Productverslag Kramse

Toelichting over het bouwen en gebruiken van het nieuwe datawarehouse/reporting systeem.

Ruben Hensen

S2084700

Contents

[Inleiding 3](#_Toc44020066)

[Data laden 4](#_Toc44020067)

[Data transformeren 5](#_Toc44020068)

[DQS 5](#_Toc44020069)

[Extractie datum 5](#_Toc44020070)

[Valuta 5](#_Toc44020071)

[Km naar Zeemijl 5](#_Toc44020072)

[Lege velden 5](#_Toc44020073)

[Notaties en datatypes 5](#_Toc44020074)

[Data verminderen 5](#_Toc44020075)

[Datum tabel 5](#_Toc44020076)

[Berekening Portdays 6](#_Toc44020077)

[Slowly Changing Dimensions 6](#_Toc44020078)

[Van relationeel naar galaxyschema 6](#_Toc44020079)

[Analyses 8](#_Toc44020080)

[Overzicht dashboard 8](#_Toc44020081)

[Detail dashboard 8](#_Toc44020082)

[Berekeningen 8](#_Toc44020083)

[Brandstofverbruik per ton per zeemijl 8](#_Toc44020084)

[Idle time 9](#_Toc44020085)

[Beladingsgraad 9](#_Toc44020086)

[Kosten per ton vracht 9](#_Toc44020087)

[Brandstofverbruik (in relatie tot optimale snelheid) in tonnen brandstof per ton vracht per zeemijl; een en ander per type boot, vaarroute en seizoen. 10](#_Toc44020088)

[Conclusie en samenvatting 10](#_Toc44020089)

[Bibliografie 11](#_Toc44020090)

# Inleiding

Het containertransport bedrijf Kramse vervoeren materiaal over de hele wereld. Ze gebruiken verschillende bestanden en databronnen om data bij te houden. Hieronder vallen containertypes, klanten, schepen en de verzendingen zelf. Kramse wil deze data analyseren, maar om dat te kunnen doen zal er een ander architectuur moeten komen voor het opslaan en verwerken van deze data. Daarom is er gekozen om een datawarehouse te bouwen. Een grote collectie van alle databronnen waar analyses op gedaan kunnen worden.

Alle keuzes in dit proces zullen worden toegelicht in dit proces. Het doel is om de vraag te beantwoorden: ‘Hoe wordt het datawarehouse opgebouwd en hoe werken de analyses?’. Om deze vraag te beantwoorden is dit verslag in meerdere kleine delen opgebouwd, namelijk:

* Hoe worden de gegevens uit de operationele systemen geladen?
* Hoe worden de gegevens getransformeerd?
* Hoe ziet de tabelstructuur er van het datawarehouse uit?
* Hoe werken de analyses in het datawarehouse?

Het rapport zelf is opgedeeld in 5 hoofdstukken. Hoofdstuk 1 bevat de inleiding, 2 tot en met 4 de deelvragen en in hoofdstuk 5 is een kleine samenvatting en discussie te vinden.

# Data laden

Een datawarehouse is een collectie van alle bruikbare data in het bedrijf waar analyses op gedaan kunnen worden. De start van een data warehouse is dan ook om alle data eerst te verzamelen op een centrale plek. Er zijn 3 verschillende bronnen van bruikbare data gevonden die verwerkt zijn in dit datawarehouse:

* Consignors.xsl
* Containers v2.xsl
* KramseTPS (MS Access database)

Er was nog een vierde databron, 2016-v3-02072019-EU MRV Publication of information, maar deze informatie was te incompleet om te gebruiken.

Om deze informatie in het datawarehouse is SSIS (SQL Server Integration Services) van Microsoft gebruikt. Elke keer als SSIS wordt uitgevoerd wordt alle data uit de drie bronnen gelezen en naar een aparte database gekopieerd zodat het weinig impact heeft op de prestaties van de operationele databases.

De systemen zijn opgezet met de ‘SSIS Import and Export wizard’. De wizard zorgt ervoor dat de datawarehouse tabellen worden verwijderd. Daarna worden er opnieuw nieuwe tabellen aangemaakt die nu helemaal leeg zijn. Vervolgens wordt alle data over gekopieerd in deze tabellen.



Figuur 1 De tabellen vol data na het laden

# Data transformeren

De data die uit de databronnen komt is verre van perfect. In de huidige staat kunnen er geen analyses mee gedaan worden. Net als de eerste stap waar de data geladen wordt, zijn er nog meer stappen waarin de data langzaam getransformeerd wordt. Hieronder staan de transformaties uitgelegd.

## DQS

DQS staat voor Data Quality Service. Dit systeem zorgt ervoor dat de kwaliteit van de data verbeterd wordt. Zo verbetert het bijvoorbeeld spellingsfouten, of transformeert alle landen in hun standaard ISO-vorm (Nederland -> NL). Deze transformatie stap gebeurt gelijk in het begin, zodat alle andere stappen in het proces betrouwbare data hebben. Het grootste voordeel van DQS tegenover andere systemen is dat DQS steeds betrouwbaarder wordt. Het systeem leert van alle verbeteringen en kan het vervolgens zelf automatisch doen.

## Extractie datum

Aan elk stuk data voegen we een extractie datum toe. Hiermee kunnen we later ervoor zorgen dat alleen nieuwe data uit de operationele database wordt gelezen en kunnen we zien hoe oud de data is sinds de laatste vernieuwing.

## Valuta

Alle verschillende valuta’s worden omgerekend naar Euro’s. Dit zorgt ervoor dat alle berekeningen makkelijk uitgevoerd kunnen worden.

## Km naar Zeemijl

Alle velden waar kilometers worden gebruikt zijn veranderd naar zeemijl.

## Lege velden

Velden waar geen informatie in staan worden verwijderd of naar een standaardwaarde wordt gezet. Bijna altijd betekent dit dat een veld op “onbekend” of 0 wordt gezet omdat we de informatie niet weten. Het beste zou zijn om deze data te verbeteren bij de bronnen. Om het tijdelijk op te lossen wordt het nu zo opgelost.

## Notaties en datatypes

Er zijn een paar kleine notatie verschillen. Zo wordt een veld veranderd van komma’s naar punten zodat de servers die we gebruiken de waardes goed interpreteren. Ook worden sommige datatypes veranderd. Zo worden stukken tekst met nummers veranderd naar daadwerkelijke nummers.

## Data verminderen

Er stond een hoop dubbele data in de bronnen. Deze zijn verwijderd zodat we geen tegenstrijdige of verwarrende informatie krijgen te zien.

## Datum tabel

Er is een datum tabel aangemaakt zodat allerlei verschillende datums met elkaar vergeleken kunnen worden. Zo kan je bijvoorbeeld alle data van het eerste kwartaal aanvragen. De database weet dan welke reizen er allemaal in die periode waren.

## Berekening Portdays

Er is een aparte view aangemaakt waar de ‘port days’ worden berekend. Hierin staat het aantal dagen dat de een boot stil ligt op een haven.

## Slowly Changing Dimensions

Bijna alle informatie is gestopt in ‘slowly changing dimensions’. Dat betekent dat alle informatie onthouden wordt, de database heeft een geschiedenis.

## Van relationeel naar galaxyschema

De manier waarop de data uit de bronnen komt is niet goed om analyses op uit te voeren. Om deze reden hebben we er een sterschema van gemaakt. Het komt erop te neer dat Shipment\_Detail, Shipment, Voyage en Voyage port in 1 grote tabel zijn gekomen. Daarbij is Portdays ook een tabel geworden. De andere tabellen zijn zogeheten dimensies geworden en zijn verbonden met deze grote tabel en Port days.



Figuur 2 Het galaxyschema

# Analyses

Kramse heeft een aantal doelen die ze willen behalen. Om die doelen te behalen hebben ze een informatiebehoefte. Deze informatie proberen we overzichtelijk te maken in de dashboards die we hebben. Aan de hand van de missie, strategie en doelen is er gekozen om 2 dashboards te maken waar al deze informatie te vinden is. Het eerste dashboard voor het overzicht. Het tweede dashboard voor meer gedetailleerde informatie. Het plan voor de dashboards is als volgt.

## Overzicht dashboard

3 KPI (key performance indicators) met de volgende doelen:

* CO2 reductie/NOx/SOx: verlagen van het brandstofverbruik met 10% (brandstofverbruik/per ton vracht per zeemijl)
* Utilisatiegraad verbeteren: Reduceren van de wachttijden in de havens, de zogenaamde idle time, met 5%.
* Beladingsgraad; Verder optimaliseren van de beladingsgraad naar 90%

## Detail dashboard

Het detail dashboard heeft de volgende informatie:

* Beladingsgraad per schip per haven.
* Wachttijden (‘idle time’) per haven en per schip inclusief trends (afgelopen jaren).
* Brandstofverbruik (in relatie tot optimale snelheid) in tonnen brandstof per ton vracht per zeemijl; een en ander per type boot, vaarroute en seizoen.
* Kosten per ton vracht per schip per route.

## Berekeningen

Alle berekeningen in het dashboard worden hieronder beschreven. Soms zijn bepaalde aannames gedaan, die staan hier ook vermeld.

### Brandstofverbruik per ton per zeemijl

Het aantal ton wordt nergens in onze data vermeld. Op boatinggeeks.com staat wel vermeld dat een standaard container tussen de 25 en 30 ton kan wegen (Stafford, 2020). We pakken hier het getal 28 ton per container, dat ligt ongeveer in het midden van deze waardes.

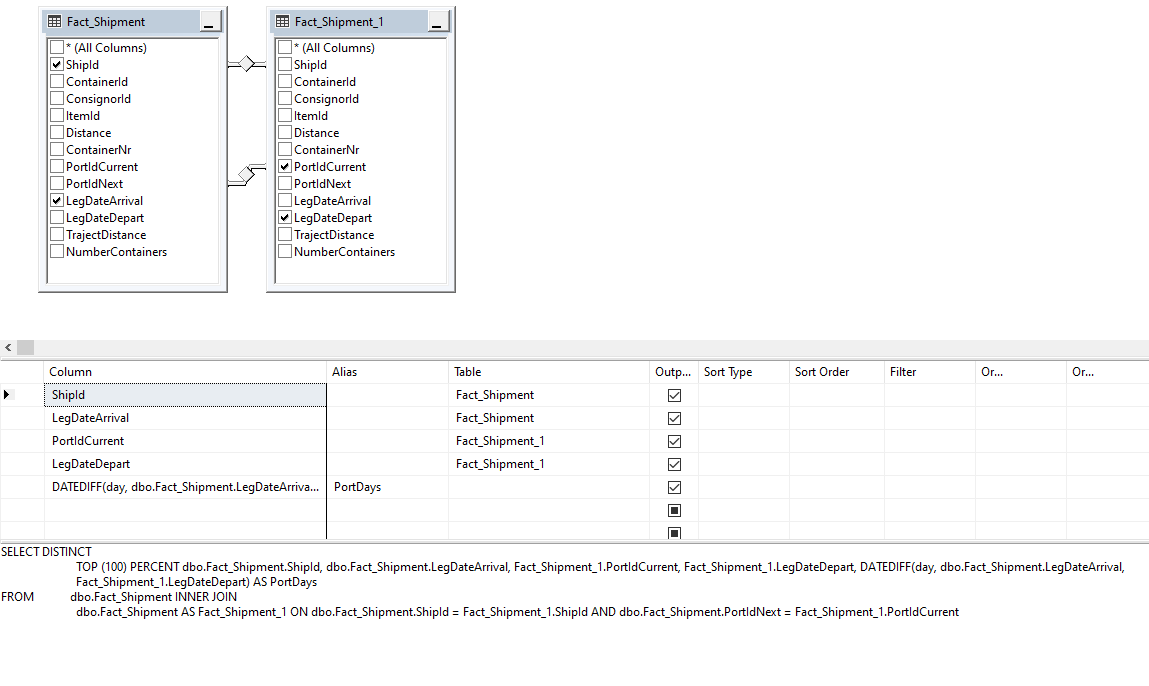
Brandstofverbruik komt uit de casus, we gaan uit van 250.000 liter per dag voor een schip dat 27 knopen vaart. Knopen staan voor zeemijl per uur.

Omdat we verschillende schepen hebben met verschillende aantal knopen berekenen we hoe lang ze gereisd hebben. Dat doen we keer het gemiddelde brandstof verbruik. we het te brandstof verbruik in te schatten alsof er een lineair verband is tussen het aantal knopen en het brandstof verbruik.

Van elke tocht weten we hoeveel containers er waren en hoeveel dagen er gevaren is. De totale berekening voor het brandstof verbruik is dus als volgt.

### Idle time

De idle time wordt berekend met behulp van een extra feiten tabel. Hier staan alle dagen dat boten stilliggen in havens. Dit is gedaan door de shipment tabel met zichzelf te joinen. De PortIdNext wordt zo verbonden met de PortIdCurrent en aan de hand daarvan kunnen de LegDateArrival en LegDateDeparture van elkaar af worden getrokken.



### Beladingsgraad

De beladingsgraad is in vergelijking redelijk makkelijk te bereken.

Toch zit deze berekening niet in het dashboard. De getallen zijn te onbetrouwbaar om analyse op uit te voeren. Zo heeft de boot OOCL Ningbo bijvoorbeeld 10400 containers bij in de tocht van Rotterdam naar Antwerpen terwijl er maar 8063 containers op de boot passen.

## Kosten per ton vracht

Kosten per ton vracht is redelijk simpel te berekenen. Het enige wat mist is de kosten van het schip jaarlijks. Dit zal verrekend moeten worden op een bepaalde manier. Containerschepen varen op stookolie (Havenkrant, 2020), de brandstofprijs is dus de prijs van stookolie per liter. Dit is nu € 0.4637/L. (Mazout, 2020)

### Brandstofverbruik (in relatie tot optimale snelheid) in tonnen brandstof per ton vracht per zeemijl; een en ander per type boot, vaarroute en seizoen.

Ik heb niet gevonden hoe ik dit het beste kan berekenen

# Conclusie en samenvatting

Kramse heeft nu een compleet datawarehouse met al hun informatie op 1 plek. Ook is er een begin gemaakt aan het maken van dashboards die passen bij hun informatiebehoefte. Met extra externe informatie (zoals seizoenen en automatisch vernieuwde brandstofprijzen) kan de informatievoorziening nog beter worden gemaakt. Toch moet er drastisch gekeken worden naar de kwaliteit van de data uit de bronnen, zo kan misschien de tegenstrijdige en foutieve data verholpen worden.

# Bibliografie

Havenkrant. (2020, june 25). *Een volle tank van zes miljoen*. Opgehaald van PortOfRotterdam: https://www.portofrotterdam.com/nl/havenkrant/havenkrant-22/een-volle-tank-van-zes-miljoen#:~:text=Als%20grote%20containerschepen%20twintig%20procent,250.000%20liter%20in%2024%20uur.

Mazout. (2020, june 25). *De officiele mazoutprijs*. Opgehaald van Informazout: https://informazout.be/nl/mazout/prijs

Stafford, T. (2020, june 25). *How much does a cargo ship weigh?* Opgehaald van Boating Geeks: https://boatinggeeks.com/how-much-does-a-cargo-ship-weigh/