Data Warehousing & Business Intelligence(IT)
3rd Year, 1st Semester

Assignment 1

Submitted to

Sri Lanka Institute of Information Technology

IT20035358

K.N.N.D.S Jayarathna.

Weekday Batch

STEP 01: Data set selection

I have selected the predicting Coupon Redemption data set as the data set. It consists of seven csv files. Among them I chose five csv files to implement my solution.

According to the assignment guidelines the data set was initiated with sufficient data. In my data set which I have selected is a transactional data type and it is used to predict whether customer will Redeem coupons such as discounts coupons or other types of coupons released in various marketing campaigns.

Customer, Campaign, Items and coupon redemption status and customer transaction are all detailed in the data set.

In predicting coupon redemption data set, there are also enough records and attributes to build complete data warehouse and perform ETL process. I can receive two sources of data from the data set that I have selected (A database and a text file) as my source data. And the data set is sufficient to build data warehouse and I was able to do ETL process with this data set.

As well as I can correctly identify hierarchies, dimensions, and aggregates in the data set. I also identified that with data set I will be able to generate appropriate reports.

- Data set link: -

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

❖ Data Description

- Train.CSV

Variable	Definition
id	Unique id for coupon customer impression
campaign_id	Unique id for a discount campaign
coupon_id	Unique id for a discount coupon
customer_id	Unique id for a customer
Redemption_status	(target) (0 - Coupon not redeemed, 1 - Coupon redeemed)

- Campaign_data.csv

Variable	Definition
campaign_id	Unique id for a discount campaign
campaign_type	Anonymised Campaign Type (X/Y
start_date	Campaign Start Date
end_date	Campaign End Date

- customer_demographics.csv

Variable	Definition
customer_id	Unique id for a customer
age_range	Age range of customer family in years
marital_status	Married/Single
rented	0 - not rented accommodation, 1 - rented accommodation
family_size	Number of family members
noOfChildren	Number of children in the family
income_bracket	Label Encoded Income Bracket (Higher income corresponds to higher number

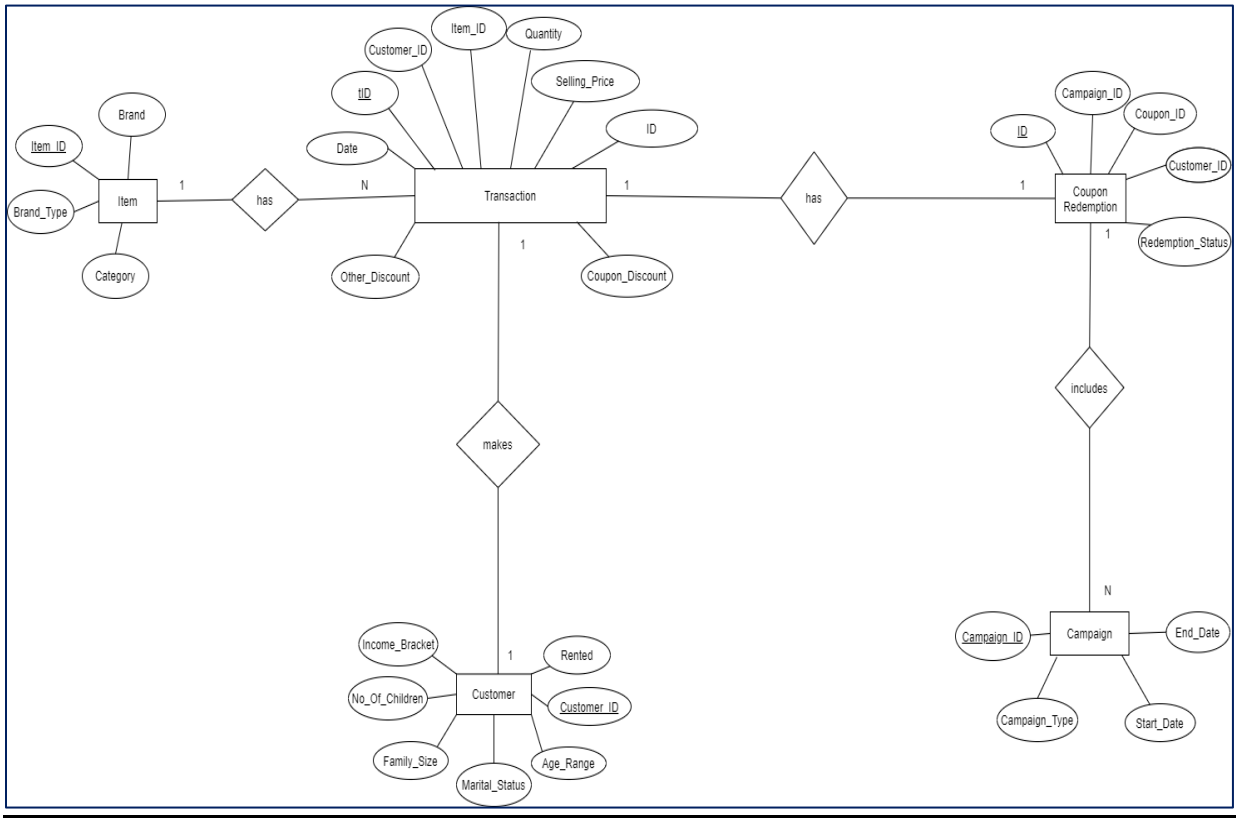
- customertransaction.csv

Variable	Definition
date	Date of Transaction
customer_id	Unique id for a customer
item_id	Unique id for item
quantity	quantity of item bought
selling_price	Sales value of the transaction
other_discount	Discount from other sources such as manufacturer coupon/loyalty card
coupon_discount	Discount availed from retailer coupon

- Item.csv

Variable	Definition
item_id	Unique id for item
brand	Unique id for item brand
brand_type	Brand Type (local/Established)
Category	Item Category

❖ ER Diagram

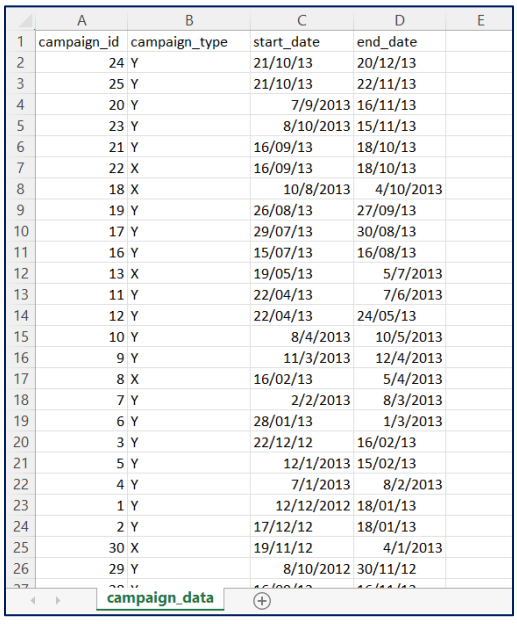


STEP 02: Preparation of Data Sources

There were seven csv files available in the data set. They are,

- Item data csv
- Customer Demographics csv
- Customer Transaction Data csv
- Campaign Data csv
- Train Data csv
- Coupon Item Mapping csv
- Test csv

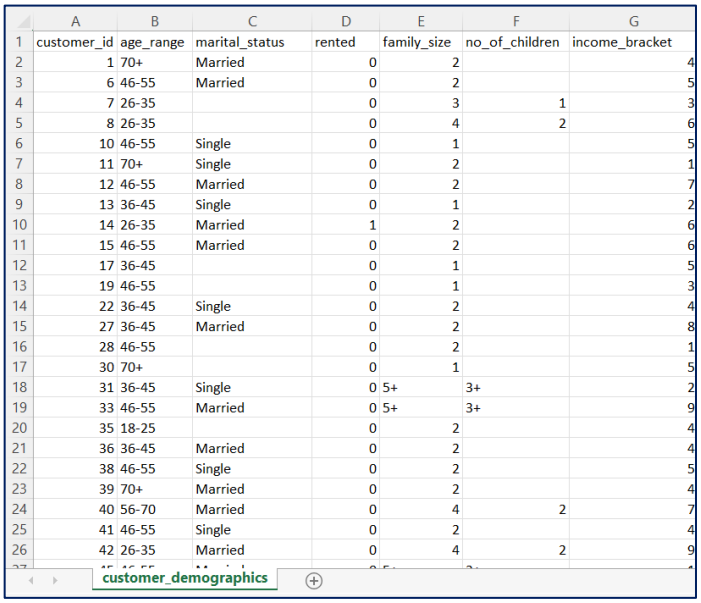
- Among these csv files Item Mapping and Test csv files were not taken to create tables.
- And I decided to convert Campaign data csv into text file format.
- Then I imported all other four csv files into my CouponRedemption_SourceDB.
- After I imported my csv files into sourceDB, I created Data warehouse named CouponRedemption_DW and created my dimension tables and fact tables inside the data warehouse.
- Here I converted train csv file data into CouponRedemption table.



	A	B	C	D	E
1	campaign_id	campaign_type	start_date	end_date	
2	24	Y	21/10/13	20/12/13	
3	25	Y	21/10/13	22/11/13	
4	20	Y	7/9/2013	16/11/13	
5	23	Y	8/10/2013	15/11/13	
6	21	Y	16/09/13	18/10/13	
7	22	X	16/09/13	18/10/13	
8	18	X	10/8/2013	4/10/2013	
9	19	Y	26/08/13	27/09/13	
10	17	Y	29/07/13	30/08/13	
11	16	Y	15/07/13	16/08/13	
12	13	X	19/05/13	5/7/2013	
13	11	Y	22/04/13	7/6/2013	
14	12	Y	22/04/13	24/05/13	
15	10	Y	8/4/2013	10/5/2013	
16	9	Y	11/3/2013	12/4/2013	
17	8	X	16/02/13	5/4/2013	
18	7	Y	2/2/2013	8/3/2013	
19	6	Y	28/01/13	1/3/2013	
20	3	Y	22/12/12	16/02/13	
21	5	Y	12/1/2013	15/02/13	
22	4	Y	7/1/2013	8/2/2013	
23	1	Y	12/12/2012	18/01/13	
24	2	Y	17/12/12	18/01/13	
25	30	X	19/11/12	4/1/2013	
26	29	Y	8/10/2012	30/11/12	

campaign_data

campaign _data.csv



	A	B	C	D	E	F	G
1	customer_id	age_range	marital_status	rented	family_size	no_of_children	income_bracket
2	1	70+	Married	0	2		4
3	6	46-55	Married	0	2		5
4	7	26-35		0	3	1	3
5	8	26-35		0	4	2	6
6	10	46-55	Single	0	1		5
7	11	70+	Single	0	2		1
8	12	46-55	Married	0	2		7
9	13	36-45	Single	0	1		2
10	14	26-35	Married	1	2		6
11	15	46-55	Married	0	2		6
12	17	36-45		0	1		5
13	19	46-55		0	1		3
14	22	36-45	Single	0	2		4
15	27	36-45	Married	0	2		8
16	28	46-55		0	2		1
17	30	70+		0	1		5
18	31	36-45	Single	0	5+	3+	2
19	33	46-55	Married	0	5+	3+	9
20	35	18-25		0	2		4
21	36	36-45	Married	0	2		4
22	38	46-55	Single	0	2		5
23	39	70+	Married	0	2		4
24	40	56-70	Married	0	4	2	7
25	41	46-55	Single	0	2		4
26	42	26-35	Married	0	4	2	9

customer_demographics

customer_demographics.csv

	A	B	C	D	E	F	G	H	I
1	date	customer_	item_id	quantity	selling_pri	other_disc	coupon_di	tID	id
2	1/2/2012	1501	26830	1	35.26	-10.69	0	18538	1
3	1/2/2012	1501	54253	1	53.43	-13.89	0	13594	2
4	1/2/2012	1501	31962	1	106.5	-14.25	0	9692	6
5	1/2/2012	1501	33647	1	67.32	0	0	16600	7
6	1/2/2012	1501	48199	1	71.24	-28.14	0	10272	9
7	1/2/2012	1501	57397	1	71.24	-28.14	0	21243	11
8	1/2/2012	857	12424	1	106.5	-14.25	0	15818	14
9	1/2/2012	857	14930	1	110.07	0	0	7282	15
10	1/2/2012	857	16657	1	89.05	-35.26	0	16283	17
11	1/2/2012	67	10537	3	32.06	0	0	21641	19
12	1/2/2012	67	19560	2	106.86	-2.85	0	12588	20
13	1/2/2012	67	21653	1	44.52	-12.11	0	24939	21
14	1/2/2012	67	32083	1	8.9	0	0	19178	22
15	1/2/2012	751	10423	1	81.57	0	0	5309	23
16	1/2/2012	751	10565	1	71.24	-35.26	0	17671	24
17	1/2/2012	751	17035	1	127.88	0	0	7035	25
18	1/2/2012	751	27801	2	71.24	-27.78	0	26329	26
19	1/2/2012	751	30645	1	127.88	0	0	26679	27
20	1/2/2012	751	30832	1	213.36	-106.86	0	10260	28
21	1/2/2012	751	34047	1	53.43	0	0	20812	29
22	1/2/2012	751	34139	1	35.62	-10.33	0	23133	30
23	1/2/2012	751	45502	1	82.99	-80.5	0	26994	32
24	1/2/2012	751	46101	1	135	-14.25	0	9229	33
25	1/2/2012	751	51615	1	142.12	-35.62	0	17838	34
26	1/2/2012	679	4408	1	63.76	0	0	28097	35

Customer_transaction_data_new.csv

	A	B	C	D
1	item_id	brand	brand_type	category
2	1	1	Established	Grocery
3	2	1	Established	Miscellaneous
4	3	56	Local	Bakery
5	4	56	Local	Grocery
6	5	56	Local	Grocery
7	6	56	Local	Grocery
8	7	56	Local	Pharmaceutical
9	8	56	Local	Bakery
10	9	11	Local	Grocery
11	10	56	Local	Grocery
12	11	21	Established	Pharmaceutical
13	12	56	Local	Grocery
14	13	1	Established	Grocery
15	14	56	Local	Grocery
16	15	56	Local	Grocery
17	16	56	Local	Grocery
18	17	56	Local	Grocery
19	18	56	Local	Grocery
20	19	56	Local	Pharmaceutical
21	20	56	Local	Grocery
22	21	56	Local	Grocery
23	22	56	Local	Grocery
24	23	56	Local	Grocery
25	24	56	Local	Grocery
26	25	56	Local	Grocery

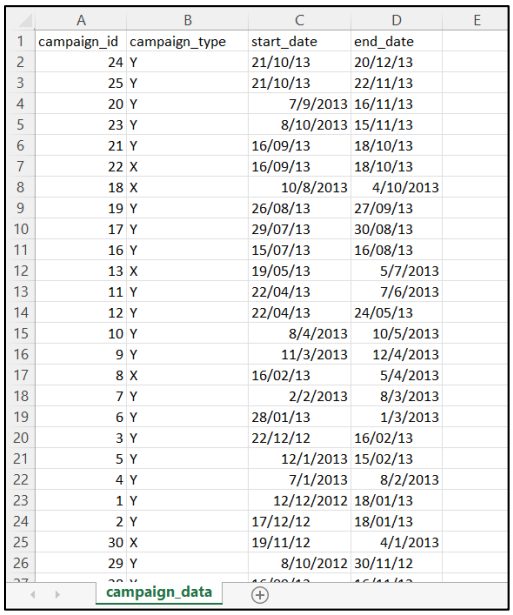
item_data

item_data.csv

	A	B	C	D	E
1	id	campaign_id	coupon_id	customer_id	redemption_status
2	1	13	27	1053	0
3	2	13	116	48	0
4	6	9	635	205	0
5	7	13	644	1050	0
6	9	8	1017	1489	0
7	11	11	795	793	0
8	14	9	444	590	0
9	15	29	538	368	0
10	17	30	857	523	0
11	19	2	559	679	0
12	20	9	575	108	0
13	21	13	1028	89	0
14	22	9	705	712	0
15	23	13	517	1067	0
16	24	5	893	761	0
17	25	13	796	248	0
18	26	11	506	1464	0
19	27	5	689	186	0
20	28	13	268	1152	0
21	29	8	8	695	0
22	30	8	674	886	0
23	32	30	320	1255	0
24	33	12	109	553	0
25	34	8	278	1398	0
26	35	13	1031	1031	0

train

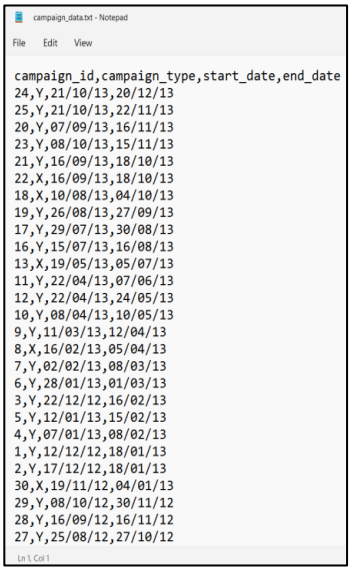
train.csv



	A	B	C	D	E
1	campaign_id	campaign_type	start_date	end_date	
2	24	Y	21/10/13	20/12/13	
3	25	Y	21/10/13	22/11/13	
4	20	Y	7/9/2013	16/11/13	
5	23	Y	8/10/2013	15/11/13	
6	21	Y	16/09/13	18/10/13	
7	22	X	16/09/13	18/10/13	
8	18	X	10/8/2013	4/10/2013	
9	19	Y	26/08/13	27/09/13	
10	17	Y	29/07/13	30/08/13	
11	16	Y	15/07/13	16/08/13	
12	13	X	19/05/13	5/7/2013	
13	11	Y	22/04/13	7/6/2013	
14	12	Y	22/04/13	24/05/13	
15	10	Y	8/4/2013	10/5/2013	
16	9	Y	11/3/2013	12/4/2013	
17	8	X	16/02/13	5/4/2013	
18	7	Y	2/2/2013	8/3/2013	
19	6	Y	28/01/13	1/3/2013	
20	3	Y	22/12/12	16/02/13	
21	5	Y	12/1/2013	15/02/13	
22	4	Y	7/1/2013	8/2/2013	
23	1	Y	12/12/2012	18/01/13	
24	2	Y	17/12/12	18/01/13	
25	30	X	19/11/12	4/1/2013	
26	29	Y	8/10/2012	30/11/12	

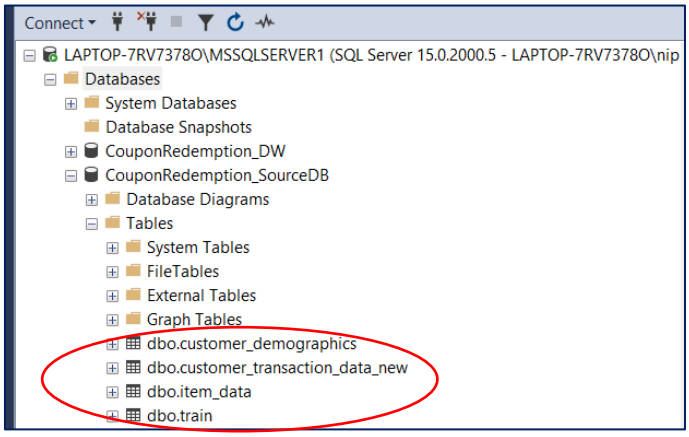
campaign_data

campaign_data.csv



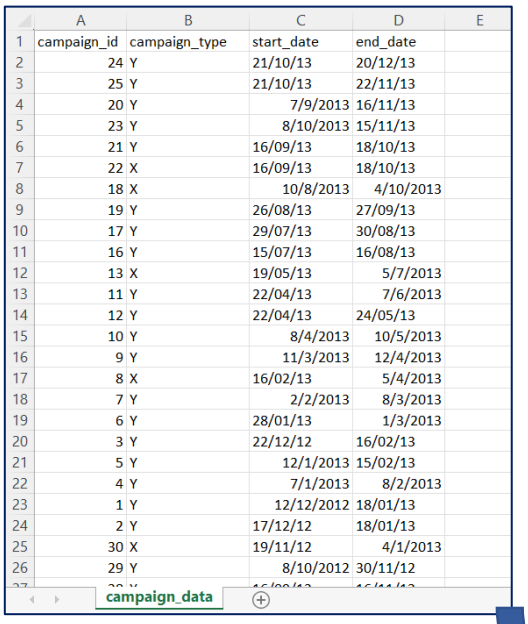
campaign_data.txt

I have loaded customer_demographics.csv, customer_transaction_dat_new.csv, item_data.csv and train.csv files to the DB called CouponRedemption_SourceDB.



CouponRedemption_SourceDB

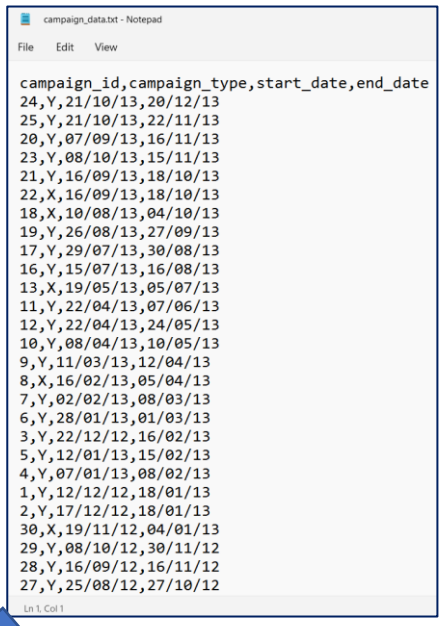
The campaign_data.csv details were taken as s text file, and I used It as another sourceType.



	A	B	C	D	E
1	campaign_id	campaign_type	start_date	end_date	
2	24	Y	21/10/13	20/12/13	
3	25	Y	21/10/13	22/11/13	
4	20	Y	7/9/2013	16/11/13	
5	23	Y	8/10/2013	15/11/13	
6	21	Y	16/09/13	18/10/13	
7	22	X	16/09/13	18/10/13	
8	18	X	10/8/2013	4/10/2013	
9	19	Y	26/08/13	27/09/13	
10	17	Y	29/07/13	30/08/13	
11	16	Y	15/07/13	16/08/13	
12	13	X	19/05/13	5/7/2013	
13	11	Y	22/04/13	7/6/2013	
14	12	Y	22/04/13	24/05/13	
15	10	Y	8/4/2013	10/5/2013	
16	9	Y	11/3/2013	12/4/2013	
17	8	X	16/02/13	5/4/2013	
18	7	Y	2/2/2013	8/3/2013	
19	6	Y	28/01/13	1/3/2013	
20	3	Y	22/12/12	16/02/13	
21	5	Y	12/1/2013	15/02/13	
22	4	Y	7/1/2013	8/2/2013	
23	1	Y	12/12/2012	18/01/13	
24	2	Y	17/12/12	18/01/13	
25	30	X	19/11/12	4/1/2013	
26	29	Y	8/10/2012	30/11/12	

campaign_data

Campaign_data.csv



campaign_data.txt - Notepad

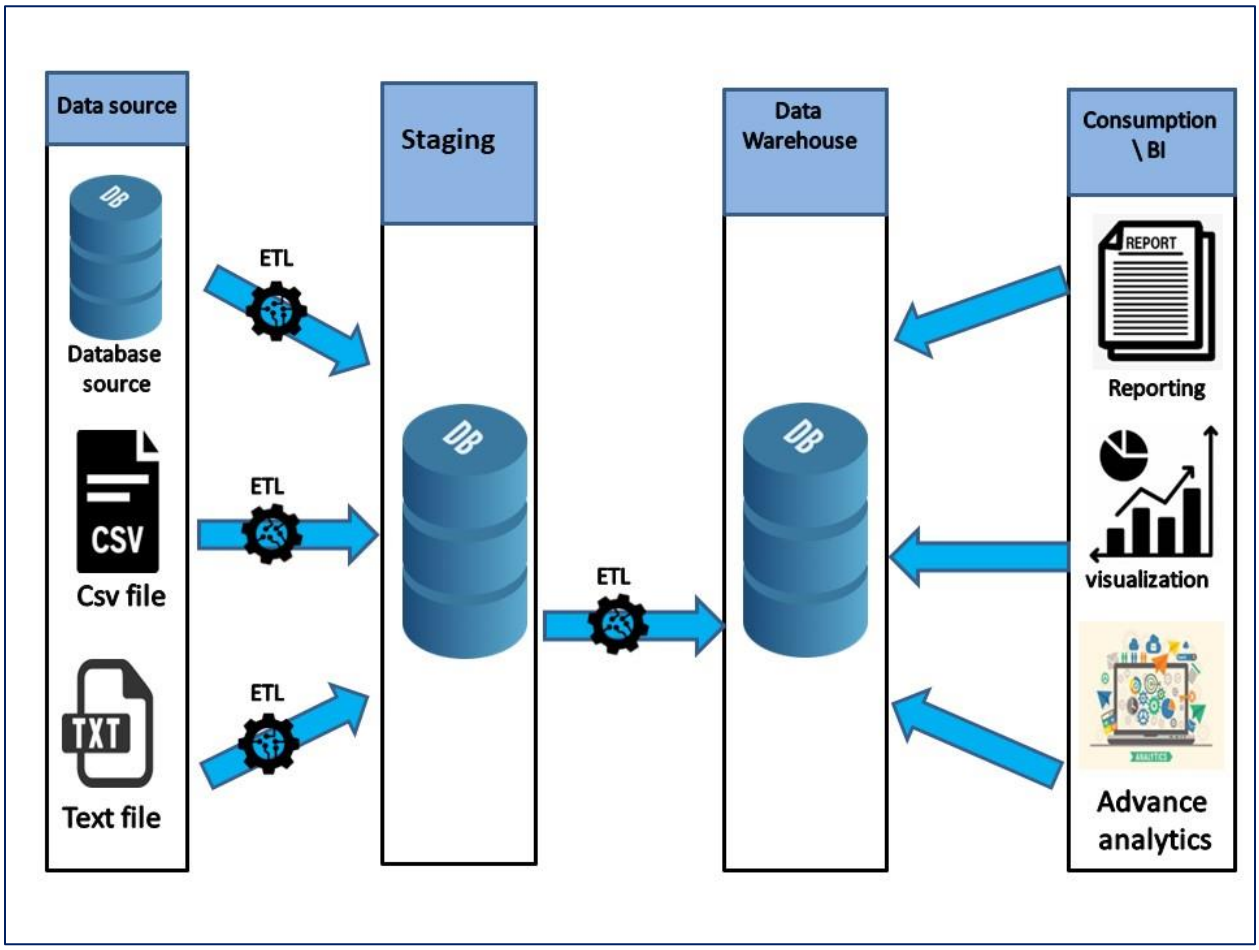
File Edit View

```
campaign_id,campaign_type,start_date,end_date
24,Y,21/10/13,20/12/13
25,Y,21/10/13,22/11/13
20,Y,07/09/13,16/11/13
23,Y,08/10/13,15/11/13
21,Y,16/09/13,18/10/13
22,X,16/09/13,18/10/13
18,X,10/08/13,04/10/13
19,Y,26/08/13,27/09/13
17,Y,29/07/13,30/08/13
16,Y,15/07/13,16/08/13
13,X,19/05/13,05/07/13
11,Y,22/04/13,07/06/13
12,Y,22/04/13,24/05/13
10,Y,08/04/13,10/05/13
9,Y,11/03/13,12/04/13
8,X,16/02/13,05/04/13
7,Y,02/02/13,08/03/13
6,Y,28/01/13,01/03/13
3,Y,22/12/12,16/02/13
5,Y,12/01/13,15/02/13
4,Y,07/01/13,08/02/13
1,Y,12/12/12,18/01/13
2,Y,17/12/12,18/01/13
30,X,19/11/12,04/01/13
29,Y,08/10/12,30/11/12
28,Y,16/09/12,16/11/12
27,Y,25/08/12,27/10/12
```

Ln 1, Col 1

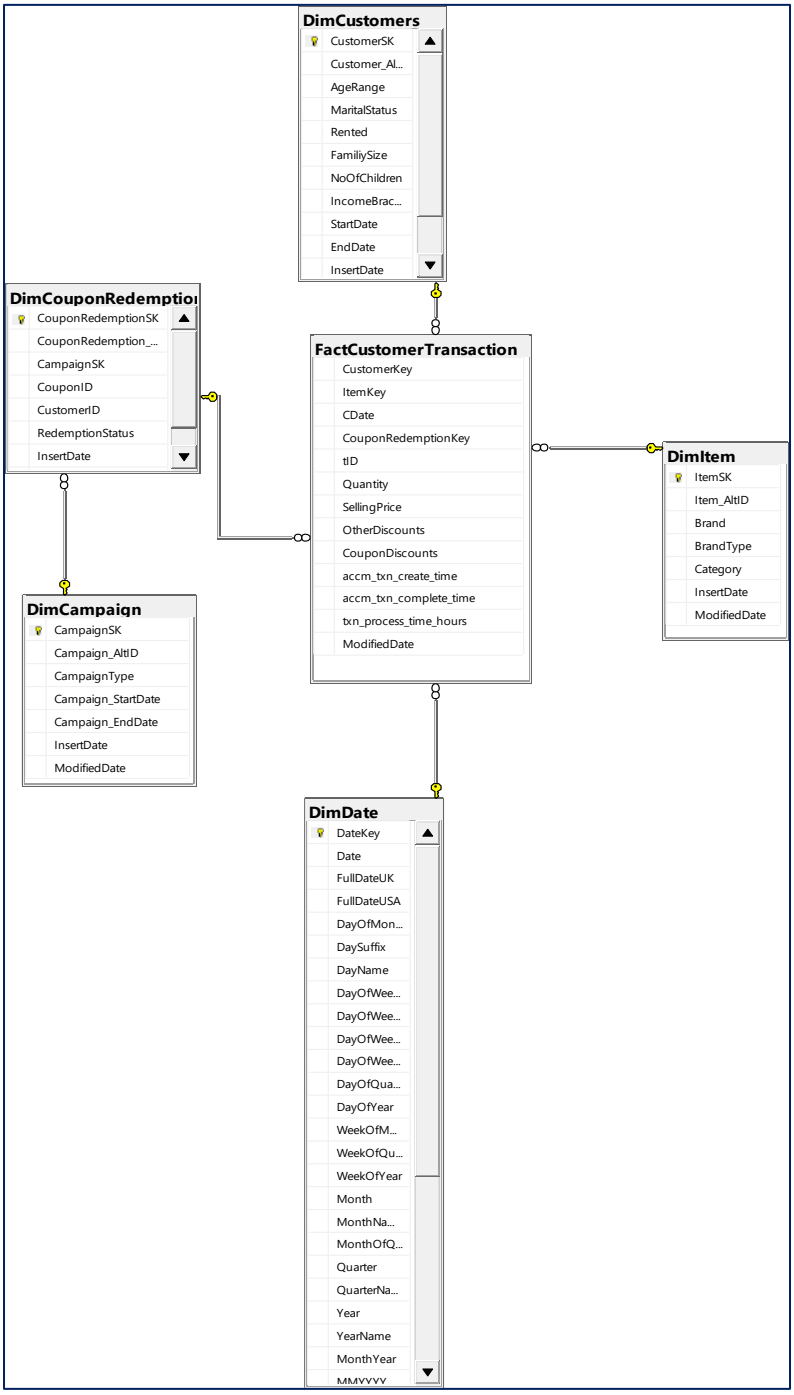
Campaign_data.txt

STEP 03: Solution Architecture



- The data warehouse is the core of the BI system. A data warehouse is a purpose of data analysis and reporting. This purpose changes the design of this database as well.
- This architecture shows the high-level BI solution to the warehouse.

STEP 04: Data warehouse Design and Development



Assumptions

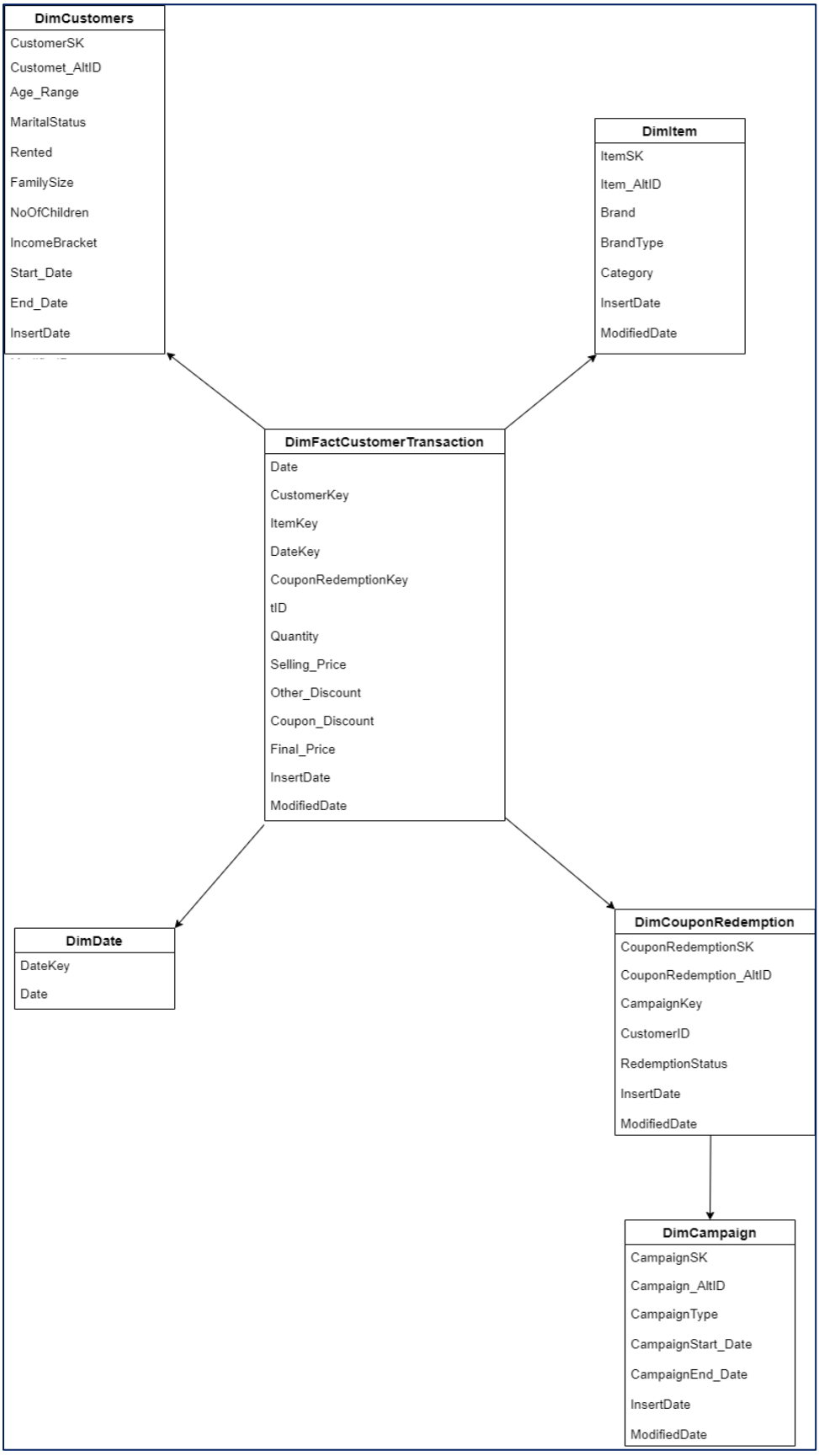
- I used a schema type of **Snowflake schema** for the Data warehouse development. At there,
- There is one Fact Table, it is **FactCustomerTransaction** table.

There are five dimensions table.

1. DimCustomers table
2. DimItem table
3. DimDate table
4. DimCouponRedemption table
5. DimCampaign table

- I have taken **DimCustomers** details as slowly changing dimension, **Income bracket** can change time to time, and we need to keep track of their historical data.

- I have implemented data warehouse tables in the SQL server and the tables that I have created shown below


DimCustomers
CustomerSK
Customet_AltID
Age_Range
MaritalStatus
Rented
FamilySize
NoOfChildren
IncomeBracket
Start_Date
End_Date
InsertDate
DimItem
ItemSK
Item_AltID
Brand
BrandType
Category
InsertDate
ModifiedDate
DimFactCustomerTransaction
Date
CustomerKey
ItemKey
DateKey
CouponRedemptionKey
tID
Quantity
Selling_Price
Other_Discount
Coupon_Discount
Final_Price
InsertDate
ModifiedDate
DimDate
DateKey
Date
DimCouponRedemption
CouponRedemptionSK
CouponRedemption_AltID
CampaignKey
CustomerID
RedemptionStatus
InsertDate
ModifiedDate
DimCampaign
CampaignSK
Campaign_AltID
CampaignType
CampaignStart_Date
CampaignEnd_Date
InsertDate
ModifiedDate

Data Warehouse Data types

- Before creating the FactTransaction fact table and other dimensios, start by creating Date dimension. For that I used the code in the “DateMaster.sql” file.

LAPTOP-7RV7378O...dbo.DimCampaign | SQLQuery45.sql - L...V7378O\nipun (53))

Column Name	Data Type	Allow Nulls
CampaignSK	int	☐
Campaign_AltID	int	☑
CampaignType	nvarchar(20)	☑
Campaign_StartDate	datetime	☑
Campaign_EndDate	datetime	☑
InsertDate	datetime	☑
ModifiedDate	datetime	☑
		☐

Campaign Dimension Table

LAPTOP-7RV7378O...ouponRedemption | SQLQuery45.sql - L...V7378O\nipun (53))

Column Name	Data Type	Allow Nulls
CouponRedemptionSK	int	☐
CouponRedemption_AltID	varchar(50)	☑
CampaignSK	int	☑
CouponID	varchar(50)	☑
CustomerID	varchar(50)	☑
RedemptionStatus	varchar(50)	☑
InsertDate	datetime	☑
ModifiedDate	datetime	☑
		☐

Coupon Redemption Dimension Table

LAPTOP-7RV7378O...dbo.DimCustomers

Column Name	Data Type	Allow Nulls
CustomerSK	int	☐
Customer_AltID	smallint	☑
AgeRange	nvarchar(50)	☑
MaritalStatus	nvarchar(50)	☑
Rented	bit	☑
FamiliySize	tinyint	☑
NoOfChildren	tinyint	☑
IncomeBracket	tinyint	☑
StartDate	datetime	☑
EndDate	datetime	☑
InsertDate	datetime	☑
ModifiedDate	datetime	☑
		☐

Customer Dimension Table

LAPTOP-7RV7378O...DW - dbo.DimItem

Column Name	Data Type	Allow Nulls
ItemSK	int	☐
Item_AltID	varchar(50)	☑
Brand	varchar(50)	☑
BrandType	varchar(50)	☑
Category	varchar(50)	☑
InsertDate	datetime	☑
ModifiedDate	datetime	☑
		☐

Item Dimension Table

LAPTOP-7RV7378O\...stomerTransaction | SQLQuery2.sql - not connected

Column Name	Data Type	Allow Nulls
CustomerKey	int	☑
ItemKey	int	☑
CDate	int	☑
CouponRedemptionKey	int	☑
tID	int	☑
Quantity	int	☑
SellingPrice	float	☑
OtherDiscounts	float	☑
CouponDiscounts	float	☑
accm_txn_create_time	datetime	☑
accm_txn_complete_time	datetime	☑
txn_process_time_hours	int	☑
ModifiedDate	datetime	☑
		☐

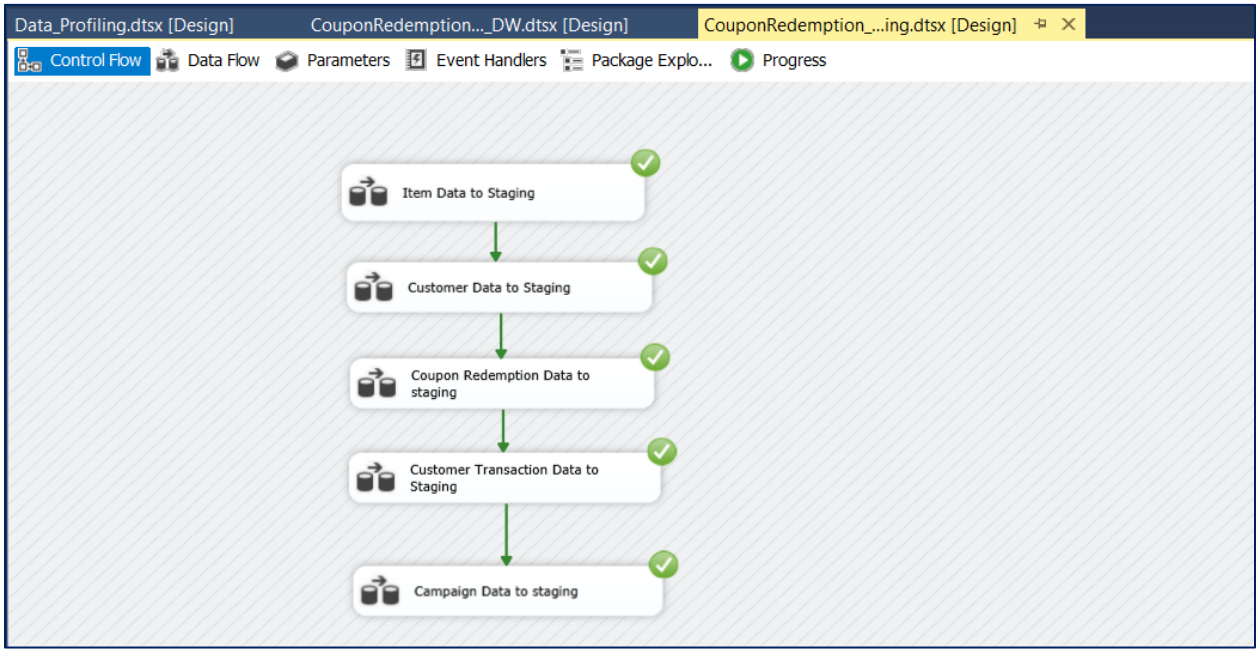
Fact Customer Transaction Table

LAPTOP-7RV7378O...DW - dbo.DimDate

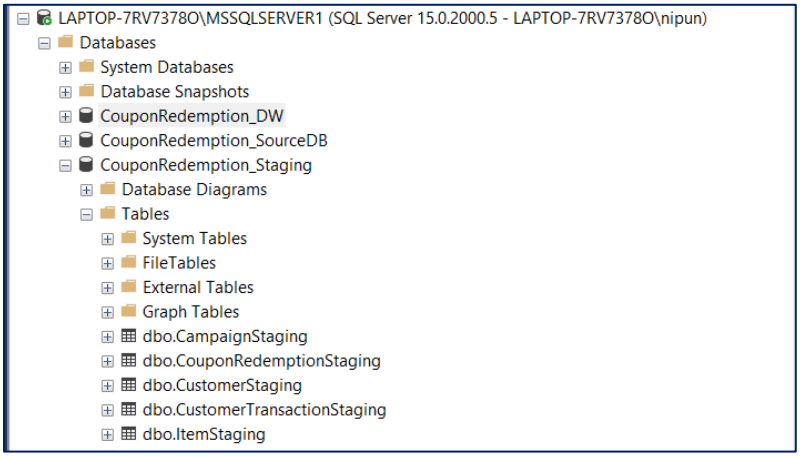
Column Name	Data Type	Allow Nulls
DateKey	int	☐
Date	datetime	☑
FullDateUK	char(10)	☑
FullDateUSA	char(10)	☑
DayOfMonth	varchar(2)	☑
DaySuffix	varchar(4)	☑
DayName	varchar(9)	☑
DayOfWeekUSA	char(1)	☑
DayOfWeekUK	char(1)	☑
DayOfWeekInMonth	varchar(2)	☑
DayOfWeekInYear	varchar(2)	☑
DayOfQuarter	varchar(3)	☑
DayOfYear	varchar(3)	☑
WeekOfMonth	varchar(1)	☑
WeekOfQuarter	varchar(2)	☑
WeekOfYear	varchar(2)	☑
Month	varchar(2)	☑
MonthName	varchar(9)	☑
MonthOfQuarter	varchar(2)	☑
Quarter	char(1)	☑
QuarterName	varchar(9)	☑
Year	char(4)	☑
YearName	char(7)	☑
MonthYear	char(10)	☑
MMYYYY	char(6)	☑
FirstDayOfMonth	date	☑
LastDayOfMonth	date	☑
FirstDayOfQuarter	date	☑

Date Dimension Table

STEP 05: ETL Development

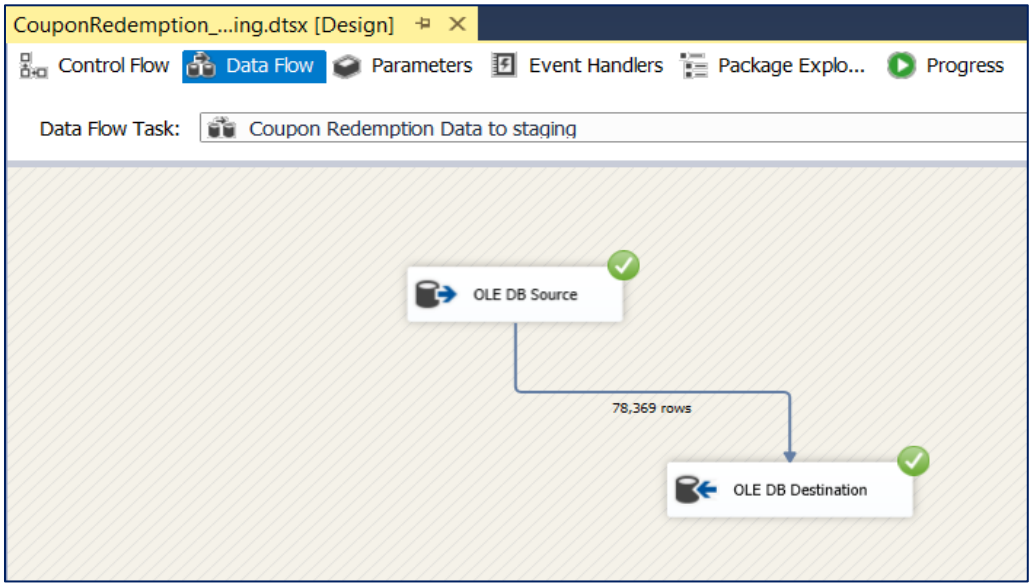


- First using the SQL Server integration service Software, I have extracted all the data from the Tables which were in the CouponRedmption_SourceDB and Extract Campaign_Data.txt to Separate DB called CouponRedemption_Staging. And also, **here I converted train.csv data into CouponRedemptionStaging table**.

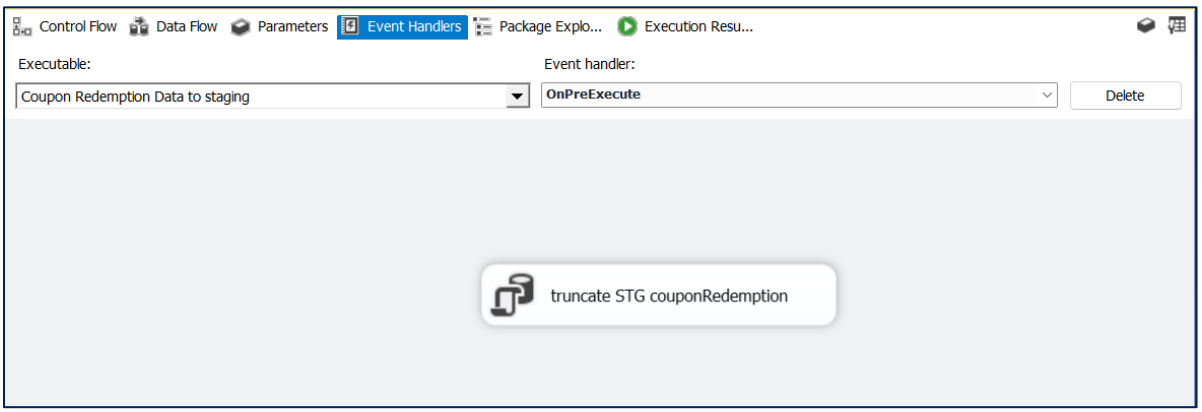


CouponRedemption_Staging DB

1. Extract Coupon Redemption Data to Staging

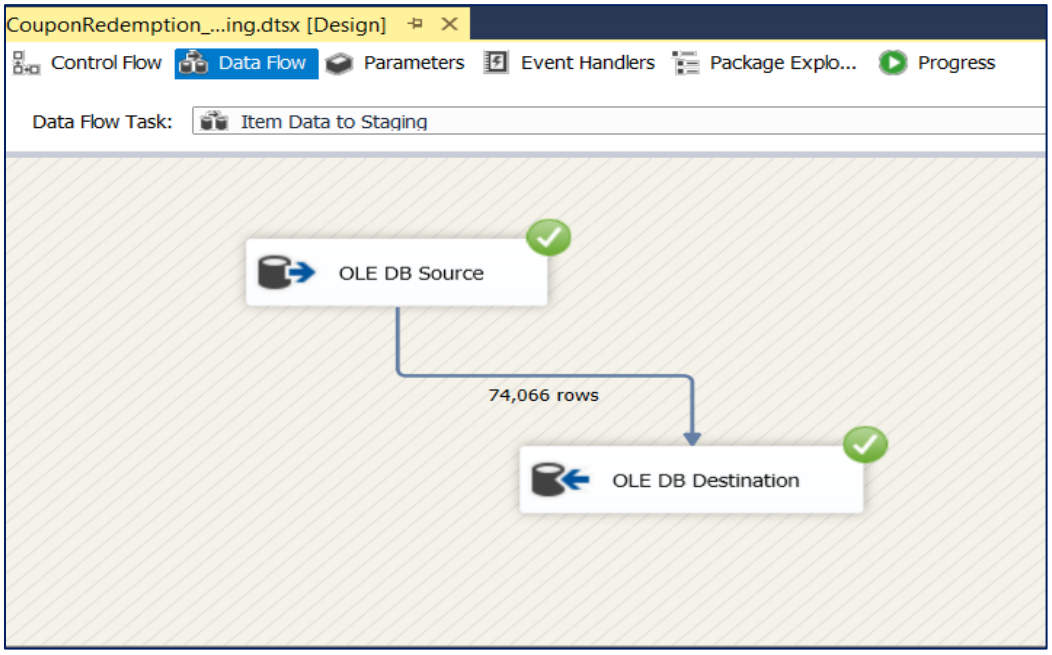


- Used OLE DB Source as dbo.train data table in CouponRedemption_SourceDB.
- OLE DB Destination for create new table CouponRedemptionStaging table in the CouponRedemption _Staging database.

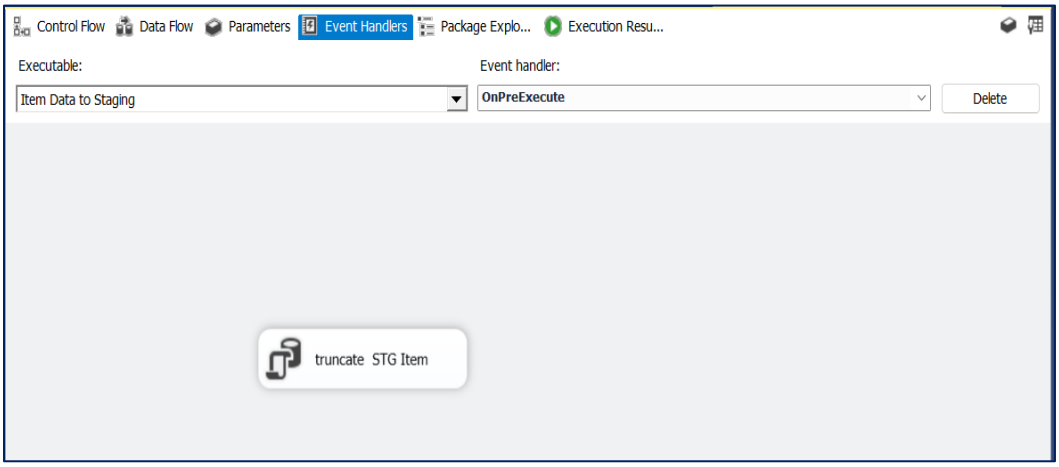


- Used to execute SQL Task SSIS tool truncate table for SQL command as truncate table CouponRedemptionStaging in CouponRedemption_Staging database.

2. Extract Item Data to Staging

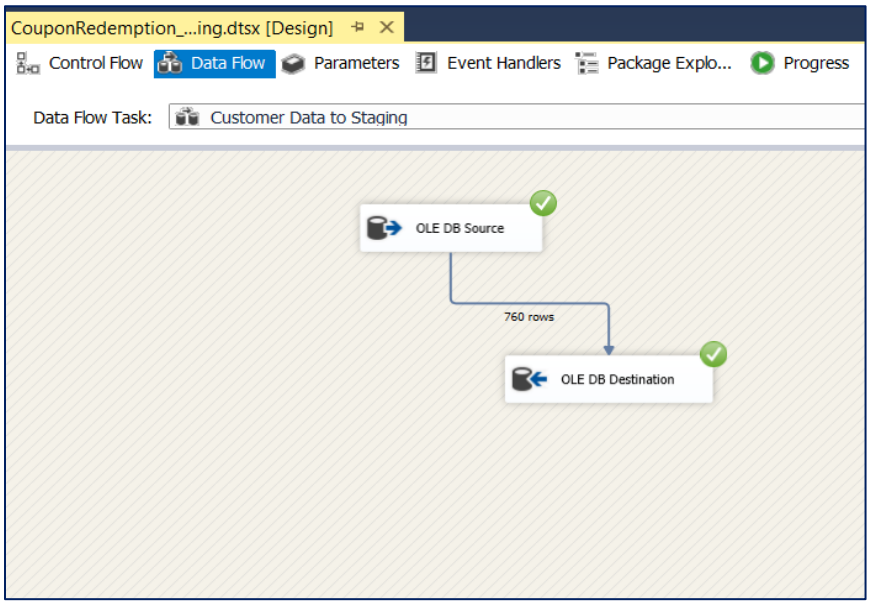


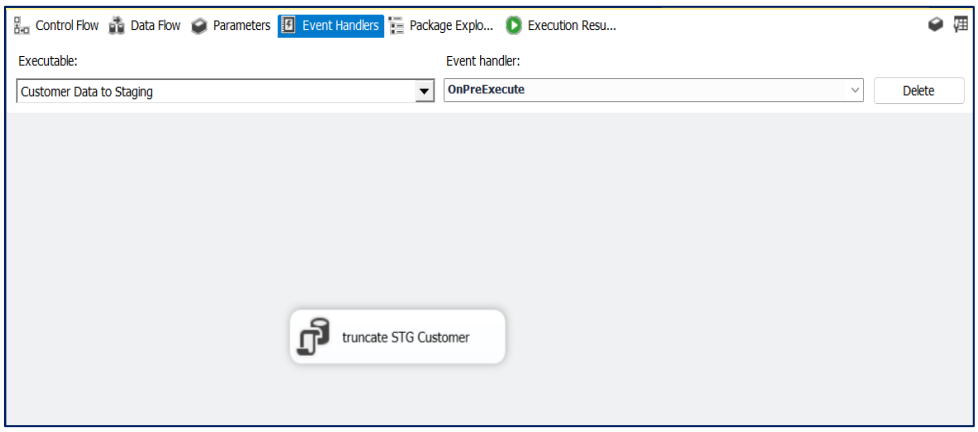
- Used OLE DB Source as dbo.item_data table in CouponRedmption_SourceDB.
- OLE DB Destination for create new table ItemStaging table in the CouponRedmption _Staging database.



- Used to execute SQL Task SSIS tool truncate table for SQL command as truncate table ItemStaging in CouponRedemption_Staging database.

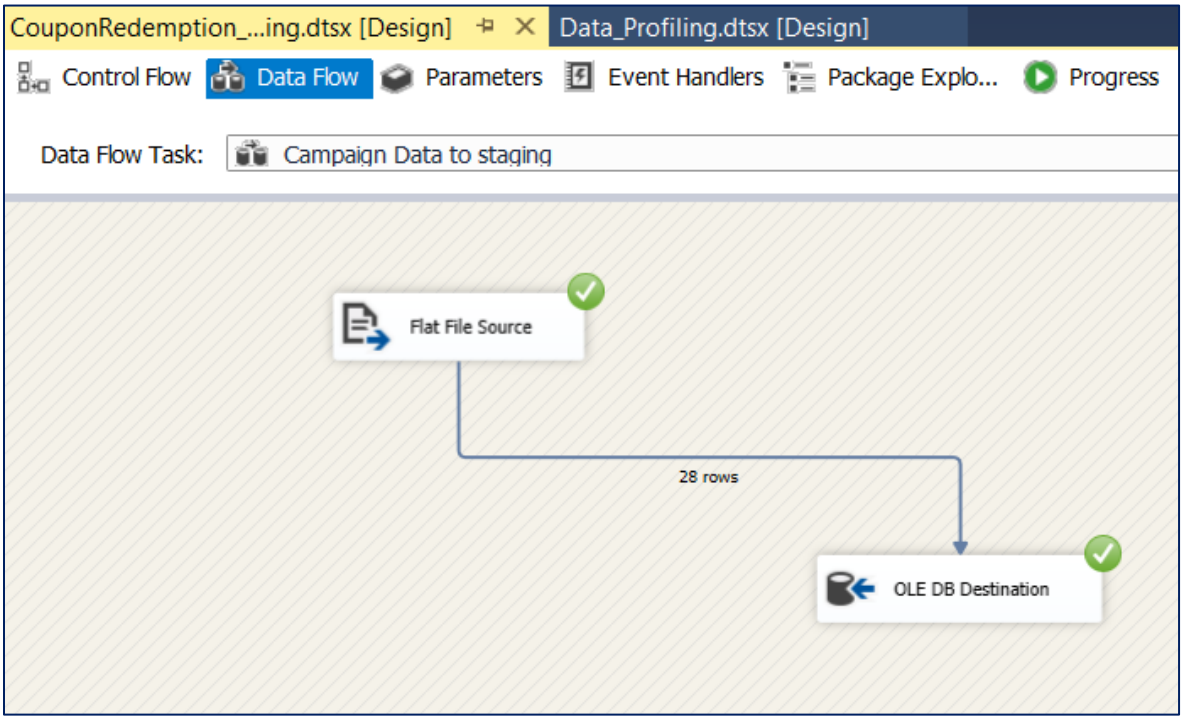
3. Extract Customer Data to Staging



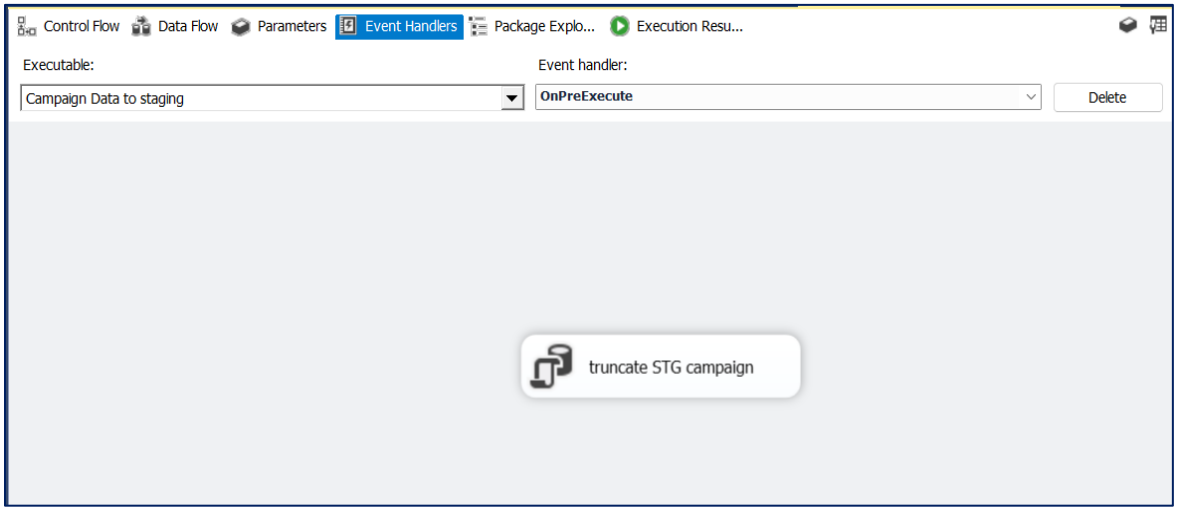


- Used OLE DB Source as dbo.Customer_demographocs table in CouponRedmption_SourceDB. OLE DB Destination for create new table CutomerStaging in the CouponRedmption _Staging database. Used to execute SQL Task SSIS tool truncate table for SQL command as truncate table CutomerStaging in CouponRedemptio_Staging database.

4. Extract Campaign Data to Staging

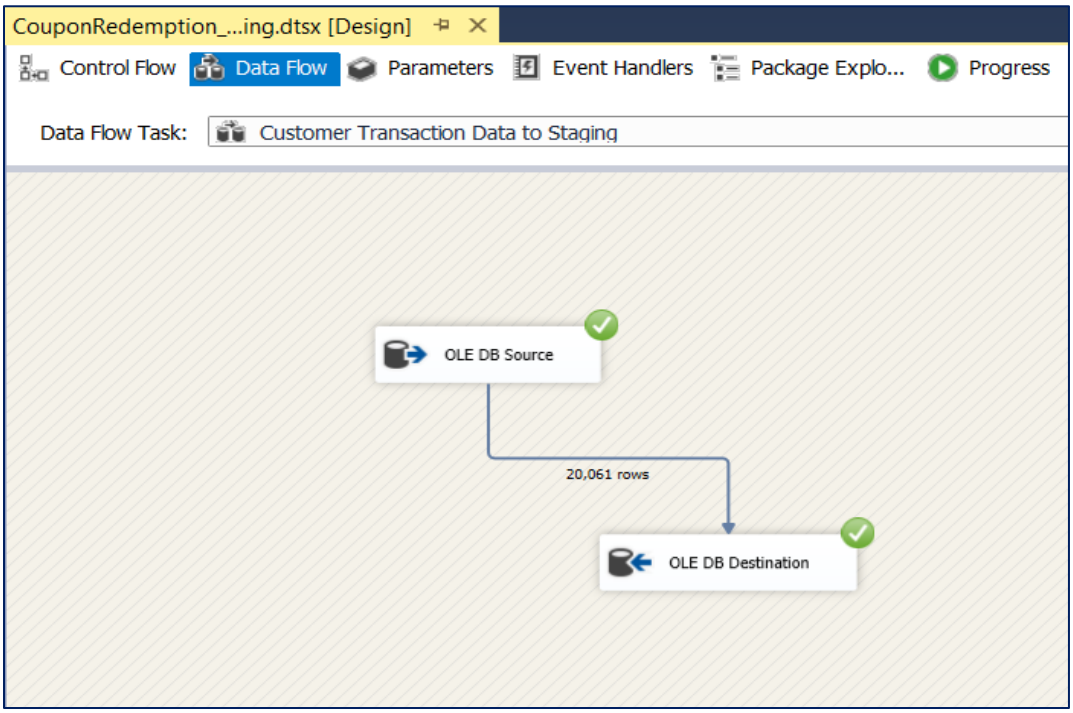


- Used OLE DB Source as **Campaign_data.txt Flat File Source.**
- OLE DB Destination for create new table CampaignStaging in the CouponRedmption_Staging database.

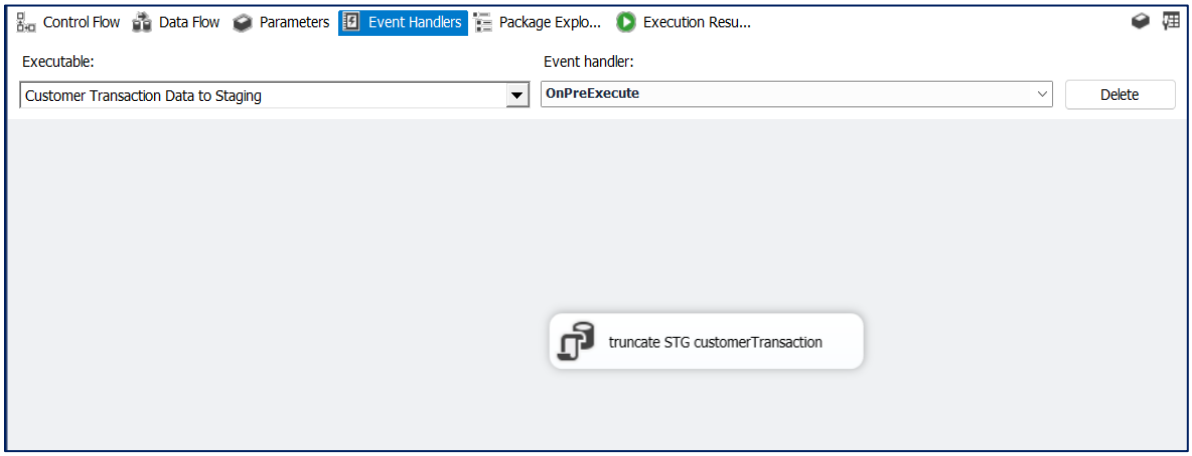


- Used to execute SQL Task SSIS tool truncate table for SQL command as truncate table CampaignStaging in CouponRedemption_Staging database.

5. Extract Customer transaction Data to Staging



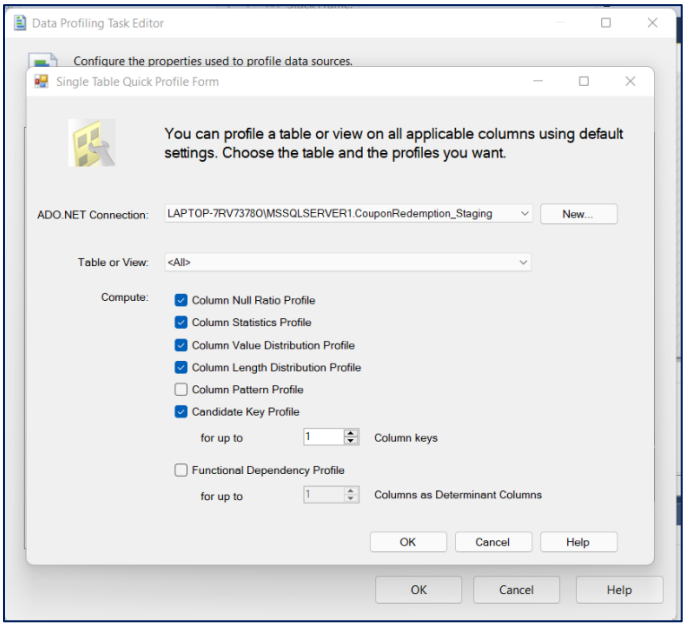
- Used OLE DB Source as dbo.customer_transaction_data_new table in CouponRedmption_SourceDB.
- OLE DB Destination for create new table CustomerTransactionStaging in the CouponRedmption _Staging database.

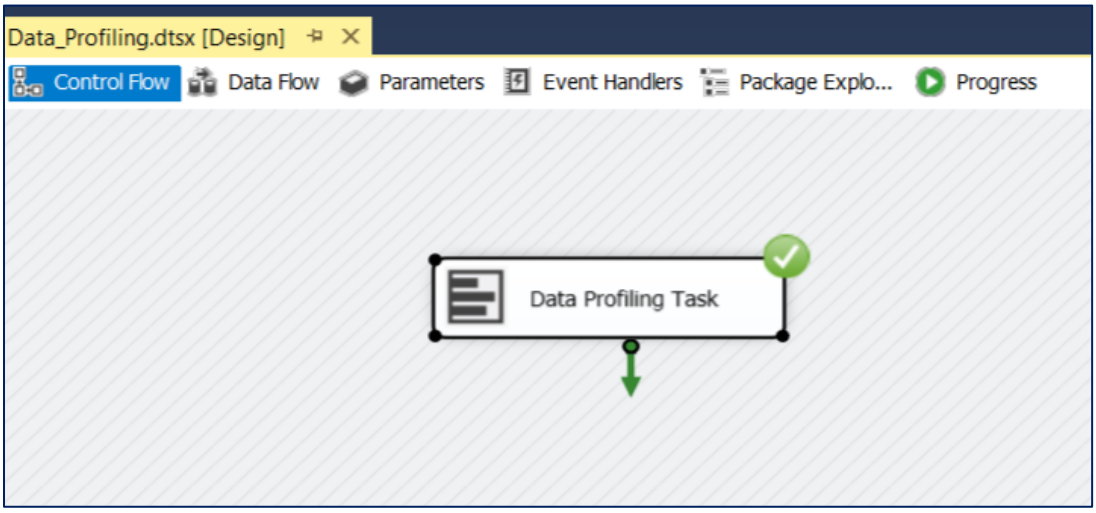


Used to execute SQL Task SSIS tool truncate table for SQL command as truncate table CustomerTransactionStaging in CouponRedemption_Staging database.

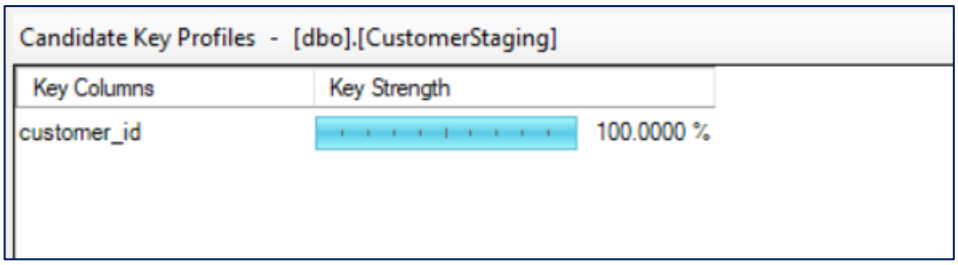
Data Profiling

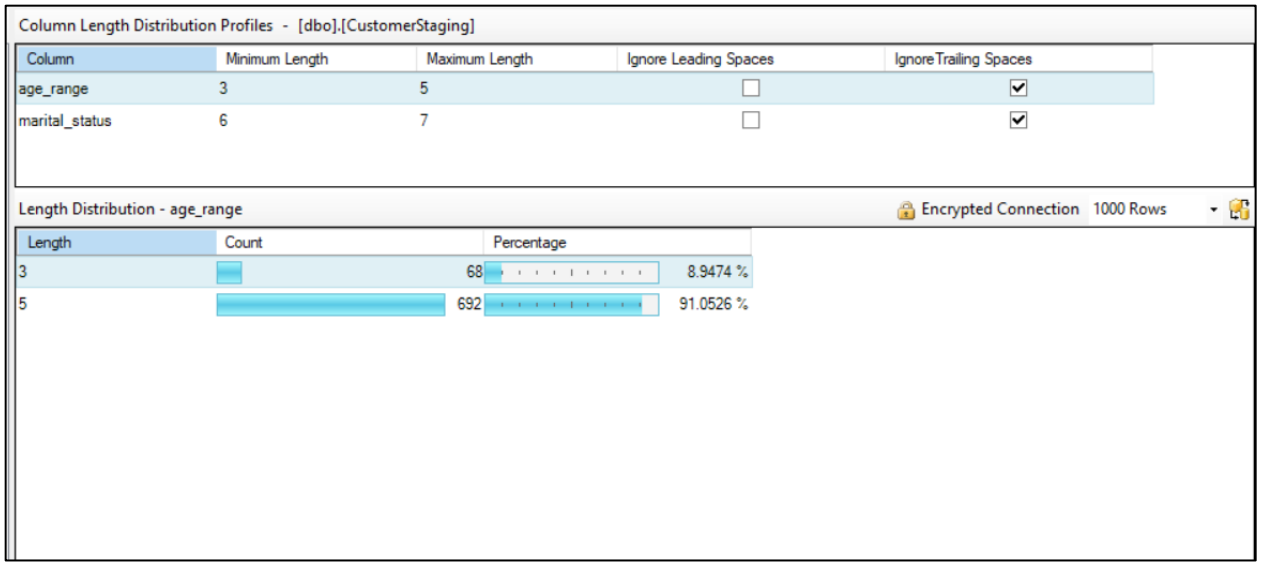
- I used the staging table data to analyze how the data looks like to determine what type of transformations I need to perform on the data.





- Data Profiling was done to all staging tables to get a better insight on the data like checking for unique values or if null values were present. CustomerStaging table data profiling shown below.





Column Null Ratio Profiles - [dbo].[CustomerStaging]

Column	Null Count	Null Percentage
age_range	0	0.0000 %
customer_id	0	0.0000 %
family_size	0	0.0000 %
income_bracket	0	0.0000 %
marital_status	329	43.2895 %
no_of_children	538	70.7895 %
rented	0	0.0000 %

Column Statistics Profiles - [dbo].[CustomerStaging]

Column	Minimum	Maximum	Mean	Standard Deviation
customer_id	1	1581	779.201315789...	459.451859053527
family_size	1	5	2.16184210526...	1.16815949538075
income_bracket	1	12	4.71578947368...	2.25733031836393
no_of_children	1	3	1.78828828828...	0.841088820147046

Column Value Distribution Profiles - [dbo].[CustomerStaging]

Column	Number Of Distinct Values
age_range	6
customer_id	760
family_size	5
income_bracket	12
marital_status	2
no_of_children	3
rented	2

Frequent Value Distribution (0.1000 %) - age_range

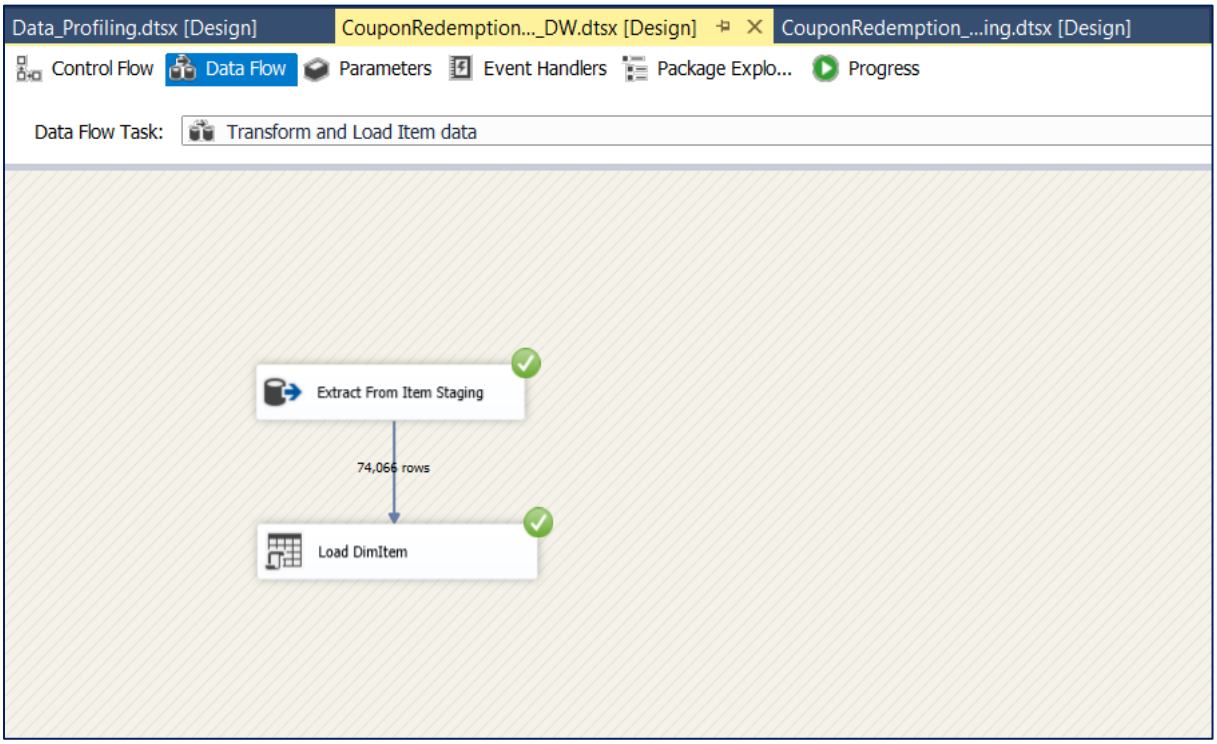
Value	Count	Percentage
36-45	187	24.6053 %
56-70	59	7.7632 %
70+	68	8.9474 %
18-25	45	5.9211 %
46-55	271	35.6579 %
26-35	130	17.1053 %

Data Transformation

1. **Transform and Load Item Data Details**

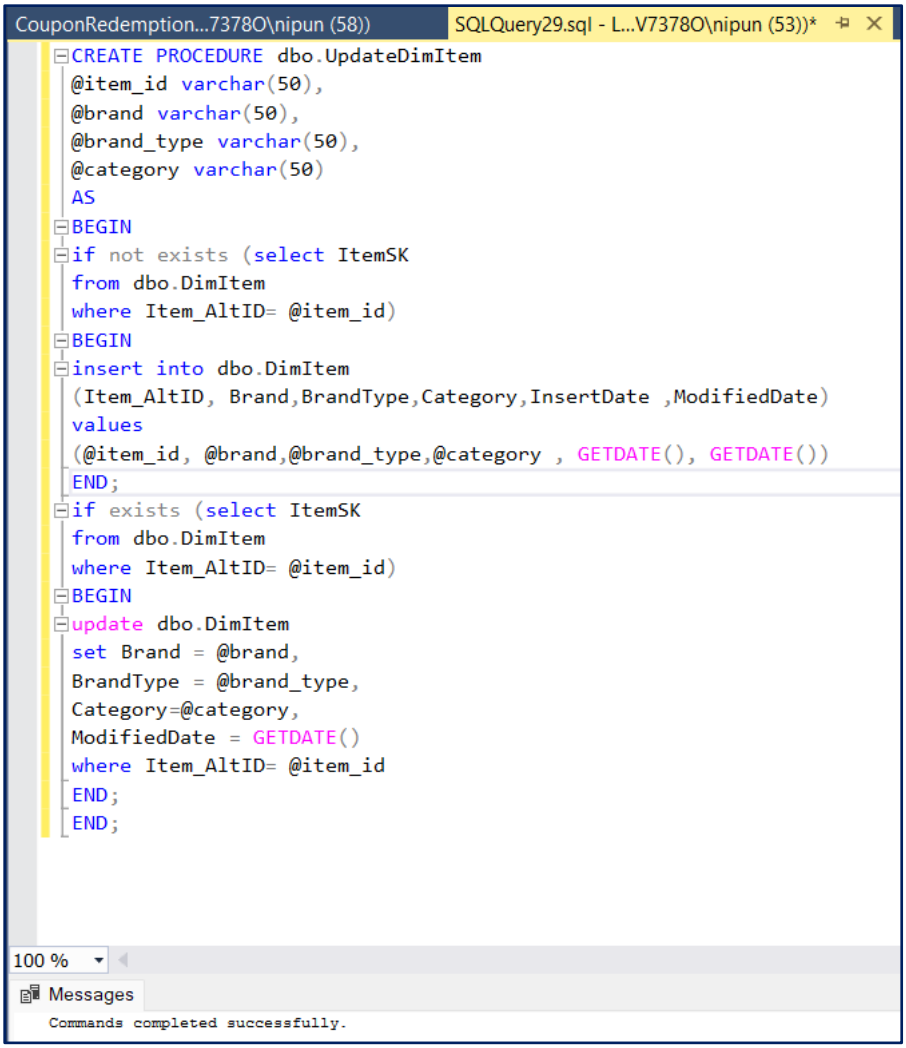
   I created Item data transformation by below mentioned steps.

   - Created new package called CouponRedemption_DW.dtsx.
   - Then dragged and dropped a data flow task, renamed it as transform and load Item Data.
   - Then dragged and dropped OLE DB source renamed as Extract from Item Staging and configure it to access the Item staging table.
   - After that I dragged and OLE DB Command and connect to the OLE DB source.



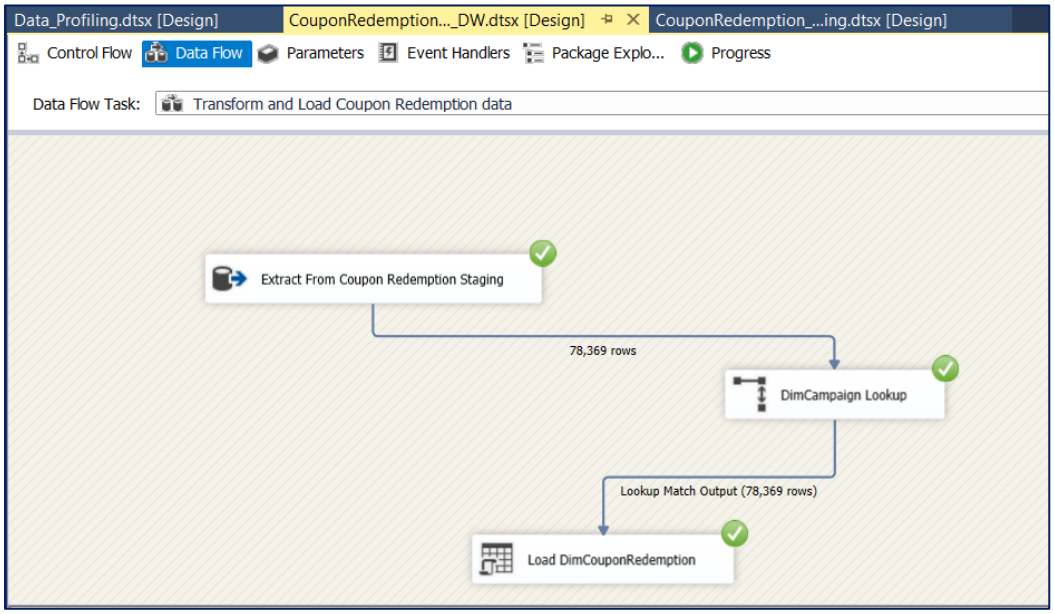
   - First, I have created a procedure called UpdateDimItem and executed in the CouponRedemption_DW database.

- **UpdateDimItem procedure**, it is used to insert data from ItemStaging to DimItem without data duplication.



- Then I did same process to DimCouponRedemption table and DimCampaign table as well.

2. Transform and Load Coupon Redemption Data Details

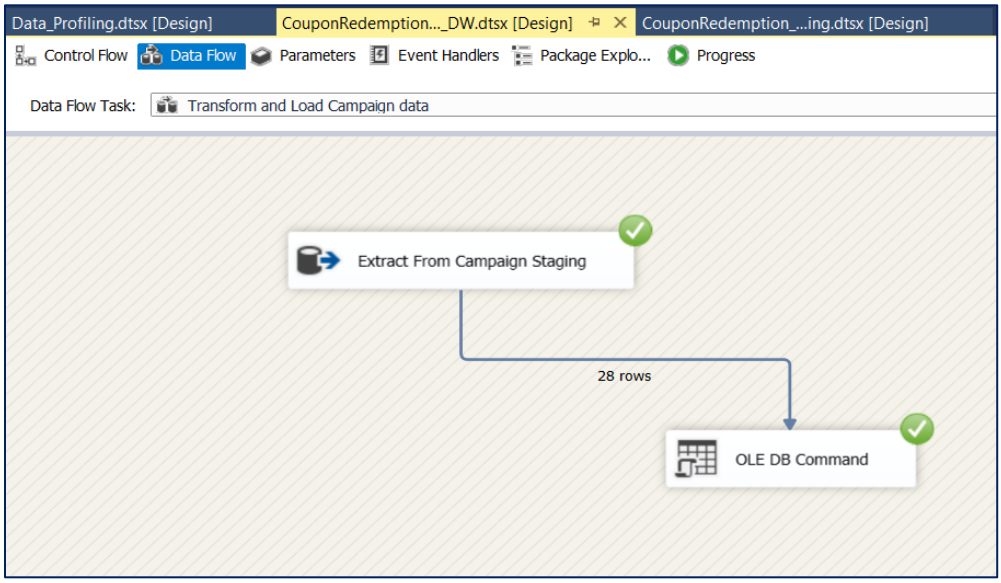


UpdateDimCouponRedemption Procedure

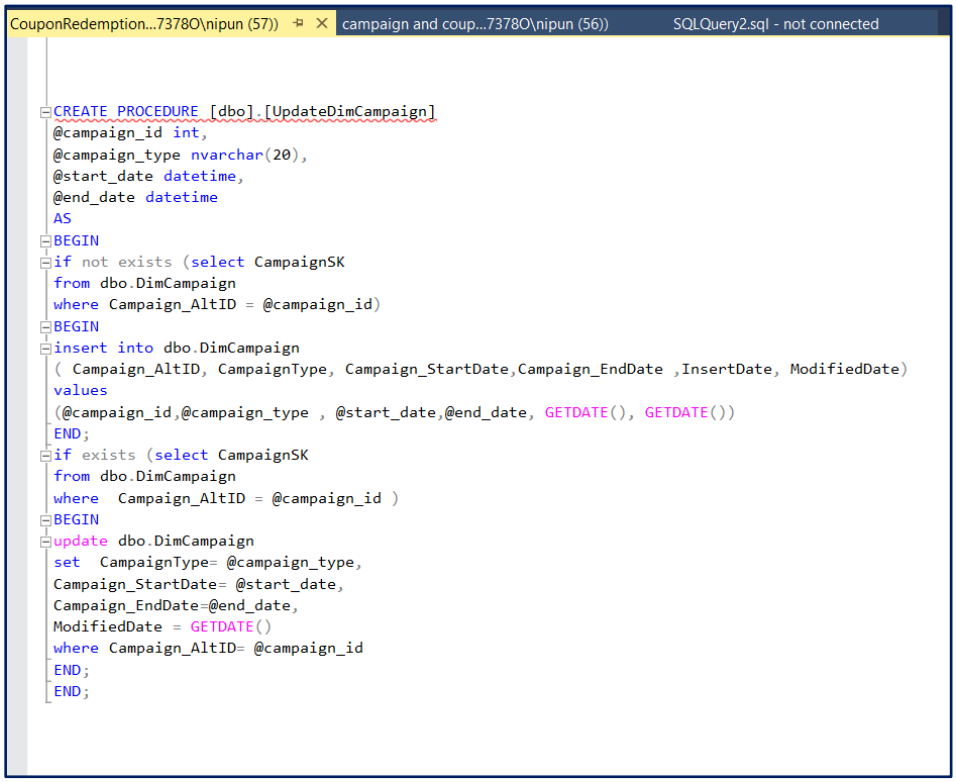
CouponRedemption...7378O\nipun (55)) | campaign and coup...7378O\nipun (54)) | LAPTOP-7RV7378O\...n_DW - Diagram_0*

```
CREATE PROCEDURE [dbo].[UpdateDimCouponRedemption]
@id varchar(50),
@Campaign_Key int,
@Coupon_id varchar(50),
@Customer_id varchar(50),
@Redemption_Status varchar(50)
AS
BEGIN
if not exists (select CouponRedemptionSK
from dbo.DimCouponRedemption
where CouponRedemption_AltID = @id)
BEGIN
insert into dbo.DimCouponRedemption
( CouponRedemption_AltID , CampaignSK, CouponID,CustomerID, RedemptionStatus,InsertDate, ModifiedDate)
values
(@id,@Campaign_Key, @Coupon_id,@Customer_id,@Redemption_Status, GETDATE(), GETDATE())
END;
if exists (select CouponRedemptionSK
from dbo.DimCouponRedemption
where CouponRedemption_AltID = @id)
BEGIN
update dbo.DimCouponRedemption
set  CampaignSK= @Campaign_Key,
CouponID= @Coupon_id,
CustomerID=@Customer_id,
RedemptionStatus=@Redemption_Status,
ModifiedDate = GETDATE()
where CouponRedemption_AltID = @id
END;
END;
```

3. Transform and Load Campaign Data Details

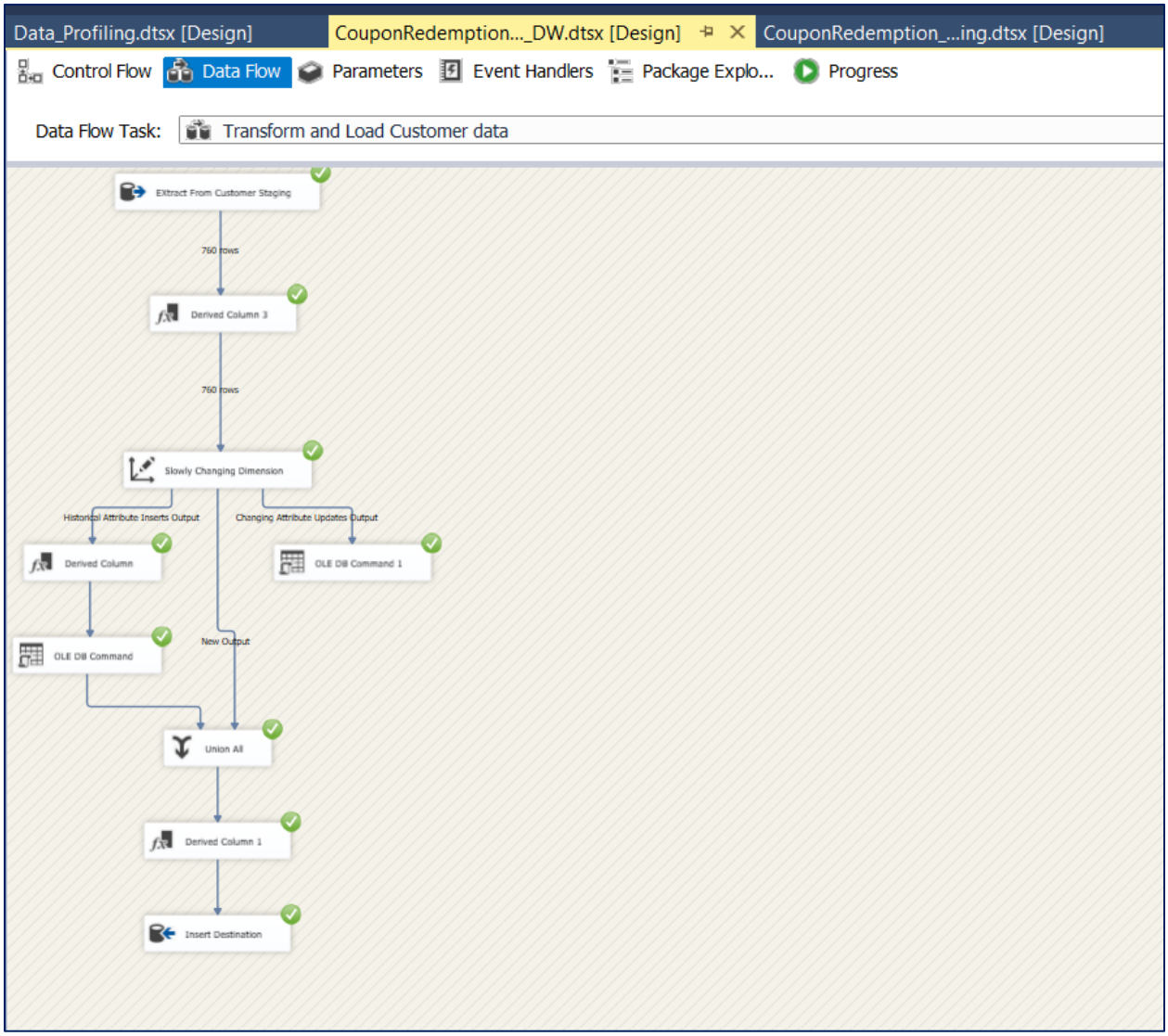


UpdateDimCampaign Procedure



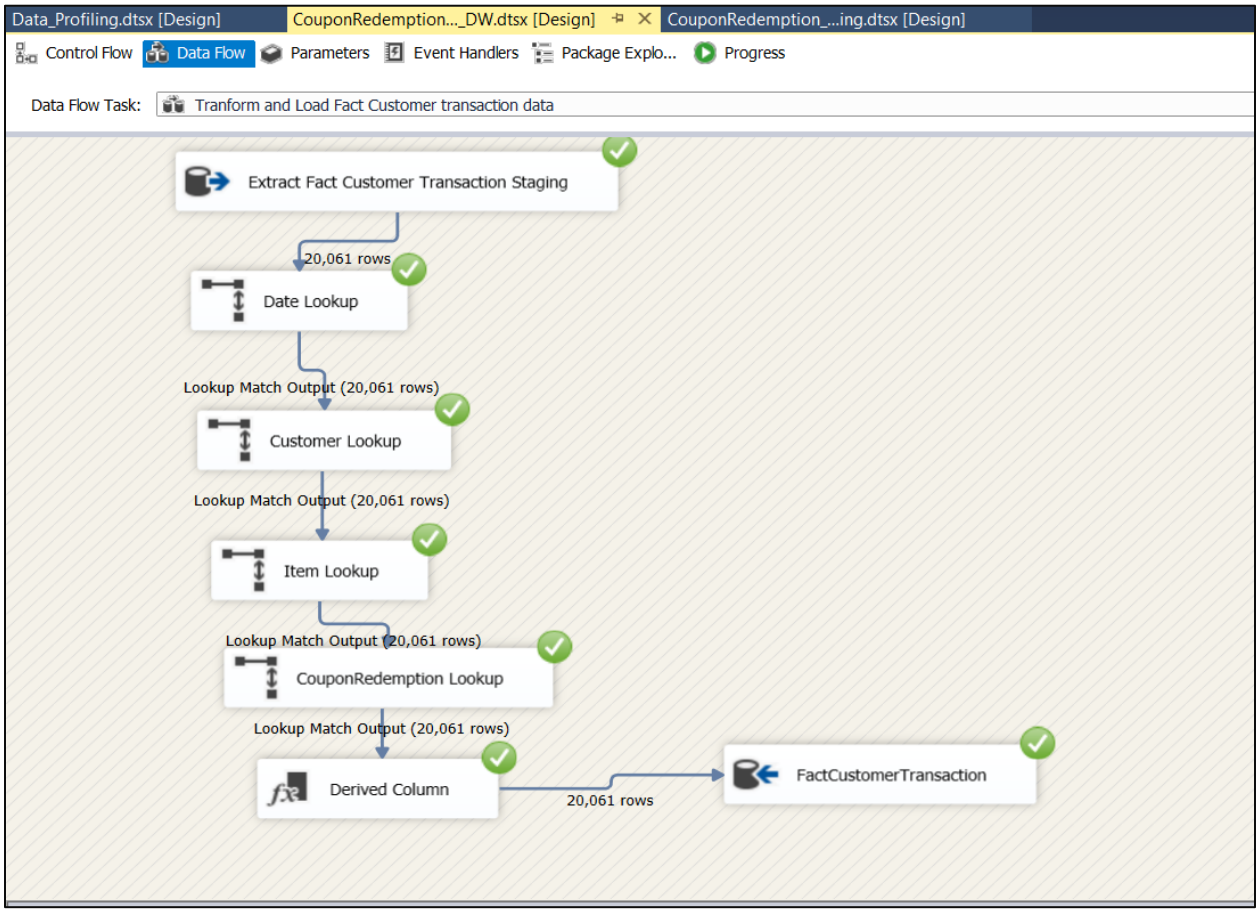

```
CREATE PROCEDURE [dbo].[UpdateDimCampaign]
@campaign_id int,
@campaign_type nvarchar(20),
@start_date datetime,
@end_date datetime
AS
BEGIN
if not exists (select CampaignSK
from dbo.DimCampaign
where Campaign_AltID = @campaign_id)
BEGIN
insert into dbo.DimCampaign
( Campaign_AltID, CampaignType, Campaign_StartDate,Campaign_EndDate ,InsertDate, ModifiedDate)
values
(@campaign_id,@campaign_type , @start_date,@end_date, GETDATE(), GETDATE())
END;
if exists (select CampaignSK
from dbo.DimCampaign
where  Campaign_AltID = @campaign_id )
BEGIN
update dbo.DimCampaign
set  CampaignType= @campaign_type,
Campaign_StartDate= @start_date,
Campaign_EndDate=@end_date,
ModifiedDate = GETDATE()
where Campaign_AltID= @campaign_id
END;
END;
```

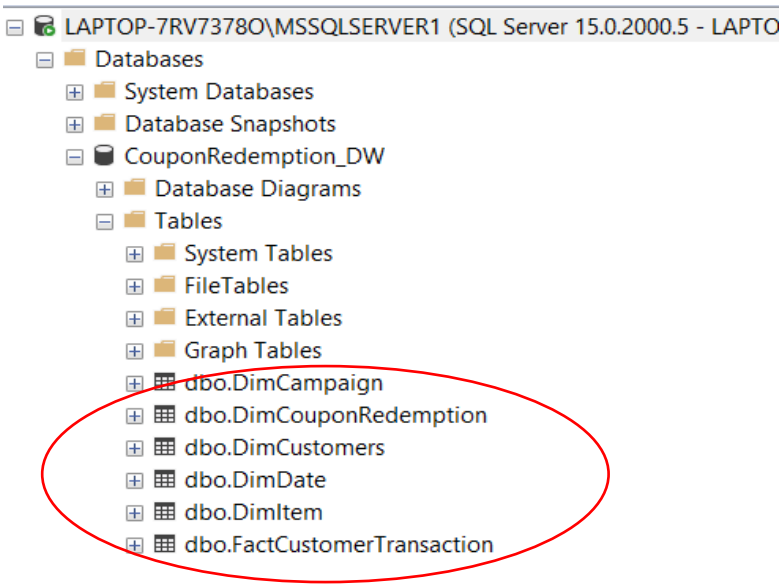
Transform and Load Customer Data Details (Slowly Changing Dimension)



- The slowly Changing Dimension was handled when creating the DimCustomers table. For the slowly changing dimension I have used a slowly changing dimension, derived columns, merge, union, and sort in order to do the ETL process of loading data to the warehouse from the staging.

Transform and Load Customer Transaction Data Details



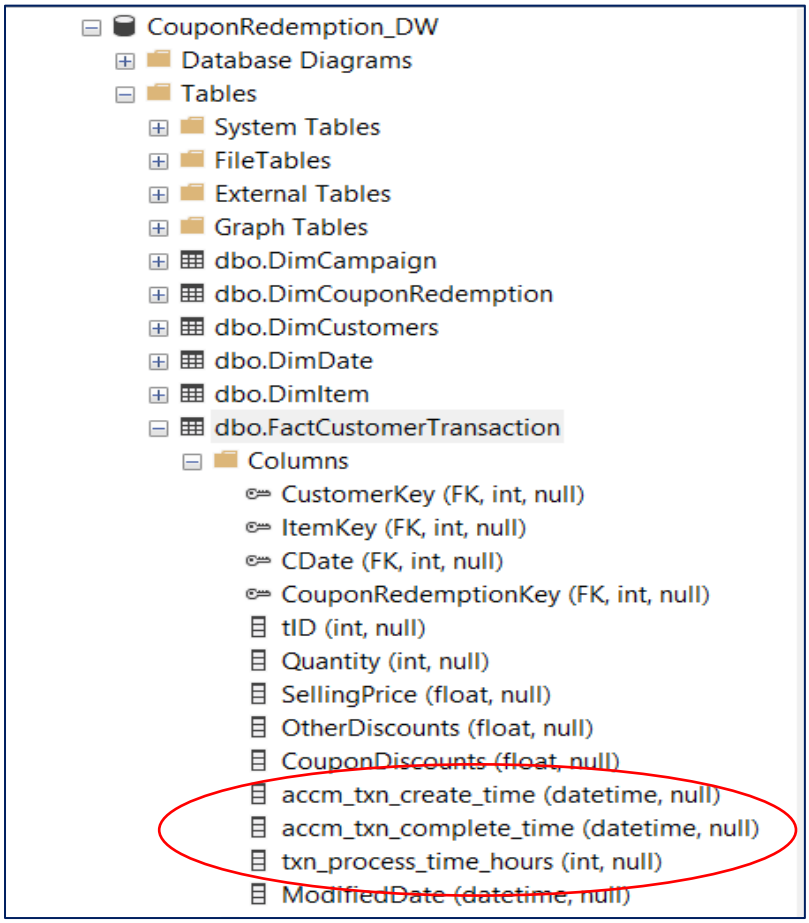


CouponRedemption_DW

STEP 06: ETL Development -Accumulating Fact Table

First, I extended my fact table (FactCustomerTransaction Table) with following 03 coloumns.

- accm_txn_create_time
- accm_txn_complete_time
- txn_process_time_hours



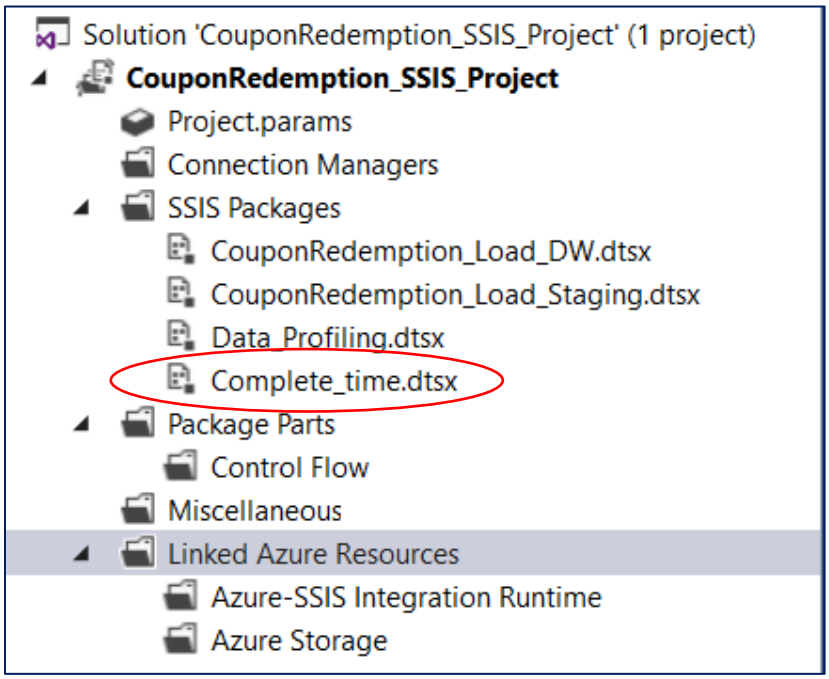
- Then I prepared a dataset which contains fact table natural key(tID) and accm_txn_complete time and saved it csv file format.

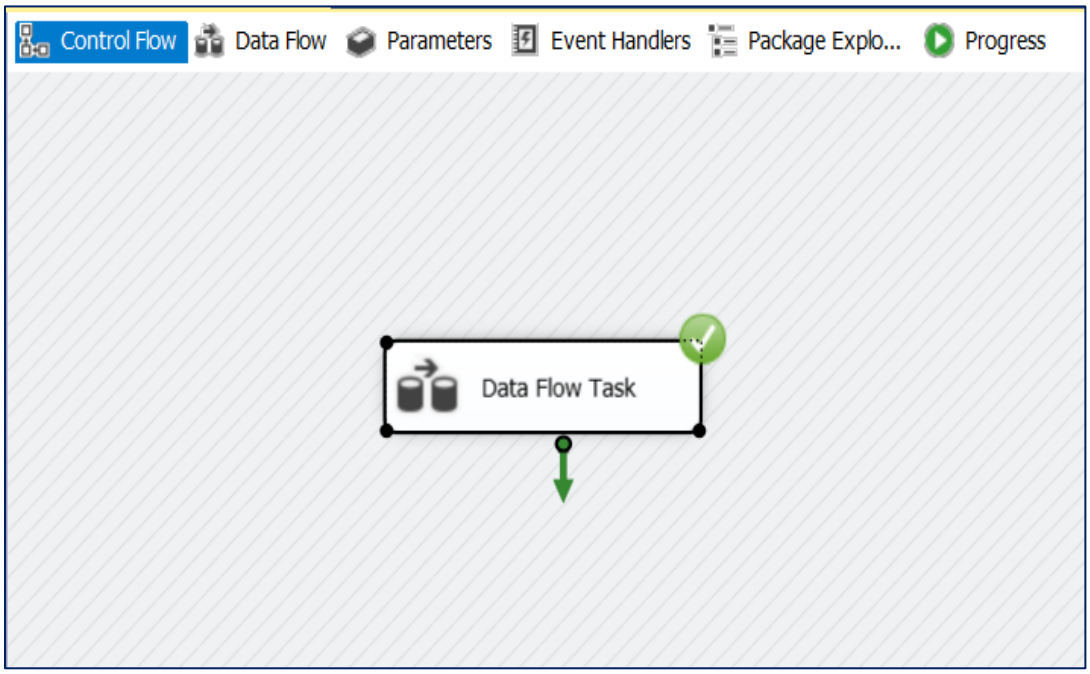
	A	B	C
1	tID	accm_txn_complete_time	
2	18538	5/11/2022 14:43	
3	13594	5/11/2022 8:55	
4	9692	5/10/2022 9:06	
5	16600	5/11/2022 9:57	
6	10272	5/10/2022 21:34	
7	21243	5/11/2022 6:43	
8	15818	5/10/2022 0:17	
9	7282	5/10/2022 22:47	
10	16283	5/10/2022 3:15	
11	21641	5/11/2022 23:49	
12	12588	5/11/2022 11:26	
13	24939	5/10/2022 23:27	
14	19178	5/11/2022 17:44	
15	5309	5/11/2022 16:35	
16	17671	5/11/2022 10:00	
17	7035	5/11/2022 9:42	
18	26329	5/11/2022 16:57	
19	26679	5/11/2022 18:07	
20	10260	5/10/2022 4:49	
21	20812	5/11/2022 3:36	
22	23133	5/10/2022 22:30	
23	26994	5/11/2022 17:02	
24	9229	5/10/2022 22:53	
25	17838	5/11/2022 18:10	
26	28097	5/10/2022 8:35	

accumulative Fact Table

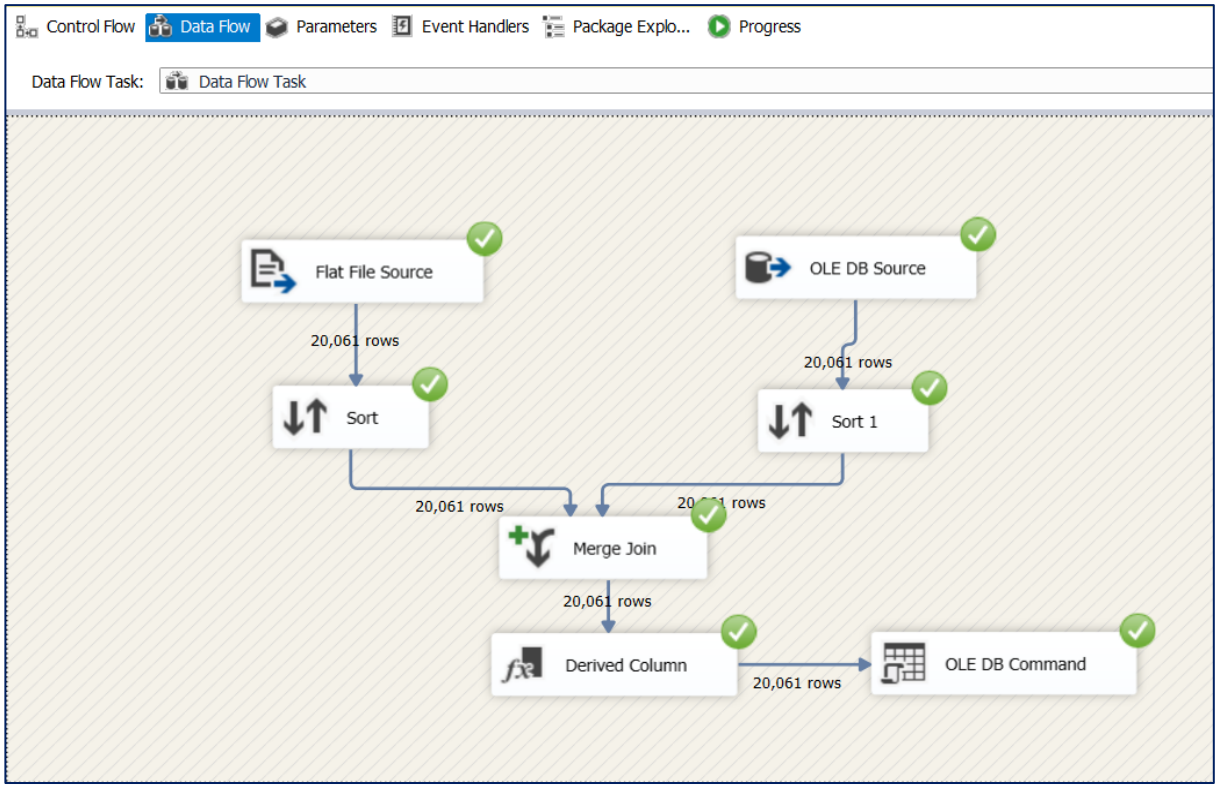
s

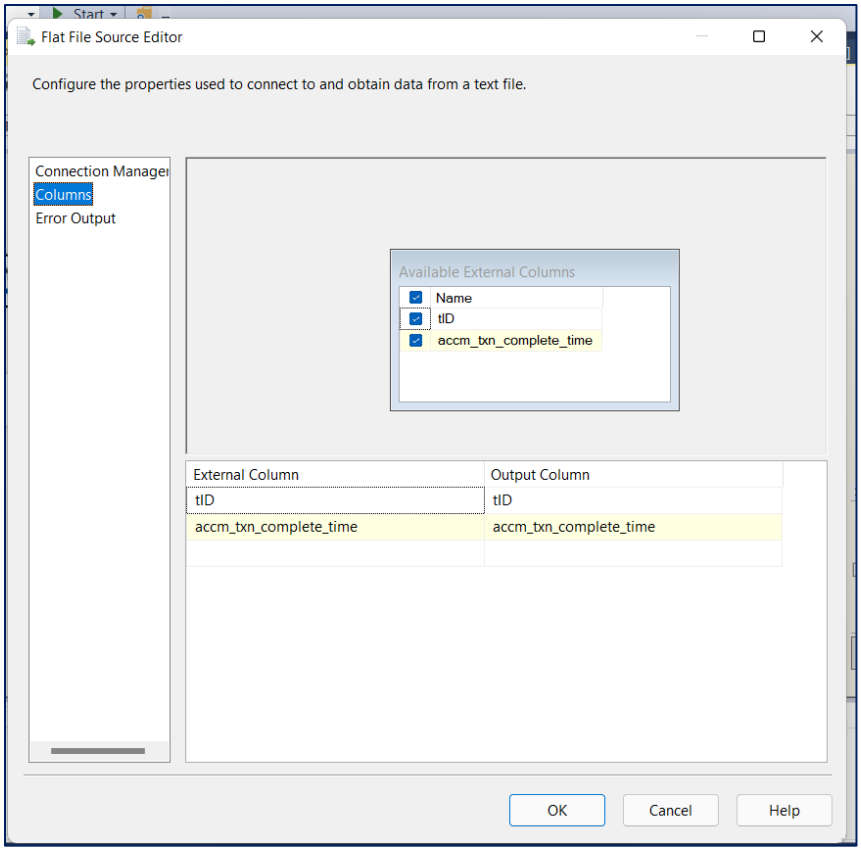
- Then I created a separate ETL SSIS package named as **Complete_time.dtsx.** Which reads data from this file and update the corresponding accm_txn_complete time in my DW Fact table.

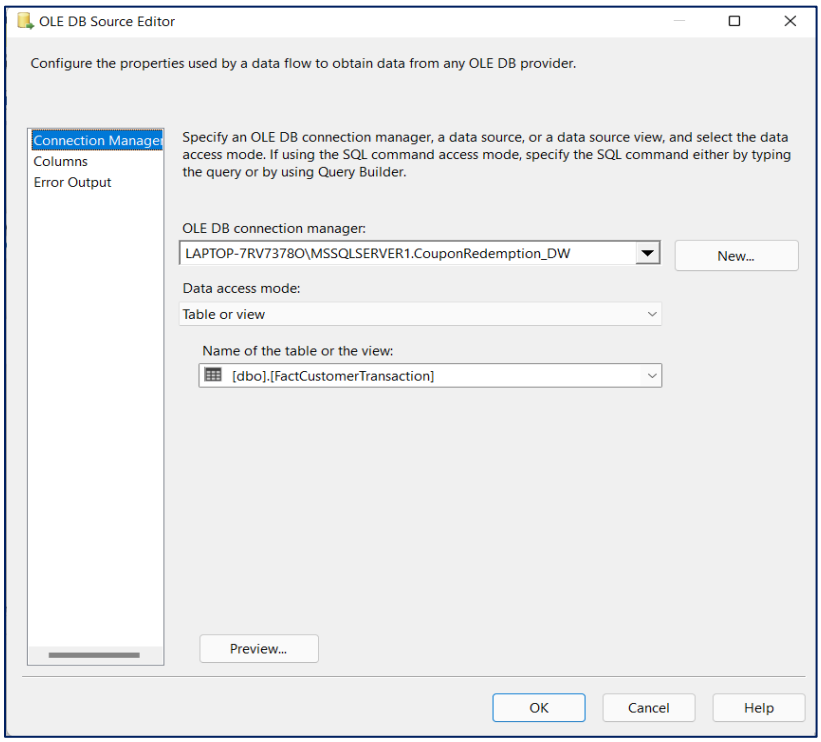

Solution 'CouponRedemption_SSIS_Project' (1 project)
CouponRedemption_SSIS_Project
Project.params
Connection Managers
SSIS Packages
CouponRedemption_Load_DW.dtsx
CouponRedemption_Load_Staging.dtsx
Data_Profiling.dtsx
Complete_time.dtsx
Package Parts
Control Flow
Miscellaneous
Linked Azure Resources
Azure-SSIS Integration Runtime
Azure Storage

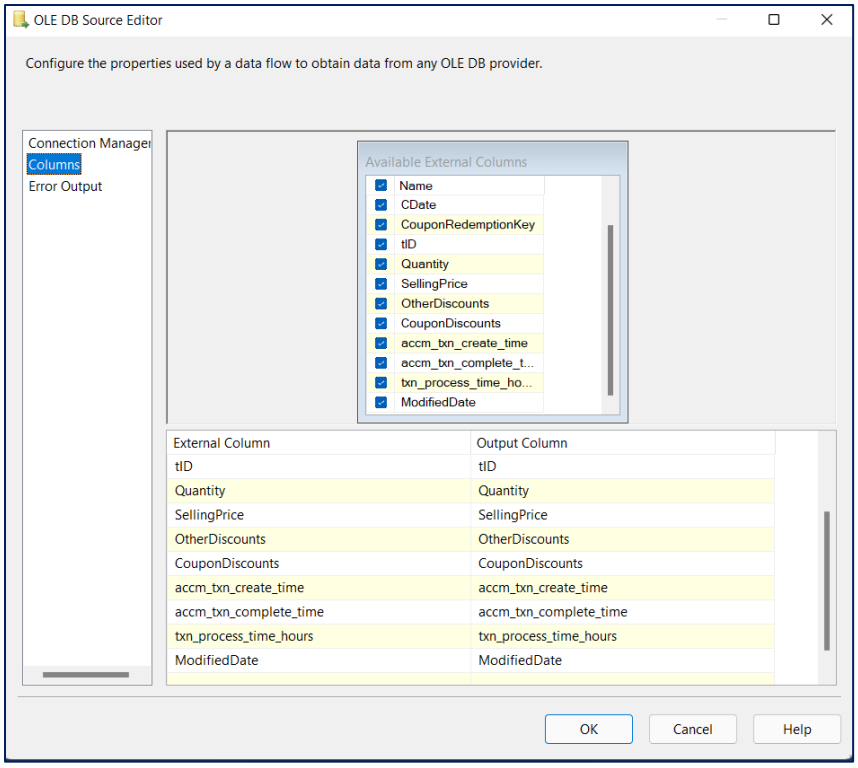

Control Flow
Data Flow
Parameters
Event Handlers
Package Explo...
Progress
Data Flow Task

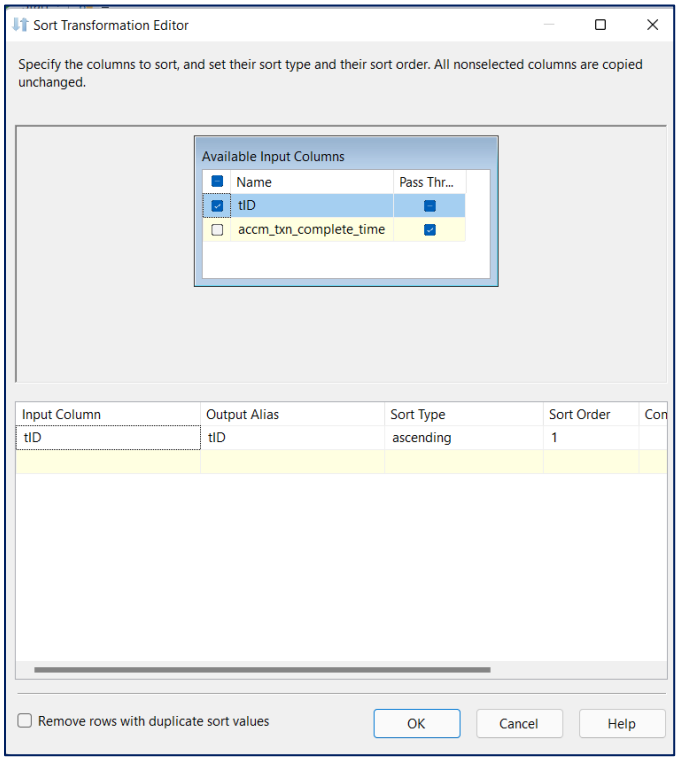
S

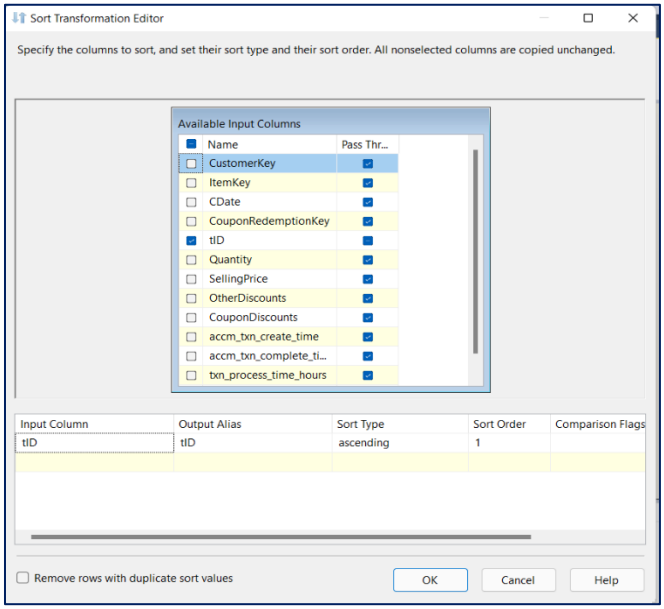

Control Flow
Data Flow
Parameters
Event Handlers
Package Explo...
Progress
Data Flow Task:
Data Flow Task
Flat File Source
OLE DB Source
20,061 rows
20,061 rows
Sort
Sort 1
20,061 rows
20,061 rows
Merge Join
20,061 rows
Derived Column
20,061 rows
OLE DB Command

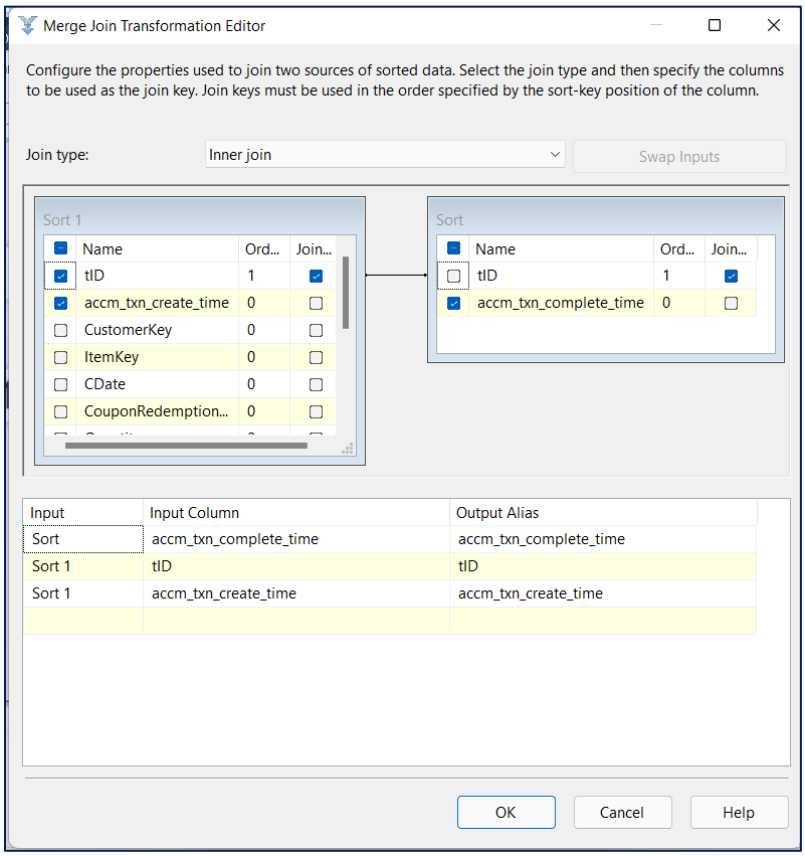

Flat File Source Editor
Configure the properties used to connect to and obtain data from a text file.
Connection Manager
Columns
Error Output
Available External Columns
Name
tID
accm_txn_complete_time
External Column
Output Column
tID
tID
accm_txn_complete_time
accm_txn_complete_time
OK
Cancel
Help

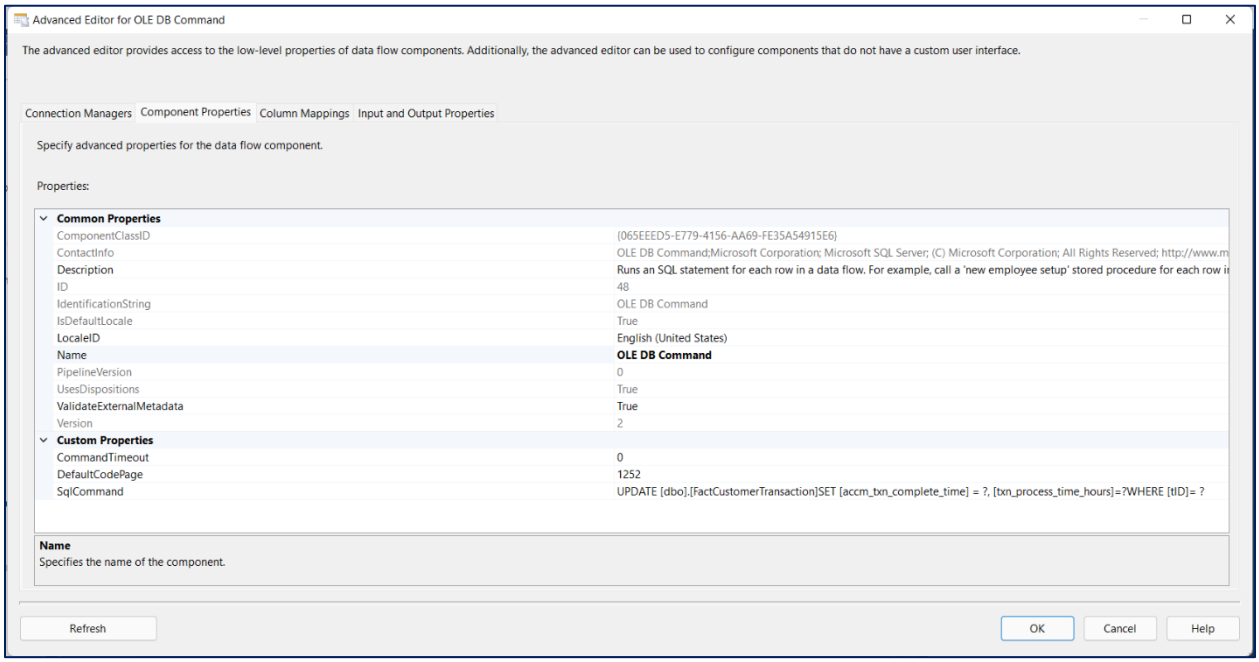

OLE DB Source Editor
Configure the properties used by a data flow to obtain data from any OLE DB provider.
Connection Manager
Columns
Error Output
Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder.
OLE DB connection manager:
LAPTOP-7RV7378O\MSSQLSERVER1.CouponRedemption_DW
New...
Data access mode:
Table or view
Name of the table or the view:
[dbo].[FactCustomerTransaction]
Preview...
OK
Cancel
Help


OLE DB Source Editor
Configure the properties used by a data flow to obtain data from any OLE DB provider.
Connection Manager
Columns
Error Output
Available External Columns
Name
CDate
CouponRedemptionKey
tID
Quantity
SellingPrice
OtherDiscounts
CouponDiscounts
accm_txn_create_time
accm_txn_complete_t...
txn_process_time_ho...
ModifiedDate
External Column
Output Column
tID
tID
Quantity
Quantity
SellingPrice
SellingPrice
OtherDiscounts
OtherDiscounts
CouponDiscounts
CouponDiscounts
accm_txn_create_time
accm_txn_create_time
accm_txn_complete_time
accm_txn_complete_time
txn_process_time_hours
txn_process_time_hours
ModifiedDate
ModifiedDate
OK
Cancel
Help


Sort Transformation Editor
Specify the columns to sort, and set their sort type and their sort order. All nonselected columns are copied unchanged.
Available Input Columns
Name
Pass Thr...
tID
accm_txn_complete_time
Input Column
Output Alias
Sort Type
Sort Order
Con
tID
tID
ascending
1
Remove rows with duplicate sort values
OK
Cancel
Help


Sort Transformation Editor
Specify the columns to sort, and set their sort type and their sort order. All nonselected columns are copied unchanged.
Available Input Columns
Name
Pass Thr...
CustomerKey
ItemKey
CDate
CouponRedemptionKey
tID
Quantity
SellingPrice
OtherDiscounts
CouponDiscounts
accm_txn_create_time
accm_txn_complete_ti...
txn_process_time_hours
Input Column
Output Alias
Sort Type
Sort Order
Comparison Flags
tID
tID
ascending
1
Remove rows with duplicate sort values
OK
Cancel
Help


Merge Join Transformation Editor
Configure the properties used to join two sources of sorted data. Select the join type and then specify the columns to be used as the join key. Join keys must be used in the order specified by the sort-key position of the column.
Join type:
Inner join
Swap Inputs
Sort 1
Name
Ord...
Join...
tID
1
accm_txn_create_time
0
CustomerKey
0
ItemKey
0
CDate
0
CouponRedemption...
0
Sort
Name
Ord...
Join...
tID
1
accm_txn_complete_time
0
Input
Input Column
Output Alias
Sort
accm_txn_complete_time
accm_txn_complete_time
Sort 1
tID
tID
Sort 1
accm_txn_create_time
accm_txn_create_time
OK
Cancel
Help


Advanced Editor for OLE DB Command
The advanced editor provides access to the low-level properties of data flow components. Additionally, the advanced editor can be used to configure components that do not have a custom user interface.
Connection Managers
Component Properties
Column Mappings
Input and Output Properties
Specify advanced properties for the data flow component.
Properties:
Common Properties
ComponentClassID
{065EEED5-E779-4156-AA69-FE35A54915E6}
ContactInfo
OLE DB Command;Microsoft Corporation; Microsoft SQL Server; (C) Microsoft Corporation; All Rights Reserved; http://www.m
Description
Runs an SQL statement for each row in a data flow. For example, call a 'new employee setup' stored procedure for each row i
ID
48
IdentificationString
OLE DB Command
IsDefaultLocale
True
LocaleID
English (United States)
Name
OLE DB Command
PipelineVersion
0
UsesDispositions
True
ValidateExternalMetadata
True
Version
2
Custom Properties
CommandTimeout
0
DefaultCodePage
1252
SqlCommand
UPDATE [dbo].[FactCustomerTransaction]SET [accm_txn_complete_time] = ?, [txn_process_time_hours]=?WHERE [tID]= ?
Name
Specifies the name of the component.
Refresh
OK
Cancel
Help

Final Control CouponRedemption_DW Control Flow.

Finally, successfully loaded data to the FactCustomerTransaction table with updated txn_process_time.

100 %

Results | Messages

	CustomerKey	ItemKey	CDate	CouponRedemptionKey	tID	Quantity	SellingPrice	OtherDiscounts	CouponDiscounts	accm_txn_create_time	accm_txn_complete_time	txn_process_time_hours	ModifiedDate
1	96	66136	20120112	53160	14739	1	71.2399978637695	-35.2599983215332	0	2022-05-09 09:17:34.387	2022-05-10 14:45:08.000	29	2022-05-09 09:17:34.387
2	123	21793	20120129	61733	14740	1	163.5	0	0	2022-05-09 09:17:34.387	2022-05-10 10:00:57.000	25	2022-05-09 09:17:34.387
3	NULL	50618	20120205	66891	14742	2	67.3199996948242	-67.3199996948242	0	2022-05-09 09:17:34.387	2022-05-11 16:41:12.000	55	2022-05-09 09:17:34.387
4	NULL	55106	20120131	62856	14744	1	78.0100021362305	0	0	2022-05-09 09:17:34.387	2022-05-11 20:11:36.000	59	2022-05-09 09:17:34.387
5	NULL	37175	20120115	54609	14745	1	71.2399978637695	-35.2599983215332	0	2022-05-09 09:17:34.387	2022-05-11 04:37:43.000	43	2022-05-09 09:17:34.387
6	NULL	6322	20120123	57972	14747	1	127.879997253418	0	0	2022-05-09 09:17:34.387	2022-05-10 12:13:14.000	27	2022-05-09 09:17:34.387
7	NULL	24187	20120202	64343	14750	1	213.360000610352	-53.4300003051758	0	2022-05-09 09:17:34.387	2022-05-10 07:51:21.000	22	2022-05-09 09:17:34.387
8	133	21275	20120124	58303	14751	1	28.1399993896484	-28.5	0	2022-05-09 09:17:34.387	2022-05-11 13:28:27.000	52	2022-05-09 09:17:34.387
9	NULL	7904	20120201	63599	14752	2	62.689998626709	0	0	2022-05-09 09:17:34.387	2022-05-11 18:57:02.000	57	2022-05-09 09:17:34.387
10	29	38195	20120203	65406	14755	1	67.3199996948242	-3.55999994277954	0	2022-05-09 09:17:34.387	2022-05-11 01:53:13.000	40	2022-05-09 09:17:34.387
11	579	32900	20120202	64246	14757	1	38.8300018310547	-21.3700008392334	0	2022-05-09 09:17:34.387	2022-05-11 07:57:46.000	46	2022-05-09 09:17:34.387
12	125	48273	20120209	50118	14759	1	31.7000007629395	-3.55999994277954	0	2022-05-09 09:17:34.387	2022-05-10 14:52:58.000	29	2022-05-09 09:17:34.387
13	580	70615	20120105	51043	14761	1	71.2399978637695	-35.2599983215332	0	2022-05-09 09:17:34.387	2022-05-11 17:36:45.000	56	2022-05-09 09:17:34.387
14	NULL	7052	20120209	50006	14764	1	117.190002441406	0	0	2022-05-09 09:17:34.387	2022-05-11 08:22:36.000	47	2022-05-09 09:17:34.387
15	394	41512	20120126	59226	14767	2	91.9000015258789	0	0	2022-05-09 09:17:34.387	2022-05-10 15:17:37.000	30	2022-05-09 09:17:34.387
16	220	41938	20120129	61952	14768	1	106.5	0	0	2022-05-09 09:17:34.387	2022-05-11 00:15:03.000	39	2022-05-09 09:17:34.387
17	NULL	32590	20120204	66238	14772	1	59.4900016784668	-39.8899993896484	0	2022-05-09 09:17:34.387	2022-05-10 11:16:59.000	26	2022-05-09 09:17:34.387
18	NULL	7547	20120120	56877	14773	8	712.400024414063	-139.630004882813	0	2022-05-09 09:17:34.387	2022-05-10 10:04:09.000	25	2022-05-09 09:17:34.387
19	NULL	64695	20120209	50009	14775	2	412.480010986328	-14.25	0	2022-05-09 09:17:34.387	2022-05-11 22:38:33.000	61	2022-05-09 09:17:34.387
20	NULL	71534	20120204	65661	14777	1	35.2599983215332	-28.5	0	2022-05-09 09:17:34.387	2022-05-11 15:57:17.000	54	2022-05-09 09:17:34.387
21	NULL	27625	20120130	62119	14778	1	106.5	-10.6899995803833	0	2022-05-09 09:17:34.387	2022-05-11 18:46:45.000	57	2022-05-09 09:17:34.387
22	NULL	5837	20120105	51119	14780	1	10.6899995803833	0	0	2022-05-09 09:17:34.387	2022-05-10 06:58:20.000	21	2022-05-09 09:17:34.387
23	707	43599	20120206	68159	14785	1	480.510009765625	-160.289993286133	0	2022-05-09 09:17:34.387	2022-05-11 03:49:50.000	42	2022-05-09 09:17:34.387
24	679	49248	20120112	53327	14786	1	53.4300003051758	-13.8900003433228	0	2022-05-09 09:17:34.387	2022-05-10 17:33:59.000	32	2022-05-09 09:17:34.387
25	679	28889	20120127	60166	14787	1	195.199996948242	0	0	2022-05-09 09:17:34.387	2022-05-11 03:07:47.000	42	2022-05-09 09:17:34.387
26	NULL	41249	20120128	61337	14789	1	67.3199996948242	0	0	2022-05-09 09:17:34.387	2022-05-10 05:46:58.000	20	2022-05-09 09:17:34.387
27	540	42079	20120209	50374	14790	1	45.9500007629395	-7.11999988555908	0	2022-05-09 09:17:34.387	2022-05-10 18:45:10.000	33	2022-05-09 09:17:34.387