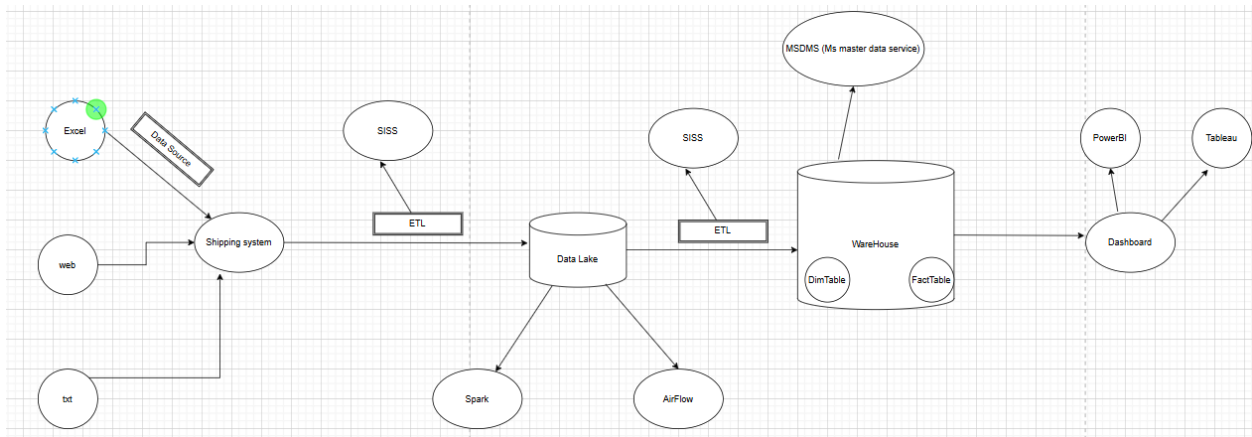


Degsin data pipeline

Model and idea

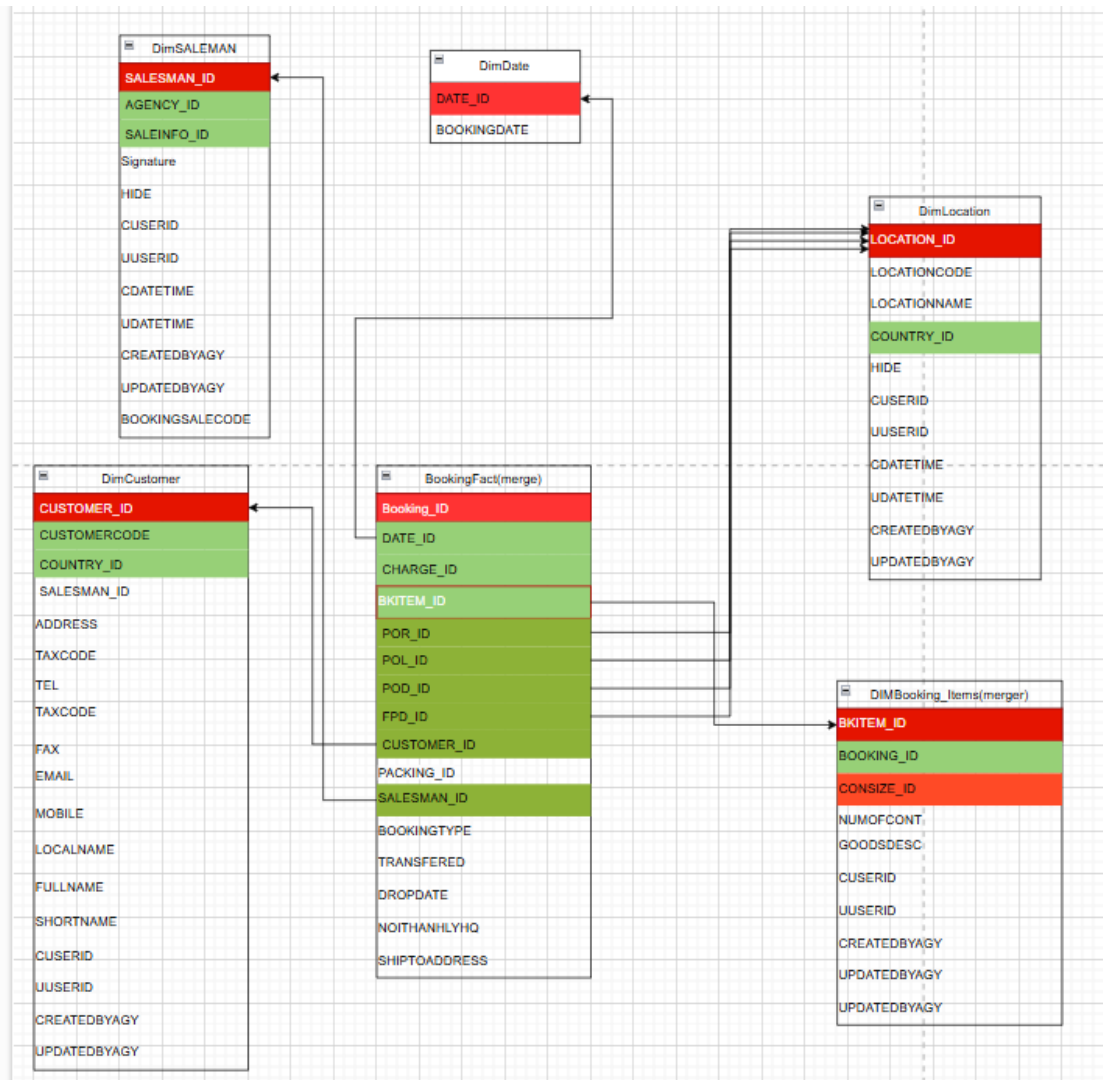


Instructor give me an idea about design data warehouse model it include:

- Get available data of company to design a shipping system
- Extract, tranform and load it to data lake
- Then ETL it into dataware house and finally visualization it on dashboard

I. Design Dim and Fact model

First i using draw.io to design dim and fact table when i have data from intructor



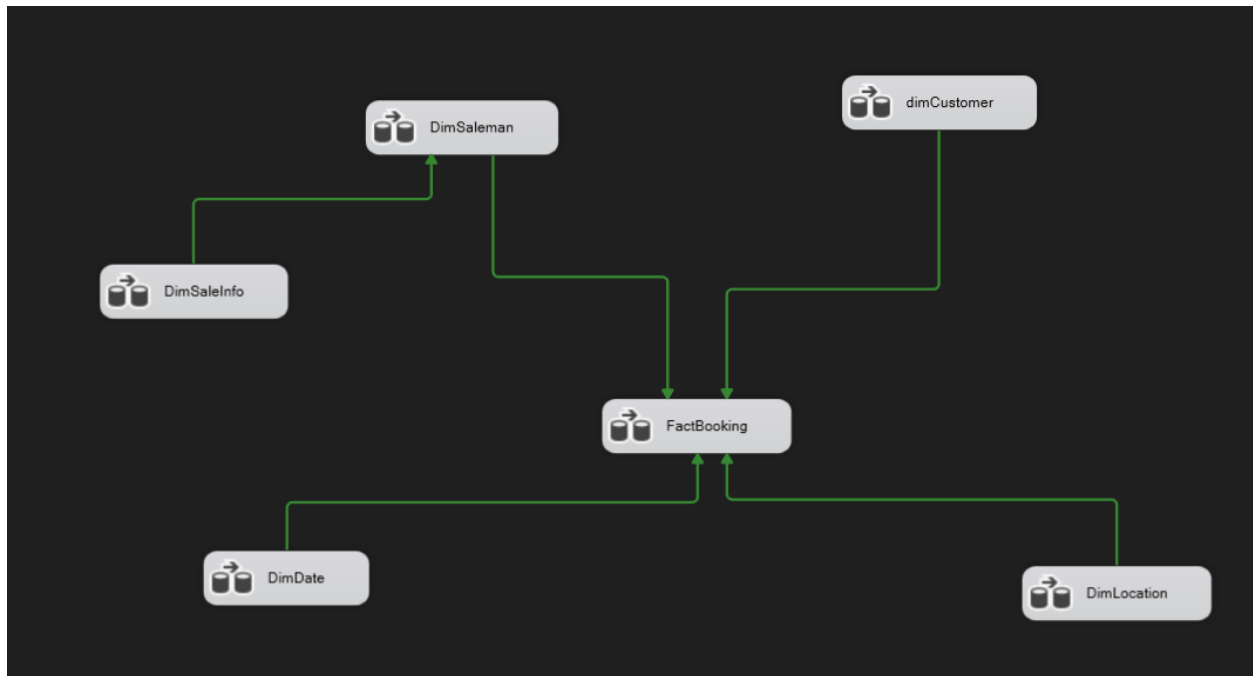
Pic1: model dim and fact table

This is Dim and Fact table

Because data very large i don't have enough time to do so instructor give some table on pic1 for me to practice.

II. Extract, tranform and load data:

- My idea is get data from source ETL it into Staging layer and check any problem with my data and finally give it into Warehouse layer
- I use Micrsoft SISS to work it's a tool of microsoft can help me etract, tranform and load data into to stagging layer.



Pic2: get data from source to staging layer

- I get data from source and transform it to staging layer and load it to sql server database
- I use some query to find any problem in data. If it do not any problem similar pic2 I will ETL it to data warehouse

III. Query in SQL

After complete ETL data from data warehouse I use sql server to query some request in that data like: Service customer, Customer count, Sum avenue, avenue saleman, name of sale man,....

```

Execute
xuly.sql - DESKTOP-J8AJODR\Asus (53)
--tong customer
select count(distinct CUSTOMER_ID) from VW_CUSTOMER
--tong so cont
select sum(NUMOFCONT) from VW_CUSTOMER
--customer nao booking cont nhieu nhât (tu 5 xuong la ít tren 5 là nhieu)
SELECT CUSTOMER_ID, FULLNAME, SHORTNAME, COUNT(NUMOFCONT) AS TotalBooking
FROM VW_CUSTOMER
GROUP BY CUSTOMER_ID, FULLNAME, SHORTNAME, CURRENCYNAME
HAVING COUNT(NUMOFCONT) <=2000
ORDER BY TotalBooking DESC;

--customer co gia tri don dat cao
select CUSTOMER_ID, FULLNAME, SHORTNAME, CURRENCYNAME,MINRATE,RATE,APPROVEDRATE,APPROVEDRATE*RATE as TotalRate,
case
when APPROVEDRATE=0 then MINRATE
end
from VW_CUSTOMER
GROUP BY CUSTOMER_ID, FULLNAME, SHORTNAME, CURRENCYNAME,MINRATE,RATE,APPROVEDRATE
ORDER BY TotalRate DESC;

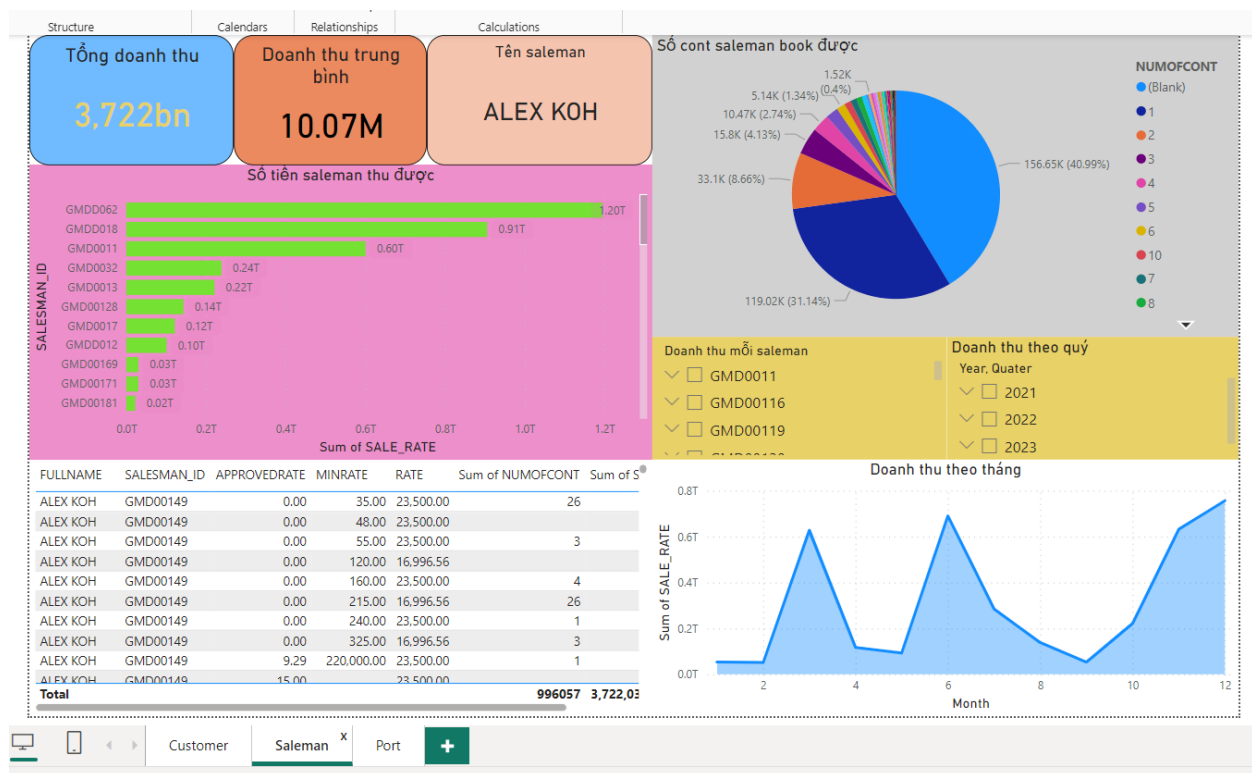
--customer co so cont là 1
select count(distinct CUSTOMER_ID),CUSTOMER_ID from VW_CUSTOMER where NUMOFCONT>=1 and NUMOFCONT<=5
group by CUSTOMER_ID
--cos so cont tu 5 den 2000
select count(distinct CUSTOMER_ID),CUSTOMER_ID from VW_CUSTOMER where NUMOFCONT<=2000 and NUMOFCONT>=5
group by CUSTOMER_ID
--customer dung dịch vụ gif
select * from VW_CUSTOMER
SELECT COUNT(DISTINCT CUSTOMER_ID), CHARGENAMEVN, SHORTNAME

```

Pic3: Example some query in sql server

IV. Visualization on dashboard

Finally when complete query I use power Bi to visualize data from dashboard



Pic4: Example dashboard of saleman