# cu.bmp

Data warehousing project

### **1- Cover page**

|  |  |
| --- | --- |
| Names | IDs |
| Farah Mohammed Hussien | 20210287 |
| Eslaam Sayed Gouda | 20211012 |

**Contents**

[1](#_Toc166869791)

[1- Cover page 1](#_Toc166869792)

[2- Physical model of the source system: 5](#_Toc166869793)

[3- Dimensional model 5](#_Toc166869794)

[a- The business processes we modelled: 5](#_Toc166869795)

[b- Grain of each fact table: 5](#_Toc166869796)

[c- The type of each fact table: 5](#_Toc166869797)

[d- The dimensions and the type of each one: 5](#_Toc166869798)

[e- The measures that will appear in the fact tables 6](#_Toc166869799)

[f- Physical model 6](#_Toc166869800)

[4- Screenshots of the data flow tasks and control Flow. 7](#_Toc166869801)

[1- Campaign dimension loading 7](#_Toc166869802)

[a- Control flow 7](#_Toc166869803)

[b- Data flow 7](#_Toc166869804)

[1- From source to stage 7](#_Toc166869805)

[2- Stage to destination 8](#_Toc166869806)

[2- Customer dimension loading 8](#_Toc166869807)

[a- Control flow 8](#_Toc166869808)

[b- Data flow 9](#_Toc166869809)

[1- From source to stage 9](#_Toc166869810)

[2. Stage to destination 9](#_Toc166869811)

[3- Product dimension loading 10](#_Toc166869812)

[a- Control flow 10](#_Toc166869813)

[b- Data flow 10](#_Toc166869814)

[1- From source to stage 10](#_Toc166869815)

[2- Stage to destination 11](#_Toc166869816)

[4- Supplier dimension loading 11](#_Toc166869817)

[a- Control flow 11](#_Toc166869818)

[b- Data flow 12](#_Toc166869819)

[1- From source to stage 12](#_Toc166869820)

[2- Stage to destination 12](#_Toc166869821)

[5- Sales Fact Table 13](#_Toc166869822)

[a- Control flow 13](#_Toc166869823)

[b- Data flow 13](#_Toc166869824)

[1- From source to stage 13](#_Toc166869825)

[2- From stage to destination 14](#_Toc166869826)

[6- Sales Fact Table 14](#_Toc166869827)

[A- Control flow 14](#_Toc166869828)

[B- Data flow 15](#_Toc166869829)

[1- From source to stage 15](#_Toc166869830)

[2- From stage to destination 15](#_Toc166869831)

[7- Sales Fact Table 16](#_Toc166869832)

[A- Control flow 16](#_Toc166869833)

[B- Data flow 16](#_Toc166869834)

[1- From source to stage 16](#_Toc166869835)

[2- From stage to destination 17](#_Toc166869836)

[8- Batch file 17](#_Toc166869837)

[5- Queries on each fact table and what insights we can get from it, and a screenshot of the result set of each query. 18](#_Toc166869838)

[Sales Fact Table Queries 18](#_Toc166869839)

[Total Sales for each country 18](#_Toc166869840)

[Total Sales for each customer we can get customer name form customer dimension 19](#_Toc166869841)

[Get number of orders for each customer 20](#_Toc166869842)

[Get the total sales based on a date 21](#_Toc166869843)

[Get total Sales for each marketing campaign 22](#_Toc166869844)

[Total Sales For each payment\_method 22](#_Toc166869845)

[Retruns Fact Table Queries 23](#_Toc166869846)

[Get amount Refunded for each customer 23](#_Toc166869847)

[Get amount Refunded for a range of date 24](#_Toc166869848)

[Get number of returns for a range of date 24](#_Toc166869849)

[Get number of returns for each customer 25](#_Toc166869850)

[Get number of returns for each product 26](#_Toc166869851)

[Product Sales Fact Table Queries 27](#_Toc166869852)

[Get Quantity sold for each product 27](#_Toc166869853)

[Get Quantity sold for each supplier 28](#_Toc166869854)

[Get Quantity sold for each subcategory 29](#_Toc166869855)

[Get Quantity sold for each customer 30](#_Toc166869856)

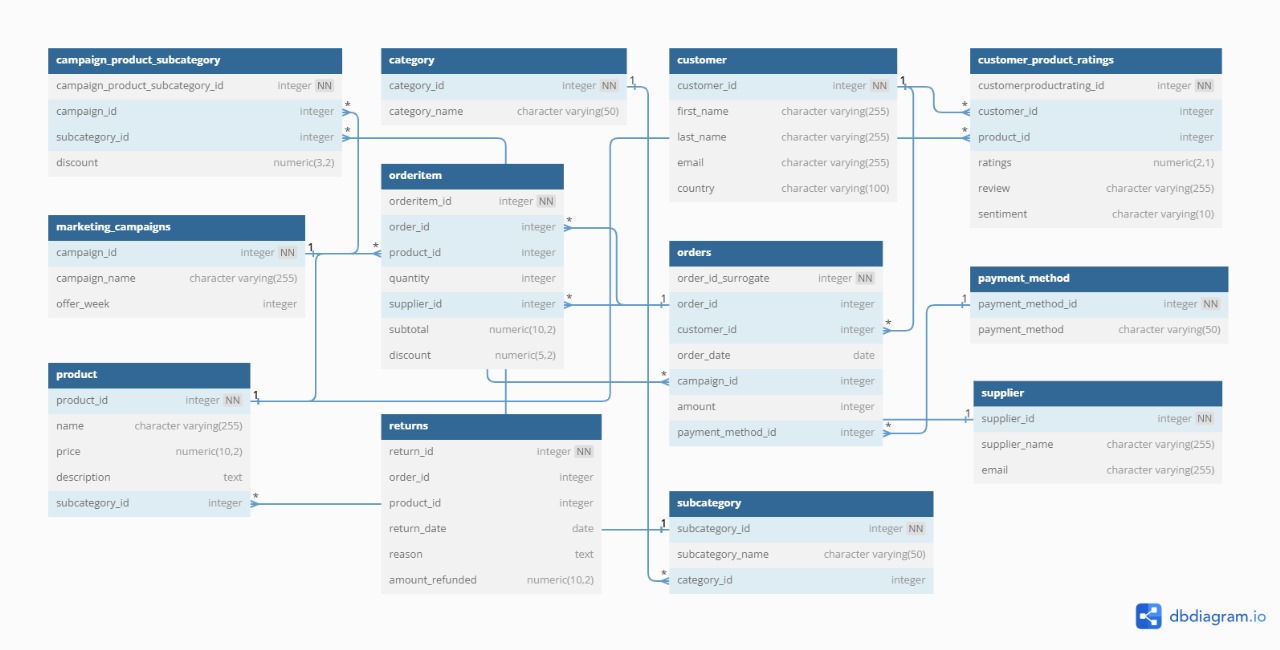
[Get total sales for each subcategory 31](#_Toc166869857)

[Get total sales for each product 32](#_Toc166869858)

[Get total sales for each supplier 33](#_Toc166869859)

[6- Screenshots of the deployed packages in SSIS with their schedule 34](#_Toc166869860)

# 2- Physical model of the source system:



---------------------------------------------------------------------------------------------------------------------

# 3- Dimensional model

## The business processes we modelled:

The business process we modelled is an ecommerce website that makes marketing campaigns and discounts on holidays (New Year's, mother's day, etc.)

## Grain of each fact table:

1. Sales for each: Customer, Marketing campaign, country, order at a certain time
2. Returns for each : product, customer, review and a sentiment at a certain time
3. Quantity sold and sales for product: supplier, subcategory, customer

## The type of each fact table:

All tables are transactional Fact Tables

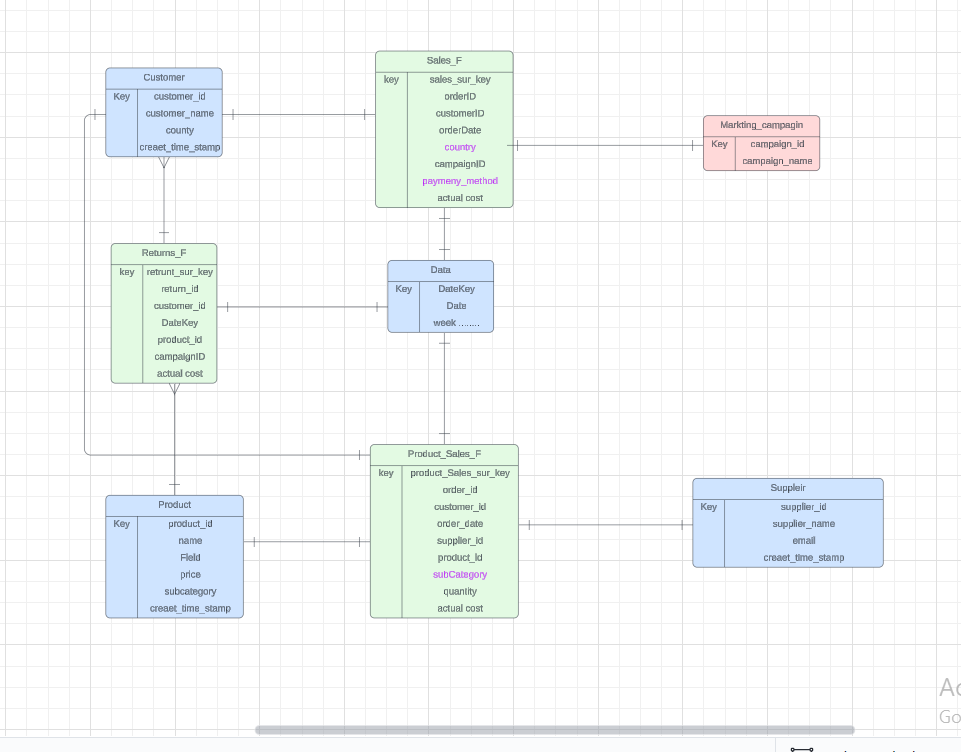
## The dimensions and the type of each one:

1. customer: Slowly changing confirmed dimension
2. country : Static degenerate dimension
3. Marketing campaign :static dimension
4. Product : slowly changing dimension
5. Subcategory: degenerate dimension
6. Supplier : slowly changing dimension
7. Payment Method: slowly changing dimension
8. DATE: confirmed dimension

## The measures that will appear in the fact tables

1. Sales \_f Actual\_cost : fully additive measure
2. Amount\_refunded : fully additive measure
3. Quantity : semi additive measure
4. product\_f Actual\_cost : fully additive measure
5. rating : non-additive
6. sentiment: non additive

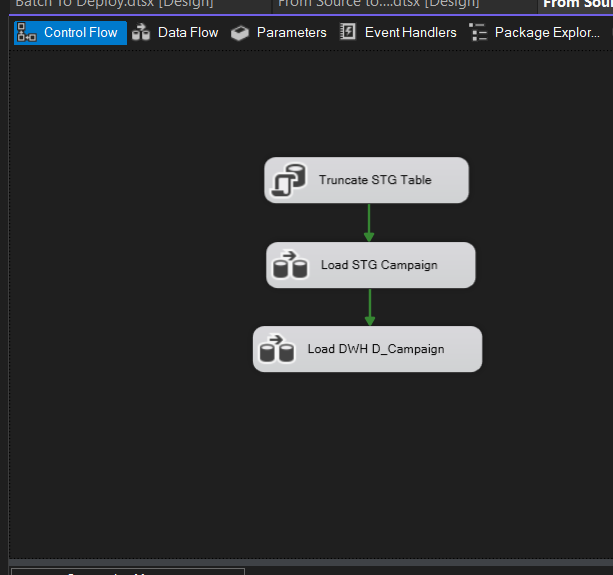
## Physical model



# 4- Screenshots of the data flow tasks and control Flow.

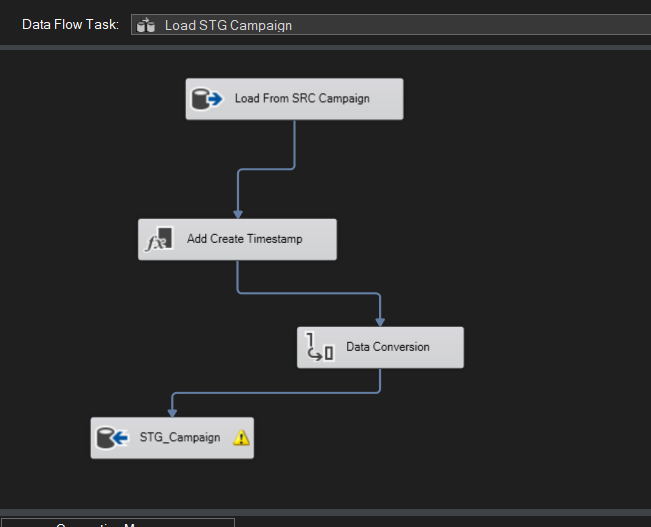
## Campaign dimension loading

### Control flow

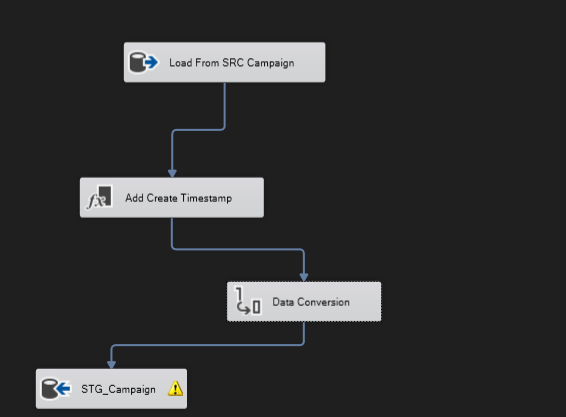


### Data flow

### From source to stage



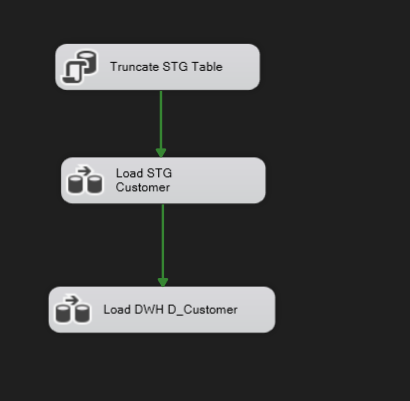
### Stage to destination



---------------------------------------------------------------------------------------------------

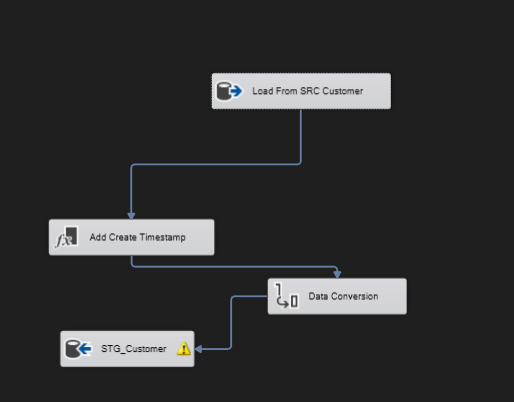
## Customer dimension loading

### Control flow

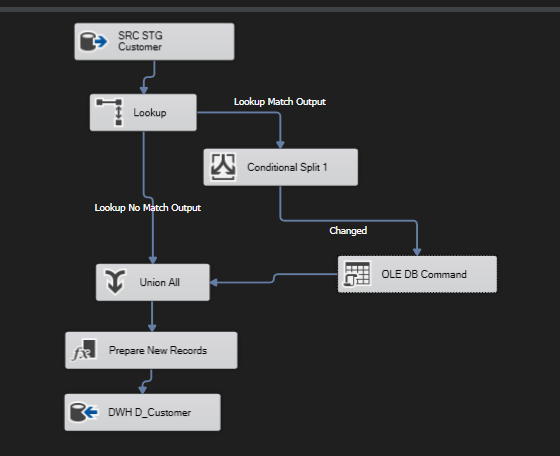


### Data flow

### From source to stage

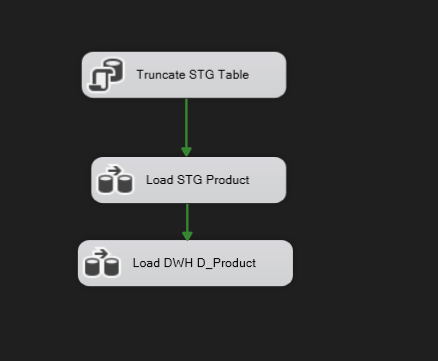


### 2. Stage to destination



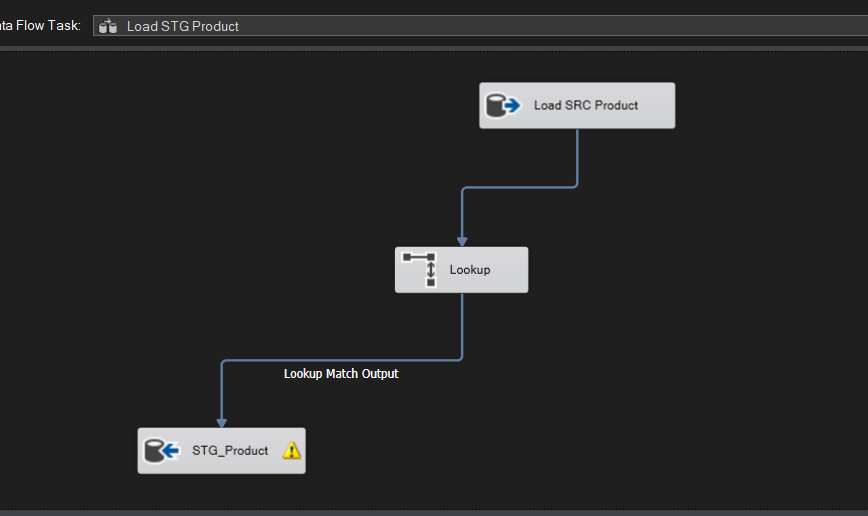
## Product dimension loading

### Control flow

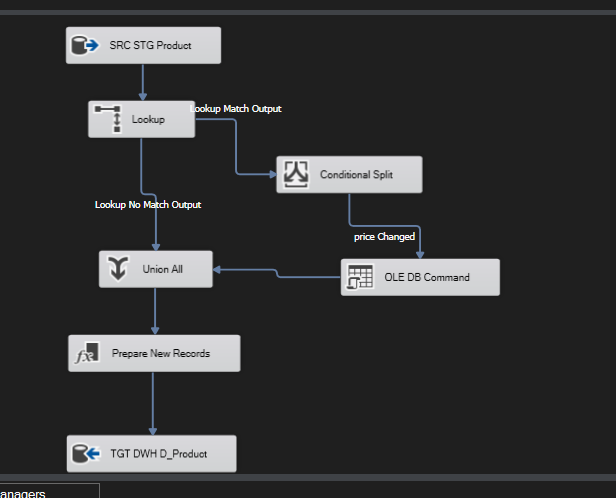


### Data flow

### From source to stage



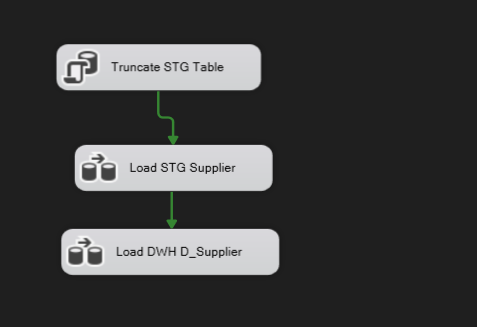
### Stage to destination



---------------------------------------------------------------------------------------------------------------------

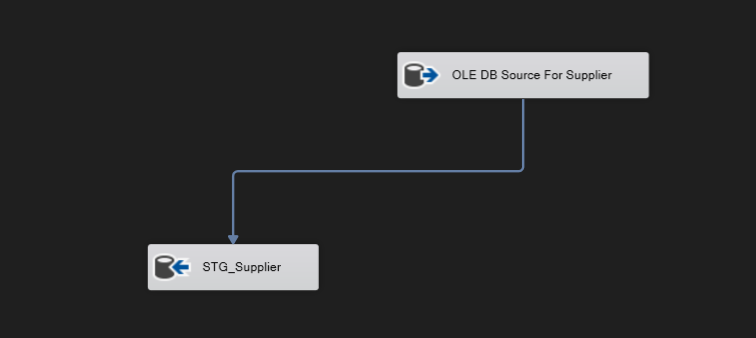
## Supplier dimension loading

### Control flow

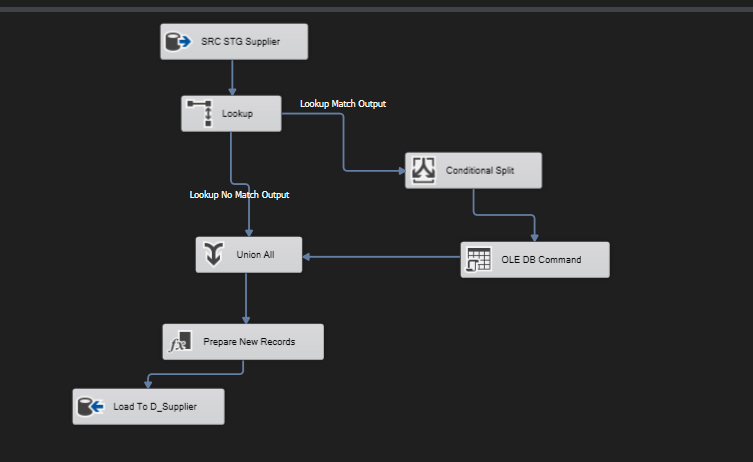


### Data flow

### From source to stage



### Stage to destination



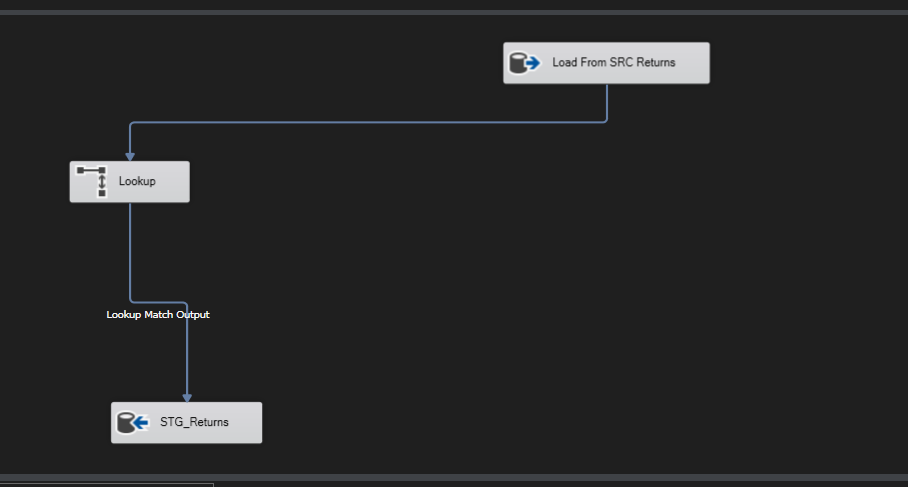
## Sales Fact Table

### Control flow

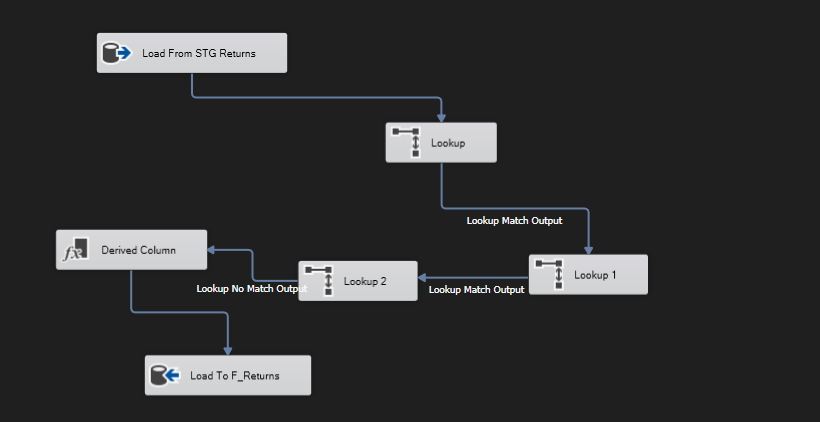


### Data flow

### From source to stage



### From stage to destination



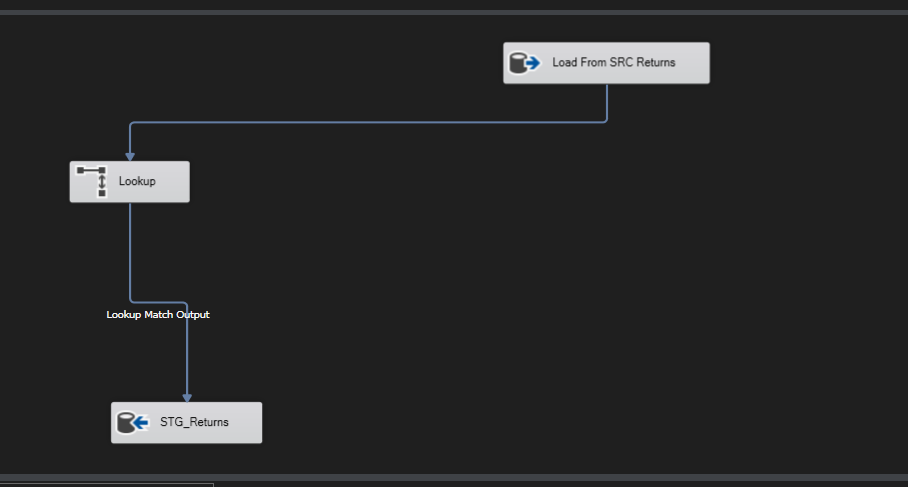
## Sales Fact Table

### Control flow

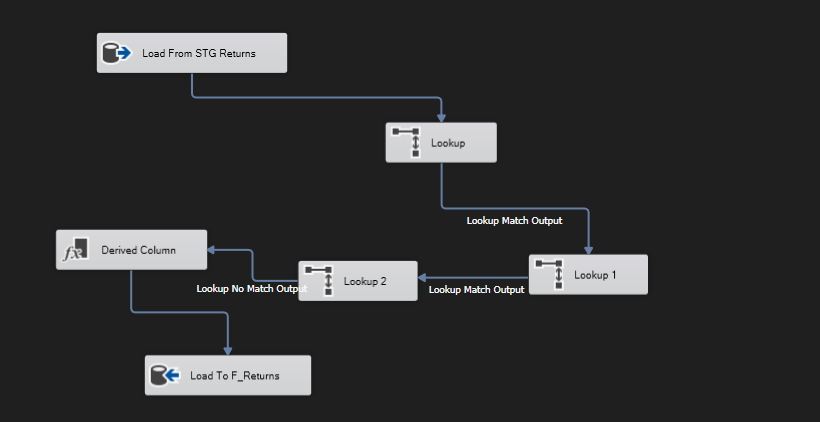


### Data flow

### From source to stage



### From stage to destination



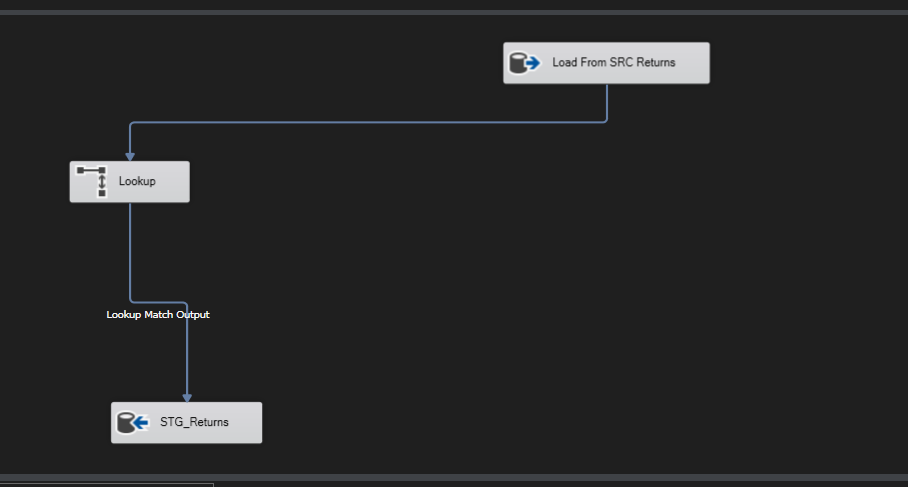
## Sales Fact Table

### Control flow

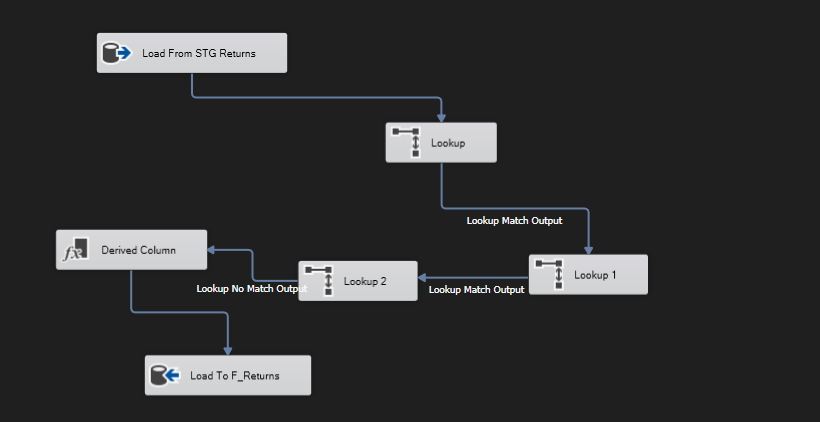


### Data flow

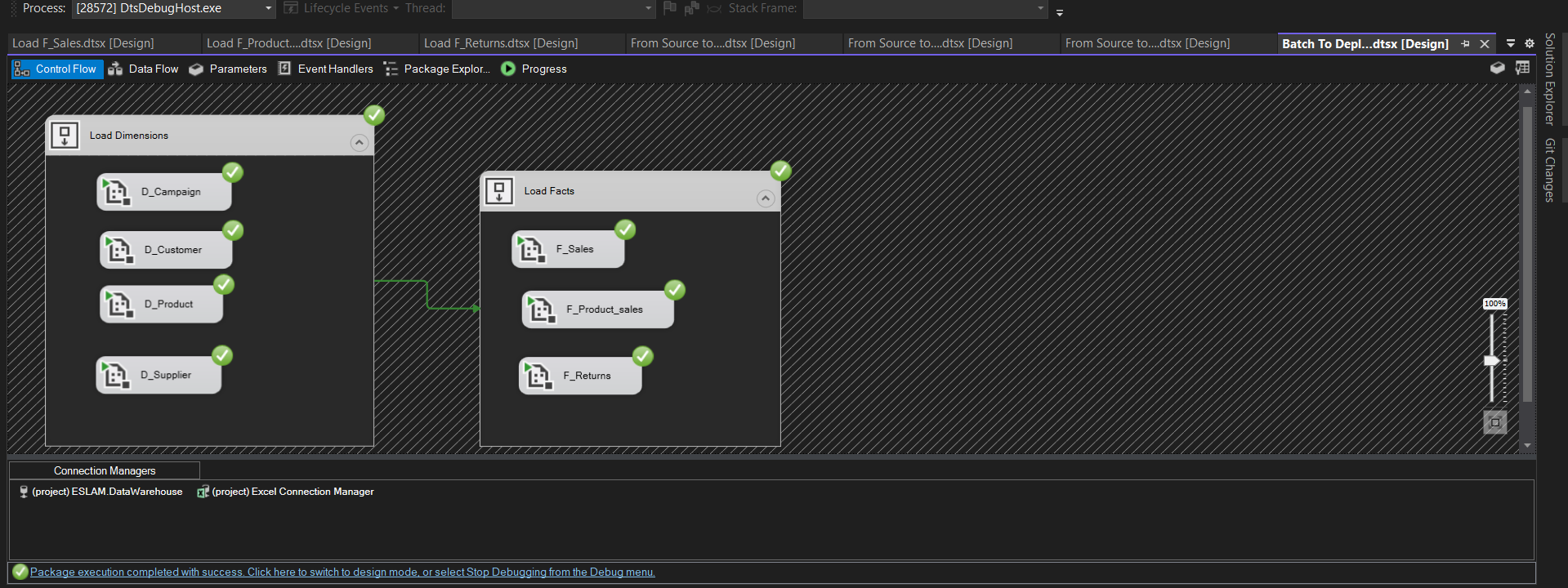
### From source to stage



### From stage to destination



## Batch file



# 5- Queries on each fact table and what insights we can get from it, and a screenshot of the result set of each query.

## Sales Fact Table Queries

### Total Sales for each country

Select country, sum(actual\_cost) as total from dwh.F\_sales

Group by country

Order by country

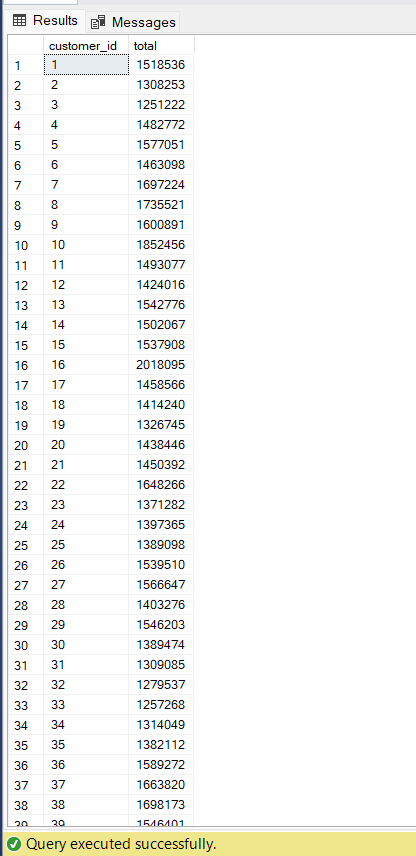


Total Sales for each customer we can get customer name form customer dimension

Select customer\_id , sum(actual\_cost) as total from dwh.F\_sales

Group by customer\_id

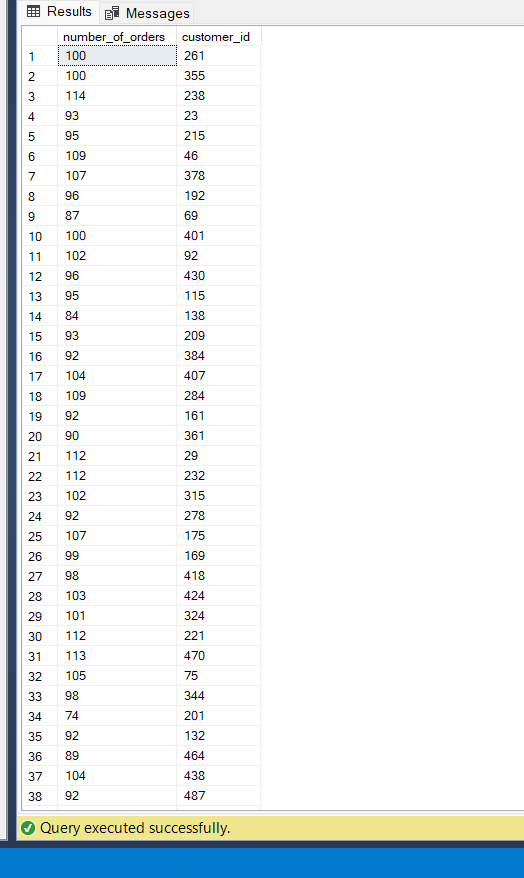
Order by customer\_id



### Get number of orders for each customer

Select count(order\_id) as number\_of\_orders , customer\_id from dwh.F\_sales

Group by customer\_id

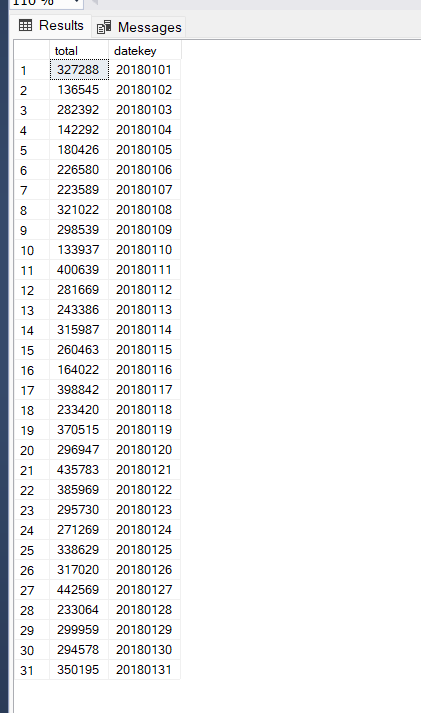


### Get the total sales based on a date

select sum(actual\_cost) as total , datekey from dwh.F\_sales

where datekey >= 20180101 and datekey < 20180201

group by datekey

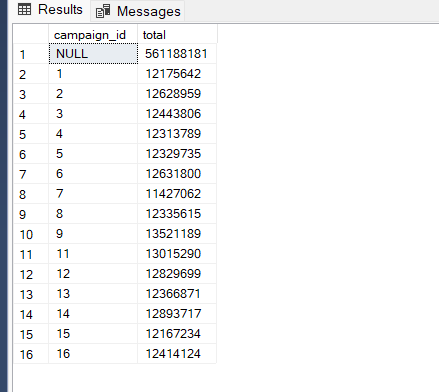


### Get total Sales for each marketing campaign

select campaign\_id , sum(actual\_cost) as total from dwh.F\_sales

group by campaign\_id

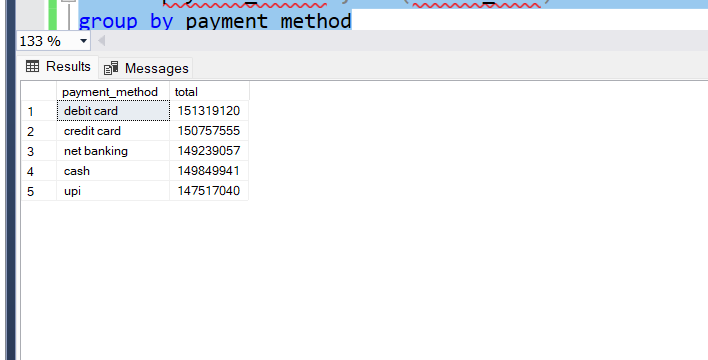
order by campaign\_id



### Total Sales For each payment\_method

select payment\_method , sum(actual\_cost) as total from dwh.F\_sales

group by payment\_method

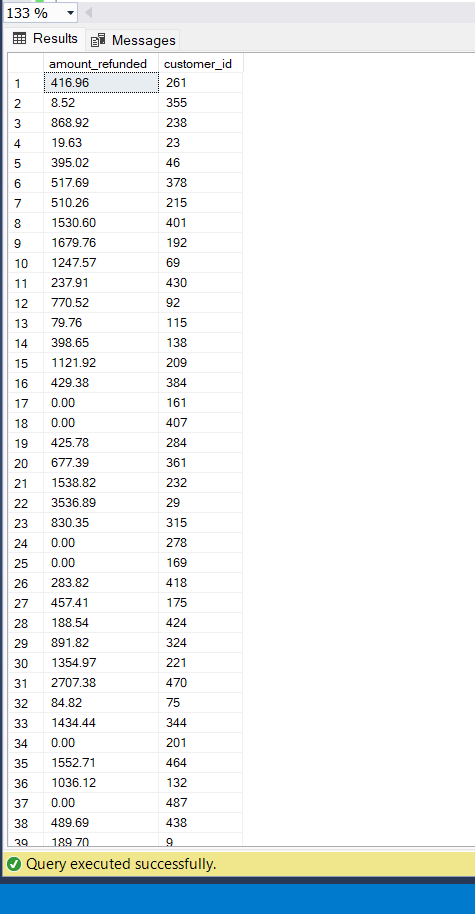


## Retruns Fact Table Queries

### Get amount Refunded for each customer

select sum(amount\_refunded) as amount\_refunded , customer\_id from dwh.F\_Returns

group by customer\_id

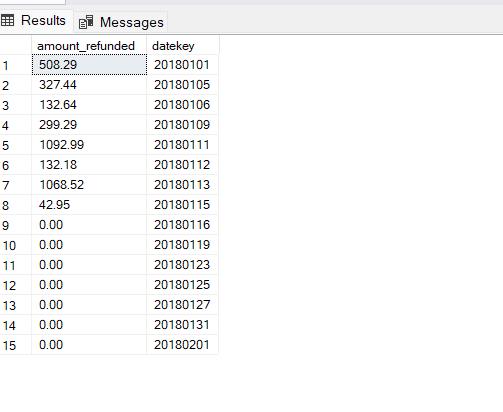


### Get amount Refunded for a range of date

select sum(amount\_refunded) as amount\_refunded , datekey from dwh.F\_Returns

where datekey >= 20180101 and datekey <= 20180202

group by datekey

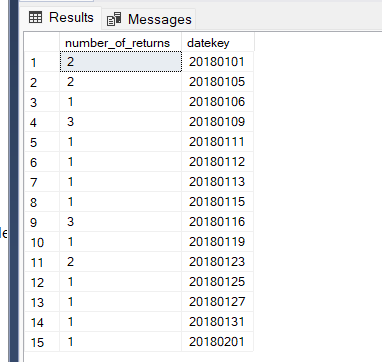


### Get number of returns for a range of date

select count(return\_sur\_key) as number\_of\_returns , datekey from dwh.F\_Returns

where datekey >= 20180101 and datekey <= 20180202

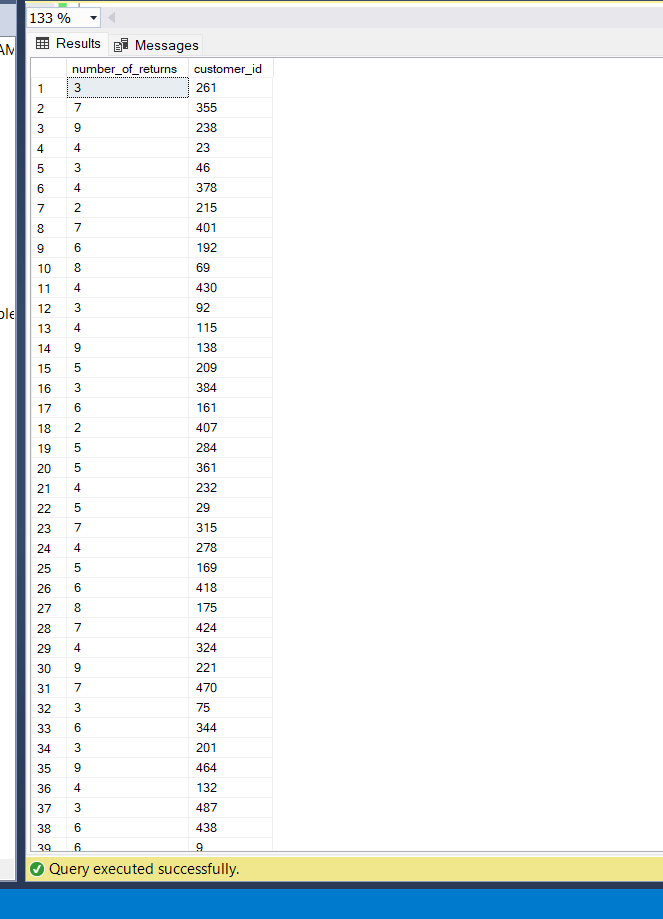
group by datekey



### Get number of returns for each customer

select count(return\_sur\_key) as number\_of\_returns , customer\_id from dwh.F\_Returns

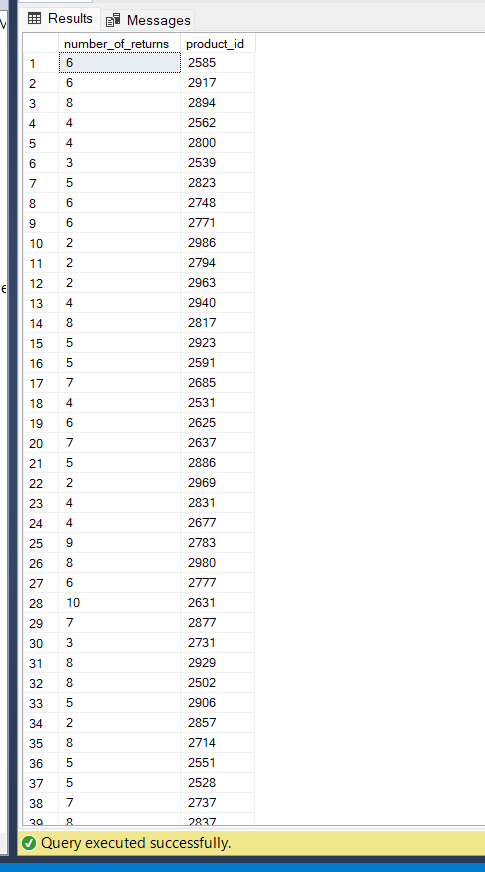
group by customer\_id



### Get number of returns for each product

select count(return\_sur\_key) as number\_of\_returns , product\_id from dwh.F\_Returns

group by product\_id

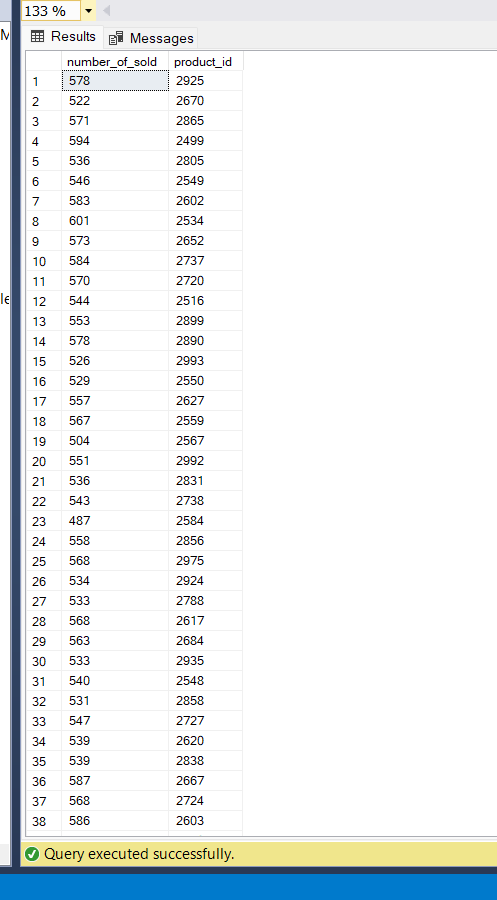


## Product Sales Fact Table Queries

### Get Quantity sold for each product

Select count(product\_sales\_sur\_key) as number\_of\_sold , product\_id from dwh.F\_product\_sales

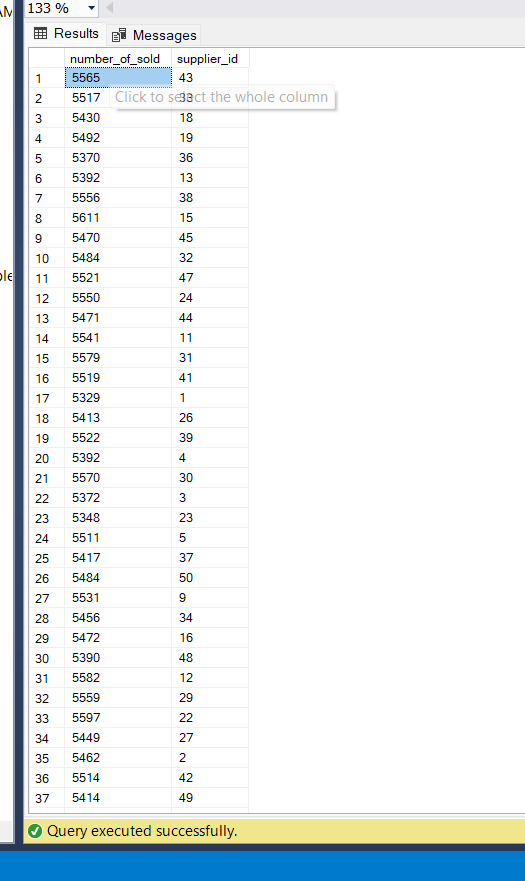
Group by product\_id



### Get Quantity sold for each supplier

Select count(product\_sales\_sur\_key) as number\_of\_sold , supplier\_id from dwh.F\_product\_sales

Group by supplier\_id



### Get Quantity sold for each subcategory

Select count(product\_sales\_sur\_key) as number\_of\_sold , subcategory from dwh.F\_product\_sales

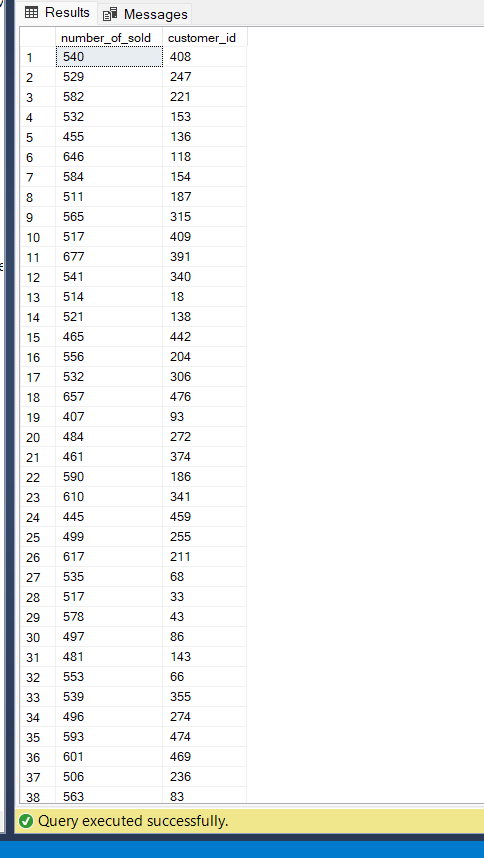
Group by subcategory



### Get Quantity sold for each customer

Select count(product\_sales\_sur\_key) as number\_of\_sold , customer\_id from dwh.F\_product\_sales

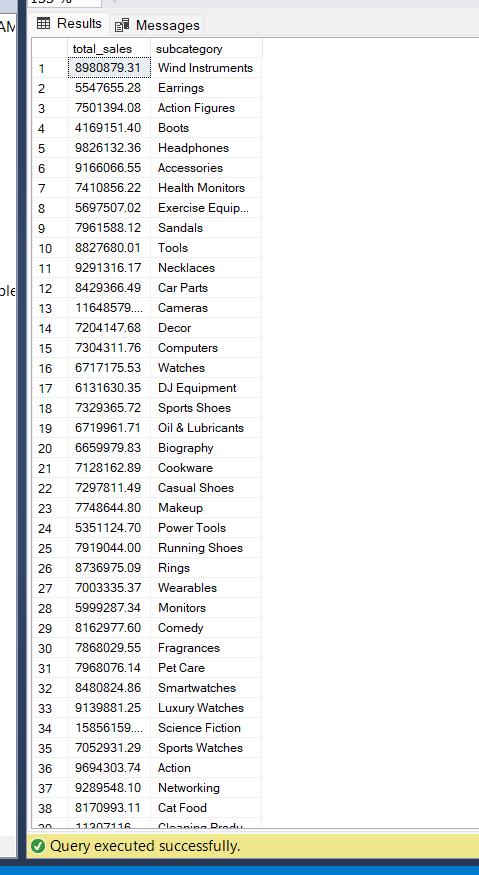
Group by customer\_id



Get total sales for each subcategory

Select sum(actual\_cost) as total\_sales , subcategory from dwh.F\_product\_sales

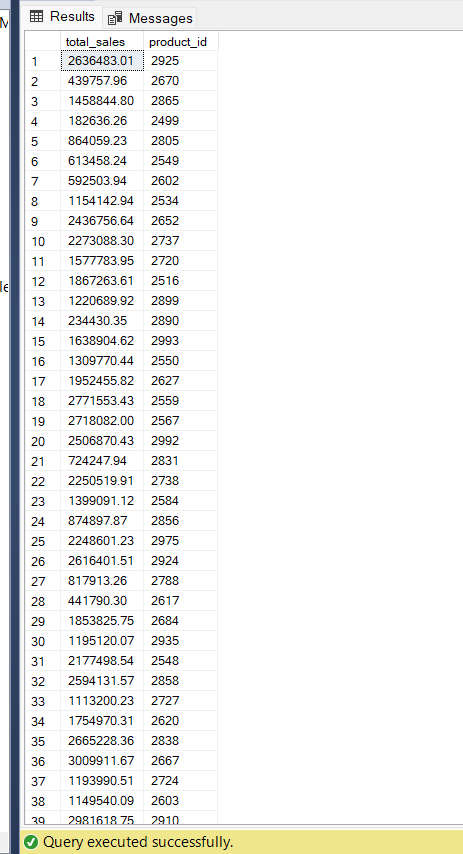
Group by subcategory



### Get total sales for each product

Select sum(actual\_cost) as total\_sales , product\_id from dwh.F\_product\_sales

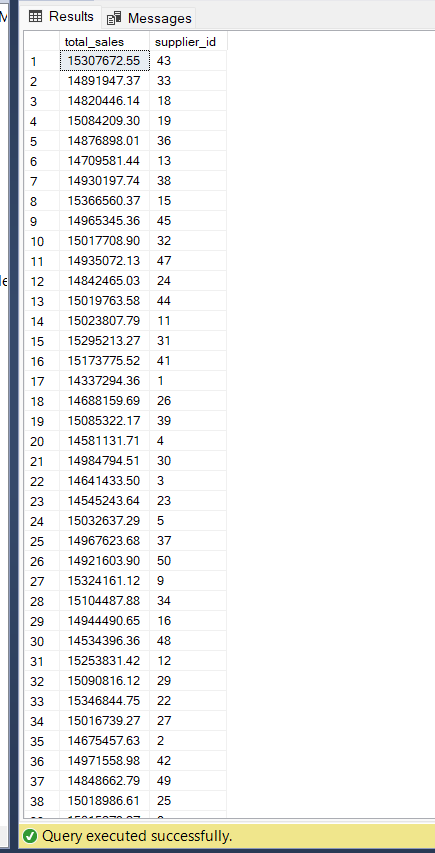
Group by product\_id



### Get total sales for each supplier

Select sum(actual\_cost) as total\_sales , supplier\_id from dwh.F\_product\_sales

Group by supplier\_id



6- Screenshots of the deployed packages in SSIS with their schedule**.**

