PROJECT: Conditional Split

Transform task on employee table to split data into multiple destinations based on the country name.

Submitted By

Akash Pal

REQUIREMENT:

We have Employee data in the AdventureWorks2017 database. We want to split data based on the CountryRegionName column in different tables.

- For employees belonging to the United Kingdom, insert data into the table [United Kingdom]
- For employees belonging to the United States, insert data into the table [United States]
- For employees belonging to Germany, insert data into the table [Germany]
- The rest of the data (other countries) will go into the [default] table

ENVIRONMENT DETAILS:

- We will use the AdventureWorksDW2017 sample database in this project.
- We should have Visual Studio 2019 with "SQL Server Integration Service Project" extension installed to prepare the SSIS package.

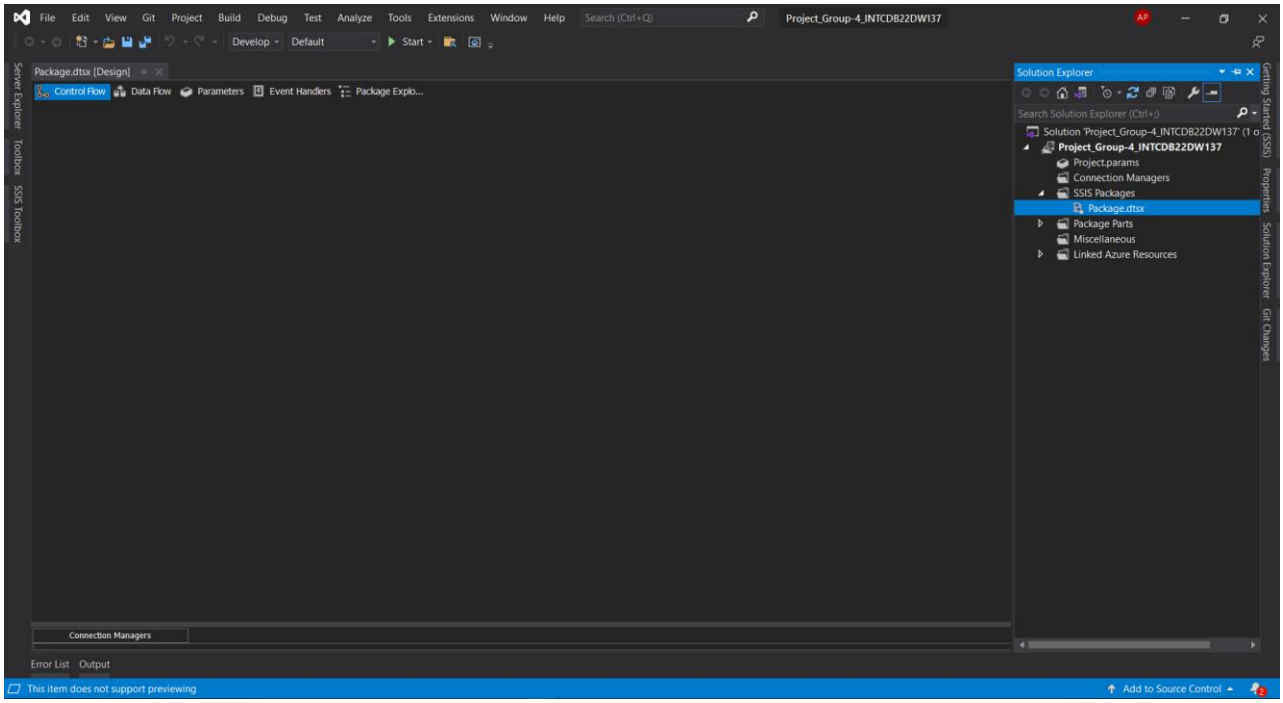
KEY LOGIC:

Our project is to split the records based on Country but we don't have Country column in DimEmployee table in AdventureWorksDW2017 database. But DimEmployee table has SalesTerritoryKey column and we have another table called DimSalesTerritory table containing SalesTerritoryKey as primary key in adventureworksdw2017 database. Also, we need to get all Employee data even if their country data in missing.

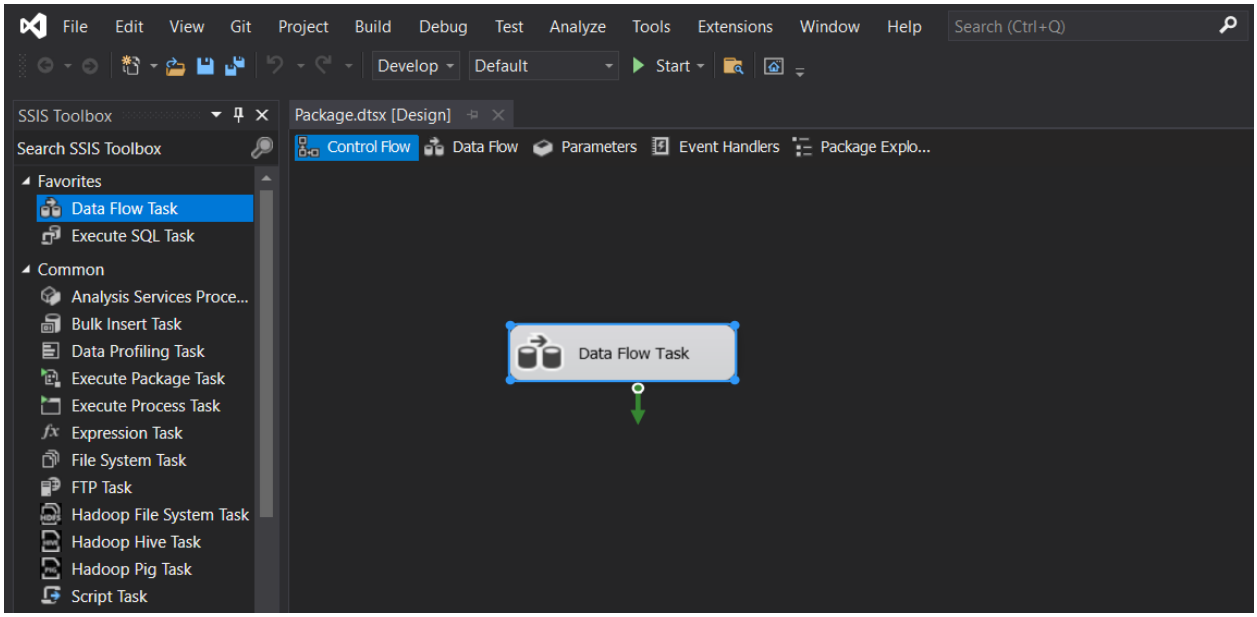
So first we need perform left outer join between DimEmployee table and DimSalesTerritory table on SalesTerritoryKey column. Then for splitting data into multiple destinations, we need to use conditional split. Then those split data will be populated in 4 different table in adventureworksdw2017 named as DimEmployee_UnitedStates, DimEmployee_UnitedKingdom, DimEmployee_Germany, DimEmployee_Others.

PROCEDURE:

STEP 1: In Visual Studio, create a new project of integration service.

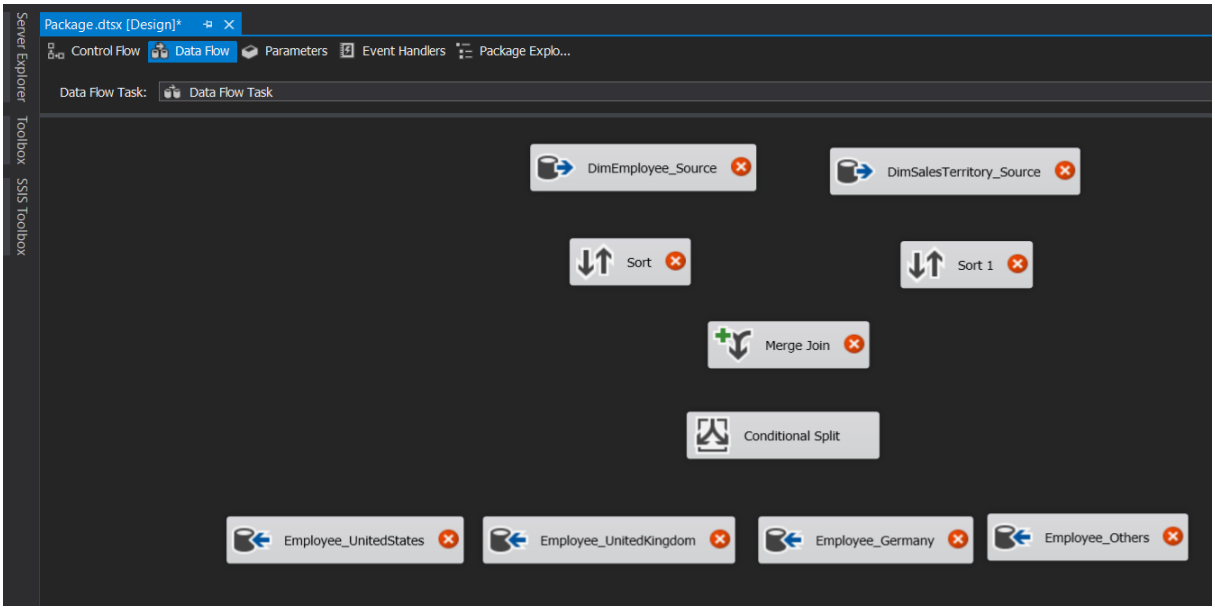


STEP 2: Use Data Flow Task SSIS tool in Control Flow section. Then we need to configure Data Flow Task.



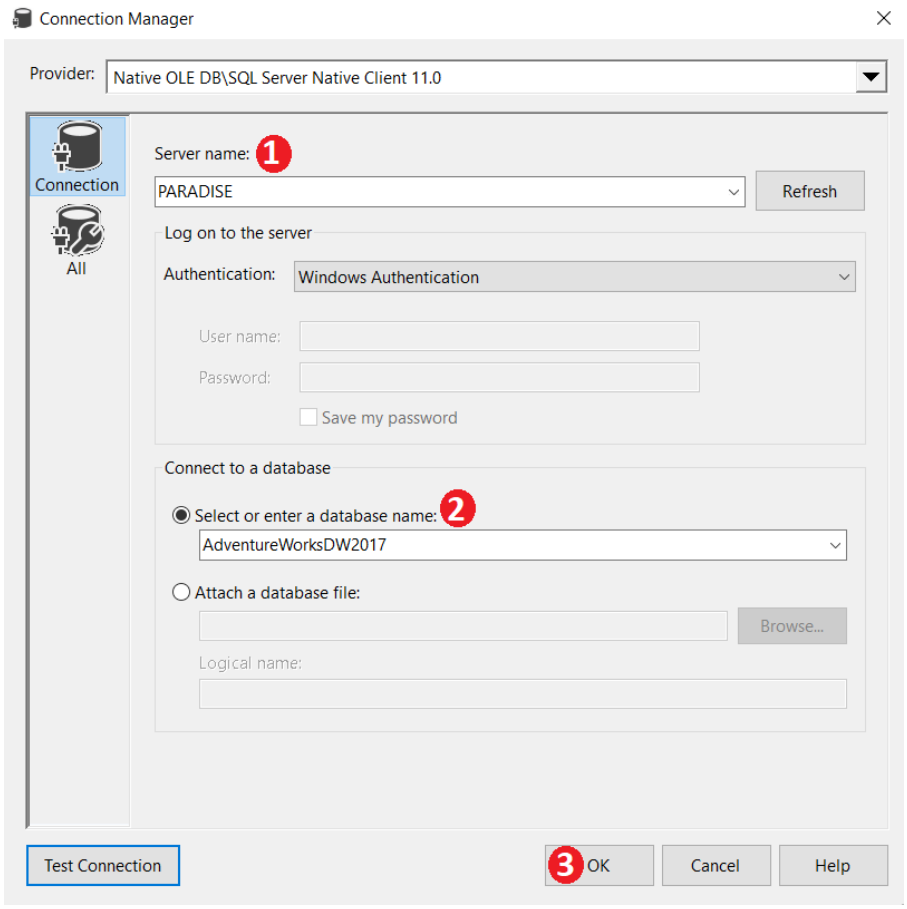
STEP 3: Now from SSIS Toolbox we add all SSIS tools inside Data Flow Task section. The SSIS tools that we have used are

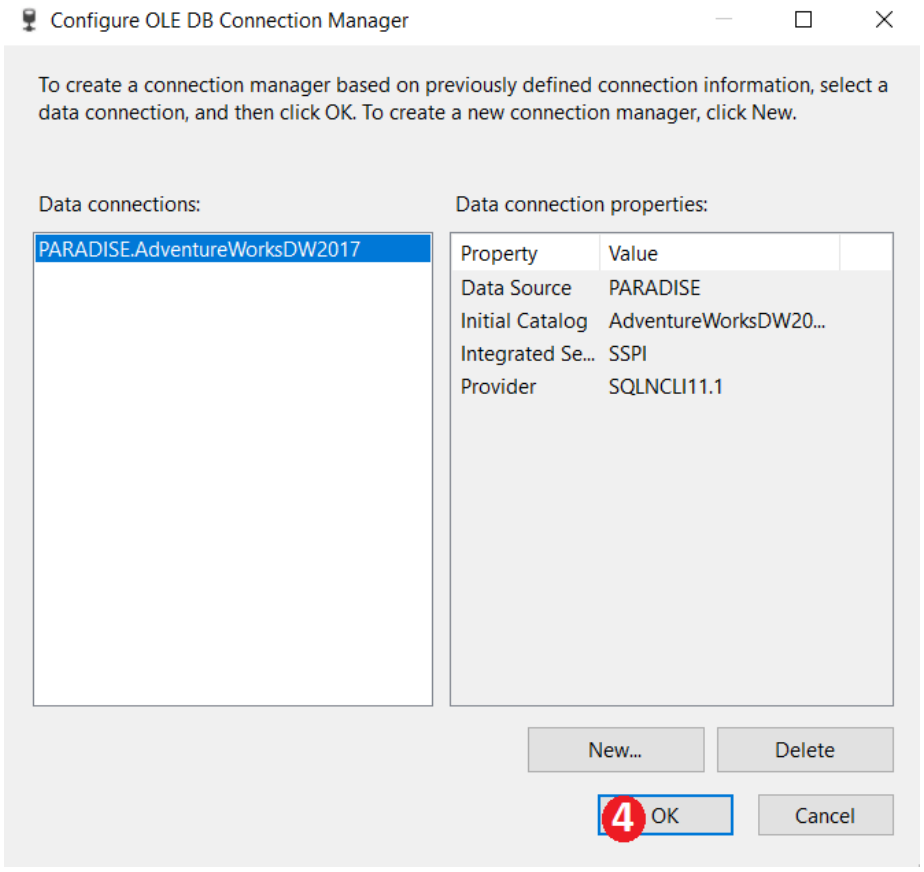
- 2 OLE DB Source tool (Named as DimEmployee_Source and DimSalesTerritory_Source for understanding) to get the source table.
- 2 Sort tool for sorting the table based on specific column.
- Merge Join tool to join two tables.
- Conditional Split tool for splitting one table to multiple destination based on conditions that we have given while configuring Conditional split.
- 4 OLE DB Destination tool (Named as Employee_UnitedStates, Employee_UnitedKingdom, Employee_Germany, Employee_Others) for populating the resultant data into SQL Server.

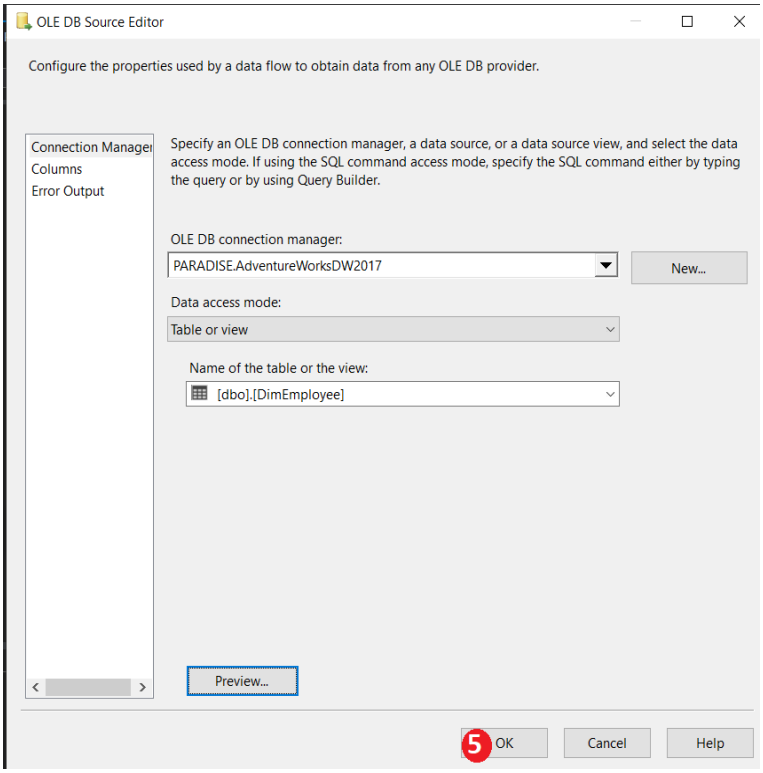


STEP 4: Then, we need to configure two OLE DB Source. Here we use OLE DB Source because our source data is in SQL Server Database table. After double clicking on the DimEmployee_Source, an OLE DB Source Editor window is opened. There we need to create a new OLE DB Connection Manager.

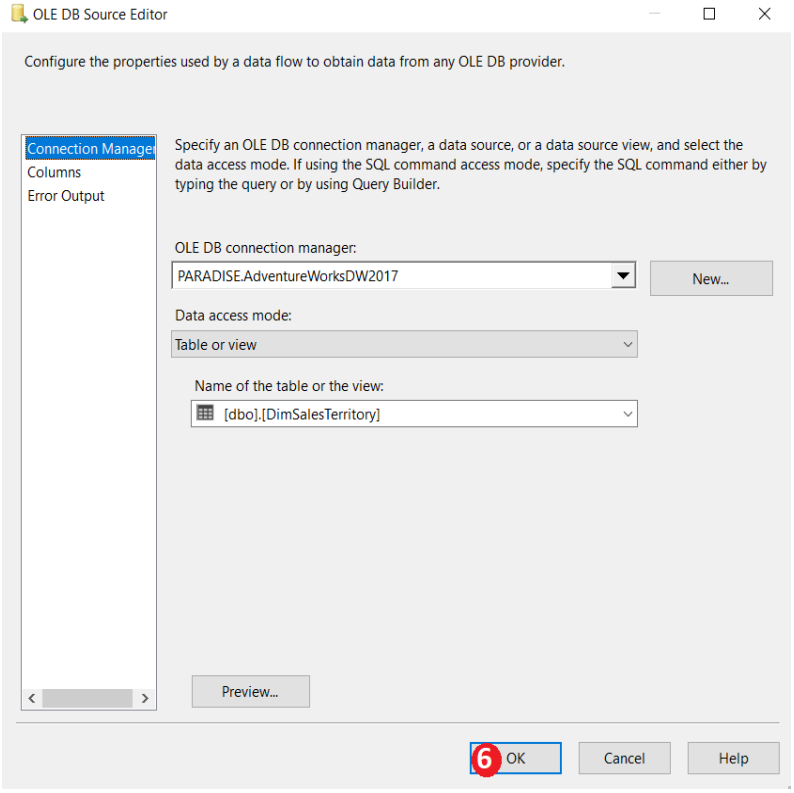
After clicking on "New..." button, a Configure OLE DB Connection Manager window is opened. There we need to provide Data Connection. Again, click on New and it opens Connection Manager window. There we need to provide our server name(1) i.e., PARADISE and connected database(2) is AdventureWorksDW2017. Then click on OK(3) and our new Data Connection is created. Again click on OK(4). Then Our OLE DB Connection manager is created. Then provide the source table i.e., DimEmployee. Then click on OK(5).


Connection Manager
Provider:
Native OLE DB\SQL Server Native Client 11.0
Connection
All
Server name:
1
PARADISE
Refresh
Log on to the server
Authentication:
Windows Authentication
User name:
Password:
Save my password
Connect to a database
Select or enter a database name:
2
AdventureWorksDW2017
Attach a database file:
Browse...
Logical name:
Test Connection
3
OK
Cancel
Help


Configure OLE DB Connection Manager
To create a connection manager based on previously defined connection information, select a data connection, and then click OK. To create a new connection manager, click New.
Data connections:
PARADISE.AdventureWorksDW2017
Data connection properties:
Property
Value
Data Source
PARADISE
Initial Catalog
AdventureWorksDW20...
Integrated Se...
SSPI
Provider
SQLNCLI11.1
New...
Delete
4
OK
Cancel



Then double click on DimSalesTerritory_Source and give the same connection manager i.e., PARADISE.AdventureWorksDW2017 and provide the source table i.e., DimSalesTerritory. Then click on OK*(6)*.



STEP 5: After configuring two OLE DB Source, connect the data flow from DimEmployee_Source to Sort. Also, connect the data flow from DimSalesTerritory_Source to Sort 1. Then configure Sort and Sort1 as following figures:

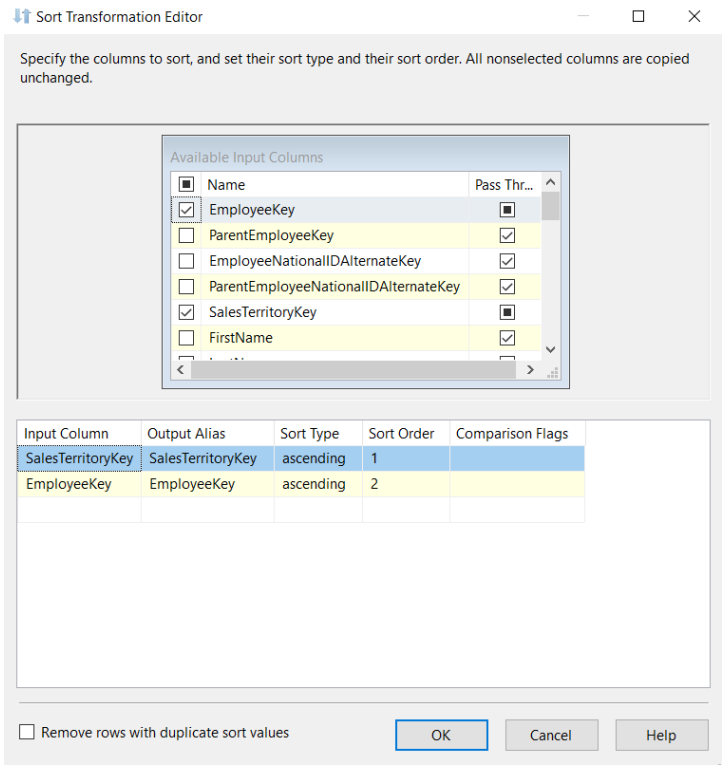


Fig: Configuration Of Sort

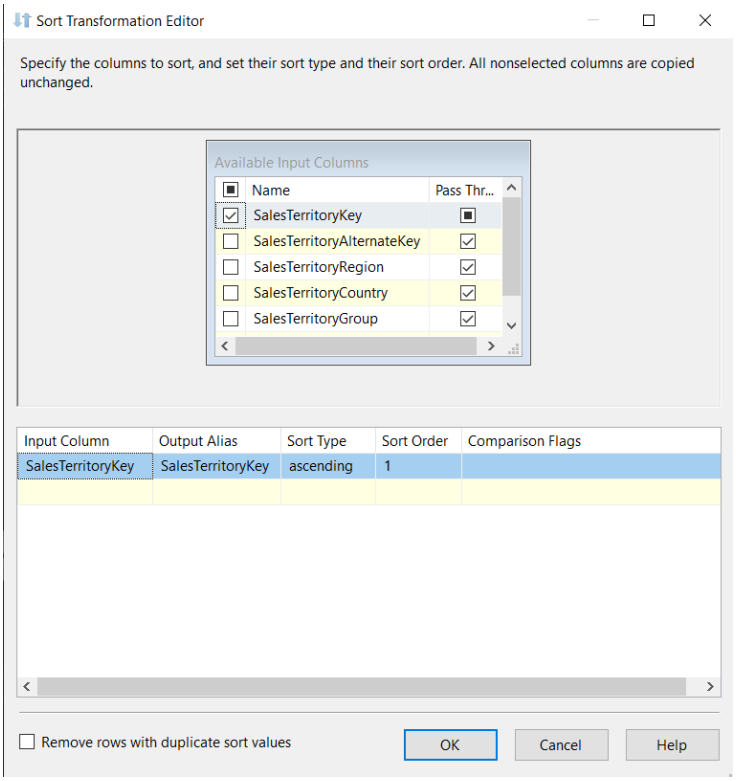
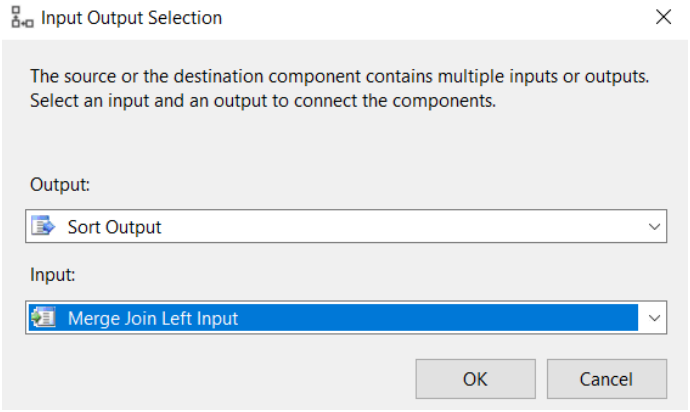
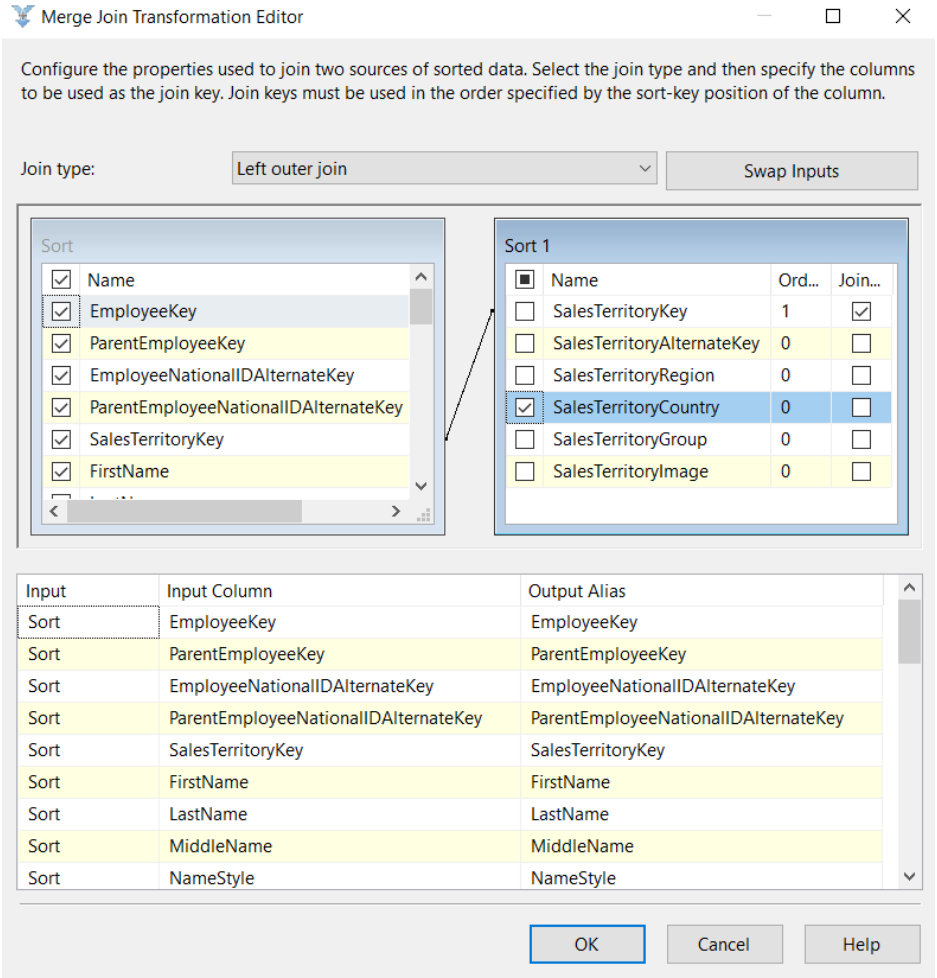


Fig: Configuration of Sort1

STEP 6: After configuring Sort and Sort1, connect the data flow from Sort to Merge Join and an Input Output Selection window is opened. Configure that as following figure:



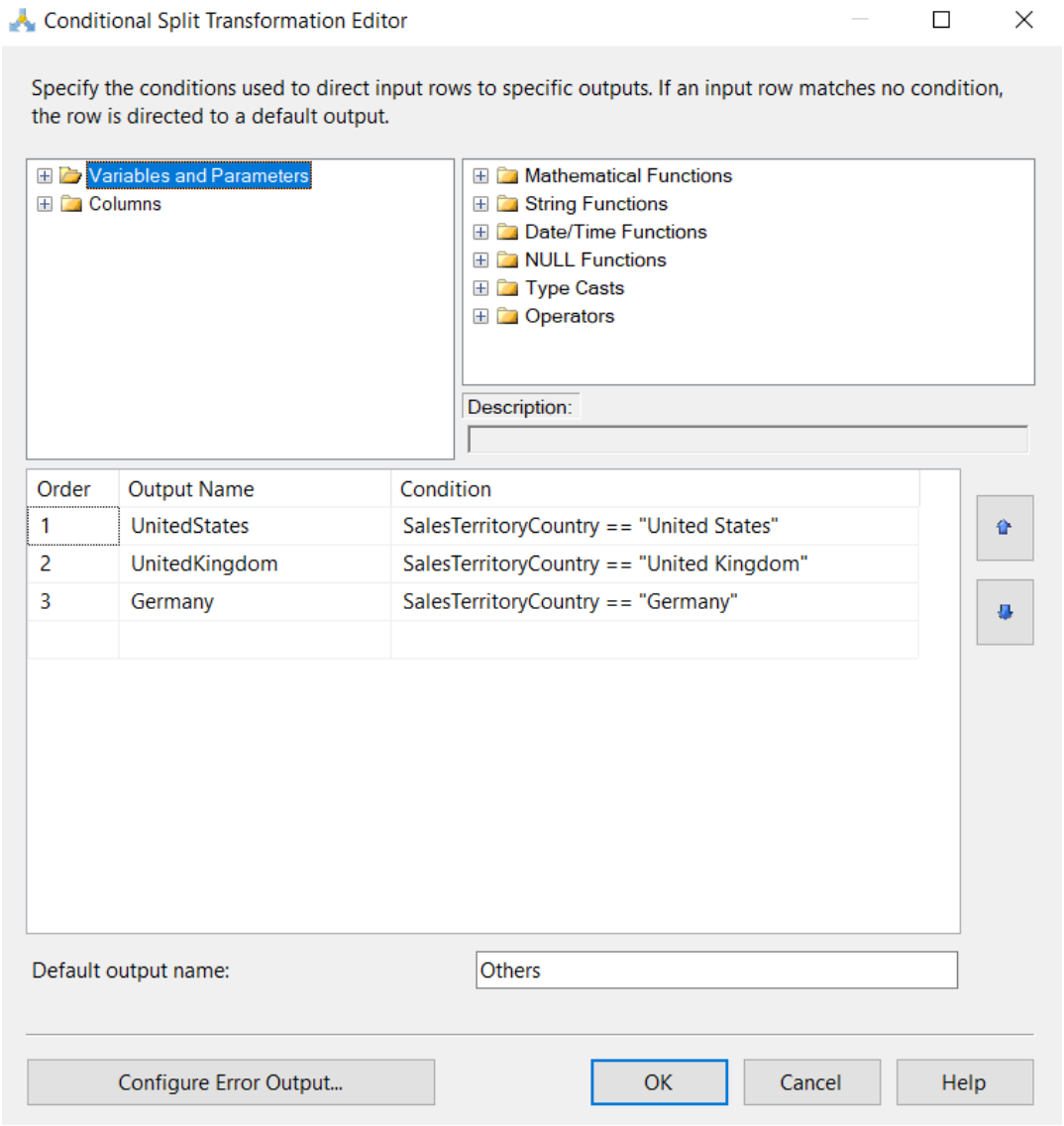
Then, connect the data flow from Sort1 to Merge Join and it is auto configured as merge join right input. After managing the data flow, double click on Merge Join to select the output columns and Join Type. Configure the Merge Join Transformation as following figure:



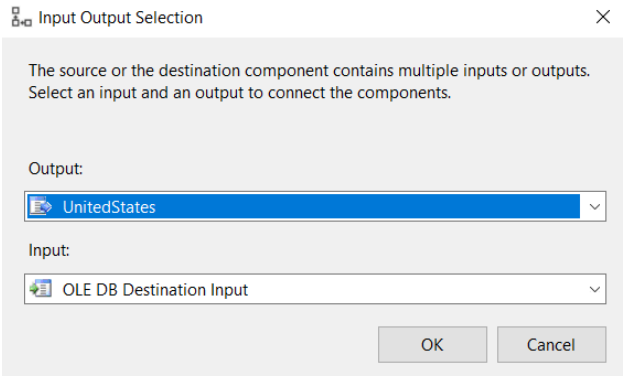
STEP 7: After configuring Merge Join, we connect the data flow from Merge Join to Conditional Split. Then configuring the Conditional Split with 3 conditions:

- SalesTerritoryCountry == "United States"
- SalesTerritoryCountry == "United Kingdom"
- SalesTerritoryCountry == "Germany"

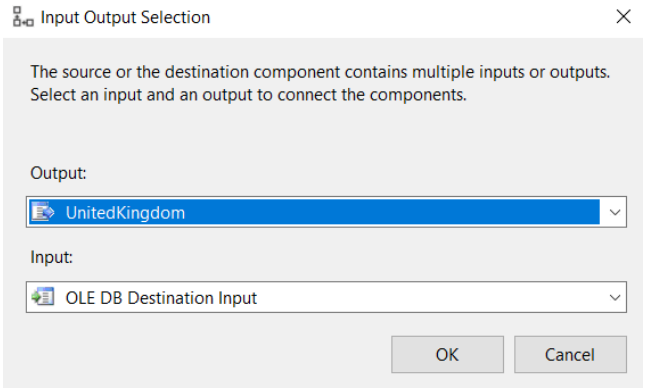
We change the Default Output Name to “Others”. The configuration of Conditional Split is as following figure:



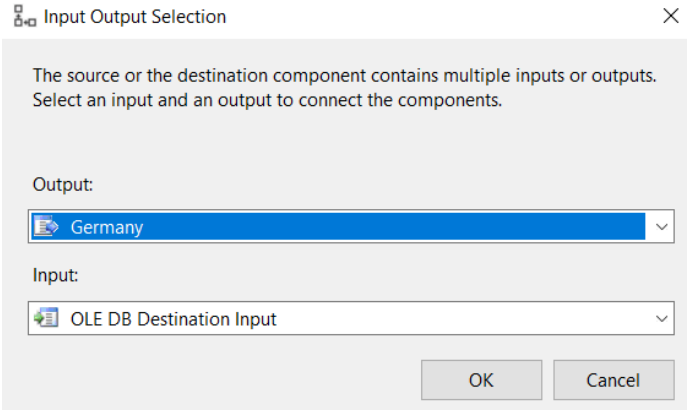
STEP 8: After configuring Conditional Split, we connect the data flow from Conditional Split to one OLE DB Destination named Employee_UnitedStates and it opens an Input Output Selection window. In Output, select the output name "UnitedStates" which is defined in the Condition Split Transformation Task earlier.



Then connect the data flow from Conditional Split to another OLE DB Destination named Employee_UnitedKingdom and it opens Input Output Selection window. In Output, select the output name "UnitedKingdom".

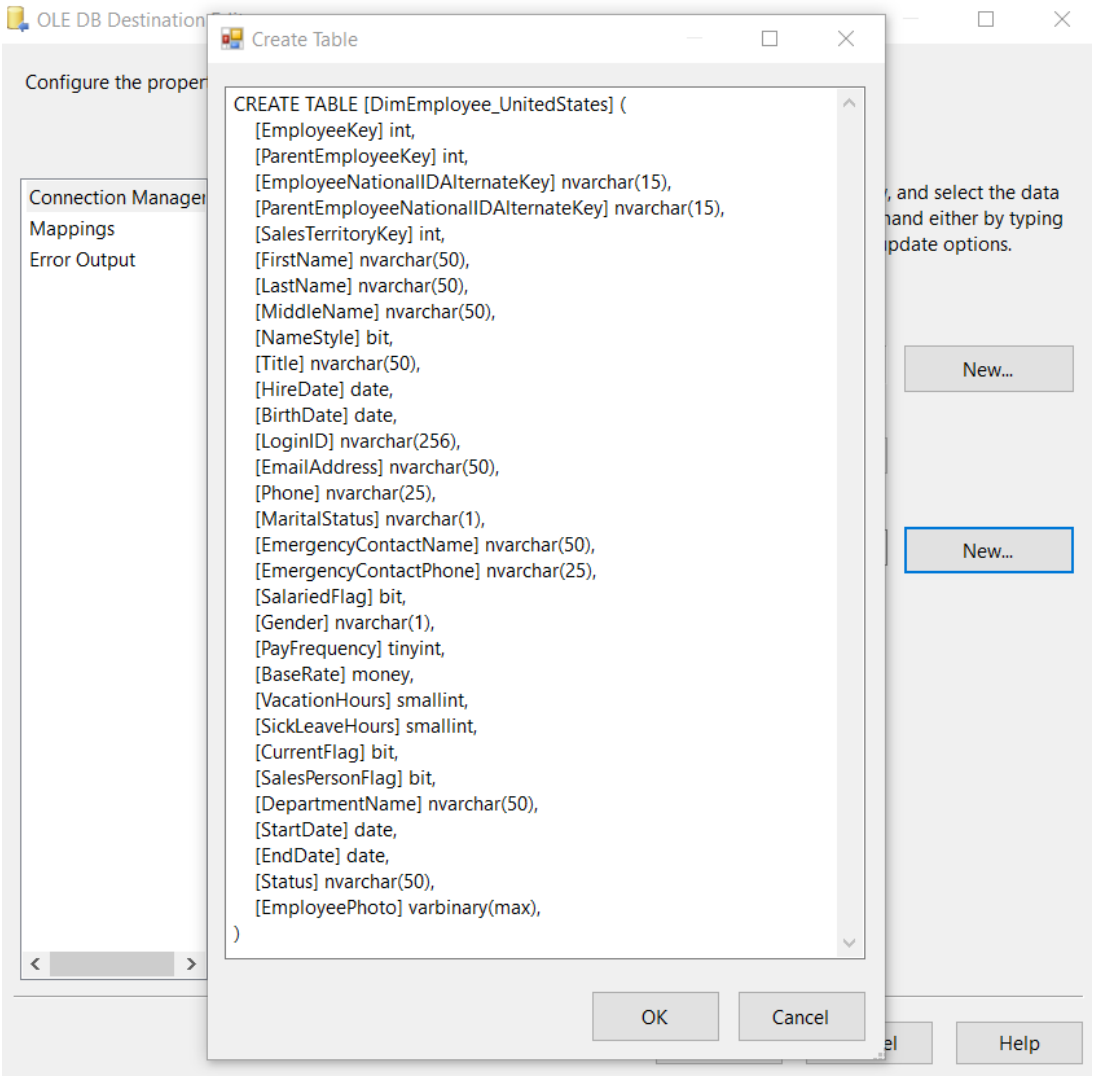


Then connect the data flow from Conditional Split to another OLE DB Destination named Employee_Germany and it opens Input Output Selection window. In Output, select the output name "Germany".

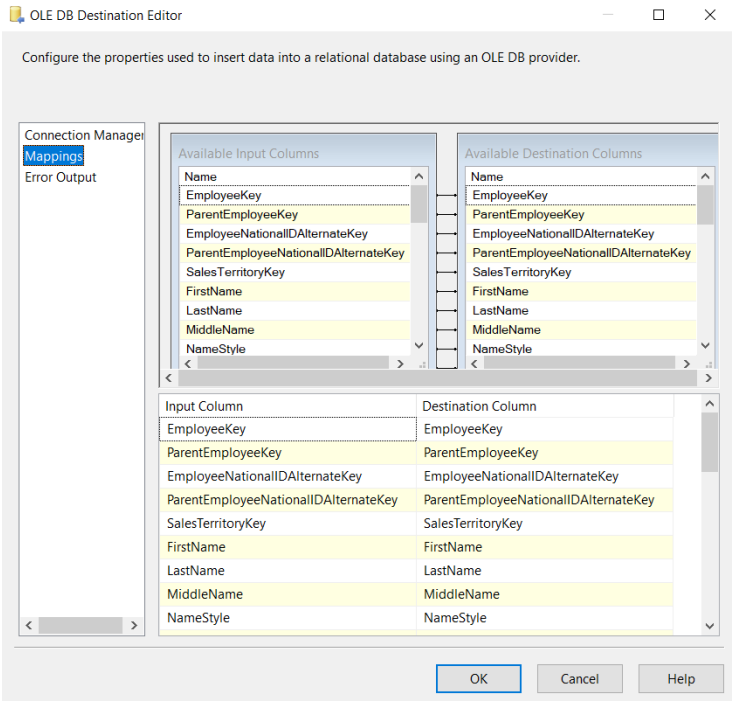


Then connect the data flow from Conditional Split to another OLE DB Destination named Employee_Others. And it is auto configured as output name “Others”.

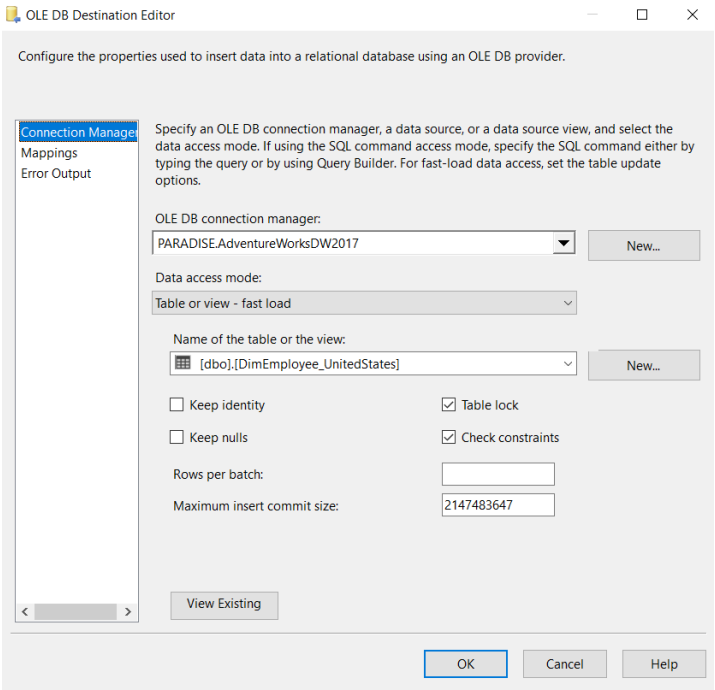
STEP 9: Now we need to configure all four OLE DB Destinations. We start with Employee_UnitedStates. Double click on that and it opens an OLE DB Destination Editor window. There we need to provide the PARADISE.AdventureWorksDW2017 as OLE DB Connection Manager and for destination table we need to create a new table. So, give the query shown in following figure.



By this query, we create a new table in AdventureWorksDw2017 for populating the details of Employees of United States. After that, change the tab to Mappings in OLE DB Destination Editor window.

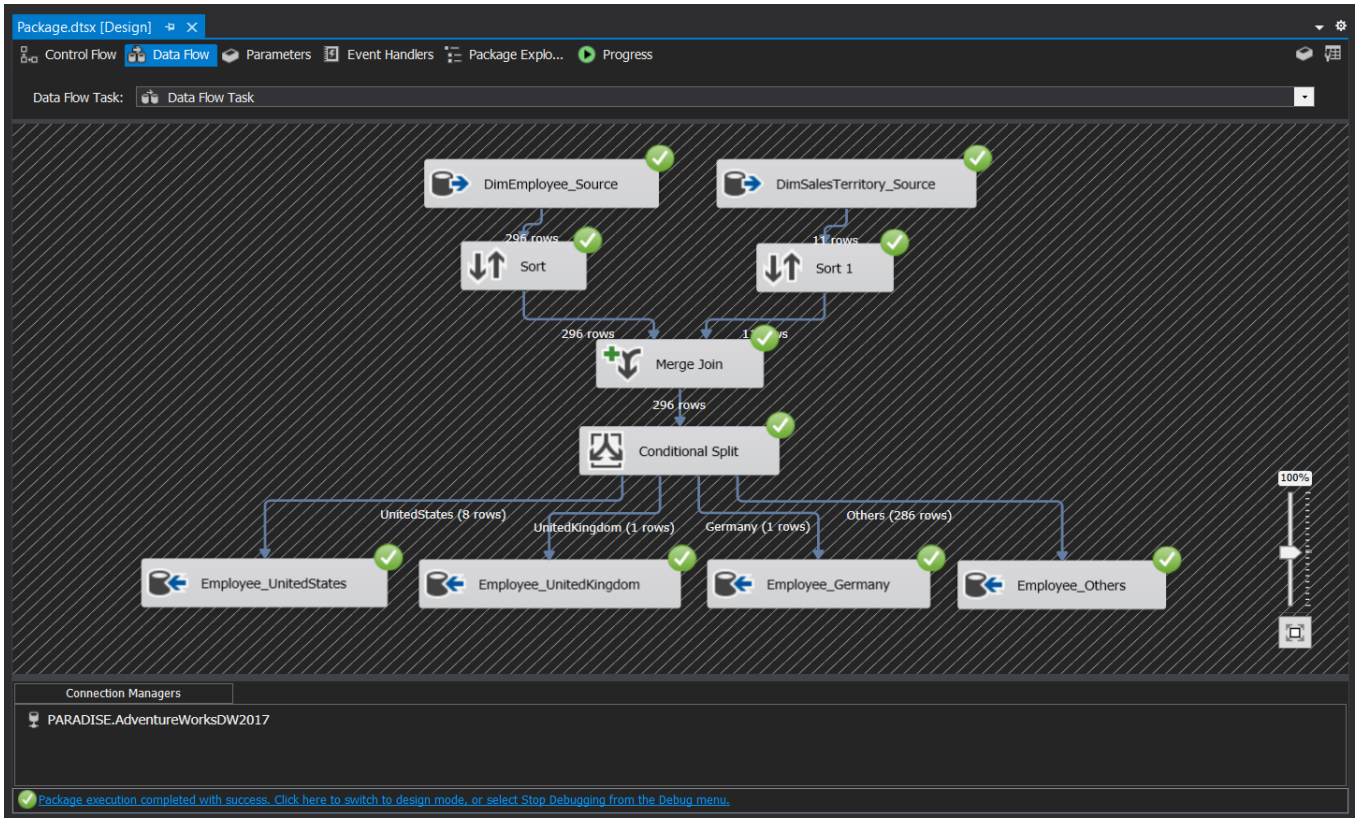


In Mappings we can see that the input columns are automatically mapped with output columns because we have used same column name in DimEmployee_UnitedStates as the output from Conditional Split. So, click on OK.

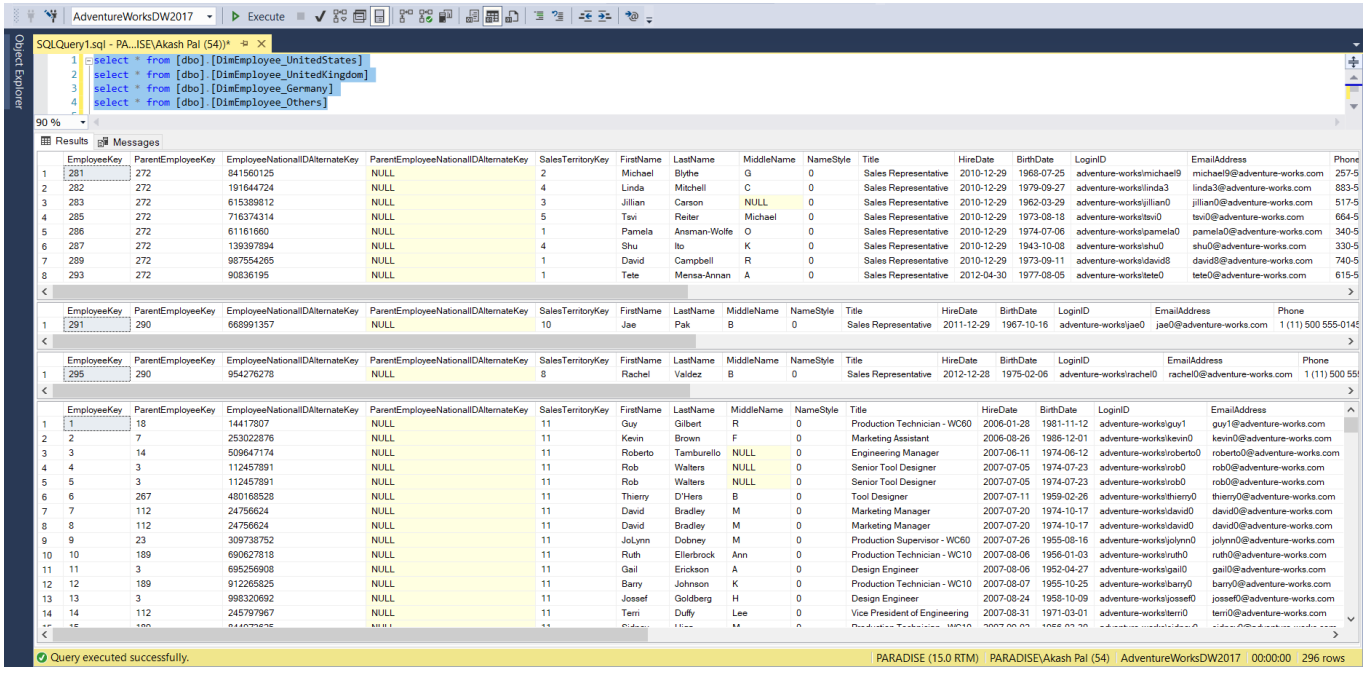


Also, configure Employee_UnitedKingdom in the similar way. For that, just provide the same OLE DB connection manager and create DimEmploye_UnitedKingdom table with same column name and datatype. Also, we follow similar steps to configure Employee_Germany and Employee_Others.

STEP 10: Now our package is ready. So, execute the package.



Then for checking the result, we run the following query in SQL Server Management Studio and we have got following result shown in the following figure:



CONCLUSION:

Our package is successfully run and working fine. The requirement is completely fulfilled by this solution package. The complete SSIS package looks as per the following screenshots.