Ik ben begonnen door de gegevens van de verschillende databronnen in een relationele database.

De volgende databronnen:

* De kerngegevens komen uit een MS Access database.
* Gegevens over opdrachtgevers komen uit een Excel-bestand; dit bestand is afkomstig van de commerciële afdeling.
* Gegevens over MRV komen uit een Excel-bestand.
* Gegevens over soorten containers komen uit een tekstbestand; dit bestand is afkomstig van een internationale organisatie.

Dit leed tot verschillende tabellen.

* Consignor
* Container
* Item
* MRV\_Publication
* Port
* Ship
* Shipment
* ShipmentDetail
* Voyage
* VoyagePort

In deze “raw” database zijn er al relaties en primary keys, deze heb ik alleen gemaakt op de Access db tabellen.

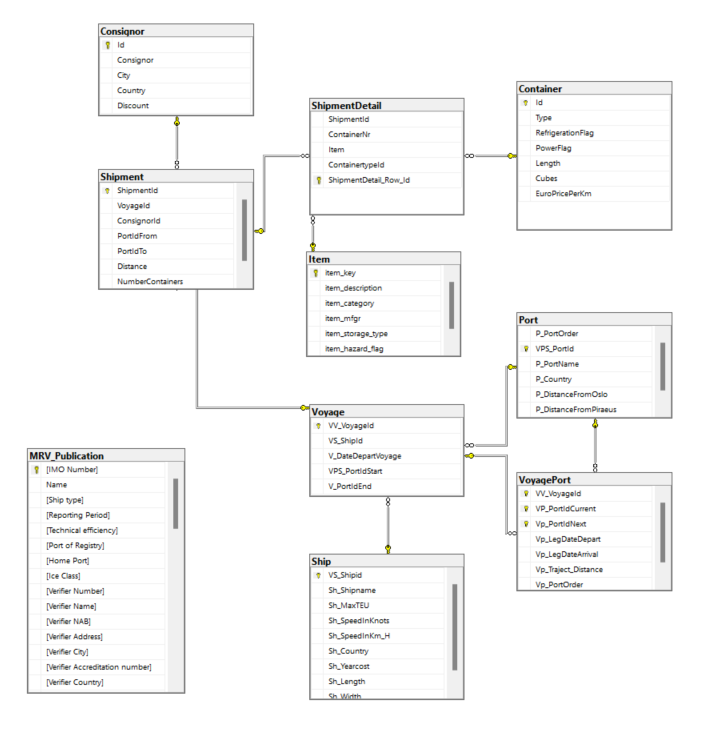
Het bedrijf heeft de volgende (sub)doelen

* 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%

Deze inzichtelijk maken met dashboards, volgende informatievragen voor deze dashboards:

* 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.

Structuur van de “raw” database:



Om tot 1 relationele database te komen heb ik relaties gemaakt tussen de Access DB tabellen, dit was vrij eenvoudig omdat dit al vast stond in de Access DB. Daarna heb ik de consignor en de Container tabel er bij betrokken

Voor de MRV table wist ik nog geen goede relatie te leggen. Uiteindelijk heb ik geen gebruik gemaakt van de MRV tabel omdat ik hier geen bruikbare data in kon vinden.

# Informatiebehoefte

Kramse heeft de volgende doelen opgesteld:

* CO2 reductie /NOx/SOx: verlagen van het brandstofverbruik met 10% (brandstofverbruik/ per ton vracht per zeemijl)
  + Benodigd: brandstofverbruik van de schepen, hoeveelheid vracht op een schip
* Ulitisatiegraad verbeteren: reduceren van de wachttijden in de havens, ze zogenaamde idle time, met 5%.
  + Tijd dat een schip stil ligt in de haven berekenen, dit kan in de VoyagePort.
* Beladingsgraad; verder optimaliseren van de beladingsgraad.
  + Ship table heeft hoeveelheid containers die er op kunnen in de Sh\_MaxTEU kolom.

Hoe ik deze berekeningen heb gemaakt en hoe dit verder inzichtelijk is, is terug te vinden op het dashboard en alle berekeningen staan hier ook tussen. Ik heb geen gebruik gemaakt van een cube in SSAS. Deze keuze heb ik gemaakt omdat om dit te maken het mij te veel moeite kostte en ik de cube zelf ook niet kon deployen naar een server, ook niet localhost.

Verder heb ik nog alle scripts voor het genereren van de staging / psa / ods databases bijgevoegd.

In het dashboard bestand heb ik 4 verschillende dashboards gemaakt.

1. Algemene informatie, voor globaal management
2. KPI’s, staat Kramse er goed voor of moeten ze nog door voor het halen van hun doelen, en hoeveel scheelt het nog.
3. Beladingsgraad, verschillende grafieken die mogelijk oorzaken kunnen zijn voor verschillende hoeveelheden beladingsgraad.
4. Staydays; waar vinden de meeste stilliggende dagen plaats. Heeft dit te maken met de grootte van de schepen?

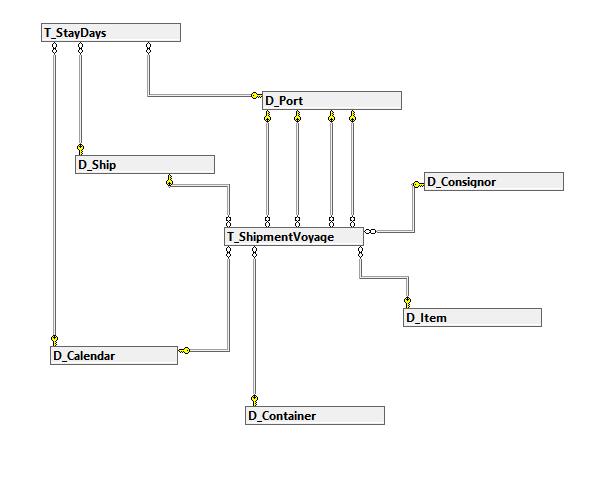
# Opdrachten

## Opdracht 1: Extraheer gegevens

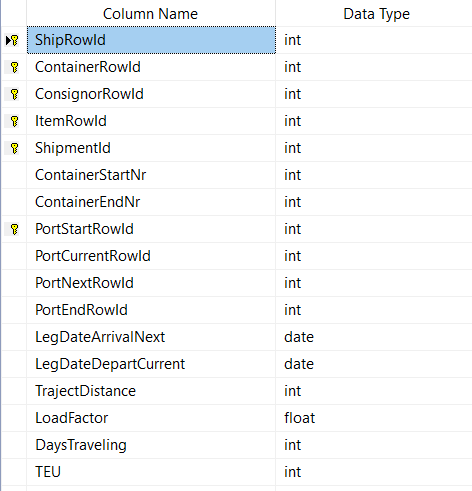
* Consignor tabel
  + Alle NULL values in de discount kolom vervangen met 0
  + Een zeer grote discount dynamisch vervangen/omgekeerd
* Container tabel
  + Lege rijen weggehaald (er was 1 lege rij aan het begin)
  + Id type veranderd in een int (was varchar)
* ShipmentDetail
  + Container Nr gesplit in twee kolommen. (bijvoorbeeld 501-2000 veranderd naar 501 en 2000 in apparte kolommen
  + Totale aantal containers berekend. Dit zou ik in het vervolg gelijk veranderen naar TEU’s, sommige containers zijn namelijk dubbel zo groot als anderen.
* VoyagePort
  + Lookup previous DateArrival ik haal de datum op van de vorige arrival zodat ik de staydays kan berekenen, als er geen vorige date is dan wordt de kolom 0

## Opdracht 2: ontwerpen galaxy schema

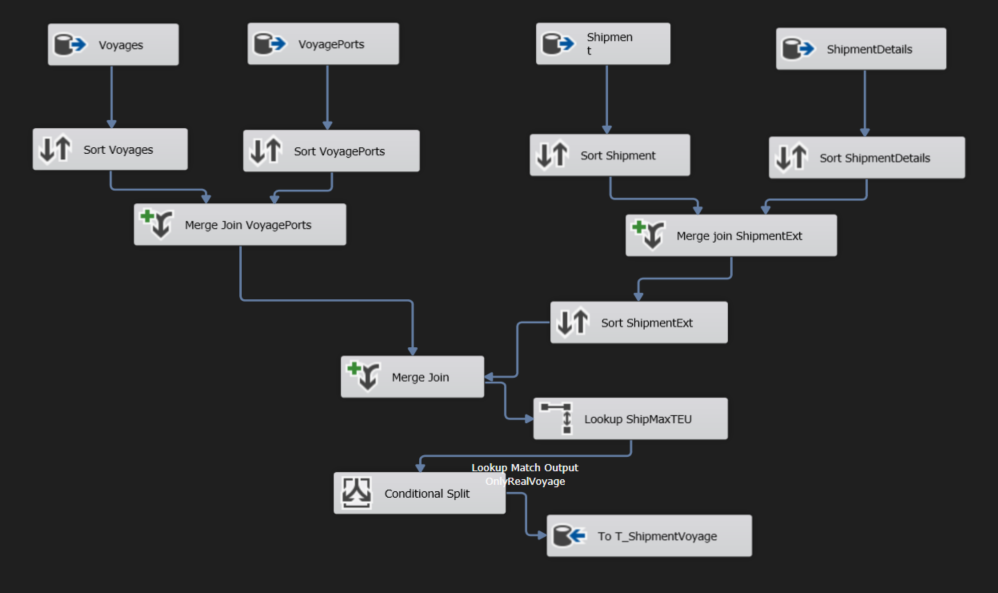
Mijn galaxy schema is er als volgt uit komen te zien: ik heb er zelf een Calendar dimensie bij gezet en ook is er een nieuwe T\_StayDays tabel waar handig in staat hoeveel staydays er zijn.



Ik heb voor de T\_ShipmentVoyage tabel 4 verschillende tabellen samengevoegd. Door te kijken welke tabellen gebeurtenissen vastleggen en welke dimensies zijn. Hieronder alle kolommen in deze ShipmentVoyage tabel



Hieronder hoe ik in PSA de feitentabel maak. ik kan deze mergen door eerst de tabellen te sorten op dezelfde kolom.



## Opdracht 3: implementeren schema

In de bijlage van dit project bevind zich het script voor het generen van alle databases

## Opdracht 4: vul het datawarehouse vanuit de staging area

Ik heb op alle dimensies slowly changing dimensions toegepast, als ik wil dat de geschiedenis van een kolom bijgehouden moet worden maak ik er historicals van en anders changing of fixed.

Bij de feitentabellen staan er lookups voor het incrementen van de data, deze geven alleen nieuwe records door zodat er geen dubbele records in de feitentabel terecht komen.

In de feitentabellen staan ook geen normale id’s maar rowId’s voor de relaties.

## Opdracht 5: Ontwerp dashboards

In het dashboard bestand heb ik 4 verschillende dashboards gemaakt.

1. Algemene informatie, voor globaal management
2. KPI’s, staat Kramse er goed voor of moeten ze nog door voor het halen van hun doelen, en hoeveel scheelt het nog.
3. Beladingsgraad, verschillende grafieken die mogelijk oorzaken kunnen zijn voor verschillende hoeveelheden beladingsgraad.
4. Staydays; waar vinden de meeste stilliggende dagen plaats. Heeft dit te maken met de grootte van de schepen?

Voor de staydays kan Kramse de conclusie trekken dat in de Bilbao en Barcelona ports veel staydays zijn, wellicht kan hier meer aandacht aan besteed worden.

Ook is te zien dat schepen met een hoge maxTEU meer staydays hebben, dit komt waarschijnlijk doordat het langer duurt om deze helemaal vol te gooien. Ook is te zien welke schepen de meeste staydays hebben. Dit kan liggen aan de crew of aan andere factoren.