



# Data Warehousing & Business Intelligence

## Assignment 1

**IT16029682**

**SILVA G.K.P**



# Catagory.csv

[illegible]

## SubCategory. csv

1	Sub-Category	ProductSubCategoryID	ProductCategoryID
2	Bookcases	1	1
3	Chairs	2	1
4	Furnishings	3	1
5	Tables	4	1
6	Appliances	5	2
7	Art	6	2
8	Binders	7	2
9	Envelopes	8	2
10	Fasteners	9	2
11	Labels	10	2
12	Paper	11	2
13	Storage	12	2
14	Supplies	13	2
15	Accessories	14	3
16	Copiers	15	3
17	Machines	16	3
18	Phones	17	3
19			
20			
21			
22			
23			
24			
25			

## OrderHeader.csv

1	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	SubTotal	TerritoryID
2	CA-2014-100006	9/7/2014	9/13/2014	Standard Class	DK-13375	377.97	2
3	CA-2014-100090	7/8/2014	7/12/2014	Standard Class	EB-13705	502.488	2
4	CA-2014-100293	3/14/2014	3/18/2014	Standard Class	FN-18475	91.056	3
5	CA-2014-100328	1/28/2014	2/3/2014	Standard Class	JC-15340	3.928	1
6	CA-2014-100363	4/8/2014	4/15/2014	Standard Class	JM-15655	2.368	2
7	CA-2014-100391	5/25/2014	5/29/2014	Standard Class	BW-11065	14.62	4
8	CA-2014-100678	4/18/2014	4/22/2014	Standard Class	KM-16720	2.688	1
9	CA-2014-100706	12/16/2014	12/18/2014	Second Class	LE-16810	99.98	1
10	CA-2014-100762	11/24/2014	11/29/2014	Standard Class	NG-18355	151.92	2
11	CA-2014-100860	3/26/2014	3/30/2014	Second Class	CS-12505	18.75	4
12	CA-2014-100867	10/19/2014	10/24/2014	Standard Class	EH-14125	321.552	3
13	CA-2014-100881	3/28/2014	4/1/2014	Standard Class	DR-12940	302.376	2
14	CA-2014-100895	6/2/2014	6/6/2014	Standard Class	SV-20785	8.56	1
15	CA-2014-100916	10/21/2014	10/26/2014	Standard Class	FH-14275	194.7	3
16	CA-2014-100972	11/19/2014	11/24/2014	Second Class	DB-13360	166.44	2
17	CA-2014-101147	12/2/2014	12/4/2014	First Class	MC-17575	2.394	2
18	CA-2014-101175	12/9/2014	12/14/2014	Standard Class	DM-12955	100.704	3
19	CA-2014-101266	8/27/2014	8/30/2014	Second Class	MM-17920	13.36	1
20	CA-2014-101364	12/22/2014	12/26/2014	Standard Class	TW-21025	296.712	1
21	CA-2014-101392	7/27/2014	12/13/2014	Standard Class	AS-10630	269.36	3
22	CA-2014-101427	12/26/2014	12/30/2014	Standard Class	AY-10555	8.016	1
23	CA-2014-101462	4/20/2014	4/25/2014	Standard Class	BP-11230	59.92	1
24	CA-2014-101476	9/12/2014	9/13/2014	First Class	SD-20485	69.99	2
25	CA-2014-101560	11/28/2014	12/1/2014	Second Class	CS-12250	397.6	4

## OrderDetail.csv

1	OrderDetail ID	Order ID	Product ID	LineTotal	Quantity	Discount	UnitPrice	Profit
2		1 CA-2016-152156	13	261.96	2	0	130.98	41.9136
3		2 CA-2016-152156	56	731.94	3	0	243.98	219.582
4		3 CA-2016-138688	947	14.62	2	0	7.31	6.8714
5		4 US-2015-108966	320	957.575	5	0.45	348.21	-383.03
6		5 US-2015-108966	1317	22.368	2	0.2	13.98	2.5164
7		6 CA-2014-115812	186	48.86	7	0	6.98	14.1694
8		7 CA-2014-115812	563	7.28	4	0	1.82	1.9656
9		8 CA-2014-115812	1762	907.152	6	0.2	188.99	90.7152
10		9 CA-2014-115812	795	18.504	3	0.2	7.71	5.7825
11		10 CA-2014-115812	438	114.9	5	0	22.98	34.47
12		11 CA-2014-115812	329	1706.184	9	0.2	236.97	85.3092
13		12 CA-2014-115812	1751	911.424	4	0.2	284.82	68.3568
14		13 CA-2017-114412	1168	15.552	3	0.2	6.48	5.4432
15		14 CA-2016-161389	782	407.976	3	0.2	169.99	132.5922
16		15 US-2015-118983	422	68.81	5	0.8	68.81	-123.858
17		16 US-2015-118983	664	2.544	3	0.8	4.24	-3.816
18		17 CA-2014-105893	1409	665.88	6	0	110.98	13.3176
19		18 CA-2014-167164	1295	55.5	2	0	27.75	9.99
20		19 CA-2014-143336	568	8.56	2	0	4.28	2.4824
21		20 CA-2014-143336	1750	213.48	3	0.2	88.95	16.011
22		21 CA-2014-143336	731	22.72	4	0.2	7.1	7.384
23		22 CA-2016-137330	478	19.46	7	0	2.78	5.0596
24		23 CA-2016-137330	409	60.34	7	0	8.62	15.6884
25		24 US-2017-156909	96	71.372	2	0.3	50.98	-1.0196

## SalePerson.csv

	A	B	C
1	Person	TerritoryID	
2	Anna Andreadi	1	
3	Chuck Magee	2	
4	Kelly Williams	3	
5	Cassandra Brandt	4	
6			
7			
8			

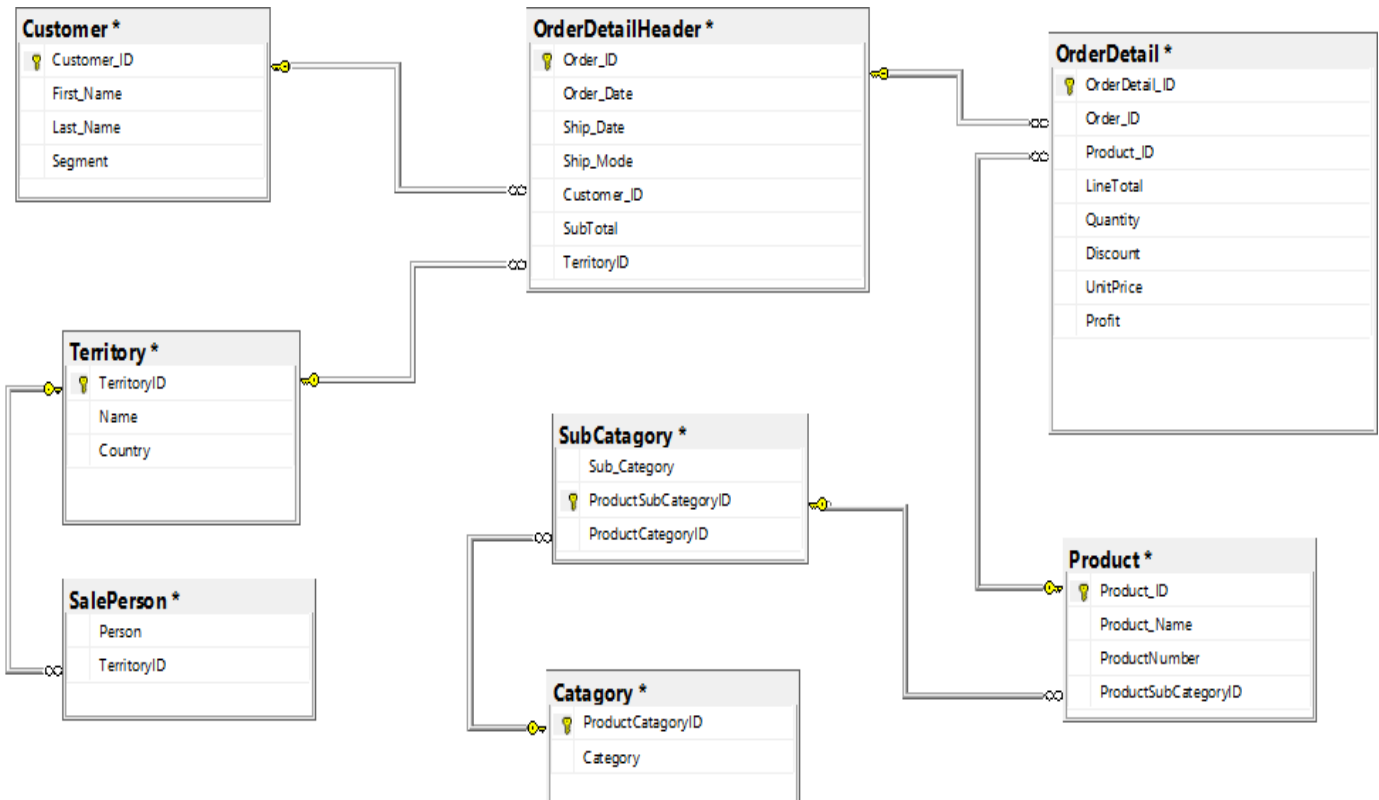
## Territory.csv

	TerritoryID	Name	Country
1	1	West	United States
2	2	East	United States
3	3	Central	United States
4	4	South	United States
5			
6			
7			
8			

# Import CSV File into SQL Server 2016 as Separates Tables

## Sources

### 1. StoreSourceDB (Database Name)



### 2. Customer Address Detail.txt (Text File Name)

Customer Address Details.txt - Notepad

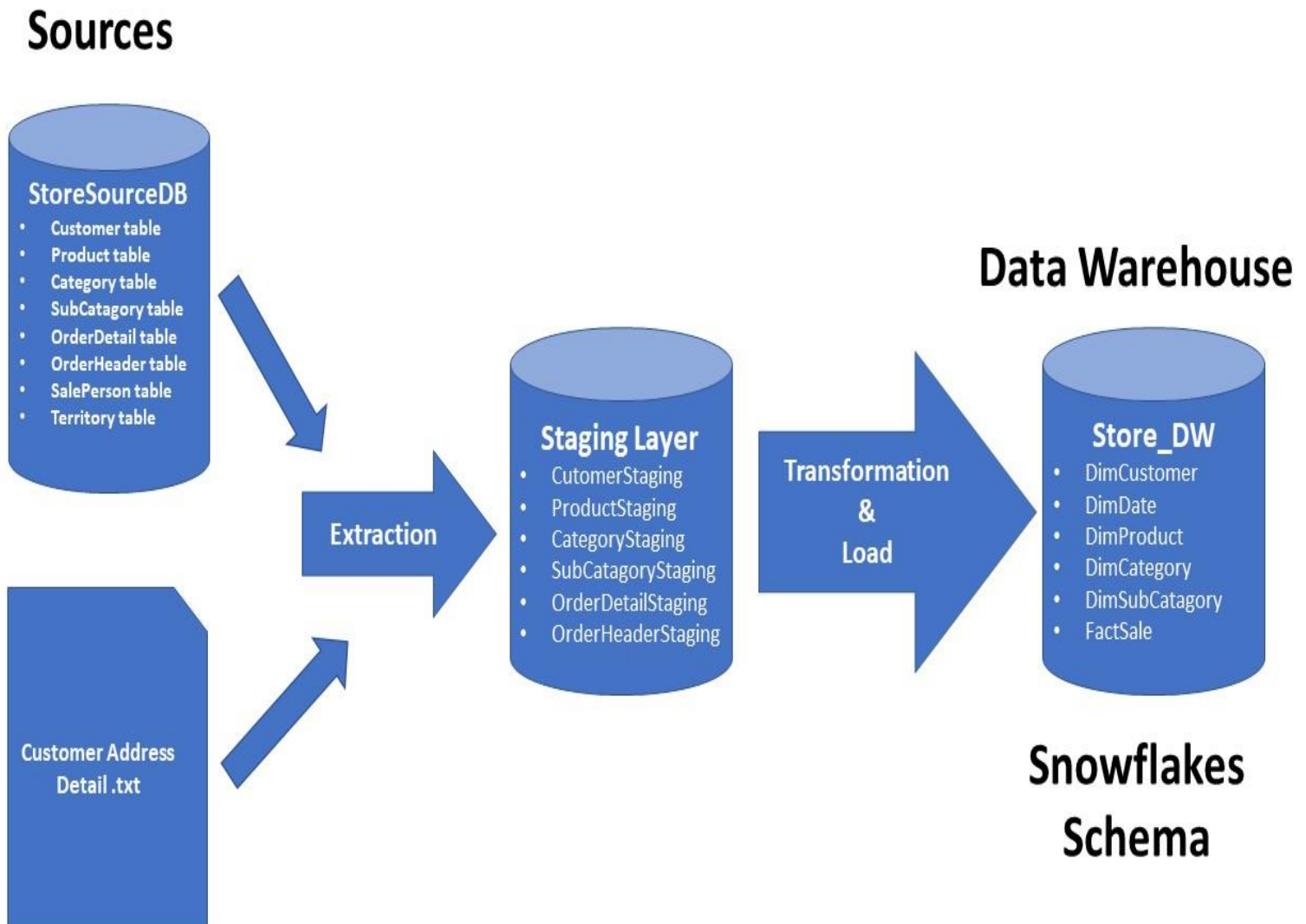
Customer ID	Country	City	State	Postal Code	Region		
CG-12520	United States	Henderson	Kentucky		42420	South	
DV-13045	United States	Los Angeles	California		90036	West	
SO-20335	United States	Fort Lauderdale	Florida	33311		South	
BH-11710	United States	Los Angeles	California		90032	West	
AA-10480	United States	Concord	North Carolina	28027		South	
IM-15070	United States	Seattle	Washington		98103	West	
HP-14815	United States	Fort Worth	Texas	76106		Central	
PK-19075	United States	Madison	Wisconsin	53711		Central	
AG-10270	United States	West Jordan	Utah	84004		West	
ZD-21925	United States	San Francisco	California		94109	West	
KB-16585	United States	Fremont	Nebraska	68025		Central	
SF-20065	United States	Philadelphia	Pennsylvania		19140	East	
EB-13070	United States	Orem	Utah	84057		West	
EH-13945	United States	Los Angeles	California		90049	West	
TB-21520	United States	Philadelphia	Pennsylvania		19140	East	
MA-17560	United States	Houston	Texas	77095		Central	
GH-14485	United States	Richardson	Texas	75080		Central	
SH-20710	United States	Houston	Texas	77041		Central	
LC-16930	United States	Naperville	Illinois		60540	Central	
RA-19885	United States	Los Angeles	California		90049	West	
ES-14080	United States	Melbourne	Florida	32935		South	
OM-18715	United States	Eagan	Minnesota	55122		Central	
PO-18865	United States	Westland	Michigan		48185	Central	
LH-16900	United States	Dover	Delaware	19901		East	
DP-13000	United States	New Albany	Indiana	47150		Central	
JH-15265	United States	New York City	New York		10024	East	
TB-21055	United States	Troy	New York	12180		East	
KM-16720	United States	Los Angeles	California		90004	West	
PS-18970	United States	Chicago	Illinois	60610		Central	
BS-11590	United States	Gilbert	Arizona	85234		West	
KD-16270	United States	Springfield	Virginia		22153	South	
HW-14980	United States	New York City	New York		10009	East	
JE-15745	United States	Memphis	Tennessee	38109		South	
KB-16600	United States	Houston	Texas	77041		Central	
SC-20770	United States	Decatur	Alabama	35601		South	
DM-13690	United States	San Francisco	California		94122	West	
JC-16105	United States	Durham	North Carolina	27707		South	
CS-12400	United States	Chicago	Illinois	60623		Central	
PG-18895	United States	Rochester	Minnesota		55901	Central	
GM-14455	United States	Houston	Texas	77095		Central	
JS-15685	United States	Los Angeles	California		90036	West	
KB-16315	United States	Minneapolis	Minnesota		55407	Central	
RB-19705	United States	Portland	Oregon	97206		West	
NU-18770	United States	New York City	New York		10000	East	

## Sources Data Type and Description

Source	Source Type	Object Name	Scheme	Object Type	Description
Store_SourceDB	SQL Database	Customer	dbo	Table	Includes all customer information
		Product	dbo	Table	Includes all product details
		ProductCategory	dbo	Table	Includes the categories of all products described in Product table
		ProductSubCategory	dbo	Table	Includes the product sub categories of all products described in product table
		OrderHeader	dbo	Table	Header table of all sales transactions.
		OrderDetails	dbo	Table	Sub table of the SalesOrderHeader. Includes the details of each sales order
		SalesPerson	dbo	Table	Includes the sales person details
CustomerAddressDetail.txt	Text File	Territory	dbo	Table	Includes the territory information of the business
					Includes the customer address details

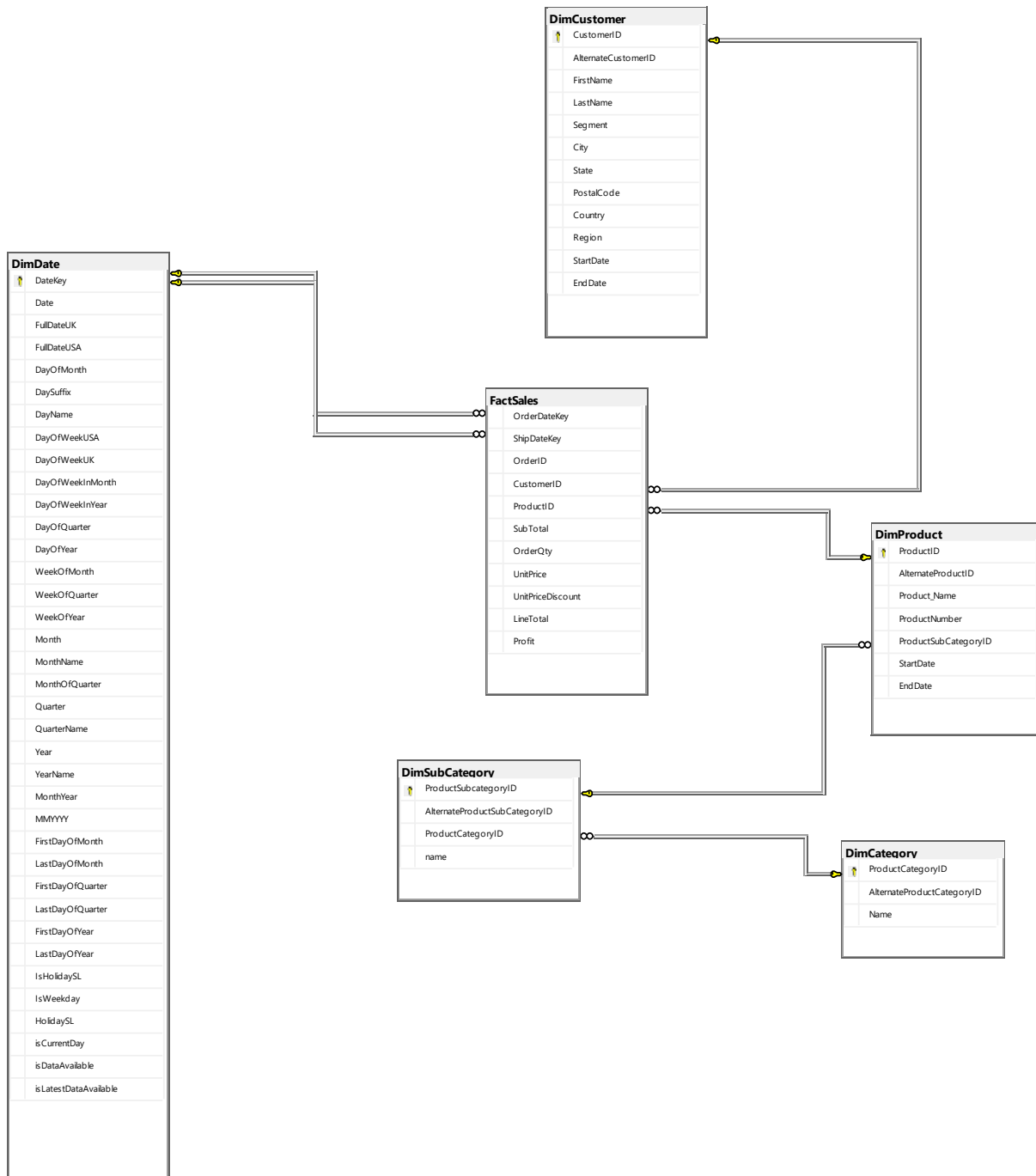
Source	Source Type	Table Name	Column Name	Data Type	Link Table	Link Column	Description
Store_SourceDB	SQL Database	Customer	CustomerID	nvarchar(50)			Unique ID.
			First_Name	nvarchar(50)			First name of the customer
			Last_Name	nvarchar(50)			Last name of the customer
			Segment	nvarchar(50)			customer Type Coparate or Consumer or Home Office
		Product	ProductID	Int			Unique ID.
			Product_Name	nvarchar(150)			Product Name
			ProductNumber	nvarchar(50)			Product Number
			ProductSubCategoryID	Int	SubCategory	ProductSubCategoryID	Product sub category
		Category	ProductCategoryID	Int			Unique ID.
			Category	nvarchar(50)			Name of the product Category
		SubCategory	ProductSubCategoryID	Int			Unique ID.
			ProductCategoryID	Int	Category	ProductCategoryID	Category of the product
			Sub_Category	nvarchar(50)			Name of the product sub category
		OrderDetail	OrderDetail_ID	Int			Unique ID.
			OrderID	nvarchar(50)	OrderHeader	OrderID	Sales order ID
			Quantity	Int			Quantity of the ordered item
			Product_ID	Int	Product	ProductID	Id of the product
			UnitPrice	Money			Unit price of the item
			Discount	float			Discount amount provided for the item
		OrderHeader	LineTotal	Money			Per product subtotal. Computed as UnitPrice * (1 - Discount) * Quantity.
			Profit	Money			Profit of Per product sub total
			Order_ID	nvarchar(50)			Unique ID.
			Order_Date	datetime2(7)			Date the order is made
			Ship_Date	datetime2(7)			Order shipping date
			Customer_ID	Int	Customer	CustomerID	Order placed customer
		SalesPerson	TerritoryID	Int	Territory	TerritoryID	Region the order took place in
			SubTotal	Money			Sub Total
			Ship_Mode	nvarchar(50)			Shipping Mode
		Territory	TerritoryID	Int			Unique ID.
			Name	nvarchar(50)			Name of the territory
			Country	nvarchar(50)			Country name
Customer Address Detail.txt	Text File	Customer Address	CustomerID	nvarchar(50)	Customer	CustomerID	Unique ID.
			City	nvarchar(50)			City of the customer
			State	nvarchar(50)			State name of the customer
			PostalCode	nvarchar(50)			Postal code of the customer
			Region	nvarchar(50)			Country Region name of the customer
			Country	nvarchar(50)			Country name of the customer

# Architectural diagram

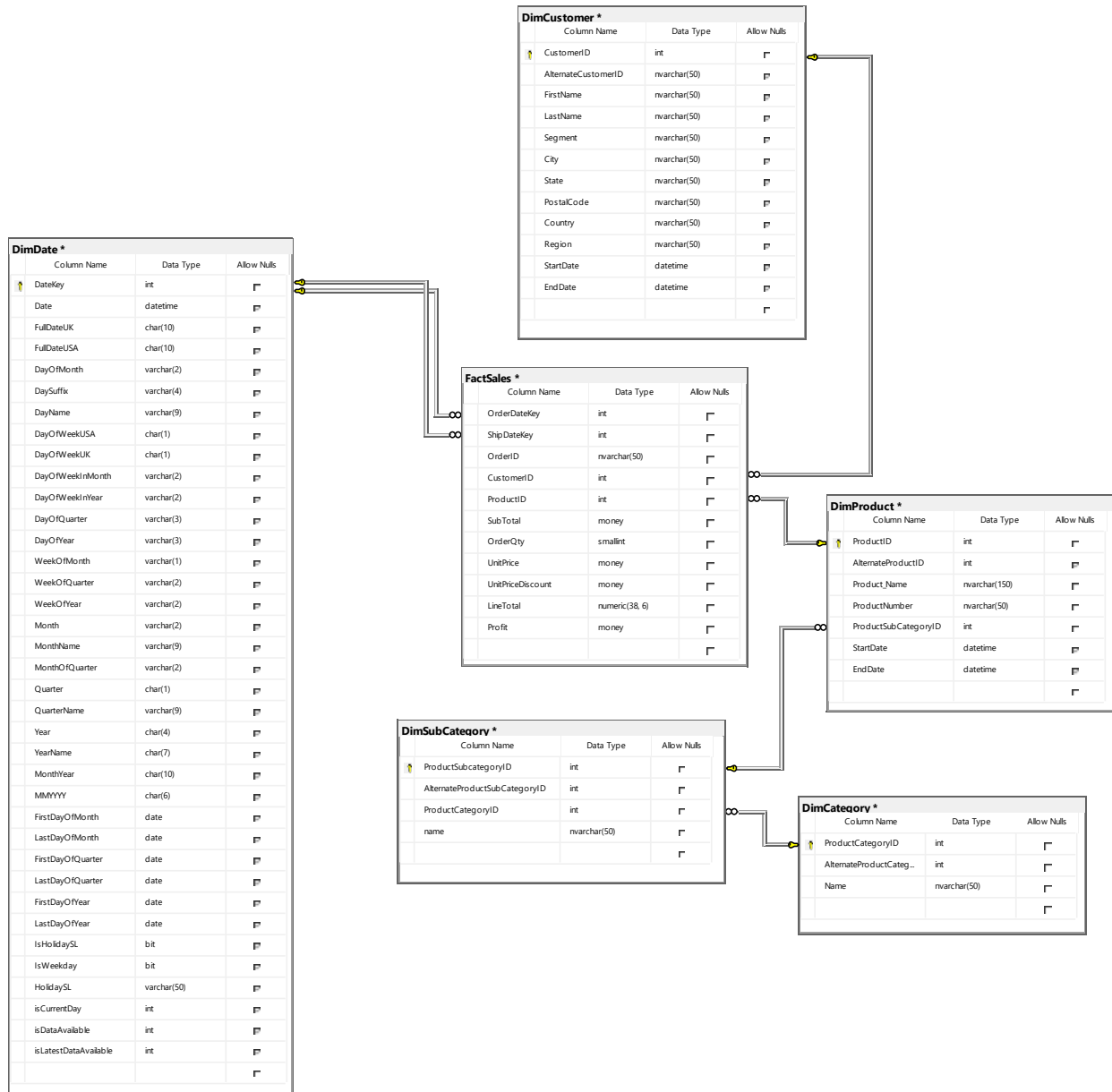


# Data Warehouse Design(Snowflakes Schema)

Store\_DW(warehouse Database name)



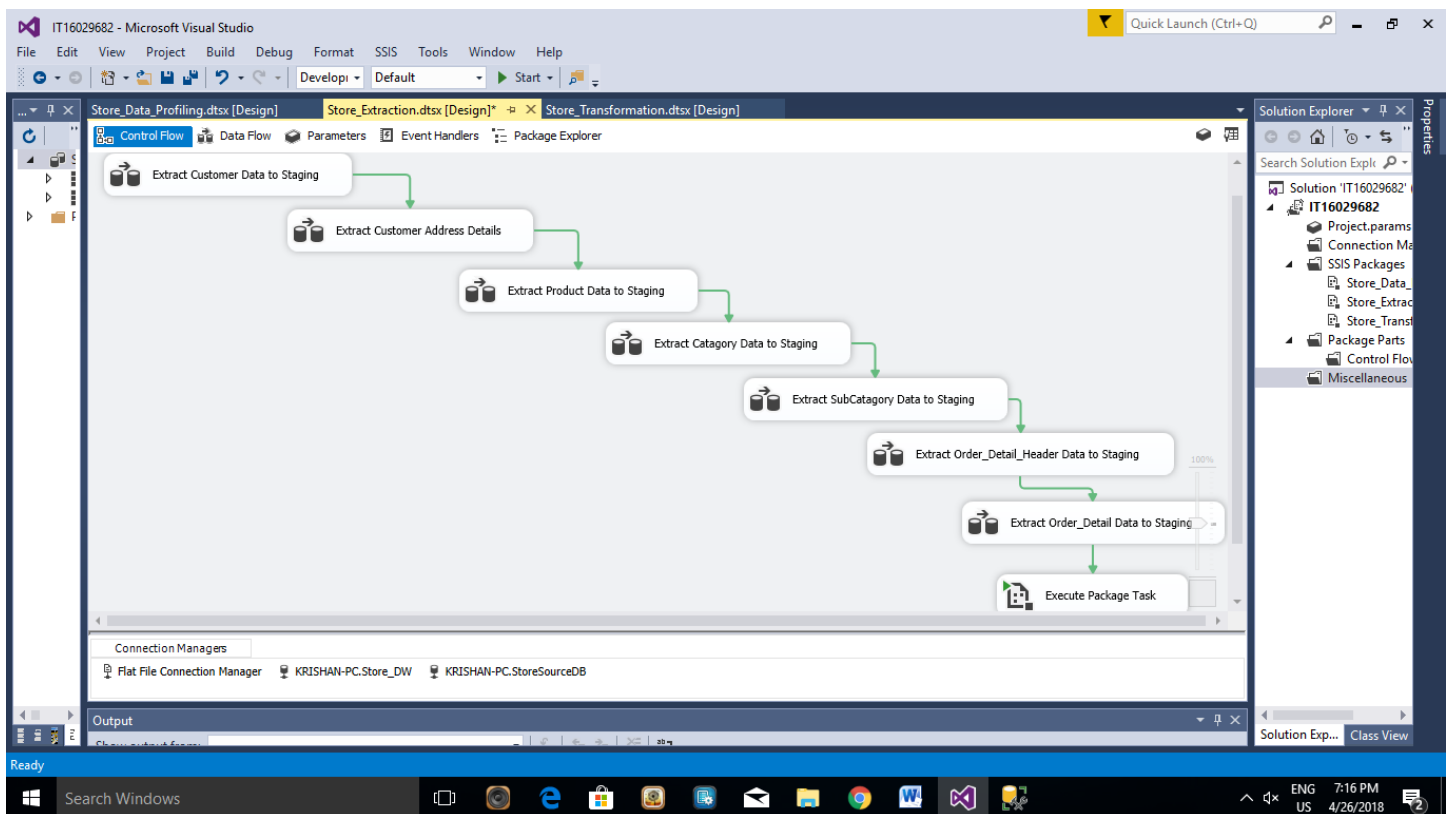
# Data Warehouse Data Type



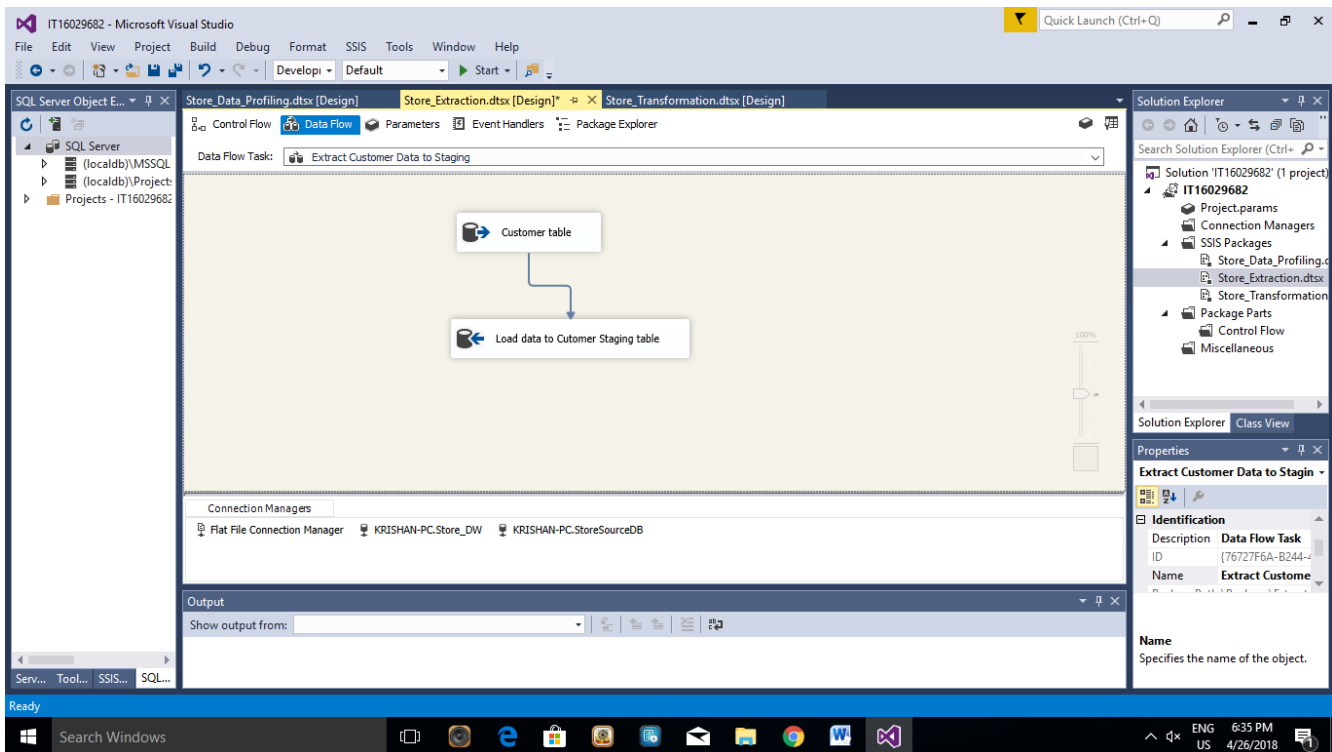


# ETL development

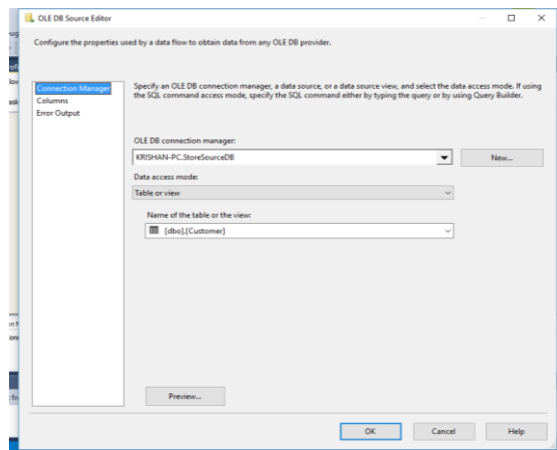
## 1.Data Extraction (Control flow)



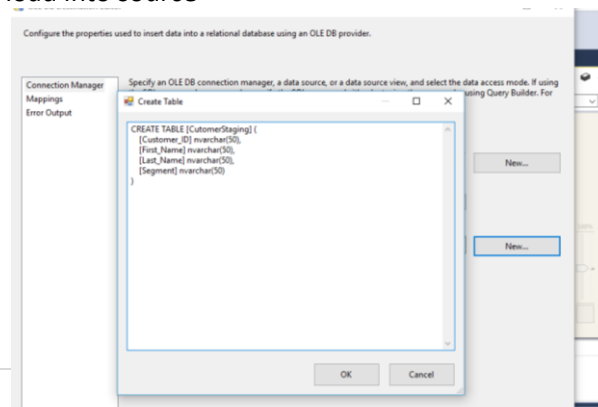
## 1.1 Customer Extraction (Data flow)



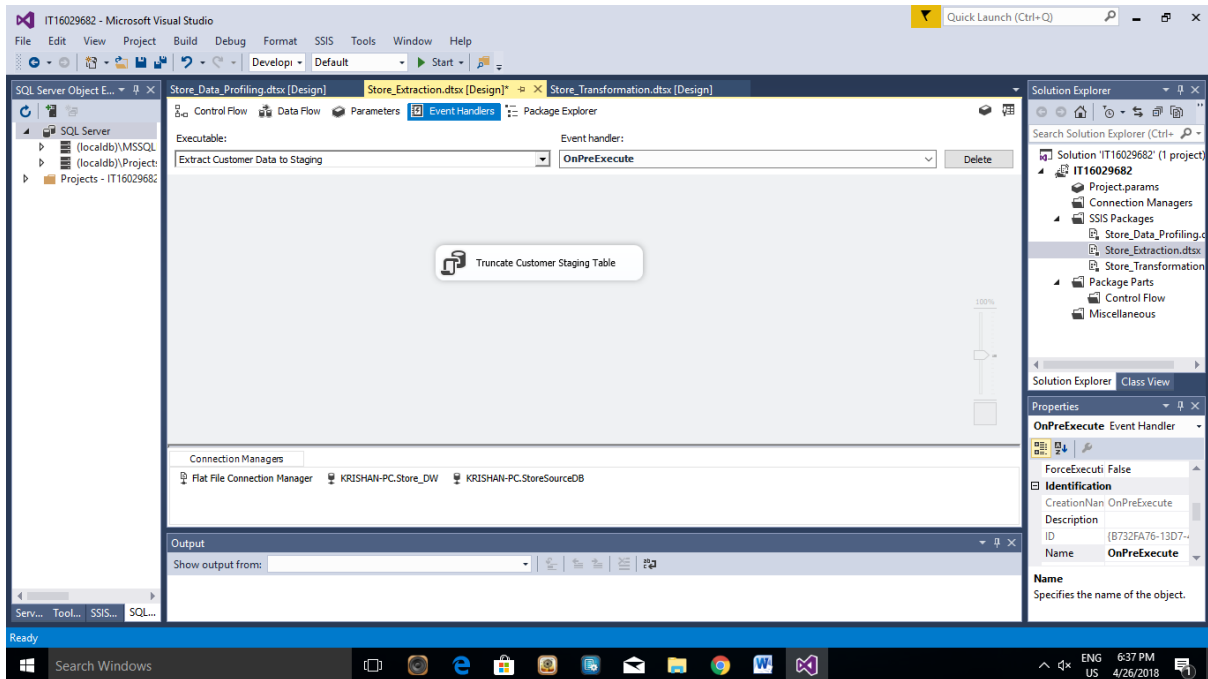
- Used OLE DB Source as **dbo.Customer** table in **Store\_SourceDB**



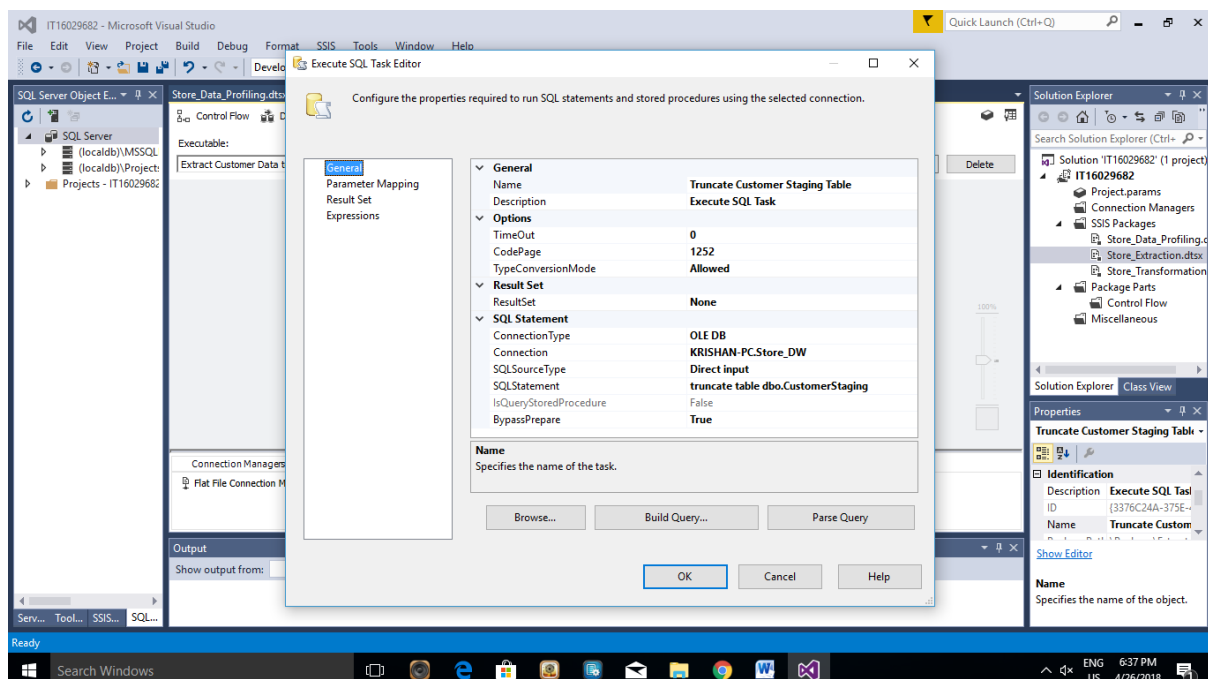
- OLE DB Destination for create new table as **CustomerStaging** in the **Store\_DW** database this table to data load into source



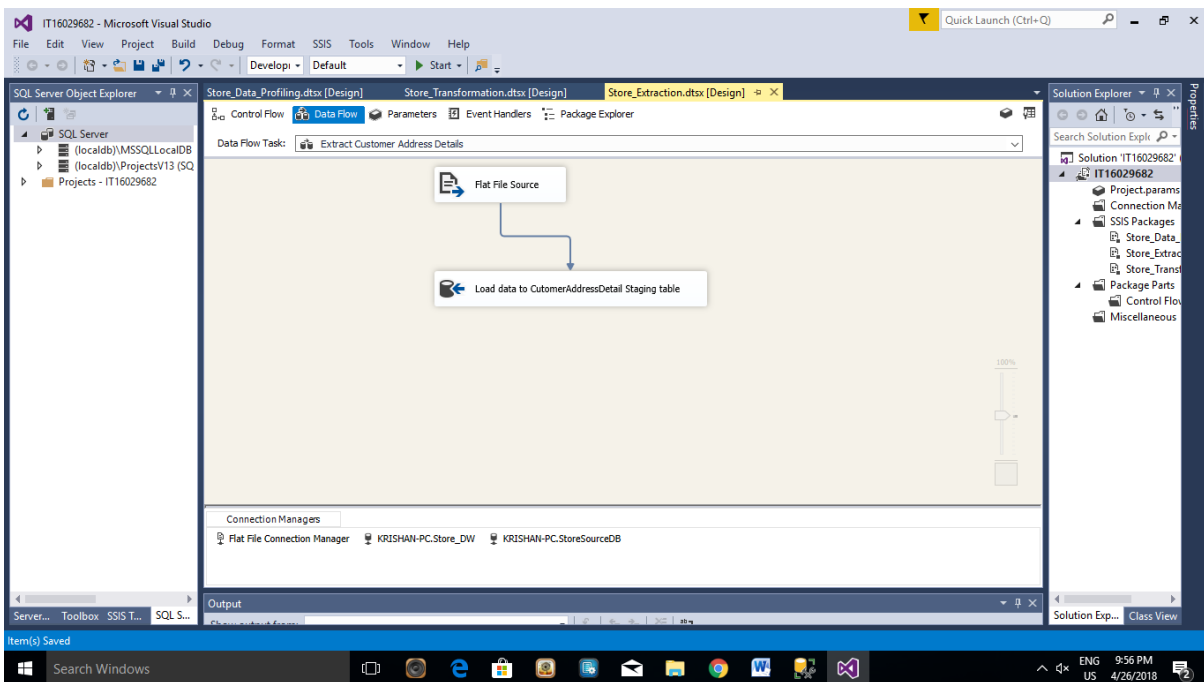
## 1.2 Customer Extraction (Event Handlers)



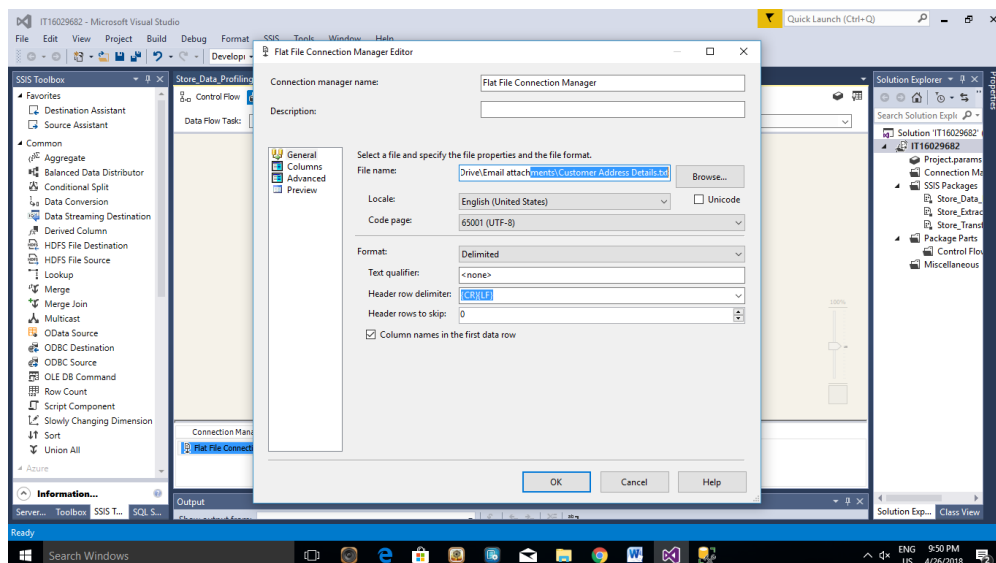
- Used **Execute SQL Task** SSIS tool Truncate table for sql command as **truncate table dbo.CustomerStaging** in **Store\_DW** Database



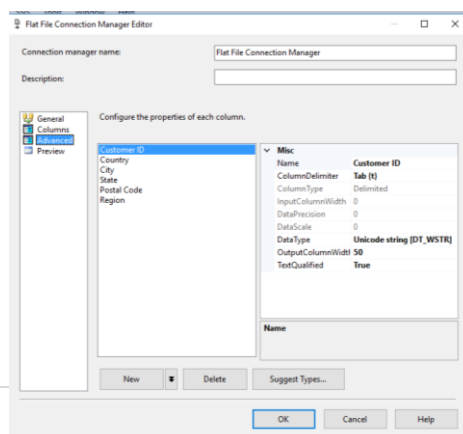
## 2.1 Customer Address Details Extraction (Data flow)



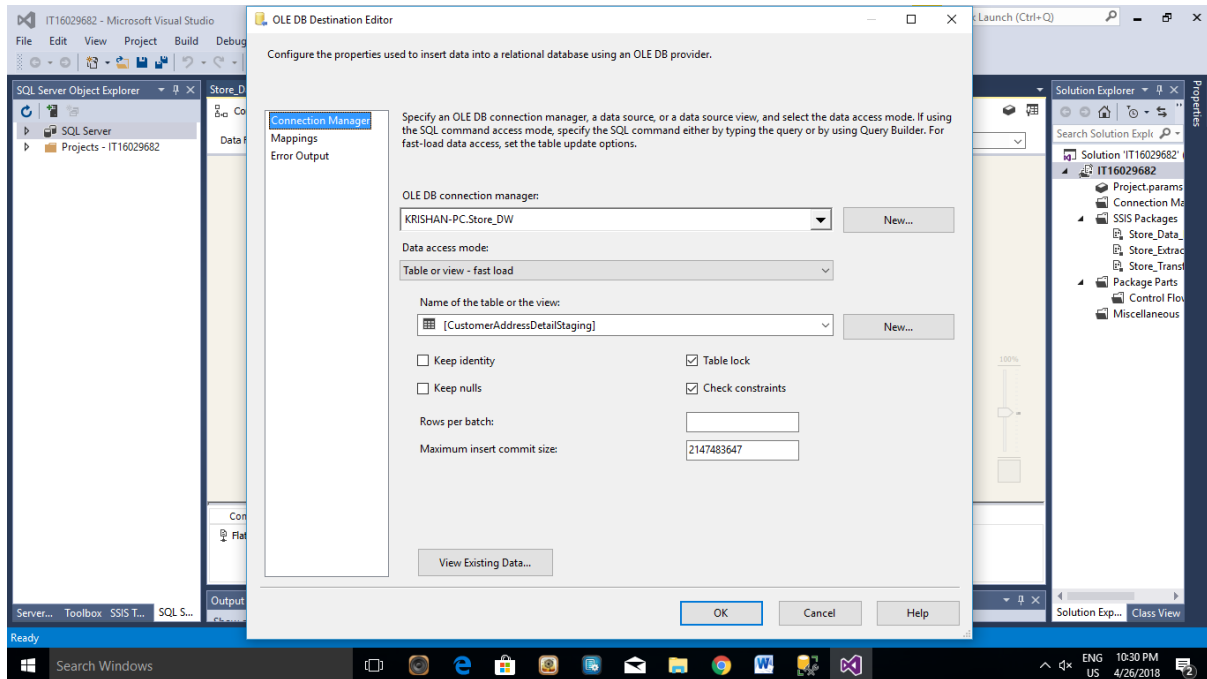
- Used **Flat file Source** SSIS tool , load data into **CustomerAddressDetail.txt** Text file



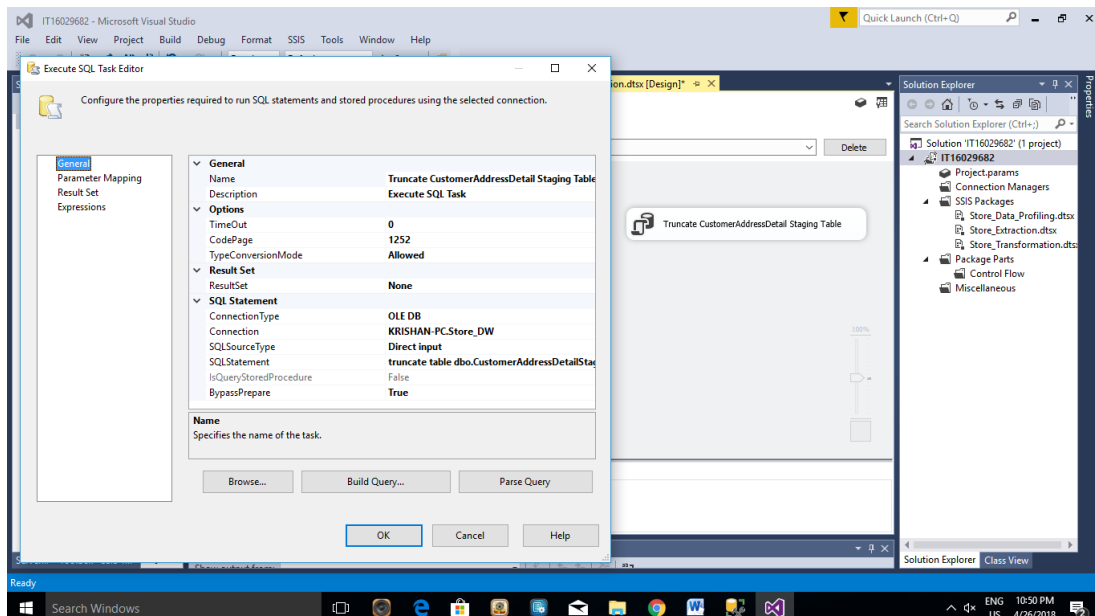
- Advanced tab, text file data has columns then every column data type select as **Unicode string[DT\_WSTR]**



- Used **OLE DB Destination** SSIS tool to create new table as **CustomerAddressDetailStaging** load text file data to **Store\_DW** Database

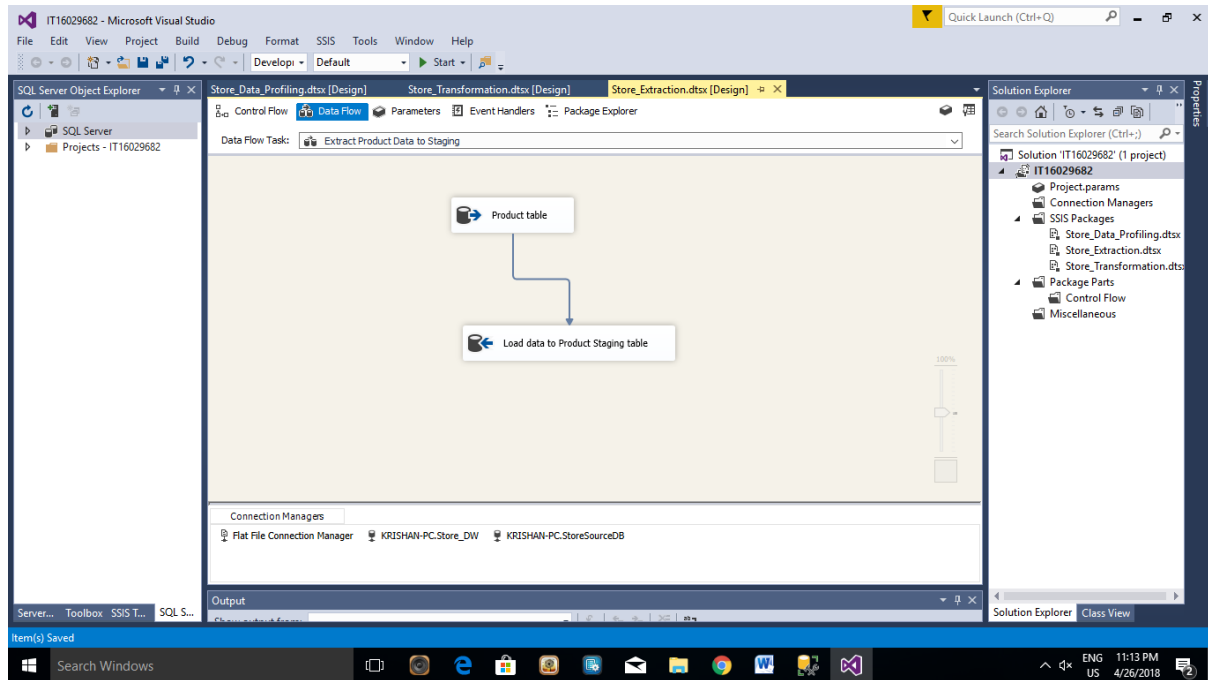


## 2.2 CustomerAddressDetail (Event Handlers)

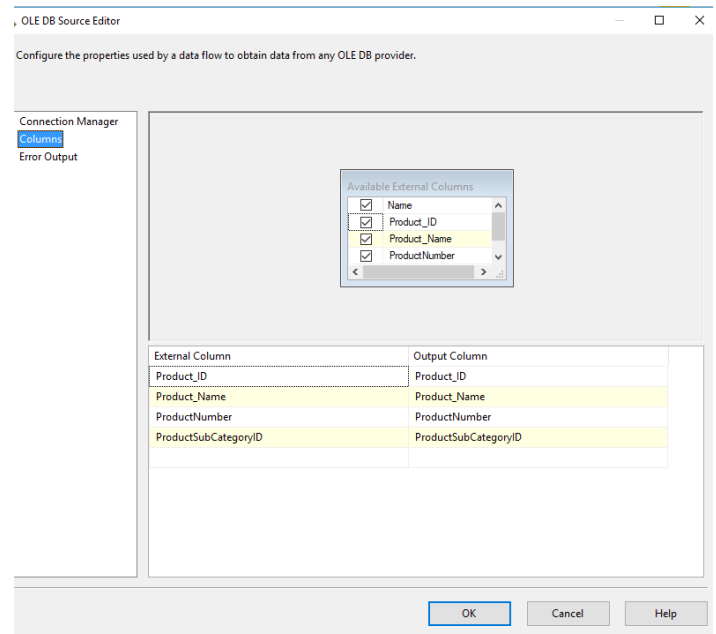
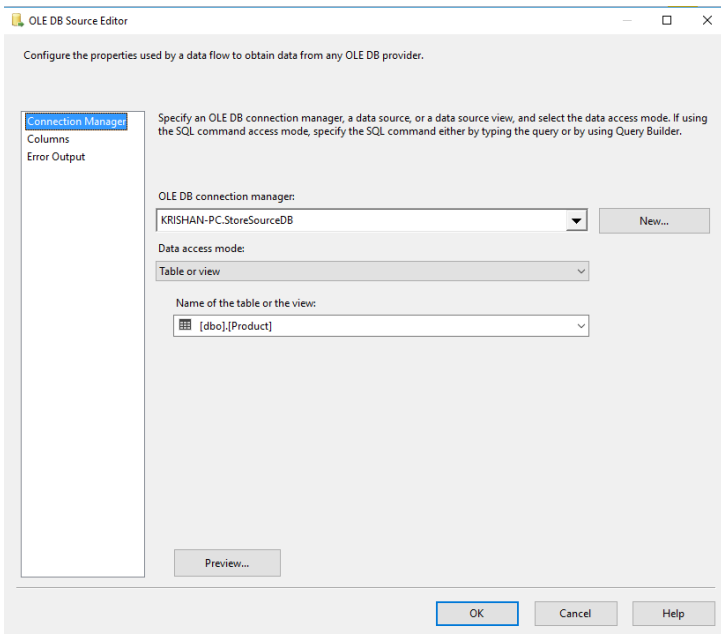


- Used **Execute SQL Task** SSIS tool Truncate table for sql command as **truncate table dbo.CustomerAddressStaging** in **Store\_DW** Database

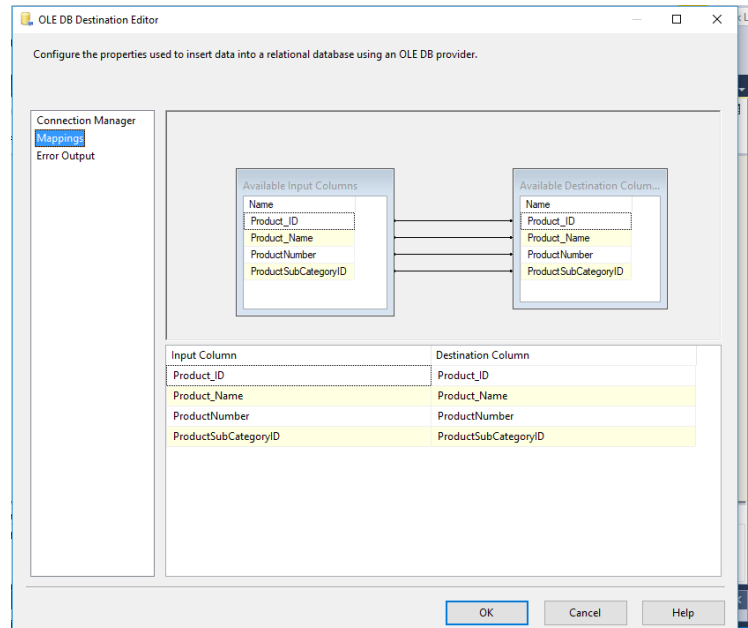
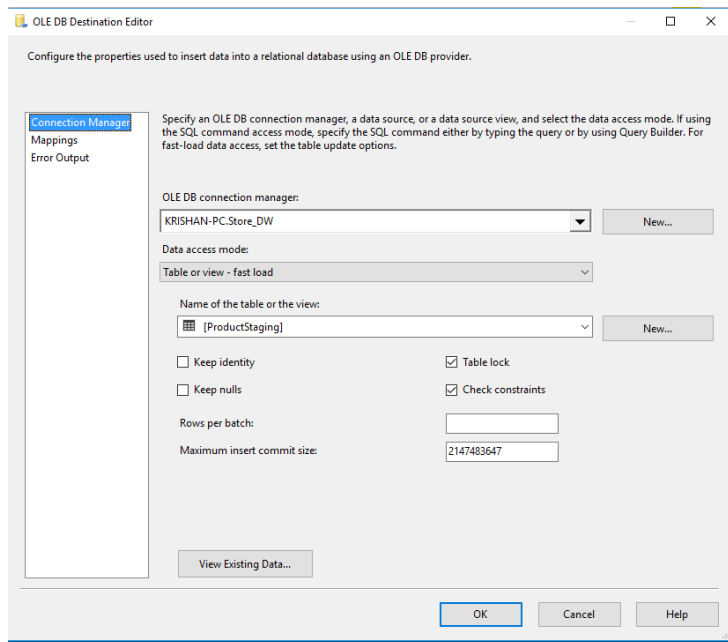
### 3.1 Product Extraction (Data flow)



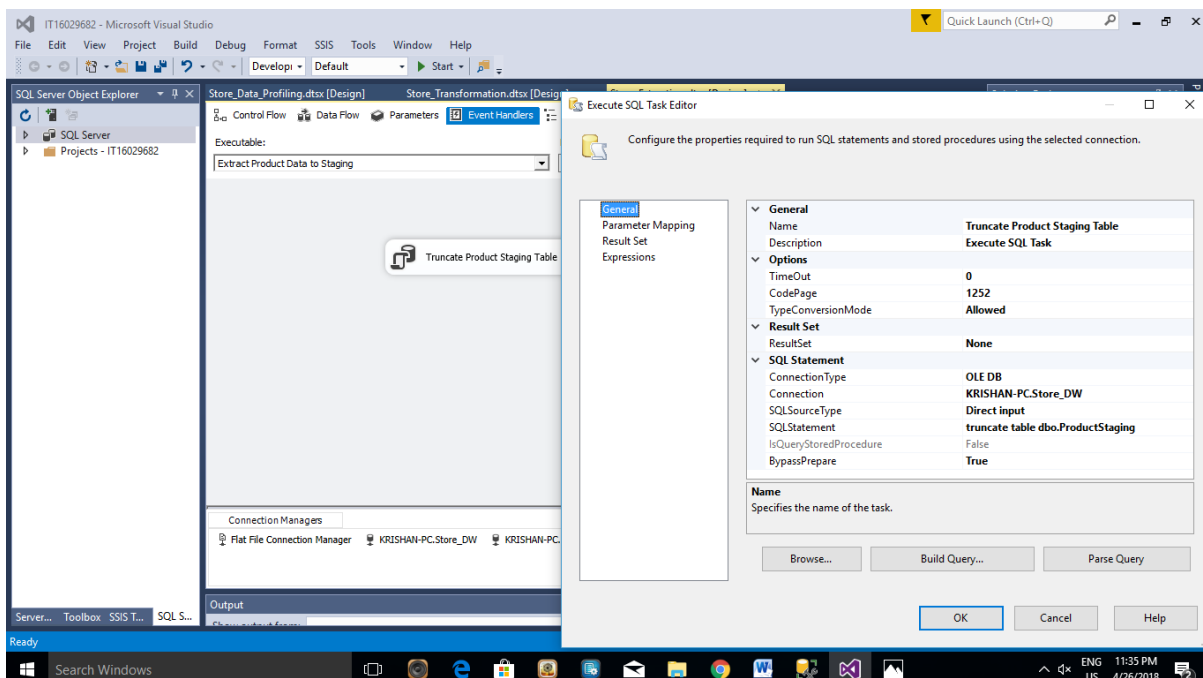
- Used OLE DB Source as **dbo.Product** table in **Store\_SourceDB**



- OLE DB Destination for create new table as **ProductStaging** in the **Store\_DW** database this table to data load into source

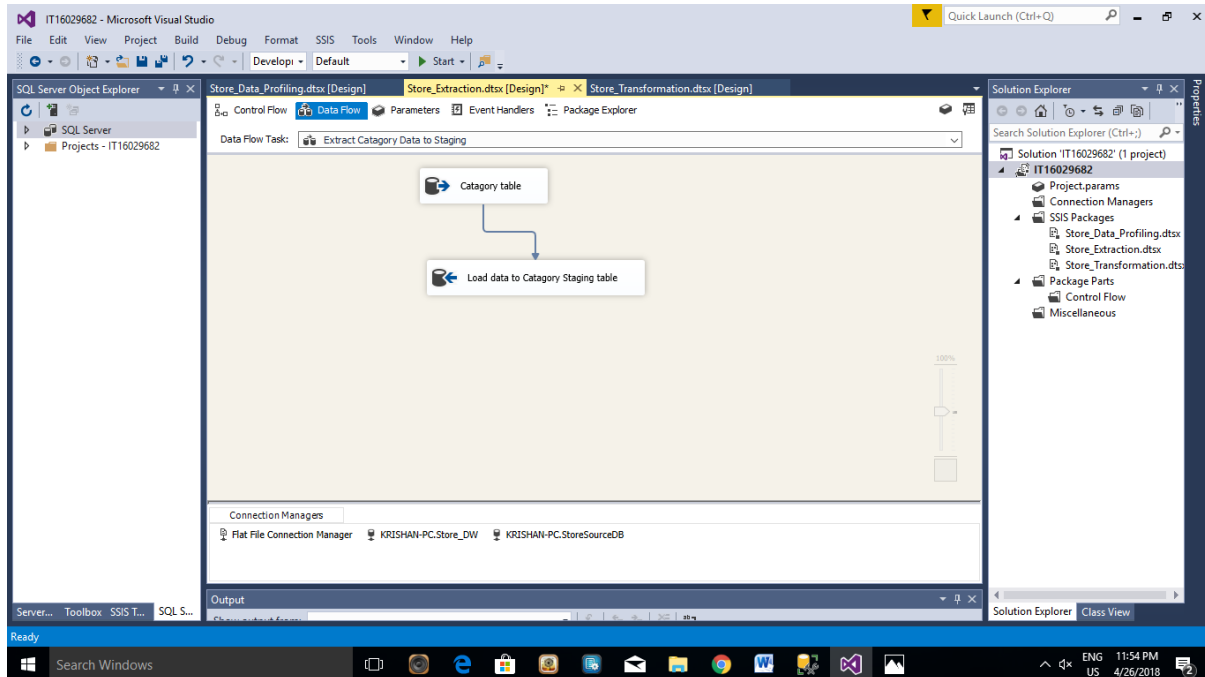


## 3.2 Product Extraction(Event Handlers)

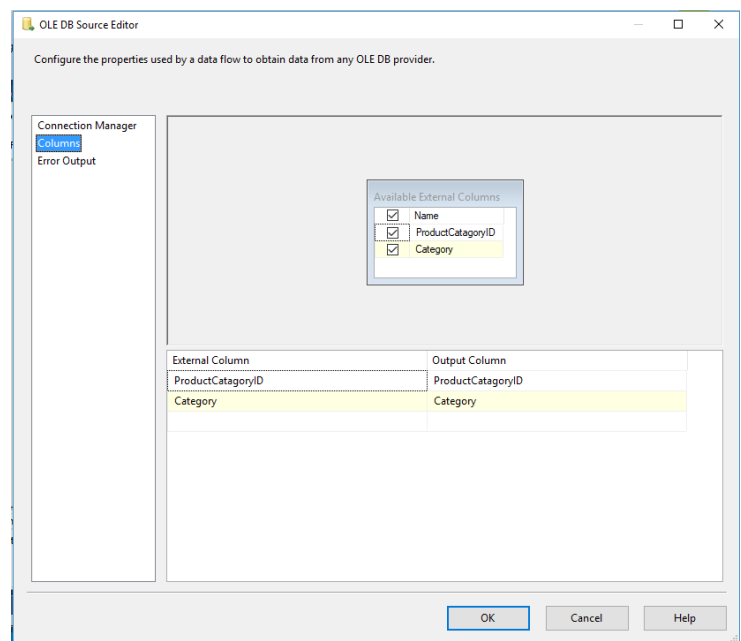
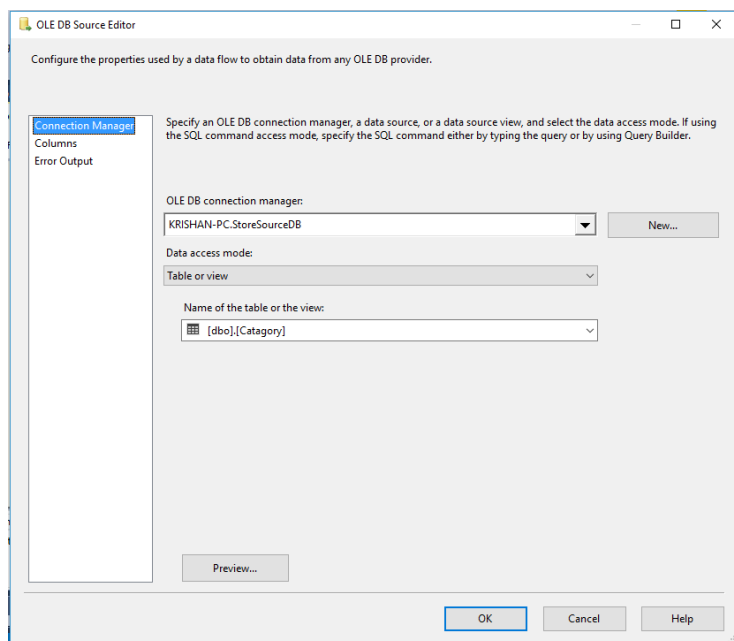


- Used **Execute SQL Task** SSIS tool Truncate table for sql command as **truncate table dbo.ProductStaging** in **Store\_DW** Database

## 4.1 Category Extraction (Data flow)

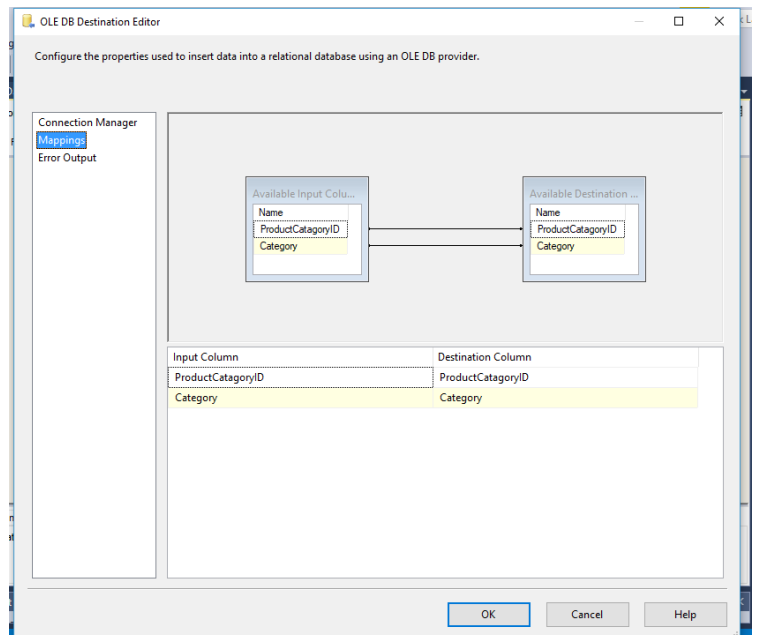
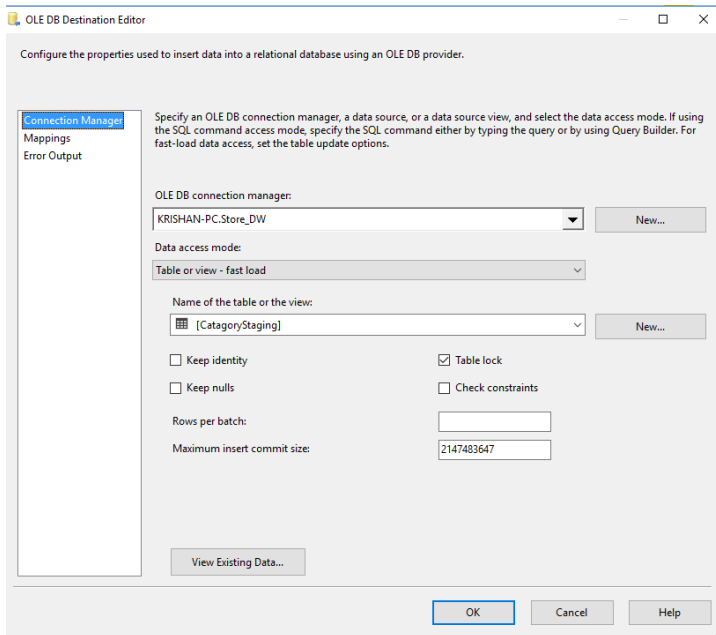


- Used OLE DB Source as `dbo.Category` table in `Store_SourceDB`

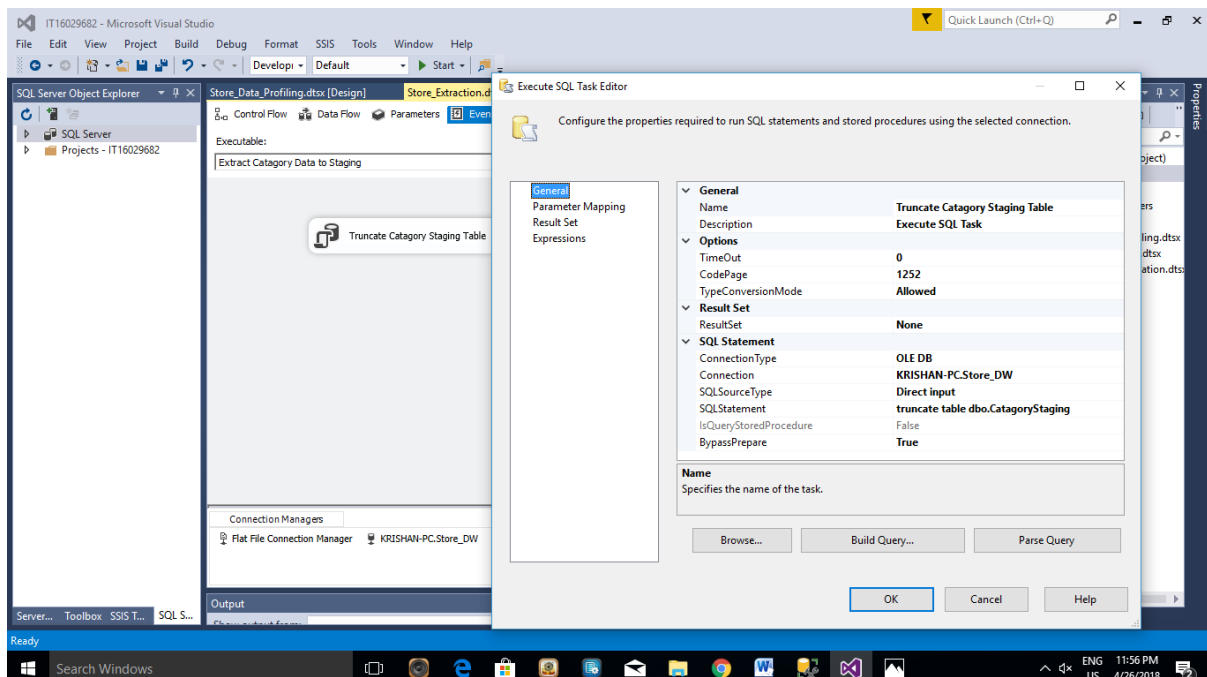




- **OLE DB Destination** for create new table as **CatagoryStaging** in the **Store\_DW** database this table to data load into source

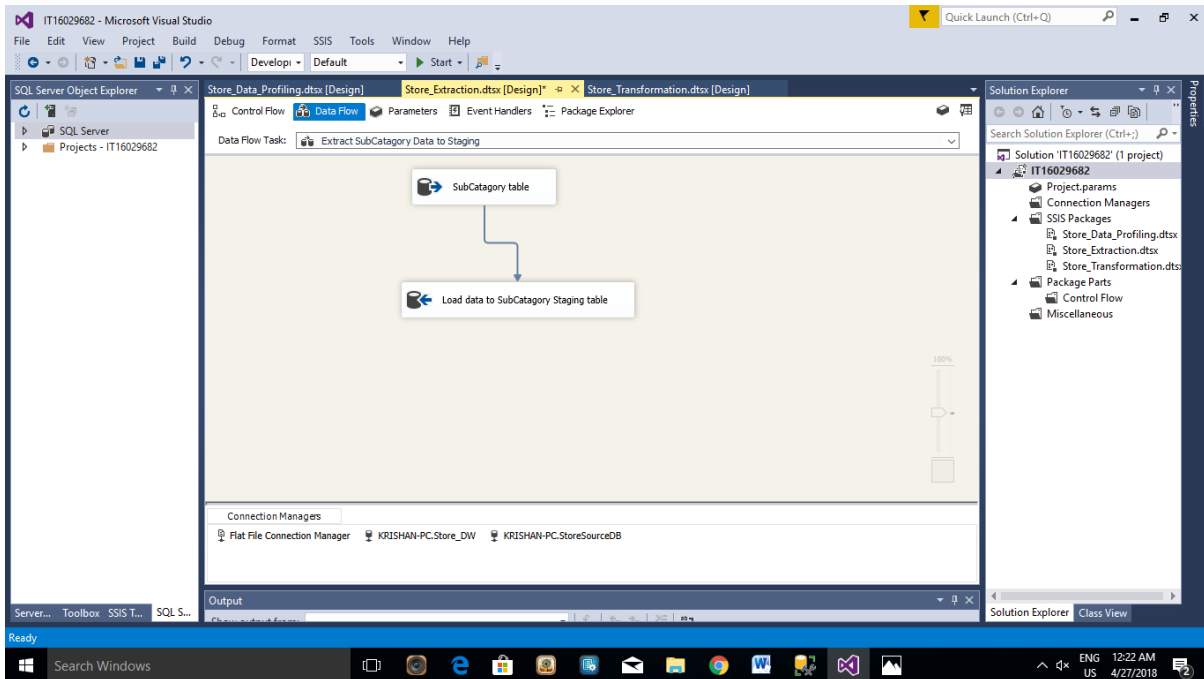


## 4.2 Catagory Extraction(Event Handlers)

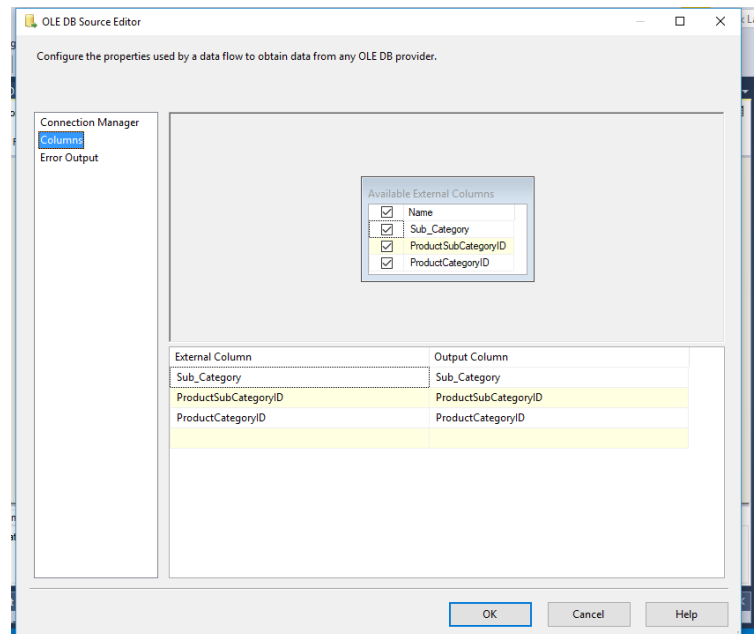
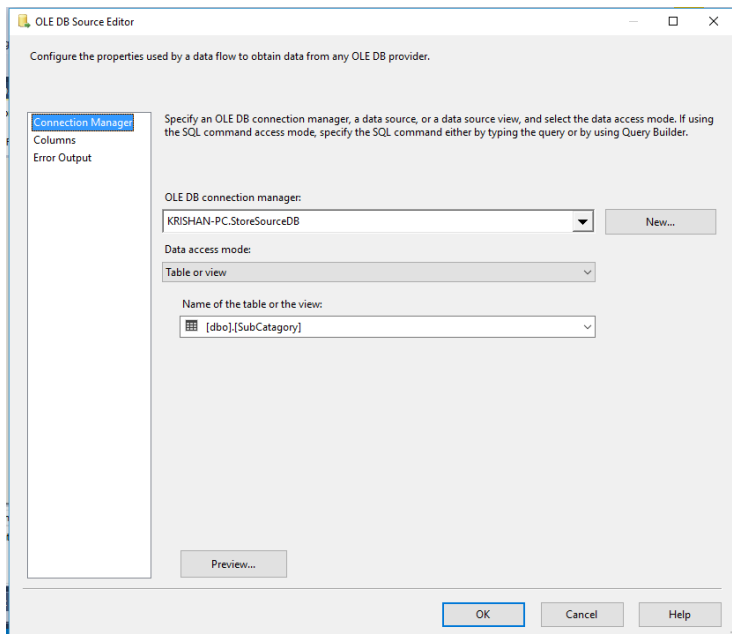


- Used **Execute SQL Task** SSIS tool Truncate table for sql command as **truncate table dbo.CatagoryStaging** in **Store\_DW** Database

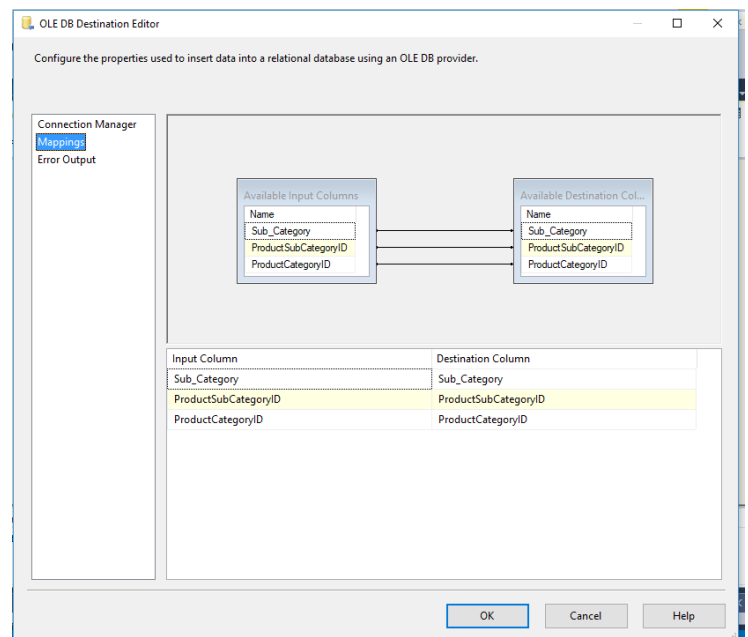
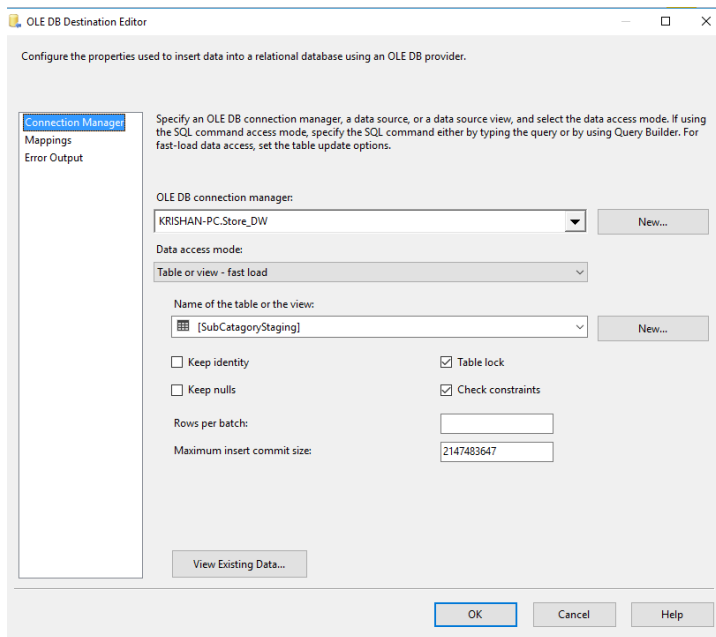
## 5.1 SubCategory Extraction (Data flow)



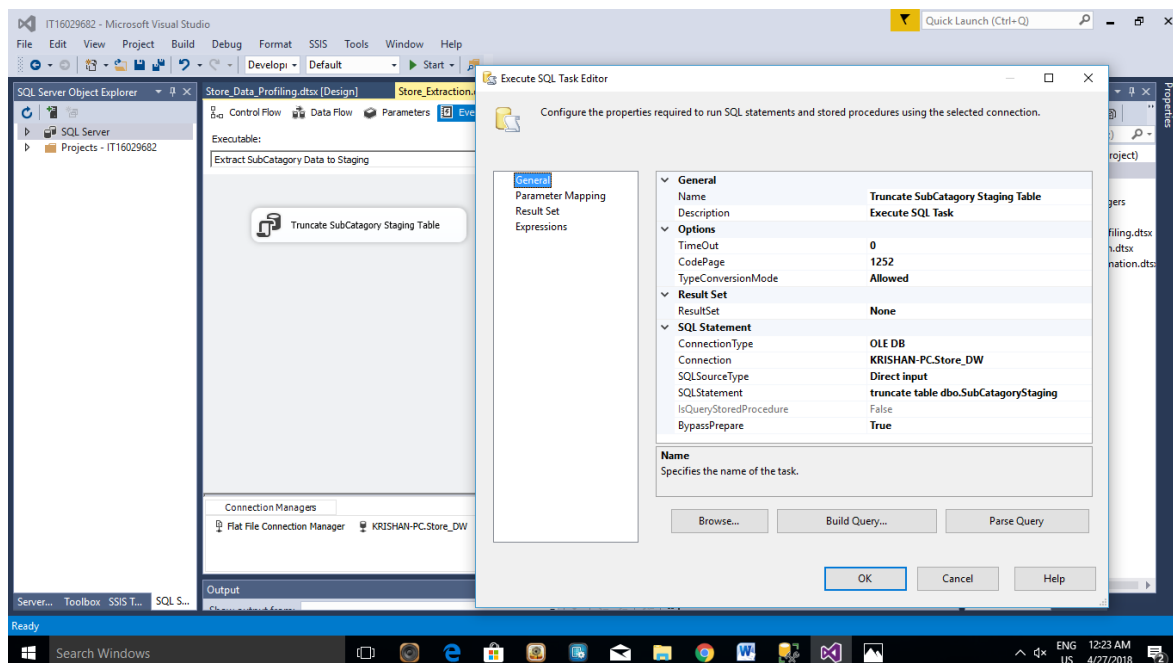
- Used OLE DB Source as `dbo.SubCategory` table in `Store_SourceDB`



- **OLE DB Destination** for create new table as **SubCatagoryStaging** in the **Store\_DW** database  
this table to data load into source

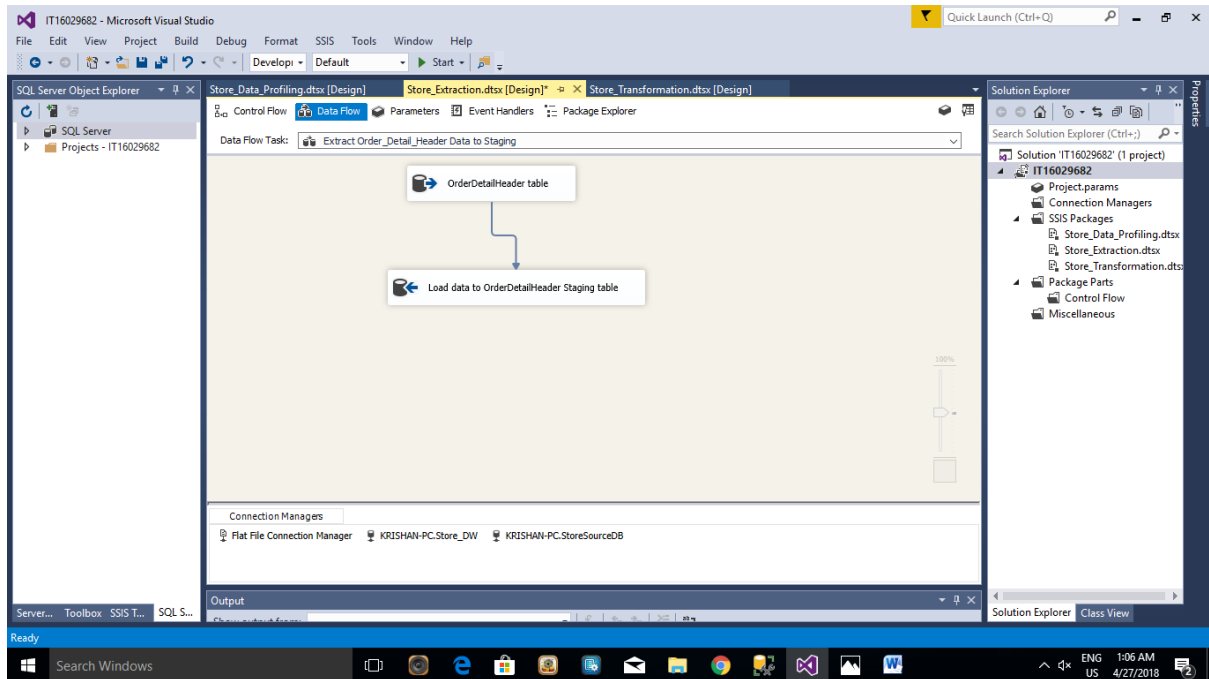


## 5.2 SubCatagory Extraction(Event Handlers)

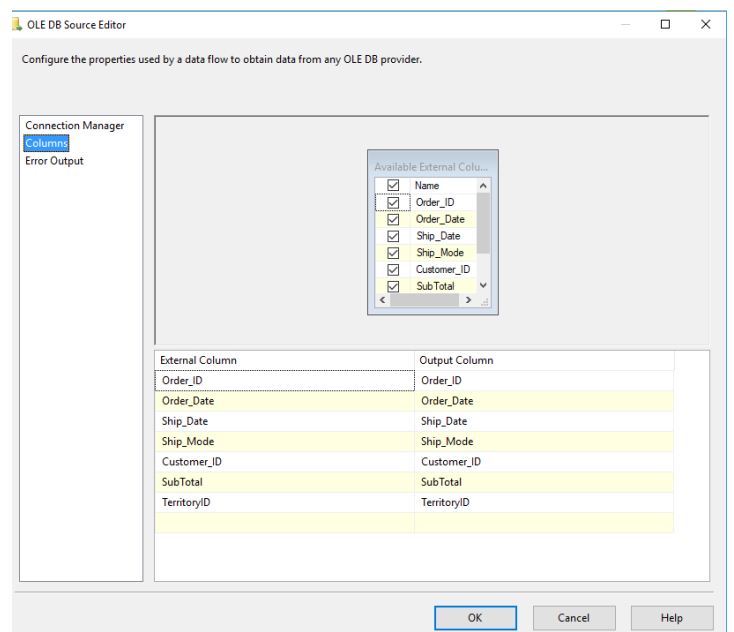
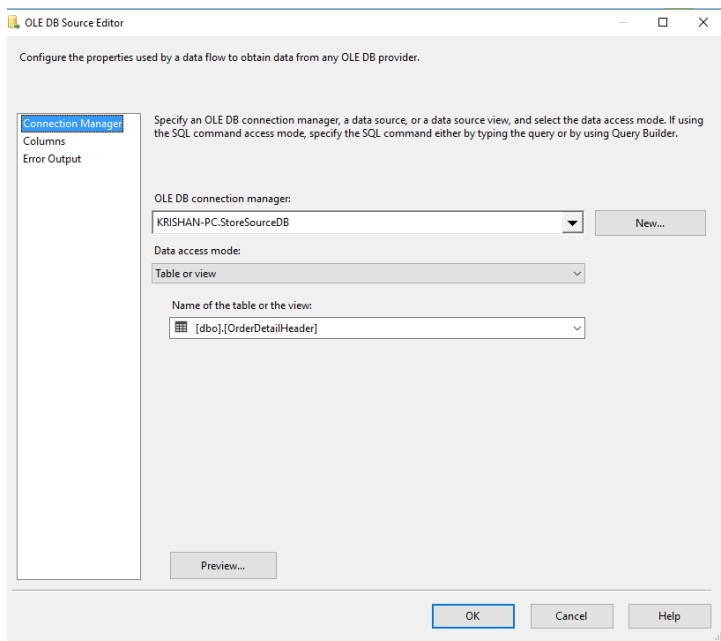


- Used **Execute SQL Task** SSIS tool Truncate table for sql command as **truncate table dbo. SubCatagoryStaging** in **Store\_DW** Database

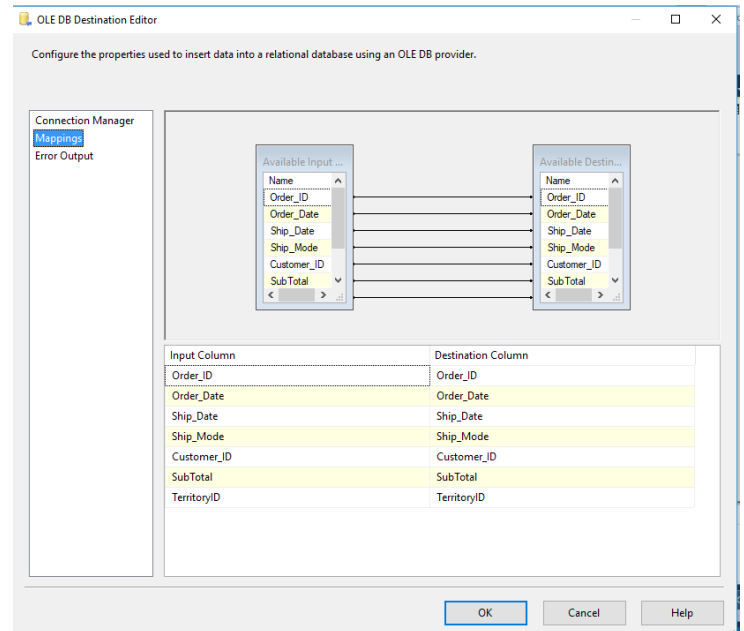
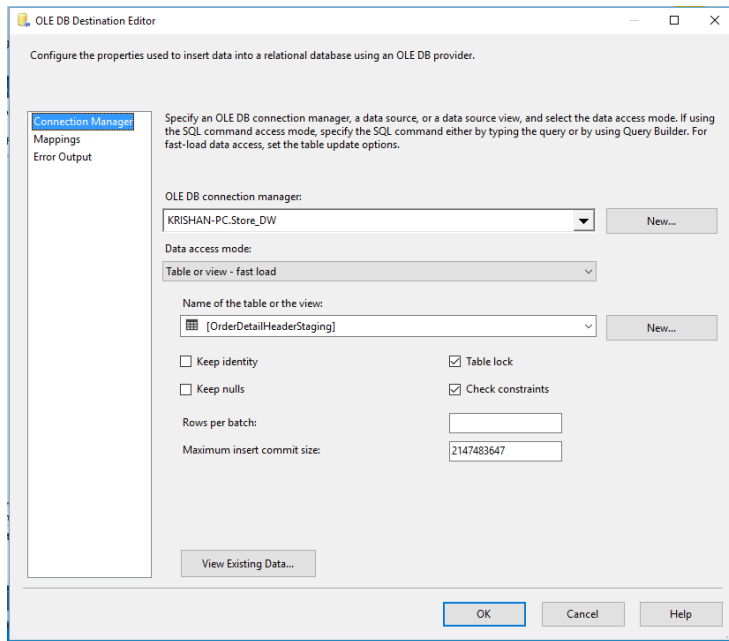
## 6.1 OrderDetailHeader Extraction (Data flow)



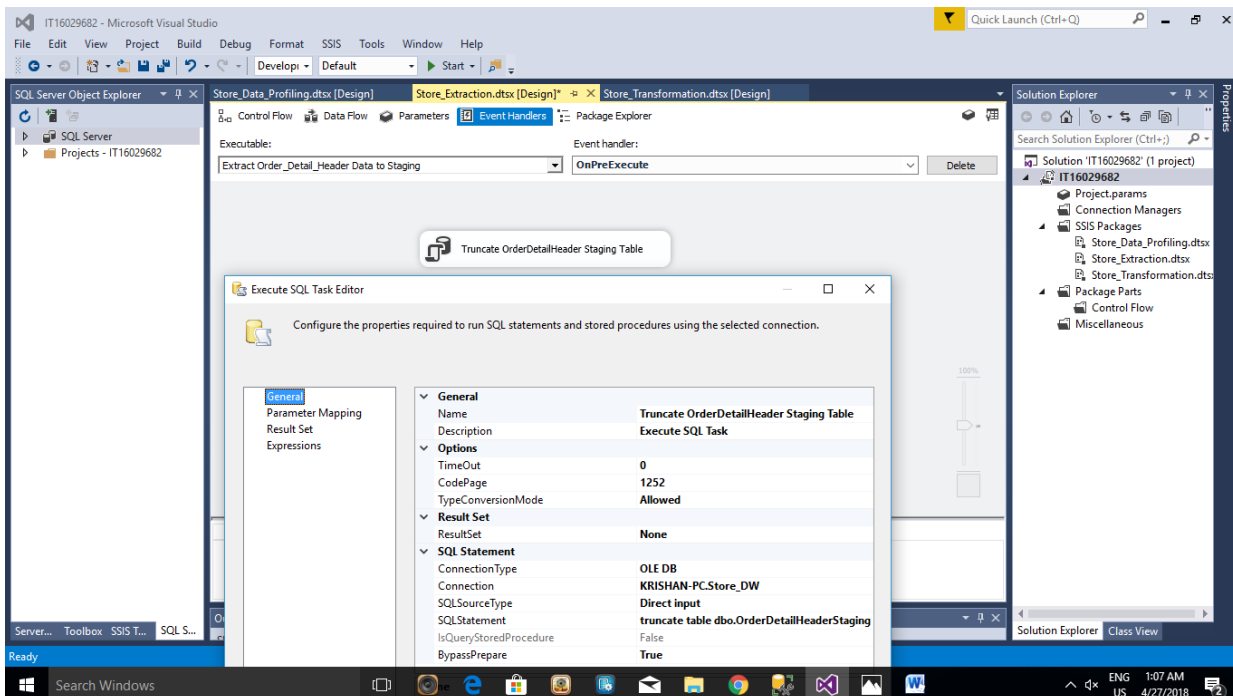
- Used OLE DB Source as **dbo.OrderDetailHeader** table in **Store\_SourceDB**



- OLE DB Destination for create new table as **OrderDetailHeaderStaging** in the **Store\_DW** database  
this table to data load into source

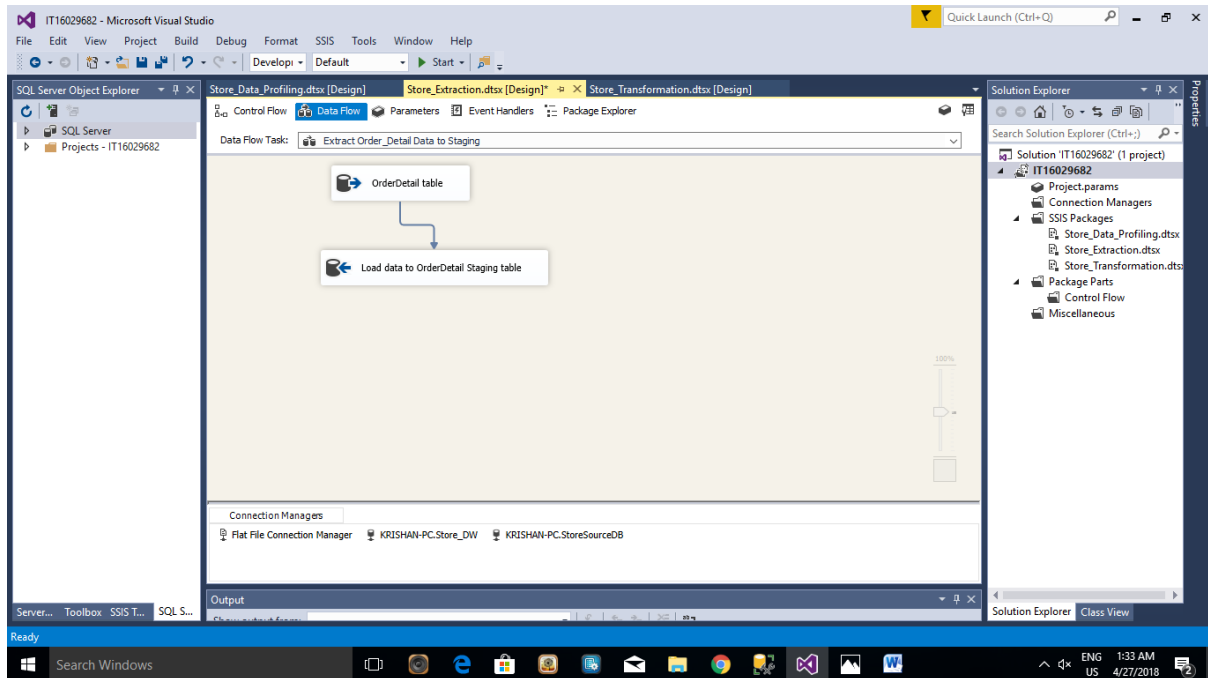


## 6.2 OrderDetailHeader Extraction(Event Handlers)

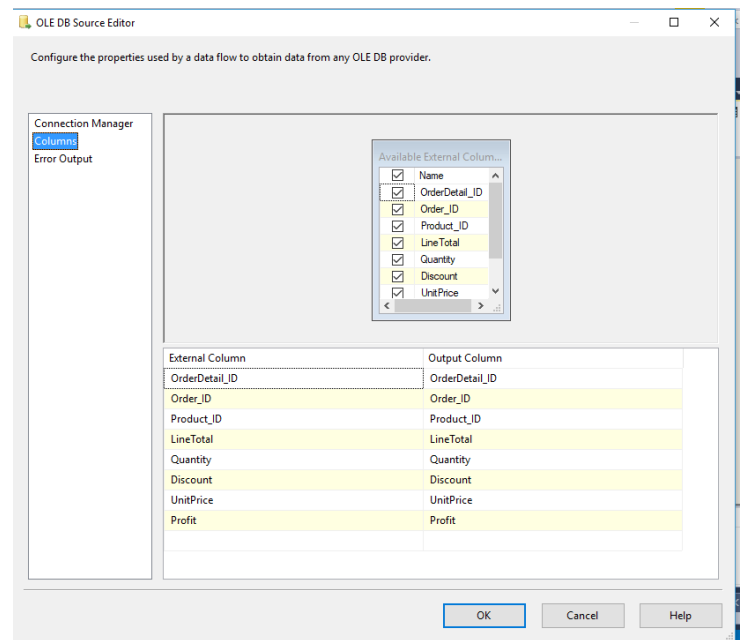
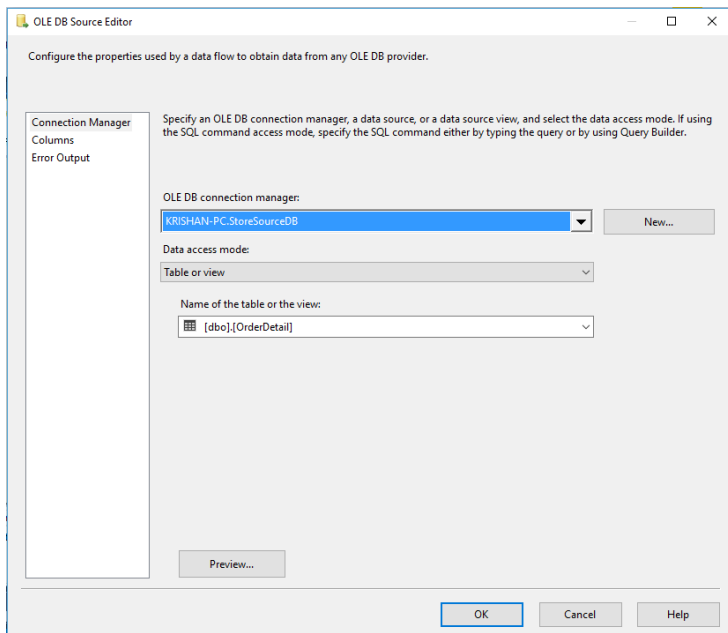


- Used **Execute SQL Task** SSIS tool Truncate table for sql command as **truncate table dbo. OrderDetailHeaderStaging** in **Store\_DW** Database

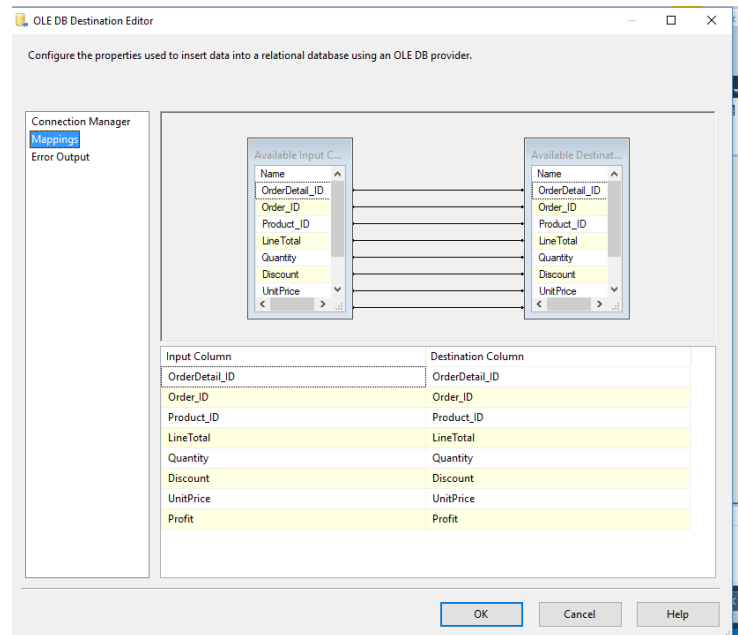
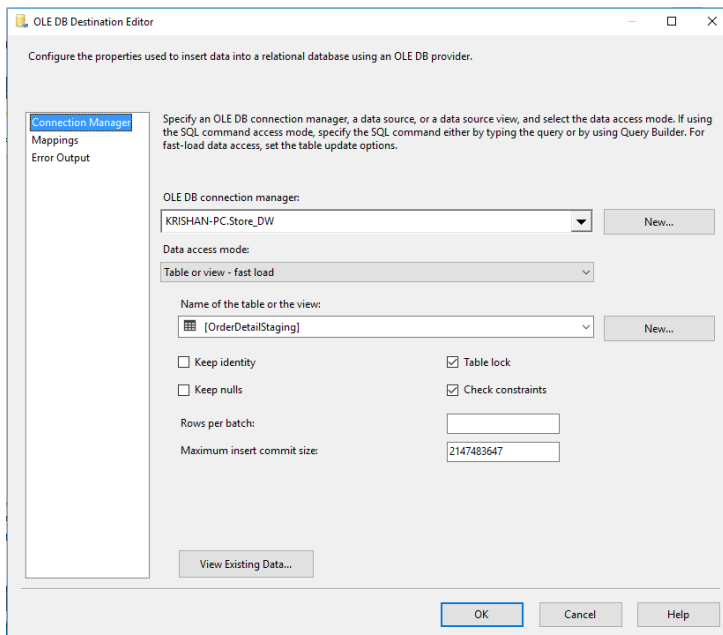
## 7.1 OrderDetail Extraction (Data flow)



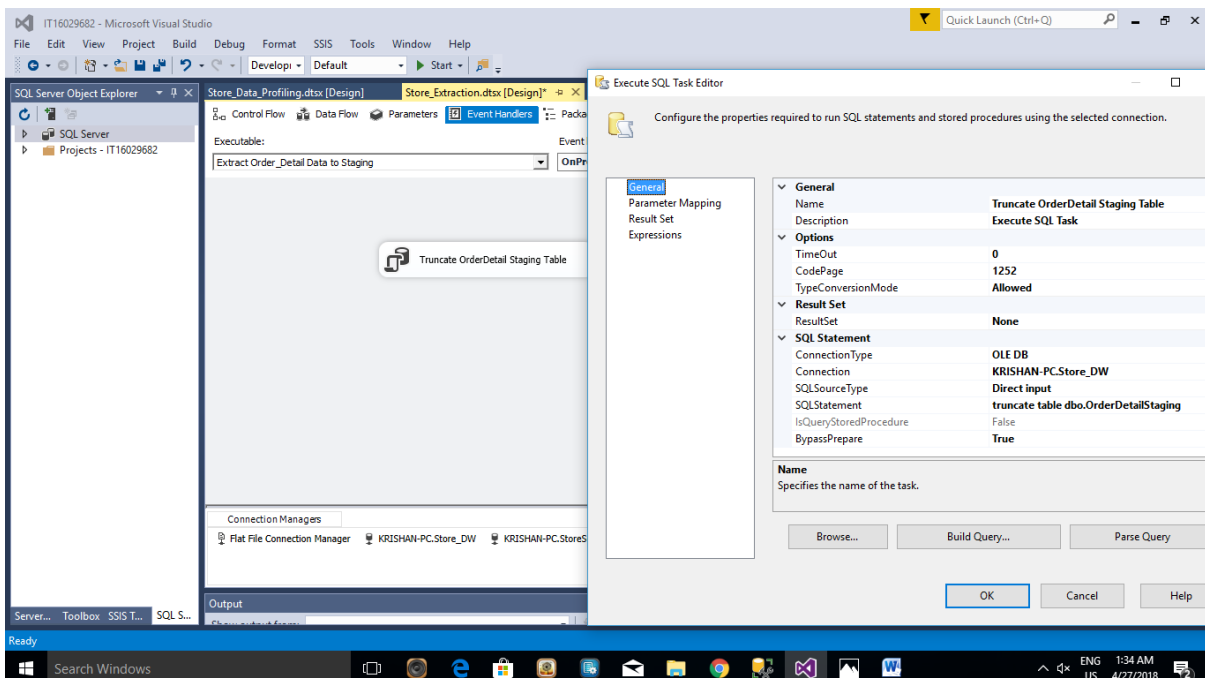
- Used OLE DB Source as **dbo.OrderDetail** table in **Store\_SourceDB**



- **OLE DB Destination** for create new table as **OrderDetailStaging** in the **Store\_DW** database this table to data load into source

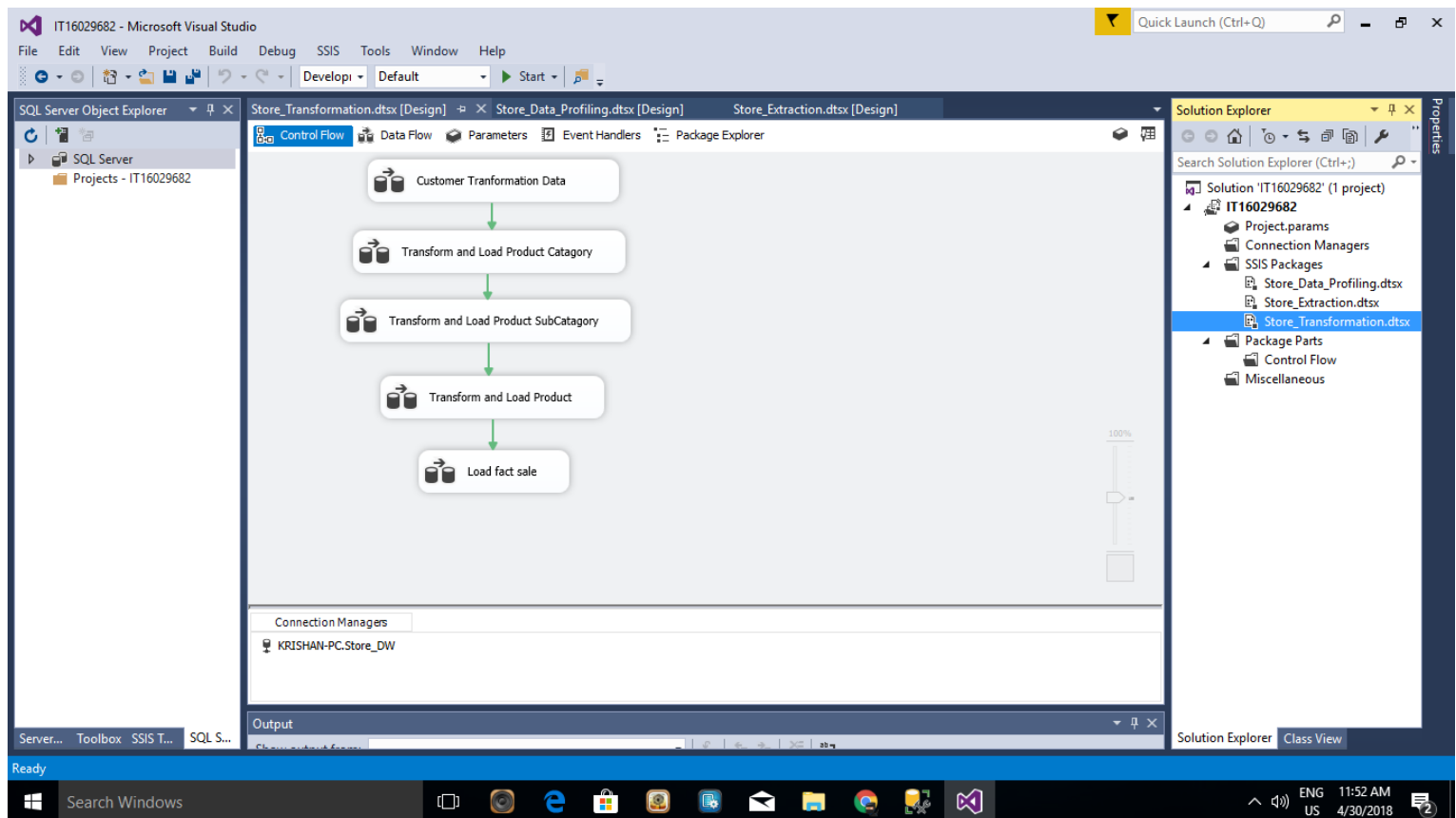


## 7.2 OrderDetail Extraction(Event Handlers)



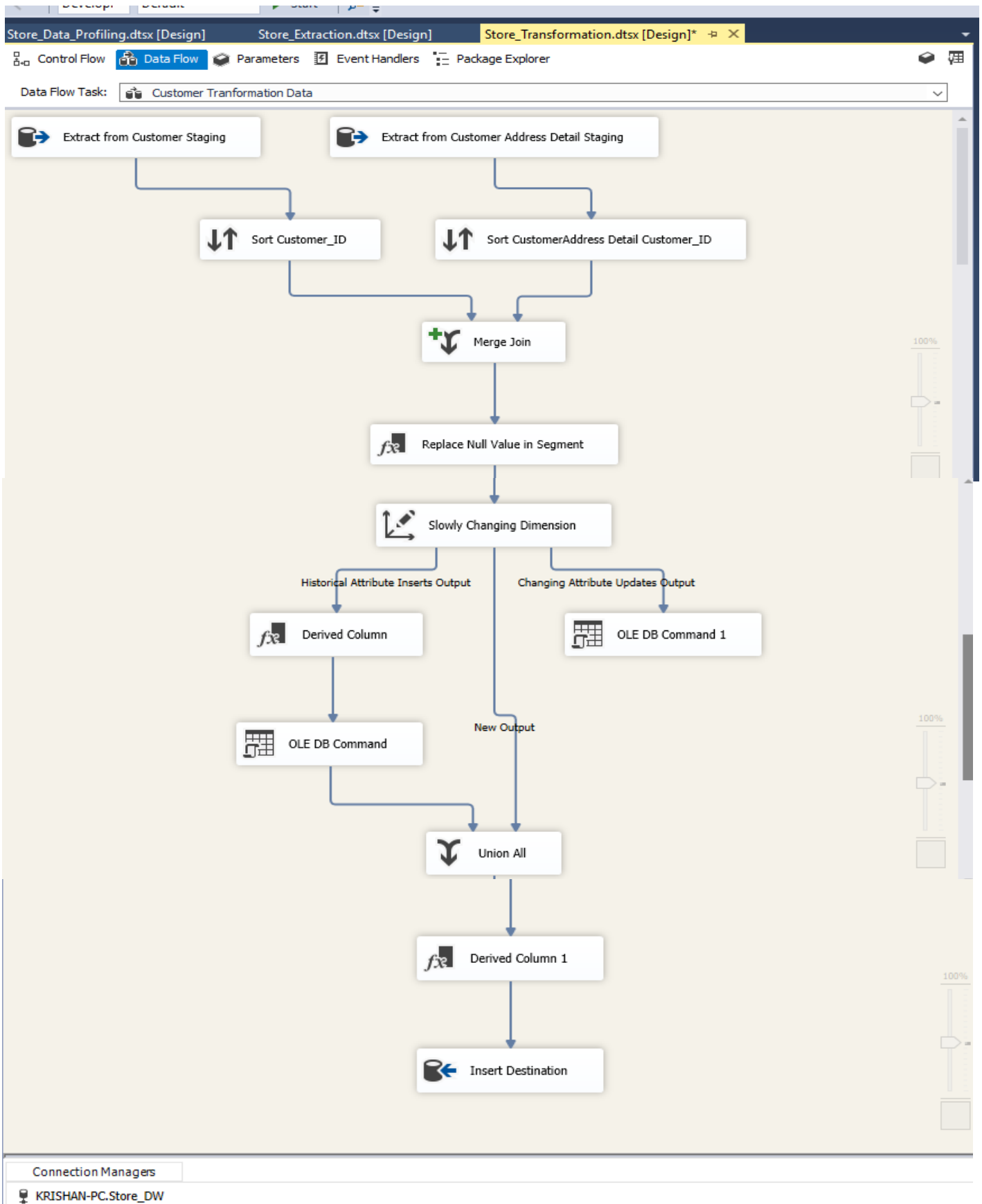
- Used **Execute SQL Task** SSIS tool Truncate table for sql command as **truncate table dbo. OrderDetailStaging** in **Store\_DW** Database

# Transformation & Loading (Control flow)

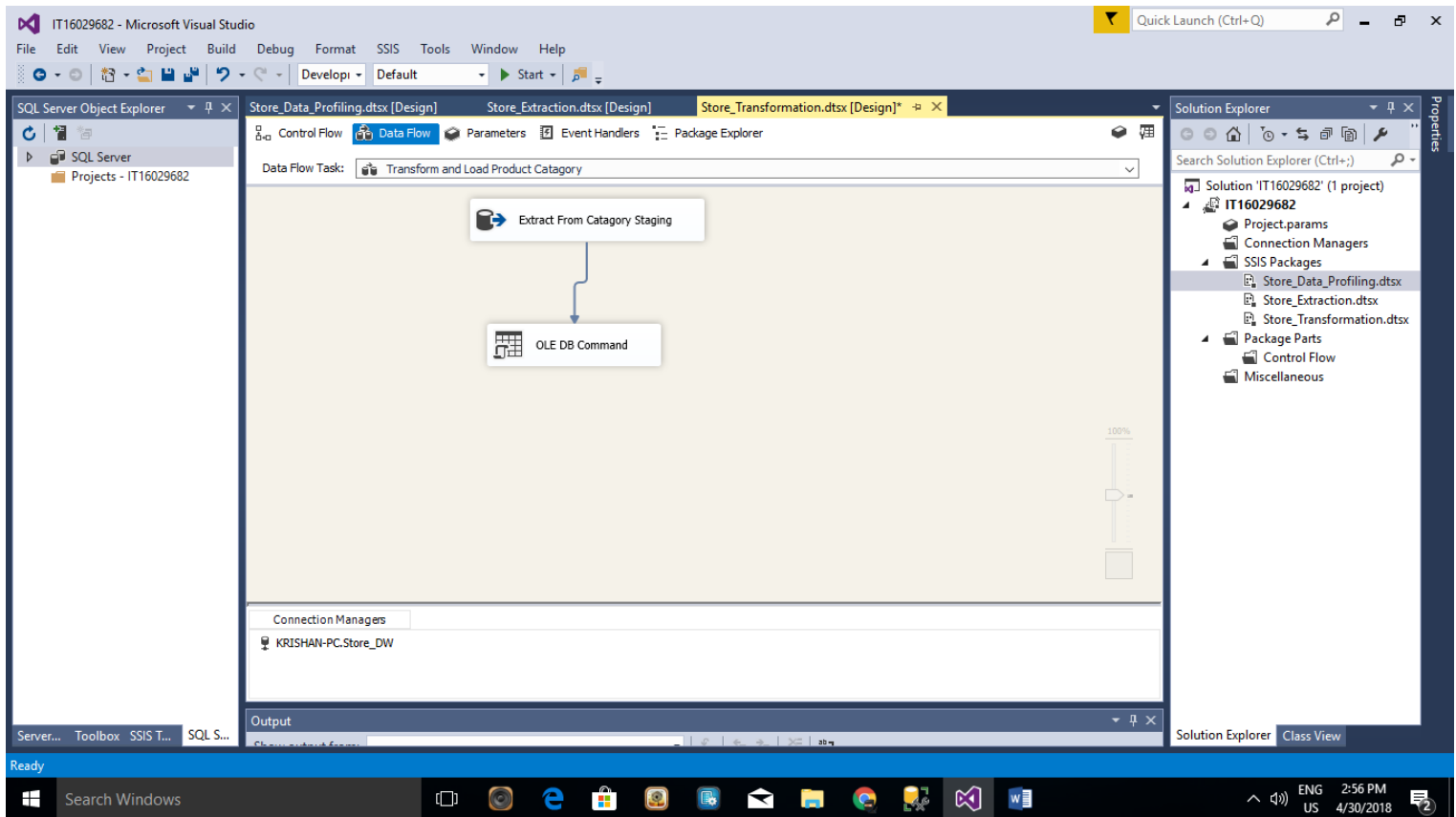




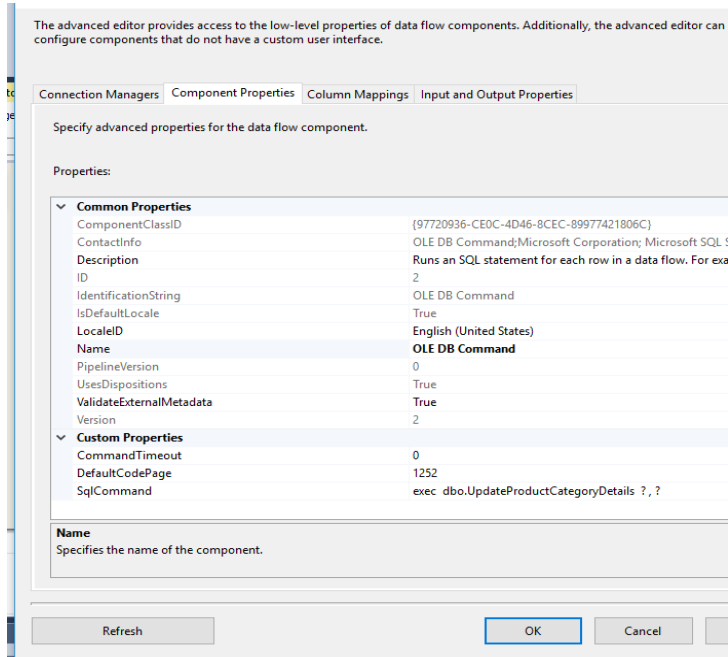
## Customer Transformation & Loading (Data flow)



# Category Transformation & Loading (Data flow)



- First have to create procedure in **Store\_DW** database as **UpdateProductCategoryDetails** in **Sql Server**.
- **OLE DB Command** SSIS tool used to **execute** , **UpdateProductCategoryDetails** procedure, it is used to insert data into Category Staging to **DimCategory** without data duplication.



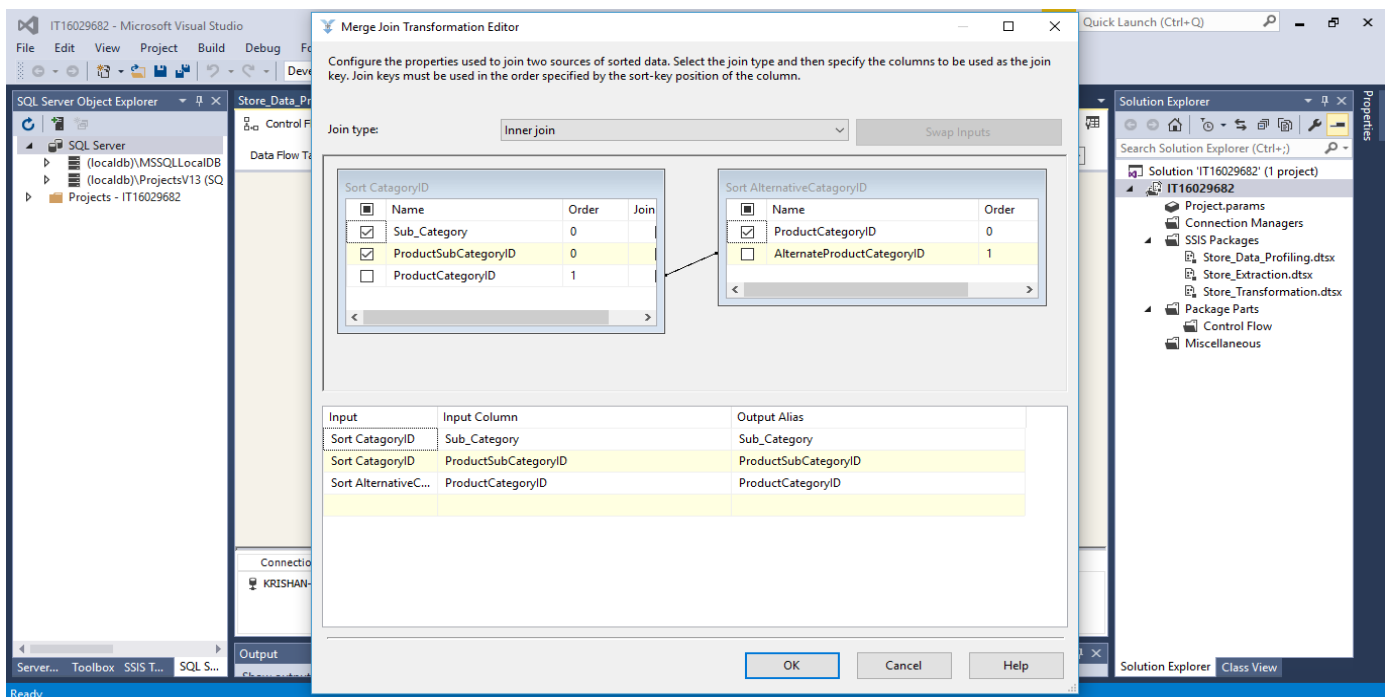
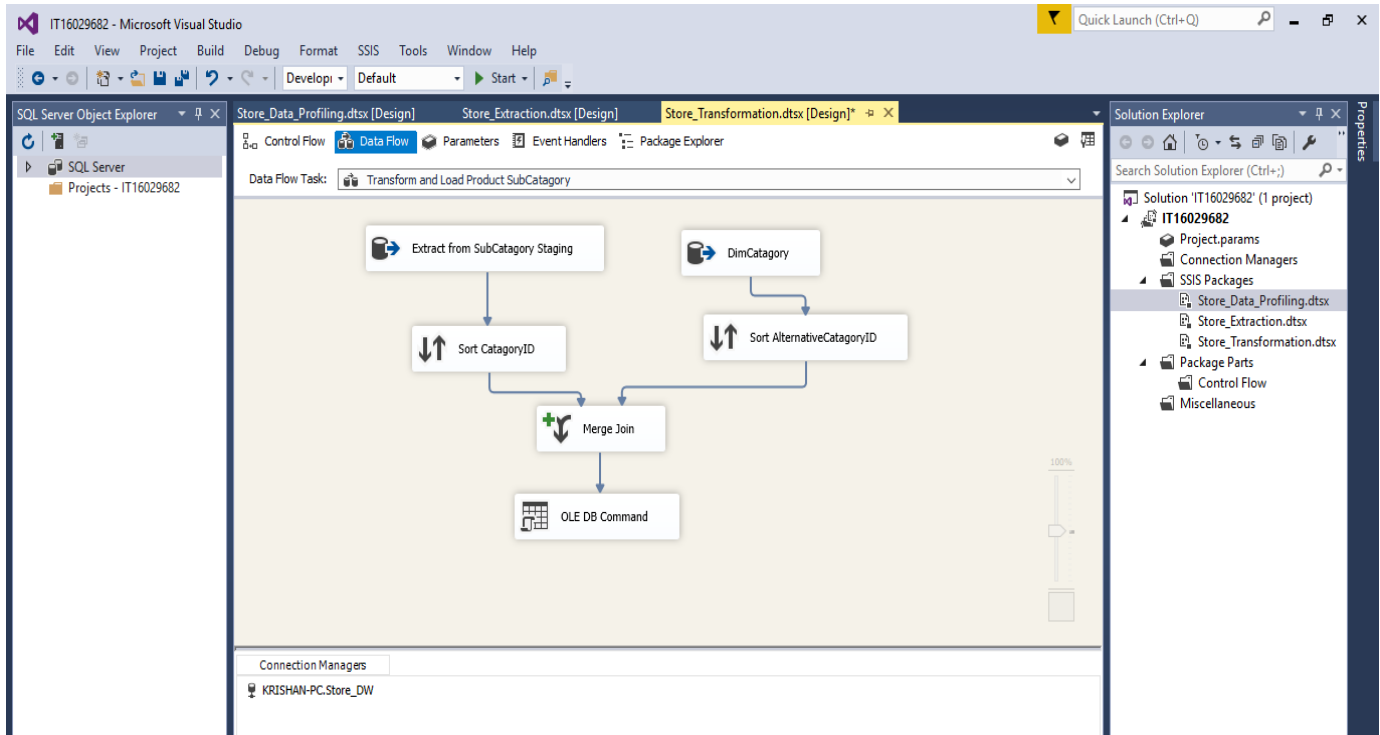
```

category.sql - KRIS...AN-PC\Krishan (60))
Create Procedure dbo.UpdateProductCategoryDetails
    @ProductCategoryID int,
    @ProductCategoryName nvarchar(50)
as
Begin
if not exists (select ProductCategoryID
from dbo.DimCategory
where AlternateProductCategoryID = @ProductCategoryID
and Name = @ProductCategoryName )
begin
insert into dbo.DimCategory
(AlternateProductCategoryID, Name)
values
(@ProductCategoryID, @ProductCategoryName)
end;
End;

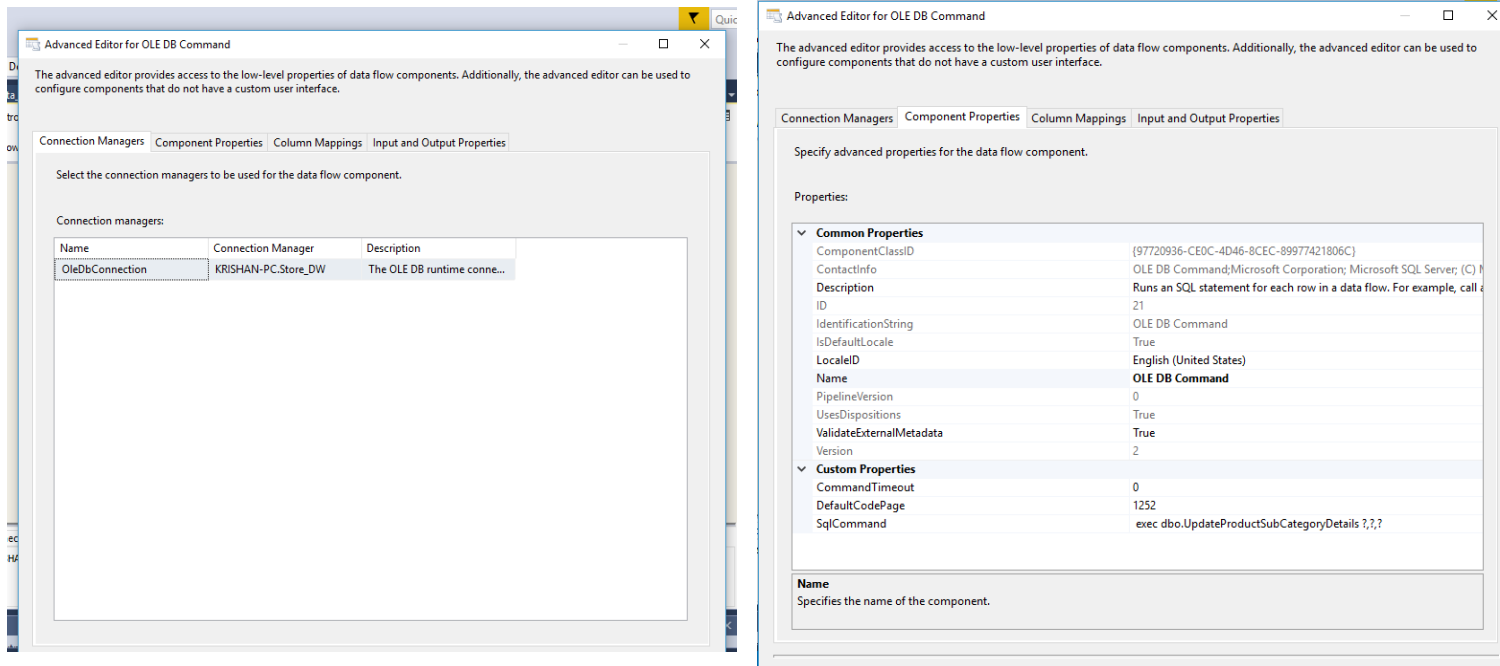
```

## SubCategory Transformation & Loading (Data flow)

- Use two **OLE DB SOURCE** to **SubCategoryStaging** and **DimCategory** in **Store\_DW** database after use each source to **Sort** ,SSIS tool and sort by **CategoryID** SubCategoryStaging table and Category Dimension table **AlternativeCategoryID**



- First have to create procedure in **Store\_DW** database as **UpdateProductSubCategoryDetails** in **Sql Server**
- **OLE DB Command** SSIS tool used to **execute** used **UpdateProductSubCategoryDetails** procedure , it used to Sorted using **CategoryID** data of **SubCategory Staging** and **DimCategory** after **Merging** insert to **DimSubCategory** without data duplication



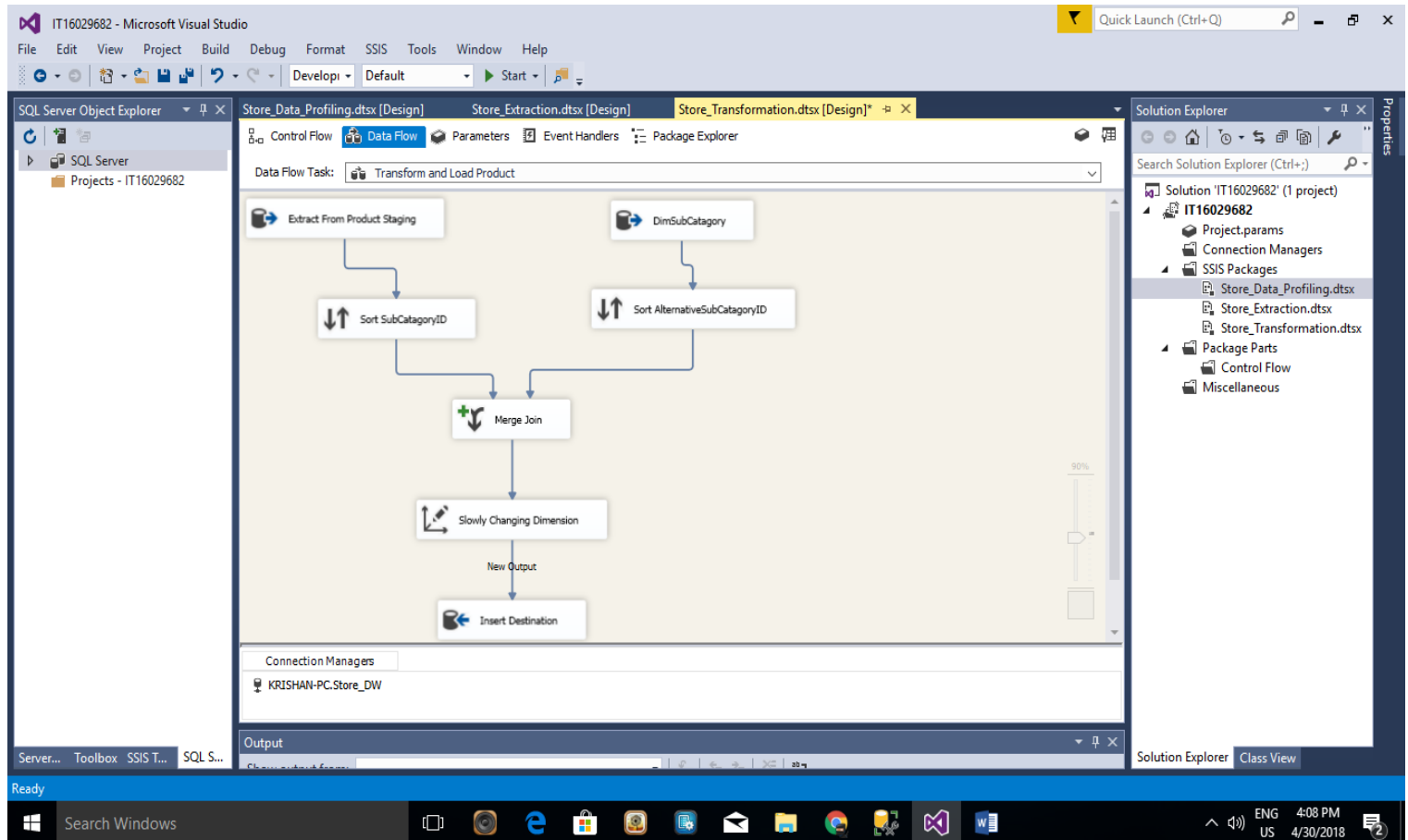
```

Create Procedure dbo.UpdateProductSubCategoryDetails
@ProductSubCategoryID int,
@ProductCategoryID int,
@ProductSubCategoryName nvarchar(50)
as
Begin
if not exists (select ProductSubcategoryID
from dbo.DimSubCategory
where AlternateProductSubCategoryID = @ProductSubCategoryID
and name = @ProductSubCategoryName
and ProductCategoryID = @ProductCategoryID )
begin
insert into dbo.DimSubCategory
(AlternateProductSubCategoryID, name, ProductCategoryID)
values
(@ProductSubCategoryID, @ProductSubCategoryName, @ProductCategoryID)
end;
End

```

## Product Transformation & Loading (Data flow)

- Use two **OLE DB SOURCE** to **ProductStaging** and **DimSubCatagory** in **Store\_DW** database after use each source to **Sort** ,SSIS tool and sort by **SubCatagoryID** ProductStaging table and SubCategory Dimension table **AlternativeSubCatagoryID**



**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

Name	Order	Join
<input checked="" type="checkbox"/> Product_ID	0	
<input checked="" type="checkbox"/> Product_Name	0	
<input checked="" type="checkbox"/> ProductNumber	0	
<input type="checkbox"/> ProductSubCategoryID	1	

Name	Order
<input checked="" type="checkbox"/> ProductSubcategoryID	0
<input type="checkbox"/> AlternateProductSubCategoryID	1
<input type="checkbox"/> ProductCategoryID	0
<input type="checkbox"/> name	0

Input	Input Column	Output Alias
Sort SubCatagor...	Product_ID	Product_ID
Sort SubCatagor...	Product_Name	Product_Name
Sort SubCatagor...	ProductNumber	ProductNumber
Sort AlternativeS...	ProductSubcategoryID	ProductSubcategoryID

**OK** **Cancel** **Help**

## FactSale Transformation & Loading (Data flow)

Data Flow Task:  Load fact sale