Sri Lanka Institute Of Information Technology

Predicting Coupon Redemption

Data Warehouse

Data Warehouse & Business Intelligence

Assignment 1

Kumaran.S

IT20211400

1.Data Preparation

This dataset gives us an idea about the Predicting Coupon Redemption.

Data Set: Predicting Coupon Redemption

Site: Kaggle

Source Link:
https://www.kaggle.com/datasets/vasudeva009/predicting-coupon-redemption-pca

Modifications were done accordingly to the data set derived from the source This data set reflects combinations between customer transactions and promotion campaigns. Customer-specific details involved in transactions, Items customers are keen to purchase, and customer participation in promotion campaigns are some of the key details included in the data set.
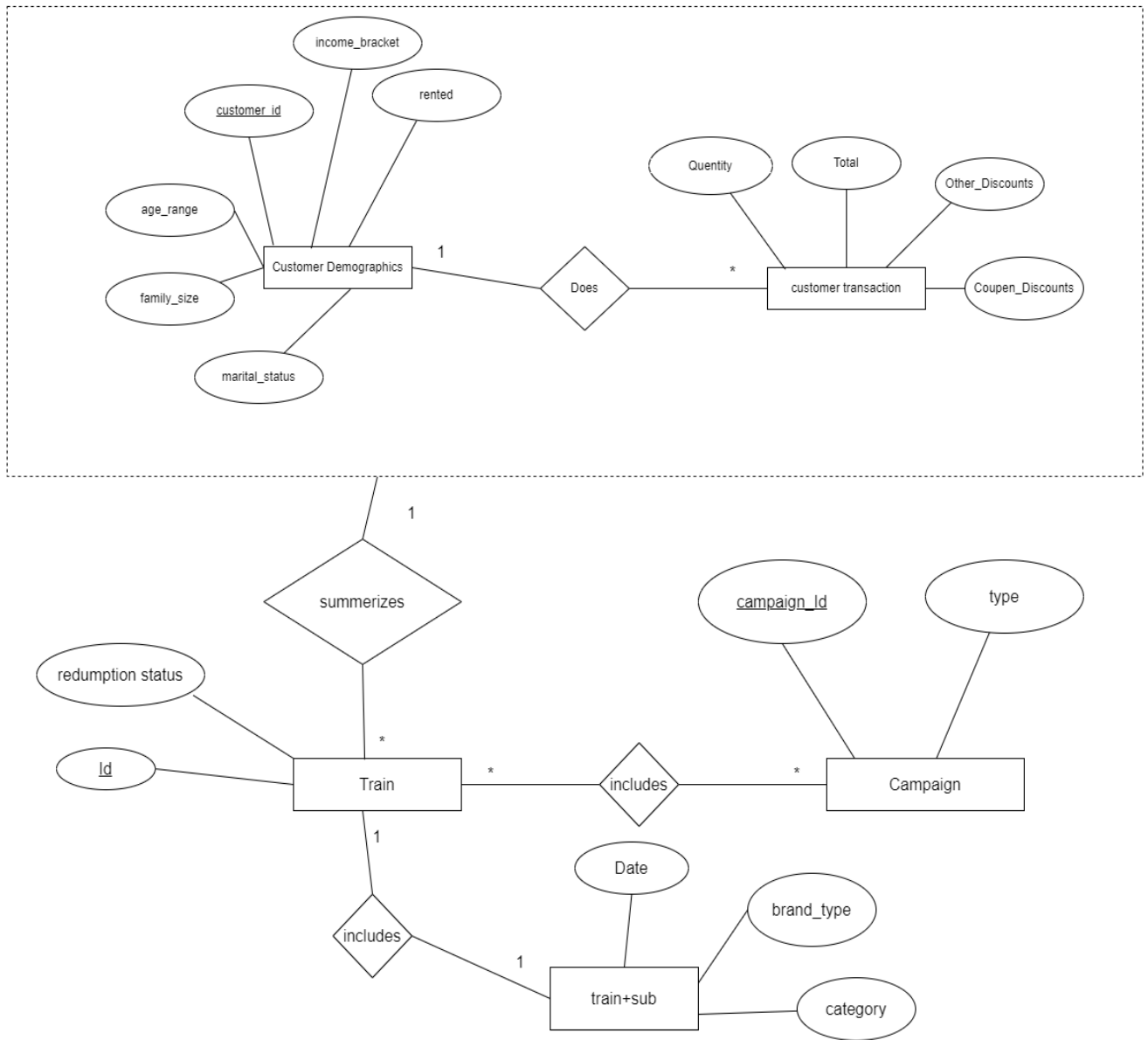
ER Diagram

Figure 1: ER Diagram

2. **Preparation of Data Sources**

There are 5 source tables in two formats (CSV & Text formats)

campaign_data.csv	5/14/2022 1:21 PM	Microsoft Excel Co...	1 KB
customer_demographics.csv	5/14/2022 11:53 AM	Microsoft Excel Co...	50 KB
customer_transaction_data.csv	5/14/2022 11:51 AM	Microsoft Excel Co...	39 KB
Train.txt	5/14/2022 2:40 PM	Text Document	1,227 KB
TrainSub.csv	5/15/2022 11:59 AM	Microsoft Excel Co...	3,041 KB

Files were imported to the SQL source database:

• Campaign_data

• Customer_demographics

• Customer_transaction_data

• TrainSub

Description of the data set
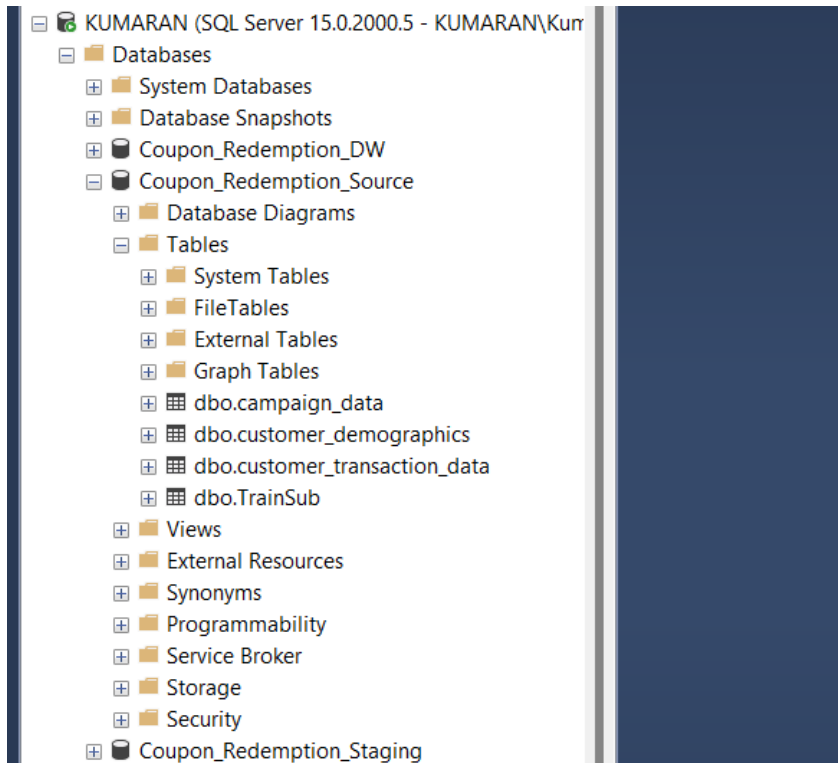
Design of Data_Source

- campaign table design
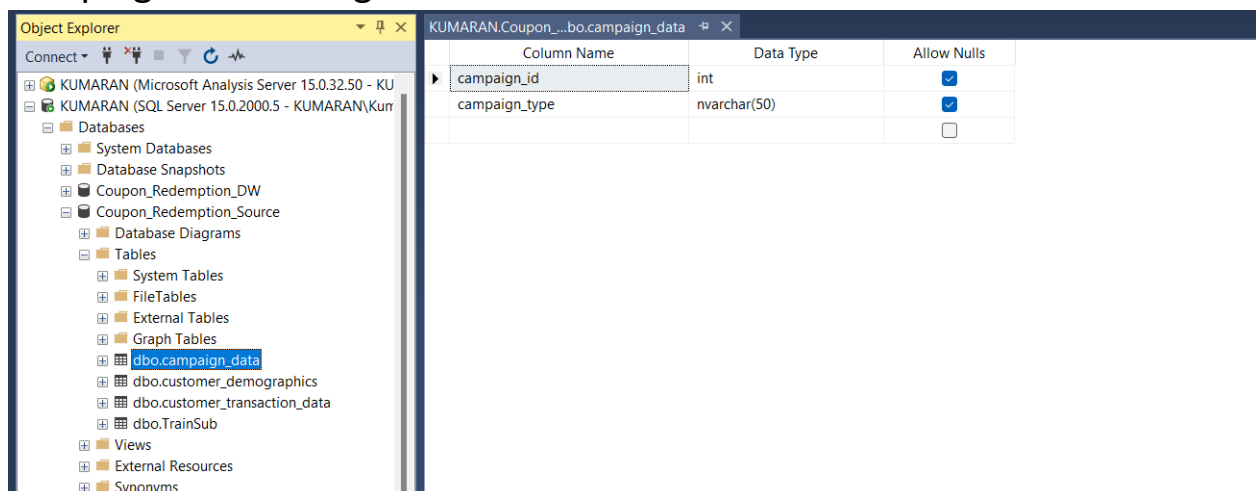
- # Customer demographics table design

- # Customer transaction table design

- # Train sub table design

Solution Architecture

The first step is staging the source data, after staging tables are created, there are:

- StgCampaignData
- StgCustomerDemographics
- StgCustomerTransactionData
- StgTrain
- StgTrainSub

Next staged tables are profiled and aggregated. Then data is transformed and loaded. Finally completing the stages and all data is tested and validated then the Data Warehouse is created.

Data warehouse design and development
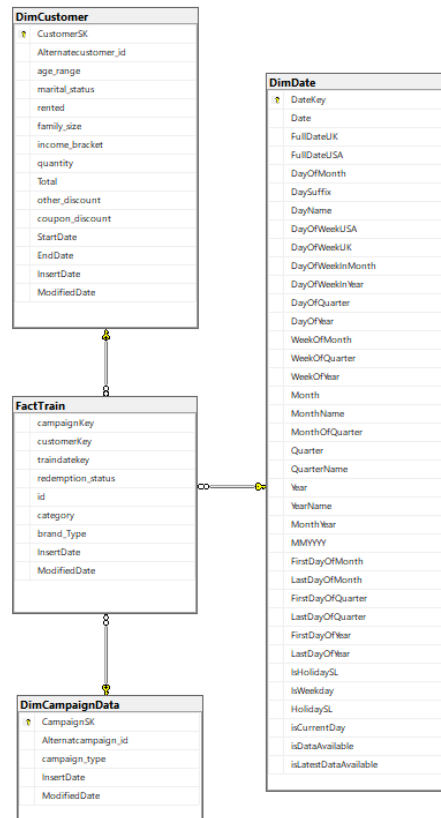
The snowflake schema is used to design the Datawarehouse design. There is one fact table as transactions and 3 dimensions.

Assumptions.

Customer Details were considered as a slowly changing dimension.

ETL development process

Steps with Screenshots will be displayed in this category.

- ## Step 01: Setting up the Environment

|			
campaign_data.csv	5/14/2022 1:21 PM	Microsoft Excel Co...	1 KB
customer_demographics.csv	5/14/2022 11:53 AM	Microsoft Excel Co...	50 KB
customer_transaction_data.csv	5/14/2022 11:51 AM	Microsoft Excel Co...	39 KB
Train.txt	5/14/2022 2:40 PM	Text Document	1,227 KB
TrainSub.csv	5/15/2022 11:59 AM	Microsoft Excel Co...	3,041 KB

Text and Excel File
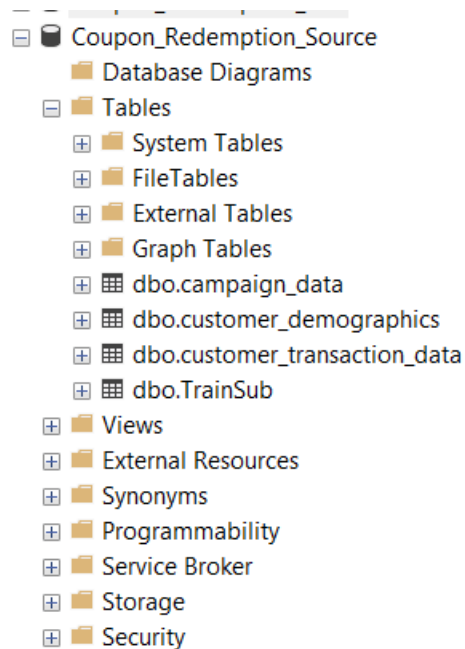
SourceDB in sql

- Coupon_Redemption_Staging
  - Database Diagrams
  - Tables
    - System Tables
    - FileTables
    - External Tables
    - Graph Tables
    - dbo.StgCampaignData
    - dbo.StgCustomerDemographics
    - dbo.StgCustomerTransactionData
    - dbo.StgTrain
    - dbo.StgTrainSub
  - Views
  - External Resources
  - Synonyms
  - Programmability
  - Service Broker
  - Storage
  - Security

Staging

- Coupon_Redemption_DW
  - Database Diagrams
  - Tables
    - System Tables
    - FileTables
    - External Tables
    - Graph Tables
    - dbo.DimCampaignData
    - dbo.DimCustomer
    - dbo.DimDate
    - dbo.FactTrain
  - Views
  - External Resources
  - Synonyms
  - Programmability
  - Service Broker
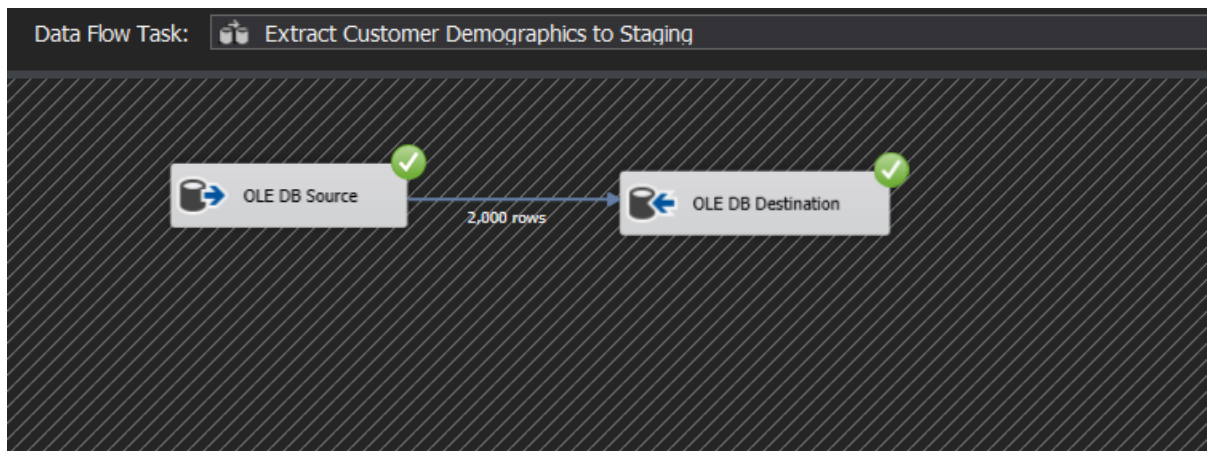 - Storage
  - Security

Data Warehouse

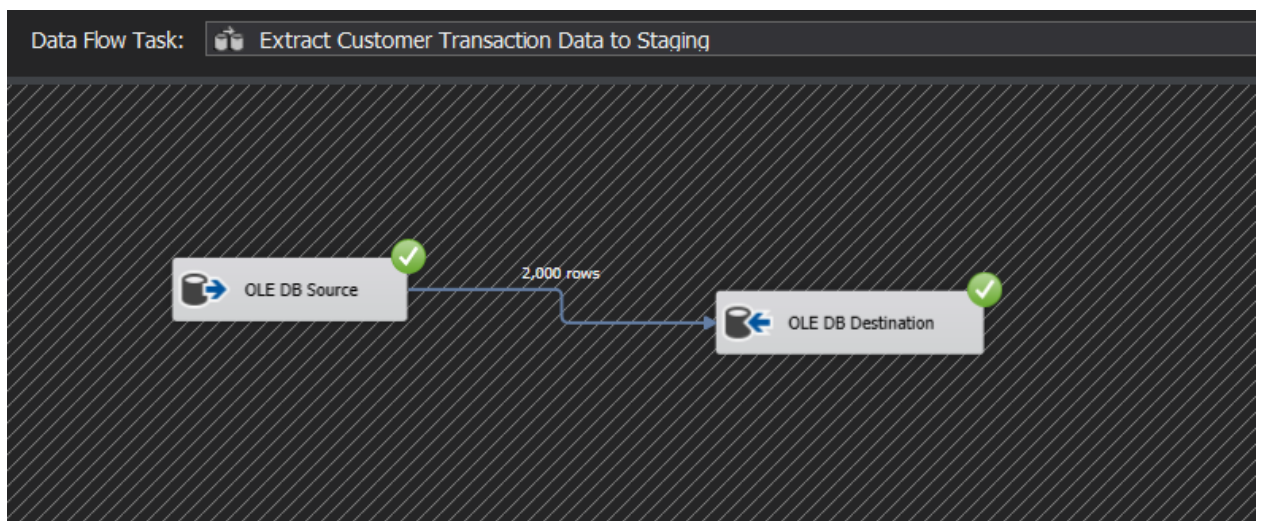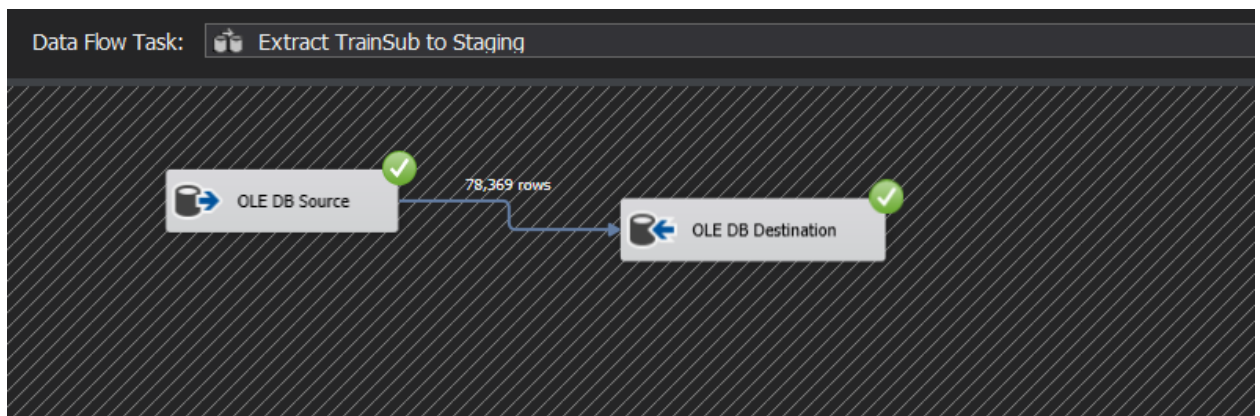- ## Step 02: Data Extracting from source to staging tables
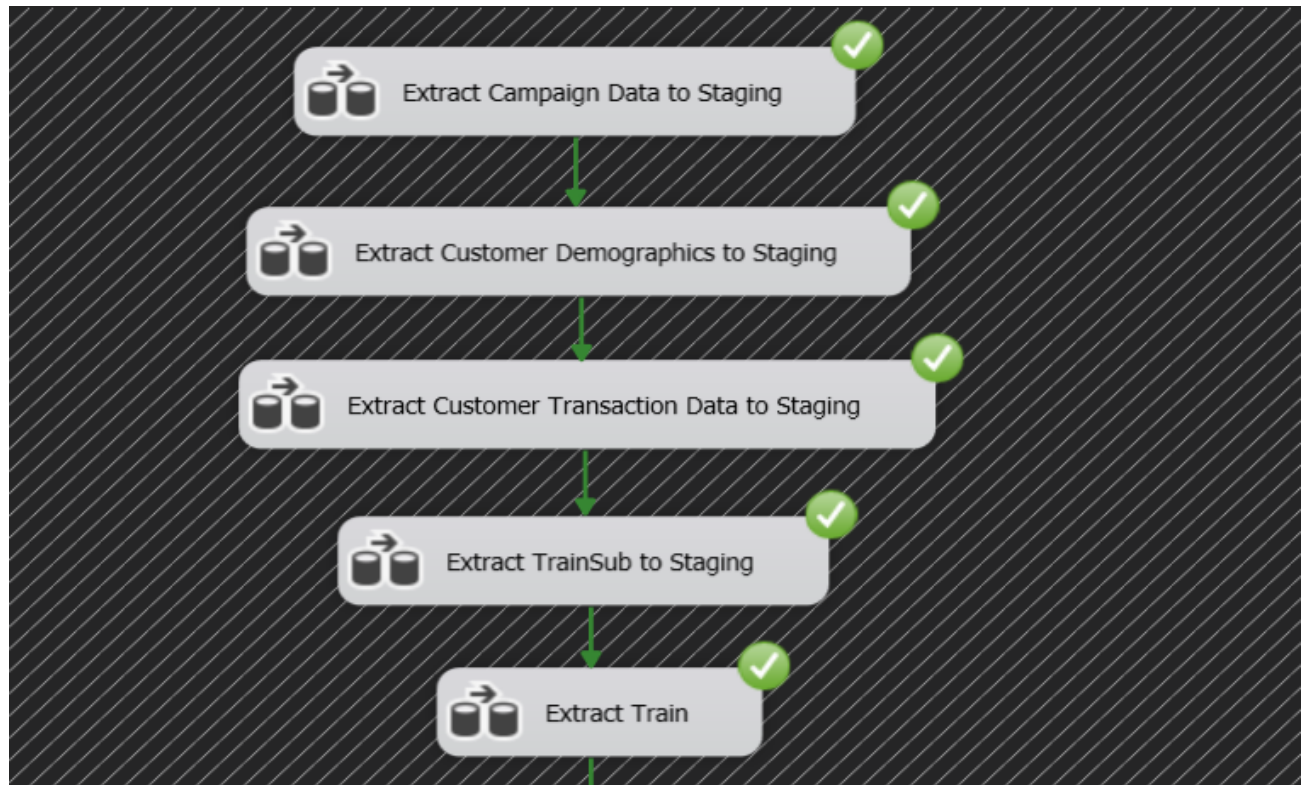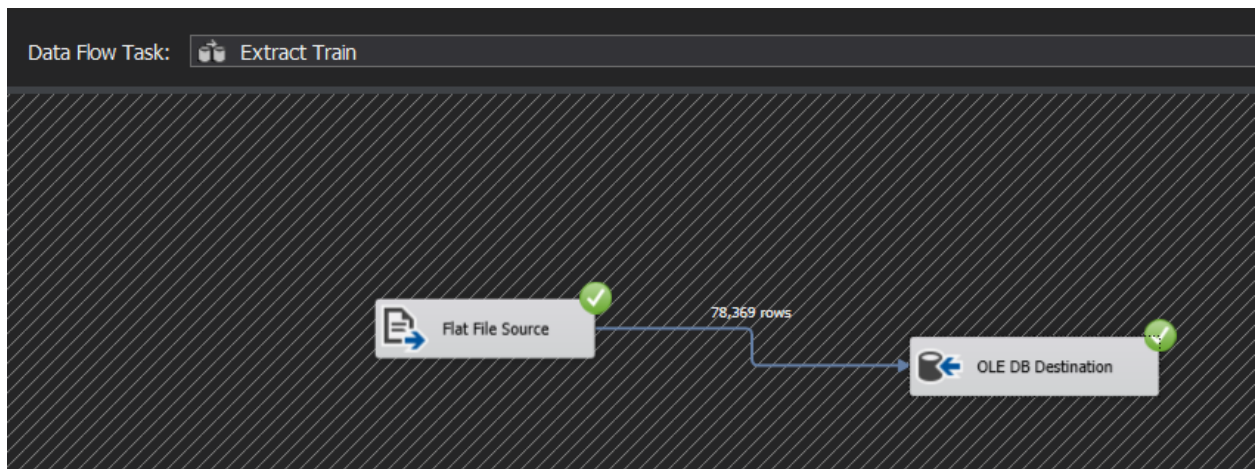

Extracting data
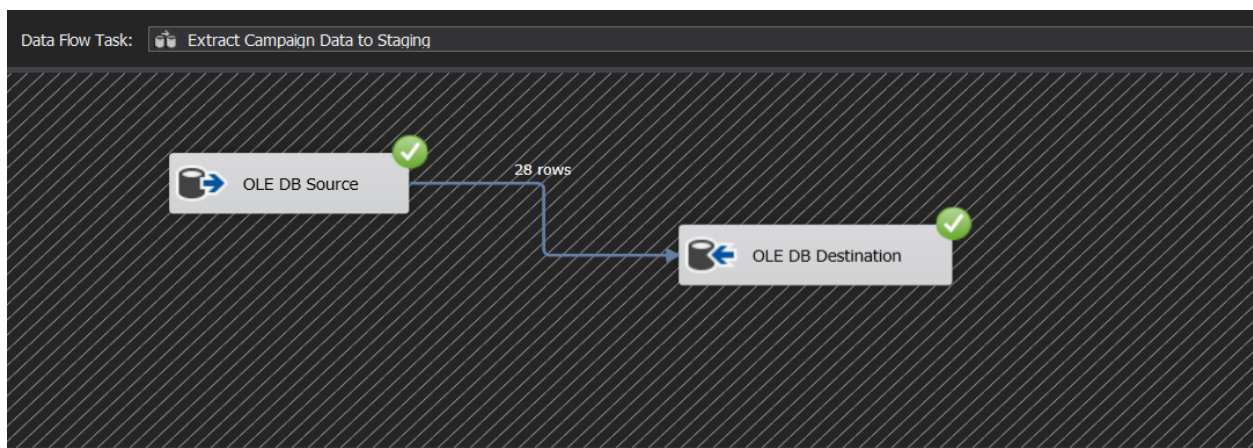
Extracting Campaign data
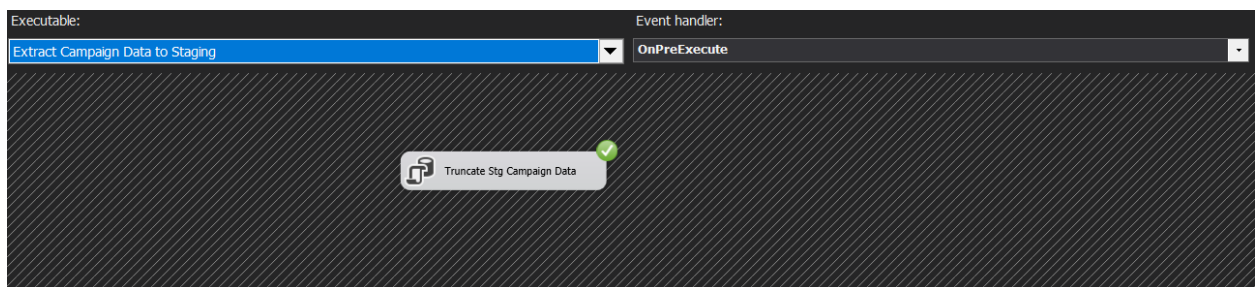
Extracting customer demographics

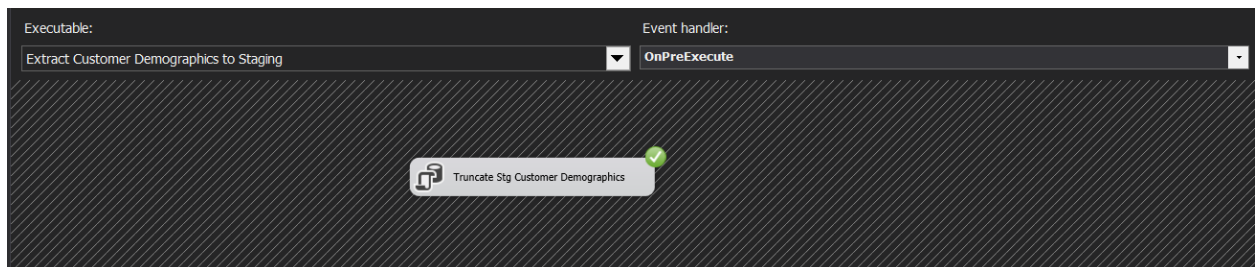Extracting customer transaction data
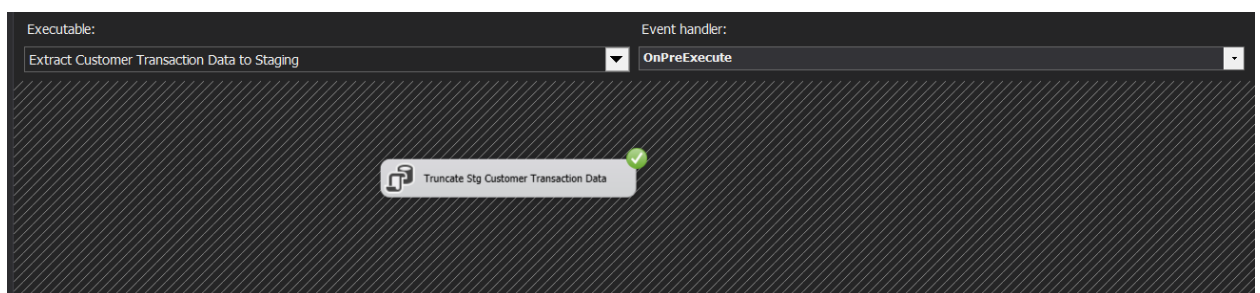

Extracting Train sub data
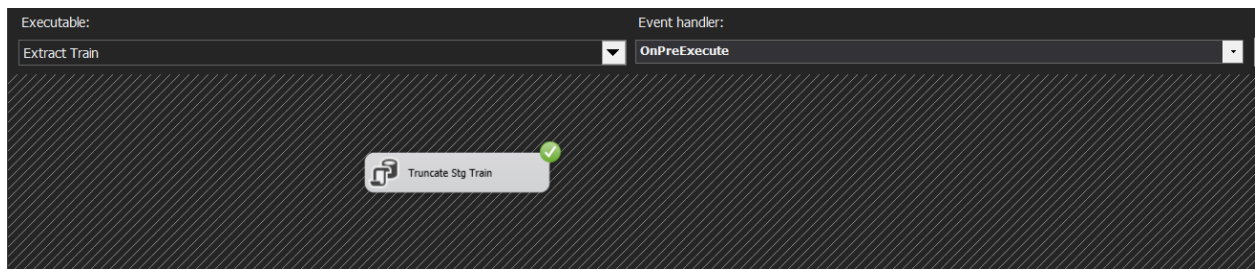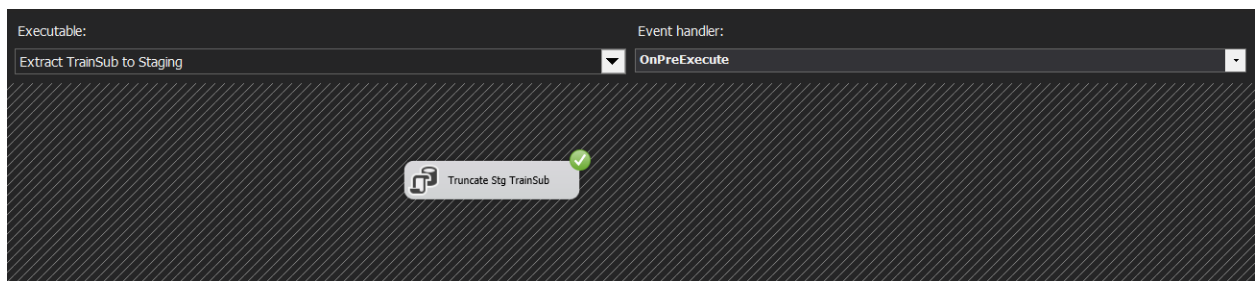
Extract train data

Truncate campaign data

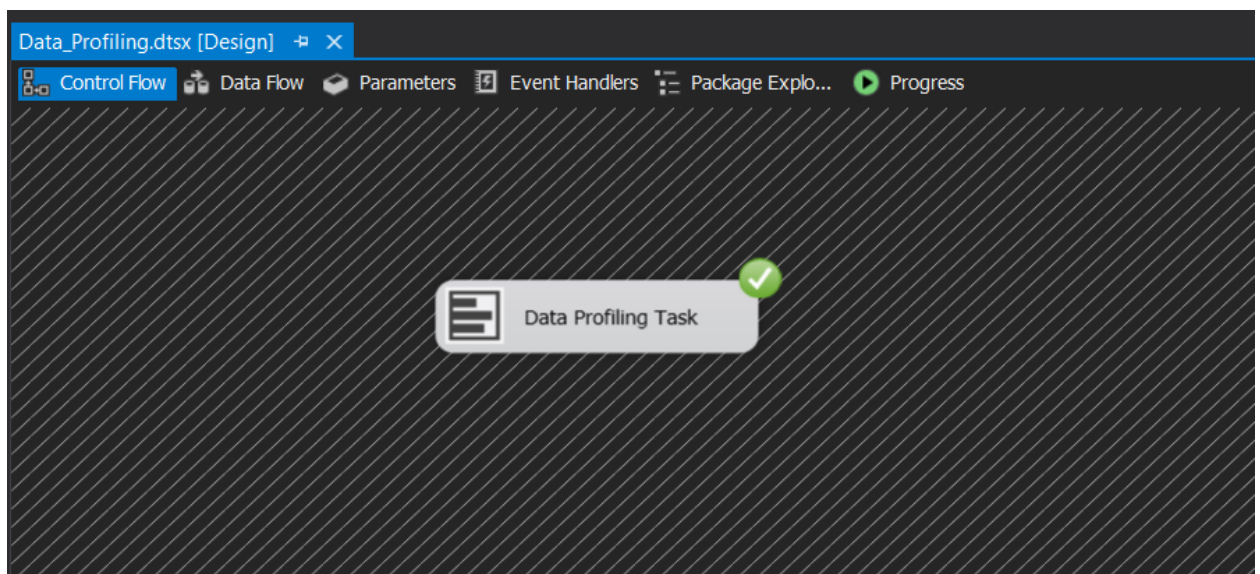Truncate customer data

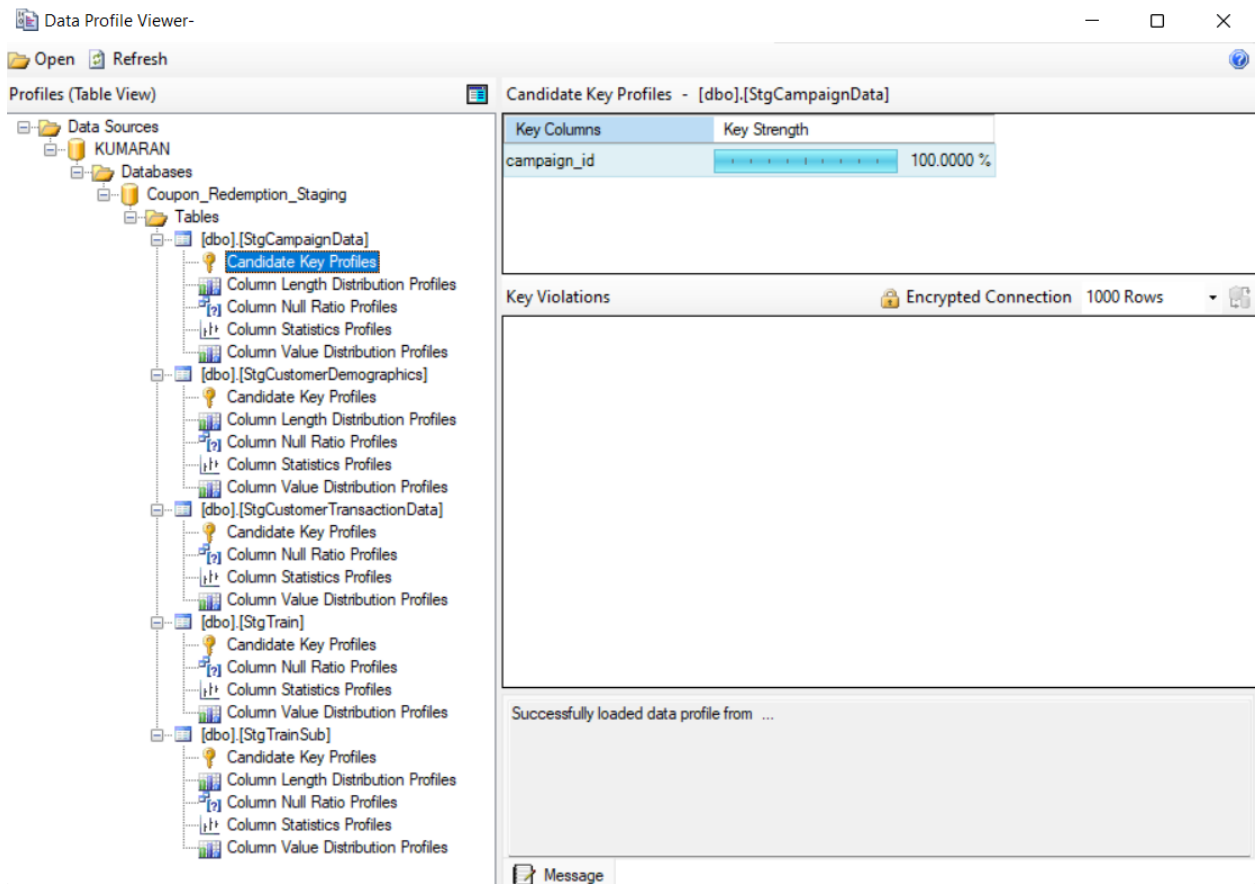
Truncate customer Transaction data

Truncate Train data
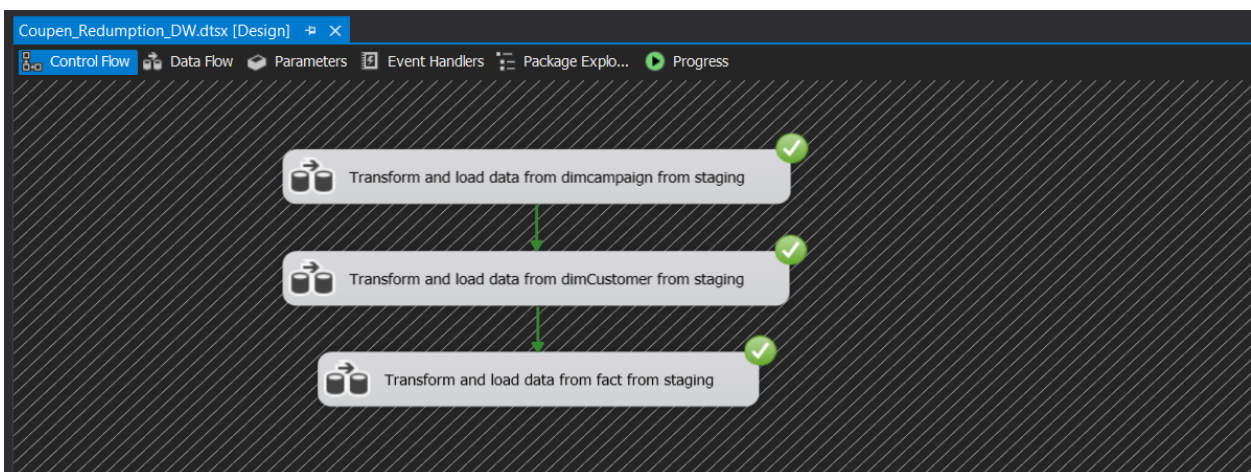

Truncate Train sub data

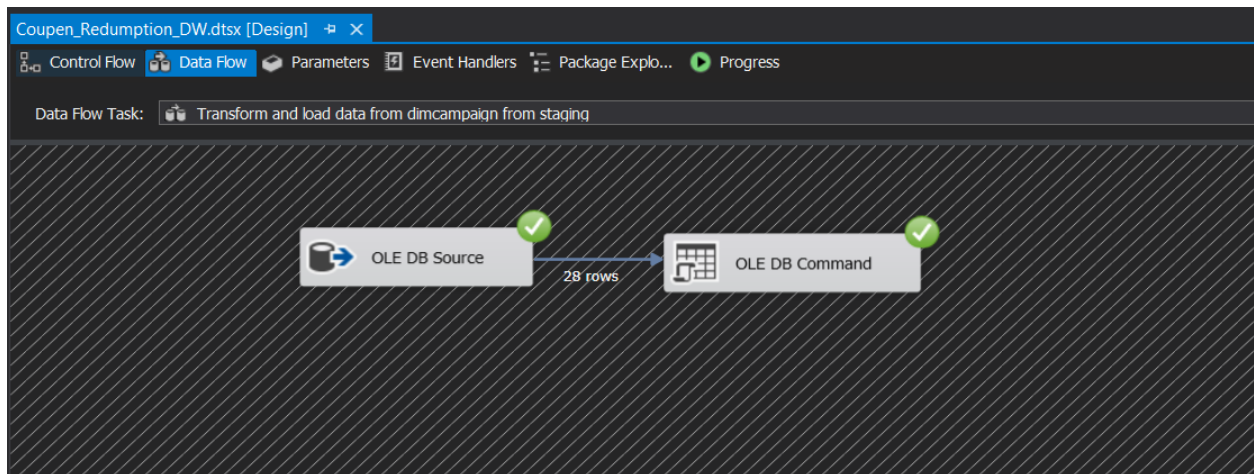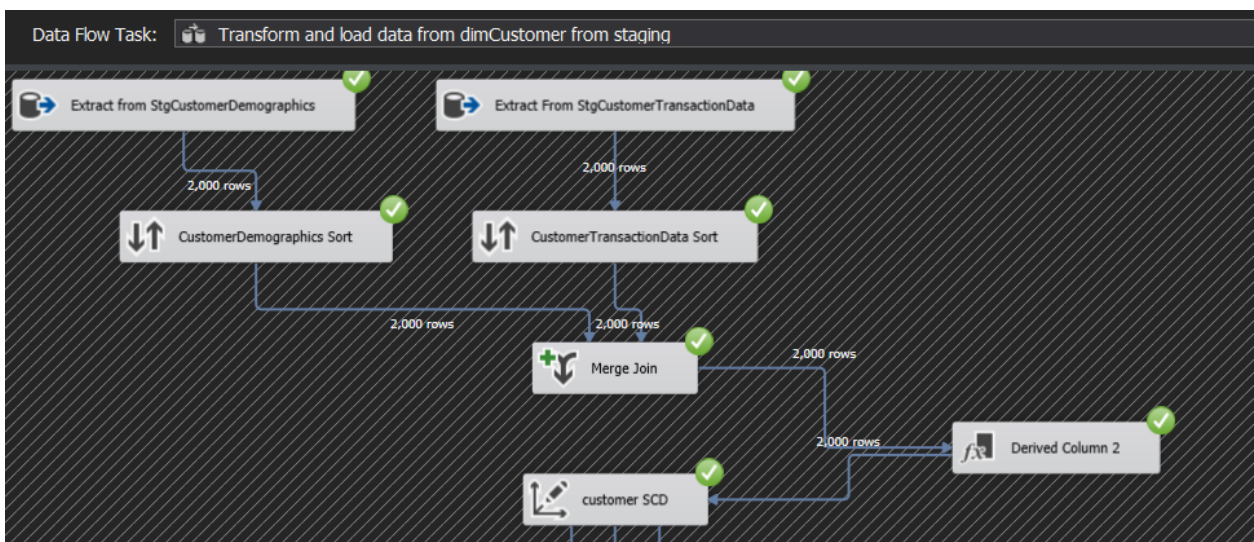- ## Step 03: Data Profiling

Data profiling

profile viewer

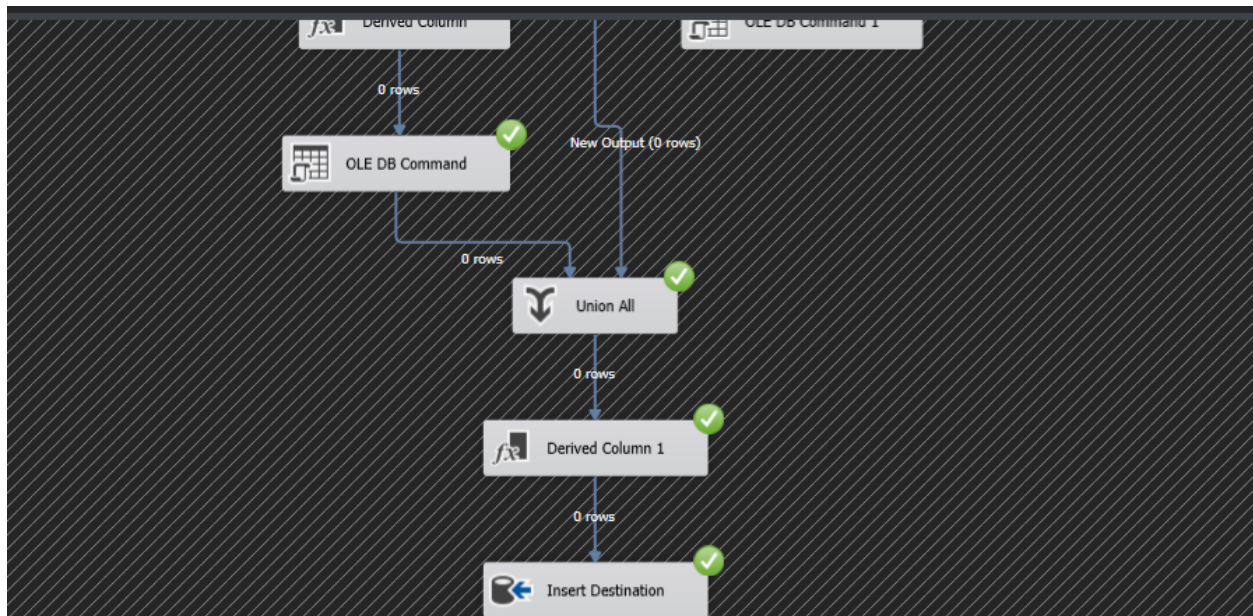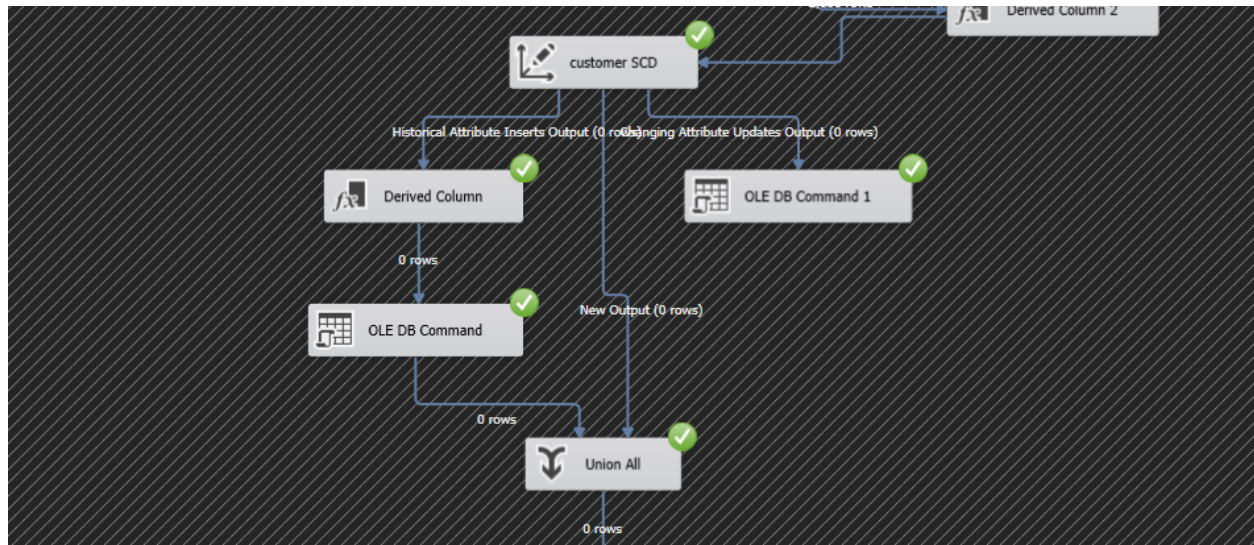- ## Step 04: Data Transformation from staging to warehouse
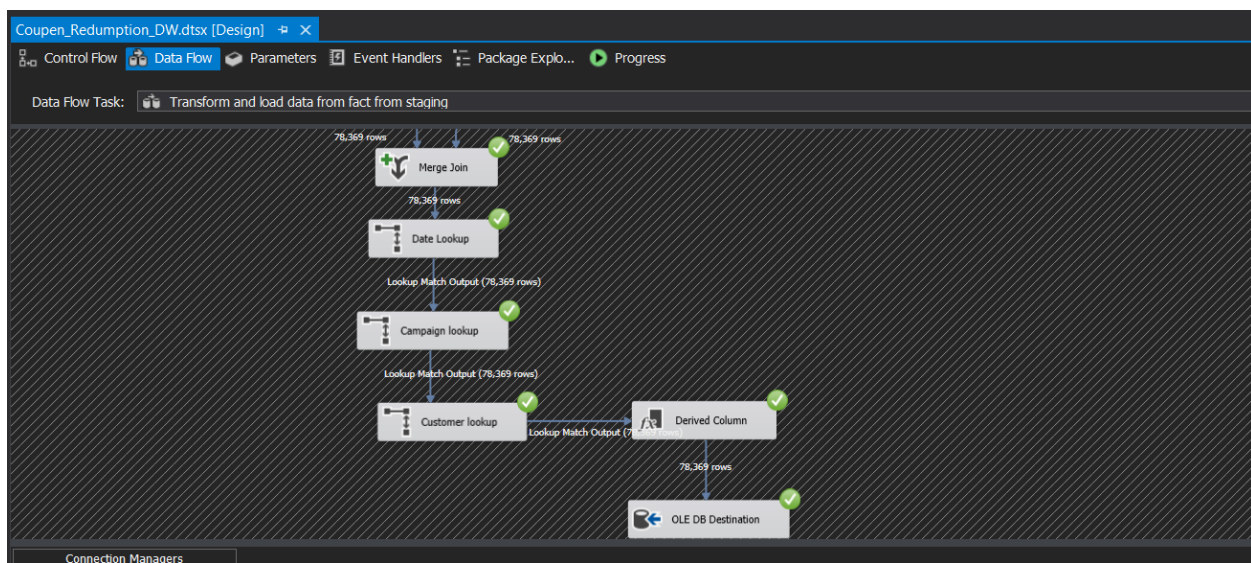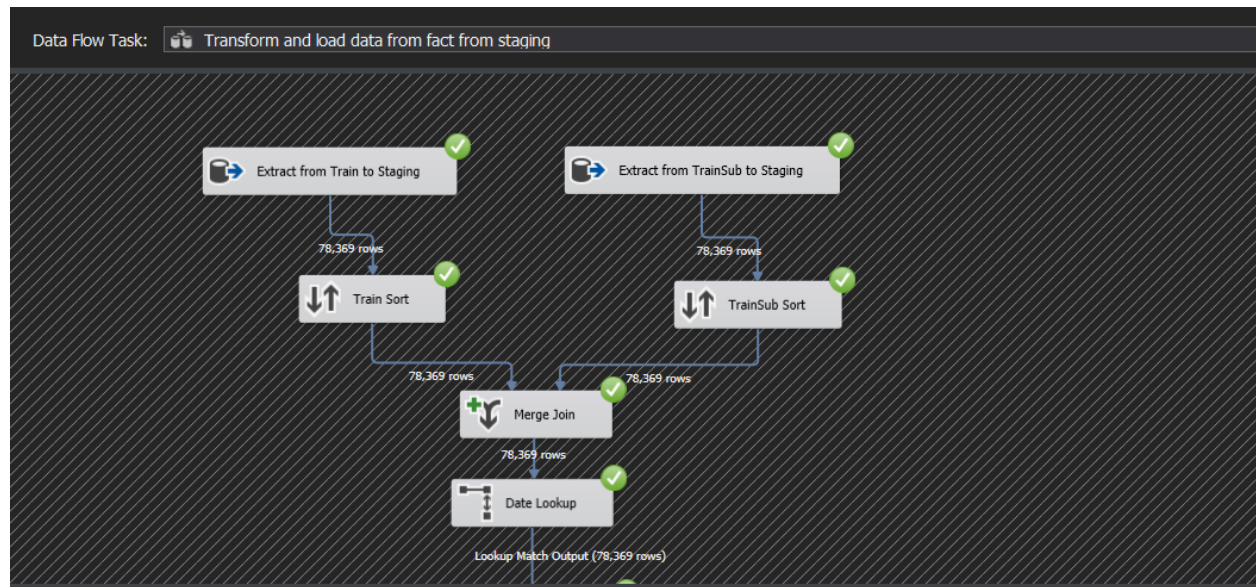

Transforming and loading

Transform and load dim campaign

Transform and load dim customer

Transform from fact train

Procedures

```sql
CREATE PROCEDURE dbo.UpdateDimCampaignData
@campaign_id int,
@campaign_type nvarchar(50)


AS
BEGIN
if not exists (select CampaignSK
from dbo.DimCampaignData
where Alternatcampaign_id = @campaign_id)
BEGIN
insert into dbo.DimCampaignData
(Alternatcampaign_id, campaign_type,InsertDate, ModifiedDate)
values
(@campaign_id, @campaign_type, GETDATE(), GETDATE())
END;
if exists (select CampaignSK
from dbo.DimCampaignData
where Alternatcampaign_id = @campaign_id)
BEGIN
update DimCampaignData
set campaign_type = @campaign_type,
ModifiedDate = GETDATE()
where Alternatcampaign_id = @campaign_id
END;
END;
```

Dim campaign procedure

Ware house tables with values

Dim campaign table

Dim customer table

Dim date table

	campaignKey	customerKey	traindatekey	redemption_status	id	category	brand_Type	InsertDate	ModifiedDate
1	11	1053	20120102	0	1	Grocery	Established	2022-05-15 13:09:40.180	2022-05-15 13:09:40.180
2	11	48	20120102	0	2	Miscellaneous	Established	2022-05-15 13:09:40.180	2022-05-15 13:09:40.180
3	15	205	20120102	0	6	Bakery	Local	2022-05-15 13:09:40.180	2022-05-15 13:09:40.180
4	11	1050	20120102	0	7	Grocery	Local	2022-05-15 13:09:40.180	2022-05-15 13:09:40.180
5	16	1489	20120102	0	9	Grocery	Local	2022-05-15 13:09:40.180	2022-05-15 13:09:40.180
6	12	793	20120102	0	11	Grocery	Local	2022-05-15 13:09:40.180	2022-05-15 13:09:40.180
7	15	590	20120102	0	14	Pharmaceutical	Local	2022-05-15 13:09:40.180	2022-05-15 13:09:40.180
8	25	368	20120102	0	15	Bakery	Local	2022-05-15 13:09:40.180	2022-05-15 13:09:40.180
9	24	523	20120102	0	17	Grocery	Local	2022-05-15 13:09:40.180	2022-05-15 13:09:40.180
10	23	679	20120102	0	19	Grocery	Local	2022-05-15 13:09:40.180	2022-05-15 13:09:40.180
11	15	108	20120102	0	20	Pharmaceutical	Established	2022-05-15 13:09:40.180	2022-05-15 13:09:40.180
12	11	89	20120102	0	21	Grocery	Local	2022-05-15 13:09:40.180	2022-05-15 13:09:40.180
13	15	712	20120102	0	22	Grocery	Established	2022-05-15 13:09:40.180	2022-05-15 13:09:40.180
14	11	1067	20120102	0	23	Grocery	Local	2022-05-15 13:09:40.180	2022-05-15 13:09:40.180
15	20	761	20120102	0	24	Grocery	Local	2022-05-15 13:09:40.180	2022-05-15 13:09:40.180
16	11	248	20120102	0	25	Grocery	Local	2022-05-15 13:09:40.180	2022-05-15 13:09:40.180
17	12	1464	20120102	0	26	Grocery	Local	2022-05-15 13:09:40.180	2022-05-15 13:09:40.180

Fact train table