**A picture containing text, sky, boat, outdoor

Description automatically generated**

**Project Kramse**

Project Business Intelligence

**Naam:** Fairys Ardhana – 2175412

**Datum:**  27-3-2022

**Beoordelaar: -**

# Inleiding

Er wordt in de afgelopen 2 maanden aan project Business Intelligence gewerkt. Hier krijgen de studenten een casus met databronnen waar wij vervolgens de missie/visie van de opdrachtgever uit gaan werken door middel van een datawarehouse en/of dashboards. In dit verslag worden de stappen en keuzes die ik heb genomen, beschreven.

Het bedrijf Kramse heeft een aantal informatiebehoeften, waarvoor het bedrijf een operationele database met enkele bijbehorende andere bestanden wil omzetten naar een datawarehouse op basis van SQL Server.

Wat het bedrijf wilt is om toeleveringsketens over de hele wereld te verbinden en te vereenvoudigen.

Voor het realiseren van de hierboven genoemde strategie heeft Kramse voor 2017 de volgende (sub)doelen geformuleerd:

* 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%

Voor een beter besluitvorming omtrent voorgenoemde doelen, wil Kramse een aantal dashboards ontwikkelen. Het management heeft daarom de volgende informatievragen op hoofdlijnen geformuleerd.

Kostenverlaging en utilisatie

* 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.

Inhoud

[Inleiding 1](#_Toc106275854)

[1. Opdracht 1 – Connection/Database 3](#_Toc106275855)

[1.1 Op welke manier heb je verbinding gemaakt met de bronbestanden? 3](#_Toc106275856)

[1.2 Welke bewerkingen/Structuur heb je toegepast om tot één geïntegreerde database te komen? 3](#_Toc106275857)

[2. Opdracht 2 – ontwerp 3](#_Toc106275858)

[Welke schema? 3](#_Toc106275859)

[2.1 Raw to ERP 4](#_Toc106275860)

[2.2 ERP TO ERP2nd 4](#_Toc106275861)

[2.3 ERP 2nd to PSA 5](#_Toc106275862)

[2.4 PSA to ODS 5](#_Toc106275863)

[3. Dashboards 5](#_Toc106275864)

# Opdracht 1 – Connection/Database

## Op welke manier heb je verbinding gemaakt met de bronbestanden?

Graphical user interface, application

Description automatically generatedText

Description automatically generatedEerst heb ik op Microsoft SQL Management studio een lege ‘raw database’ gemaakt door middel van een script (Zie bijlage). Vervolgens heb ik een SSIS project gemaakt op Visual Studio 2019 waar ik connecties met elke informatie bronnen en het database gemaakt heb.

## Diagram Description automatically generatedWelke bewerkingen/Structuur heb je toegepast om tot één geïntegreerde database te komen?

Ik heb vervolgens een dataflow gemaakt voor Containers, Consignor, MRV, Country en TPS om het raw database met data te vullen.

Het bron van container komt van een tekst document. Hierdoor heb ik flat file source gekozen en vervolgens het datatype met data conversion verandert. Daarna heb ik bij Refrigeration Flag het type naar Boolean gemaakt omdat de kolom “Yes” of “No” bevatten.

Het bron Consignor, MRV en Country heb ik met excel source verbindt en vervolgens de data type aangepast met data conversion voordat hij naar de destination(raw database) gestuurd wordt.

Als laatste bestaat TPS bron uit meerdere tabellen. Ik heb OLE DB Source gekozen omdat het bron een database is en vervolgens heb ik bij elke tabellen een OLE DB source -> destination.

# Opdracht 2 – ontwerp

## Welke schema?

Graphical user interface, text, application, chat or text message

Description automatically generatedNa het bestuderen van de casus en de data die ze willen heb ik besloten om de data naar een ster-schema te werken omdat ster-schema simpeler en overzichtelijker is. Vervolgens heb je niet alle kolommen/data van de data sources nodig om de vragen het opdrachtgever te beantwoorden.

## Raw to ERP

Bij dit proces wordt er relaties gemaakt tussen de tabellen. Eerst heb ik een database gemaakt voor ERP en vervolgens heb ik de data van rauw database naar het ERP geladen. Zoals op het afbeelding te zien is het de structuur van mijn dataflow naar het ERP database.

Timeline

Description automatically generatedBij Country, Container en Item is het nog een simpele source to destination flow. Bij het laden van Consignor heb ik lookup naar Country names gedaan. Het reden is omdat de landen namen heel veel vormen bevatten; nederlandse namen, engelse namen, 2 alpha code namen, etc. Hierdoor heb ik vier lookups gedaan en vervolgens met Union All ze verbonden met elkaar.

Bij Port en Ship vind hetzelfde proces plaats net als bij Consignor. Vervolgens is bij MRV een lookup naar shipname gedaan om deze twee tabellen te verbinden. Daarna wordt de data van calender opgeladen, ook het Depart Date en Arrival Date in de Voyage Port.

Bij het laden van Voyage moet gekeken zijn naar Ship, start van het port en het einde van het port. Bij het VoyagePort wordt er verbonden naar Voyage, Port current en next, datum voor LegDateDepart en arrival. Bij Shipment is er gekeken naar Voyage, Consigner, Port From en Port To. En als laatste is er Bij het laden van ShipmentDetail gekeken naar Shipment, Item, Container en ContainerNr.

## Diagram Description automatically generatedERP TO ERP2nd

Dit is het proces waar het meest tijd besteed is omdat ik dan keuzes moet maken om mijn data te verkleinen en ook te cleanen. Zoals op het afbeelding te zien heb ik nu veel minder tabellen. Van de niet meegenomen tabellen heb ik een paar kolommen meegenomen wat belangrijk kan zijn om de vragen van de opdracht gever te beantwoorden.

Diagram

Description automatically generatedHier heb ik elke tabel al dimensie of feit genoemd. Bij dimensie Port heb ik lookup naar de country gedaan. Bij dimensie Calendar is het een simpele source destination.

Bij dimensie Ship heb ik een lookup naar MRV gedaan en neem ik de kolommen totale CO2 uitstoot, totale brandstof verbruik, jaarlijkse gemiddelde brandstof verbruik per afstand, CO2 uitstoot en brandstof verbruik per transport. Uit onderzoek heb ik besloten dat deze 4 kolommen genoeg zijn om de vragen te beantwoorden. Vervolgens heb ik nog lookup naar de country naam gedaan.

Bij het feit Voyage wordt verbonden naar Ship en wordt lookup gedaan naar ERP Voyage om de kolommen over te nemen. Als laatste is Feit Shipment geladen. De kolommen worden aggregate met groupby en sum.

## Diagram Description automatically generatedERP 2nd to PSA

Bij PSA wordt er eindelijk een ster model gemaakt met 3 dimensie tabellen en 1 feit tabel. Dit model heb ik gekozen omdat het overzichtelijker is in vergelijken met een galaxy model.

Bij het laden van de dimensies zijn het een simpele source en destination. Bij het laden van de feit Voyage heb ik de kolommen van ERP\_Shipment meegenomen naar de tabel Voyage. Vervolgens heb ik een nieuwe kolom gemaakt voor het delay in days.

## PSA to ODS

Bij dimensie Calendar heb ik alle kolommen een fixed attribute van gemaakt. Ook bij het dimensie Port. Bij het dimensie Ship heb ik voor het jaarlijkse gemiddelde kolommen, totale CO2 uitstoot en brandstof verbruik en het jaarlijkse kosten van het scheep een historical attribute van gemaakt en de rest fixed attribute.

Chart, pie chart

Description automatically generatedBij het feit tabel Voyage heb ik naar de Ship een lookup gedaan om de shiprowId mee tenemen, dit komt omdat sommige schepen meerdere voyages hebben. Vervolgens heb ik lookup naar ODS Voyage gedaan.

# Dashboards

Mijn eerste dashboard is de beladingsgraad voor elke varen en hier kun je de schepen ook zien. Vervolgens is er ook een brandstof verbruik en CO2 uitstoot te zien. Je kunt ook bij elke varen de traject afstand analyseren.

Chart, line chart

Description automatically generatedVervolgens zijn er delay in days voor de schepen bij elke haven en een tabel van CO2 uitstoot en brandstof verbruik in elke maanden te zien.

Met deze tabellen kunnen de opdracht gever zien welke de beste scheep zijn en het info van die scheep van bijvoorbeeld het jaarlijkse cost, brandstofverbruik en Co2 uitstoot en hierbij beslissingen maken.

Chart, bar chart

Description automatically generatedJe kunt bijvoorbeeld zien dat het scheep Cosco het minste beladingsgraad hebben en vervolgens gebruiken ze bij elke varen meer brandstof in vergelijken met de andere schepen. Ook stoot Cosco gemiddelde meer CO2 uit dan de rest van de schepen.



# Bijlage (Scripts)

## Zie folder