTSHIP
WHERE DEPARTUREHUBID not in
(
    SELECT RECID
    FROM dbo.CTWINTSHIPHUB
    WHERE HUBTYPE = '0'
)
AND ARRIVALHUBID not IN
(
    SELECT RECID
    FROM dbo.CTWINTSHIPHUB
    WHERE HUBTYPE = '1'
)
```

*Script 1 Controle semantiek hubs*

## Test cases geschiedenis

Hier onder vindt u een aantal testcases die betrekking op het testen van gegevensgeschiedenis.

Testcase 1	
<b>Actie:</b>	Invoeren van een nieuwe record in de database. Controle of het record is ingevoerd in de database.
<b>Verwacht resultaat:</b>	In de database is een nieuwe record toegevoegd. Het record attribuut; startdate heeft de invoeringsdatum van record als waarde. De attributen 'endDate' en 'previous_version' zijn leeg. De overige attributen bevatten de waarden zoals is ingevoerd
<b>Geslaagd (Ja / Nee)</b>	Ja

Testcase 2	
<b>Actie:</b>	Updaten van een bestaand record waarvan één of meerdere attributen van het type SCD 1 zijn. Vervolgens controleren of het betreffende record is geüpdatet in de database
<b>Verwacht resultaat:</b>	Het attribuut; startdate heeft de invoeringsdatum van record als waarde. De attributen 'endDate' en 'previous_version' zijn leeg. De attributen van het type SCD 1 zijn geüpdatet.
<b>Geslaagd (Ja / Nee)</b>	Ja

Testcase 3	
<b>Actie:</b>	Updaten van een bestaand record waarvan één of meerdere attributen van het type SCD 2 zijn. Vervolgens controleren of het betreffende record is geüpdatet in de database
<b>Verwacht resultaat:</b>	<p>In de database is een nieuwe record toegevoegd. Het attribuut 'startDate' van het nieuwe record heeft als waarde de wijzigingsdatum van het record. Het attribuut 'endDate' is leeg. Het attribuut 'previous_version' verwijst naar de primary key van de vorige record versie. De overige attributen van het SCD type 1 of 2 bevatten de gewijzigde waarden.</p> <p>De waarden van de attributen van de vorige record versie zijn onveranderd. Het attribuut 'endDate' heeft als waarde de datum waarop het record is geüpdatet. Het attribuut 'previous_version' blijft leeg.</p>
<b>Geslaagd (Ja / Nee)</b>	Ja

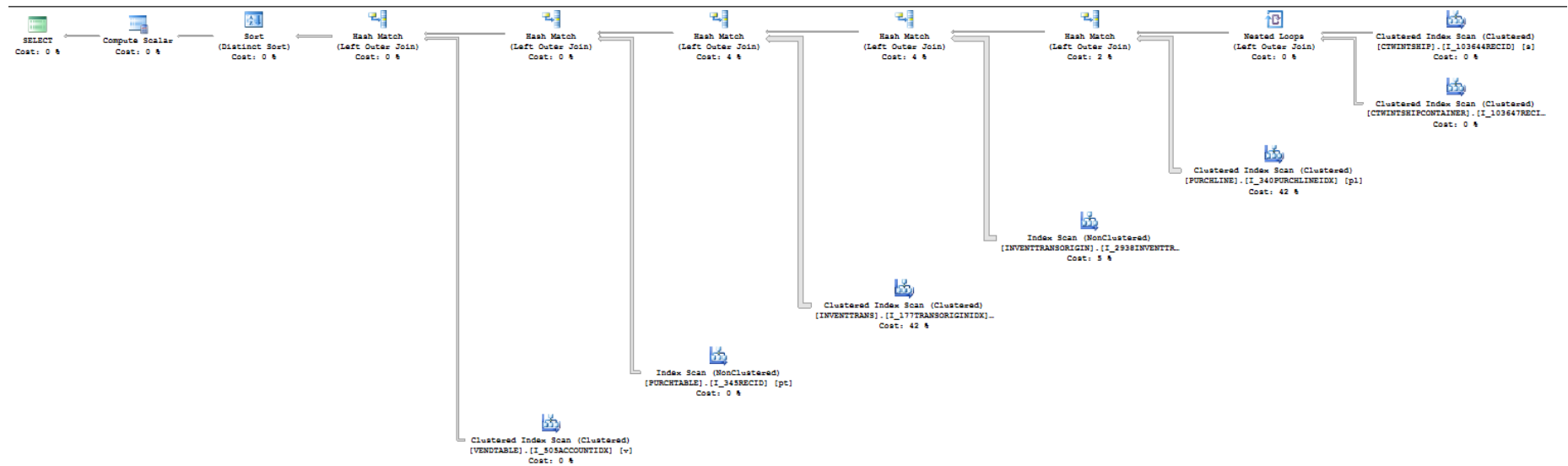
Testcase 4	
<b>Actie:</b>	Verwijderen van een bestaand record uit de database. Vervolgens controleren of het betreffende record is gewijzigd of verwijderd.
<b>Verwacht resultaat:</b>	Het betreffende record is niet verwijderd uit de database. Het attribuut 'endDate' heeft als waarde de datum waarop het record is gedelete. De overige attributen behouden hun oorspronkelijke waarde.
<b>Geslaagd (Ja / Nee)</b>	Ja



Testcase 5	
<b>Actie:</b>	Wijzigen van de business of de surrogaat sleutel binnen een dimensie. Deze sleutels zijn van type SCD type 0.
<b>Verwacht resultaat:</b>	Het betreffende record wordt niet gewijzigd. Het wijzigen van de business key resulteert in een nieuw record in de database. Het vorige record blijft nog steeds valide. Het attribuut 'endDate' blijft bij beide records leeg.
<b>Geslaagd (Ja / Nee)</b>	Ja

## Performance rapport

In Figuur 16 ziet u het grafische queryplan van de query op de staging database. Bijgevoegd vindt u de statistieken van Sql Server.



Figuur 16 Queryplan op staging database

```
(52 row(s) affected)
Table 'CTWINTSHIPCONTAINER'. Scan count 1, logical reads 2, physical reads 0, read-ahead reads 0, lob logical reads 0, lob
physical reads 0, lob read-ahead reads 0.
Table 'Worktable'. Scan count 0, logical reads 0, physical reads 0, read-ahead reads 0, lob logical reads 0, lob physical reads 0,
lob read-ahead reads 0.
Table 'Worktable'. Scan count 0, logical reads 0, physical reads 0, read-ahead reads 0, lob logical reads 0, lob physical reads 0,
lob read-ahead reads 0.
Table 'PURCHLINE'. Scan count 5, logical reads 32316, physical reads 0, read-ahead reads 0, lob logical reads 0, lob physical
reads 0, lob read-ahead reads 0.
Table 'VENDTABLE'. Scan count 1, logical reads 880, physical reads 0, read-ahead reads 0, lob logical reads 0, lob physical reads
0, lob read-ahead reads 0.
Table 'CTWINTSHIP'. Scan count 0, logical reads 88, physical reads 0, read-ahead reads 0, lob logical reads 0, lob physical reads
0, lob read-ahead reads 0.
Table 'PURCHTABLE'. Scan count 1, logical reads 21, physical reads 0, read-ahead reads 0, lob logical reads 0, lob physical reads
0, lob read-ahead reads 0.
Table 'INVENTTRANSORIGINPURCHLINE'. Scan count 5, logical reads 1186, physical reads 0, read-ahead reads 0, lob logical reads 0,
lob physical reads 0, lob read-ahead reads 0.
Table 'INVENTTRANSORIGIN'. Scan count 1, logical reads 2572, physical reads 0, read-ahead reads 0, lob logical reads 0, lob
physical reads 0, lob read-ahead reads 0.
Table 'INVENTTRANS'. Scan count 5, logical reads 32336, physical reads 0, read-ahead reads 0, lob logical reads 0, lob physical
reads 0, lob read-ahead reads 0.
```

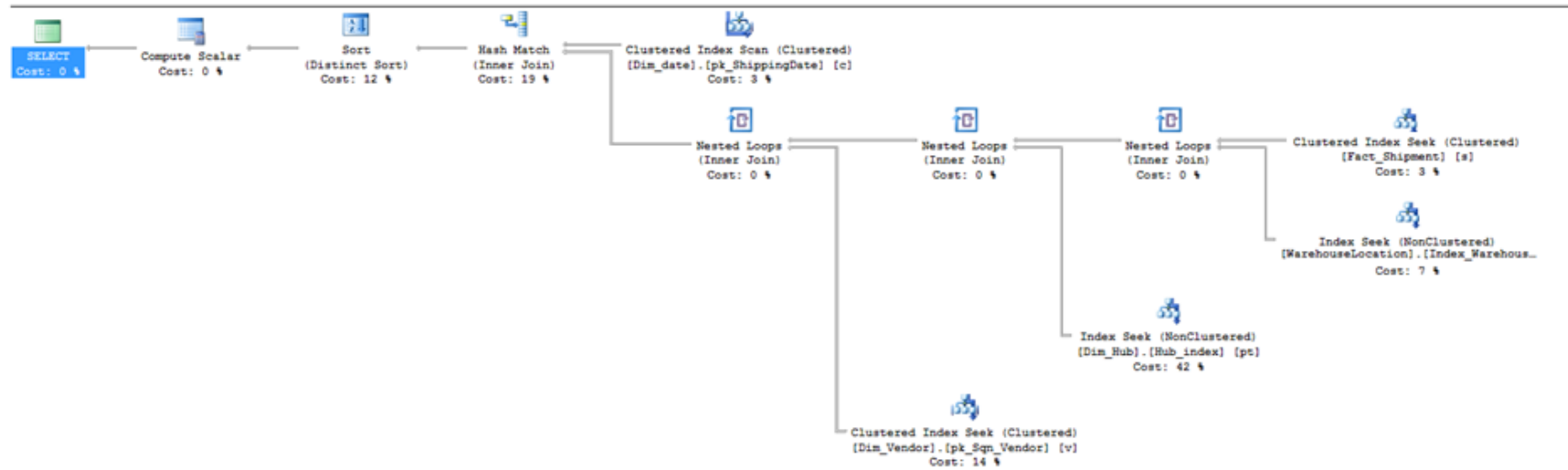
```
(1 row(s) affected)
```

```
SQL Server Execution Times:
```

```
  CPU time = 626 ms,  elapsed time = 554 ms.
```

*Figuur 17 log van queryplan op staging database*

In Figuur 18 ziet u het grafische queryplan van de query op de datawarehouse.



*Figuur 18 Queryplan op datawarehouse*

```
(28 row(s) affected)
Table 'Worktable'. Scan count 0, logical reads 0, physical reads 0, read-ahead reads 0, lob logical reads 0, lob physical reads 0,
lob read-ahead reads 0.
Table 'Dim_Vendor'. Scan count 32, logical reads 64, physical reads 0, read-ahead reads 0, lob logical reads 0, lob physical reads
0, lob read-ahead reads 0.
Table 'Dim_Hub'. Scan count 32, logical reads 86, physical reads 0, read-ahead reads 0, lob logical reads 0, lob physical reads 0,
lob read-ahead reads 0.
Table 'WarehouseLocation'. Scan count 25, logical reads 76, physical reads 0, read-ahead reads 0, lob logical reads 0, lob
physical reads 0, lob read-ahead reads 0.
Table 'Fact_Shipment'. Scan count 25, logical reads 1, physical reads 0, read-ahead reads 0, lob logical reads 0, lob physical
reads 0, lob read-ahead reads 0.
Table 'Dim_date'. Scan count 1, logical reads 2, physical reads 0, read-ahead reads 0, lob logical reads 0, lob physical reads 0,
lob read-ahead reads 0.
```

```
(1 row(s) affected)
```

```
SQL Server Execution Times:
```

```
  CPU time = 0 ms,  elapsed time = 131 ms.
```

*Figuur 19 Log queryplan op datawarehouse*