MIS 587 Guide to Data Warehouse Creation using SSIS Package

(For more materials and examples, go to shared link)

Overview:

In this exercise, we are going to build a data warehouse for the given data sample. We will be using SQL Server Integration Services tool to build a data warehouse. By the end of this exercise you should be able to design a data warehouse schema, implement the design using SSIS tool and perform analysis/ reporting using SSAS/ SSRS.

Problem Statement:

A physical warehouse near Tucson has inventory belonging to various categories which is shipped to many countries around the world by three shippers. A data warehouse needs to be built to maintain a single source of truth relating to the information about the goods shipped. They want to use this data to make business decisions.

Solution:

- Analyze the data sample
- Design appropriate schema for data warehouse
- Build Data warehouse
- Design dashboards and reports to help Inventory/ Superstore make their business decisions (Out of Scope of this Lab session)

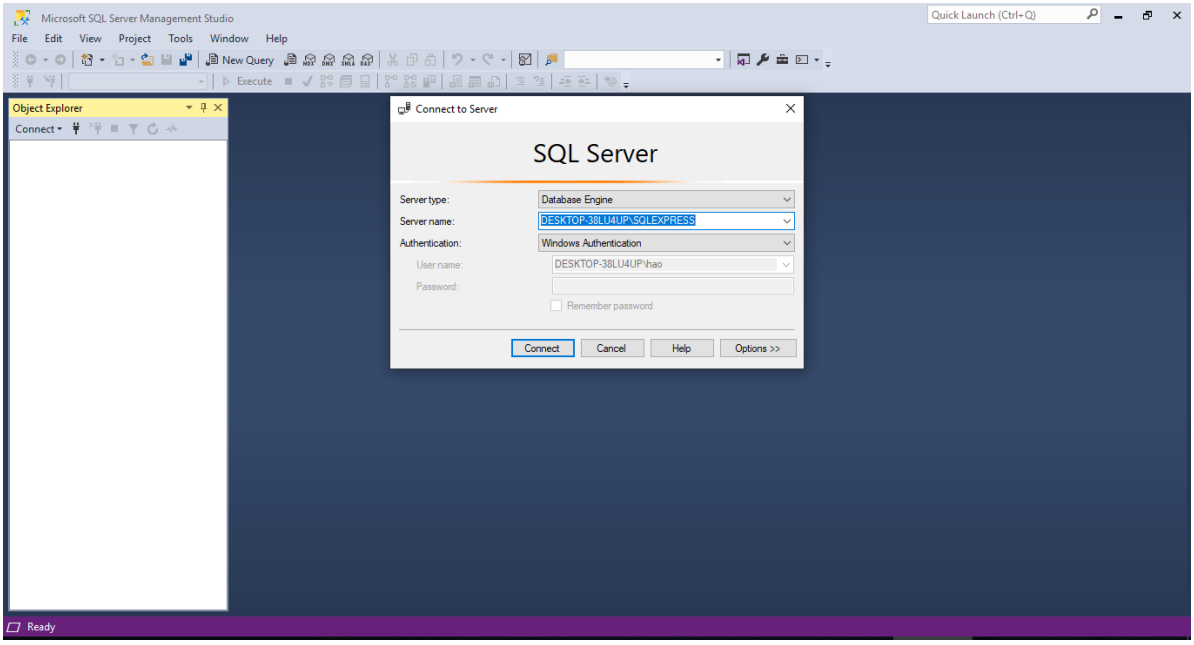
Pre-requisite:

Install SQL Server Express, SQL Server Management studio, SQL Server Data Tools, Visual Studio community following the Installation tutorial

Description step by step:

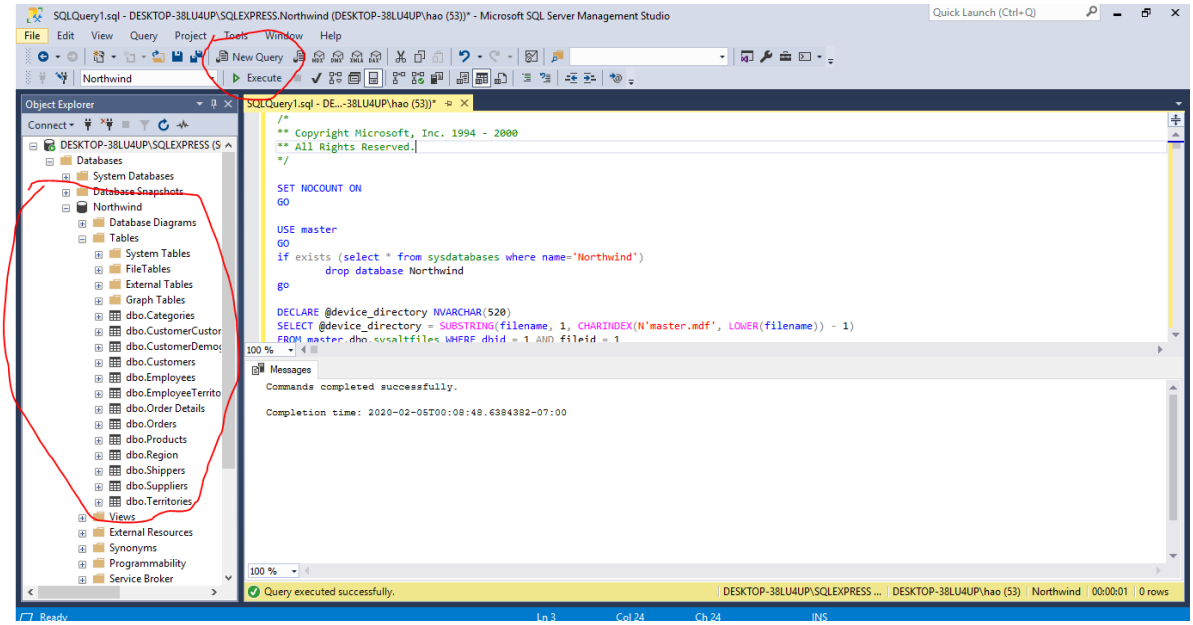
We will extract tables from Northwind database, import data into `Northwind_DW` database and build star schema.

1. Open Microsoft SQL server Management Studio.

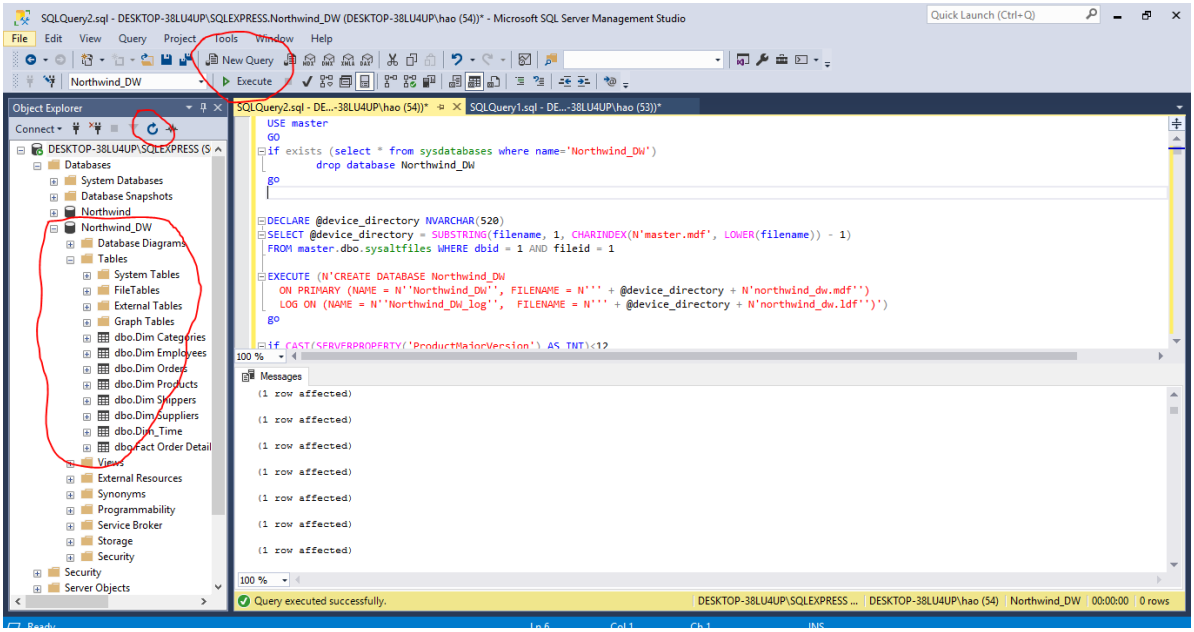


2. Build Northwind database as data source:

click `New Query`, copy sql code from the link, paste the code into query editor, click `Execute`, you will find a database `Northwind` in left sidebar. This contains Inventory source tables with data. Relevant attributes for each entity are present in the tables.

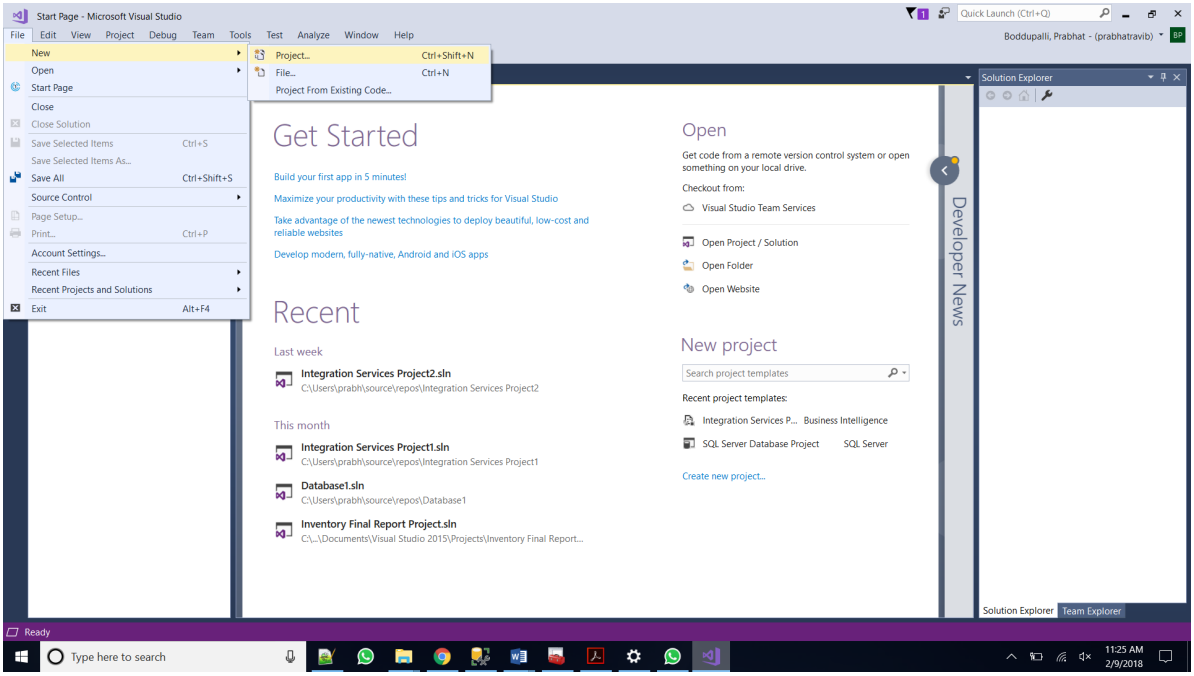


3. Build a database named `Northwind_DW` with empty tables for the `Northwind` Data warehouse using sql code like previous step. Refresh database list, you will find a new database `Northwind_DW`.

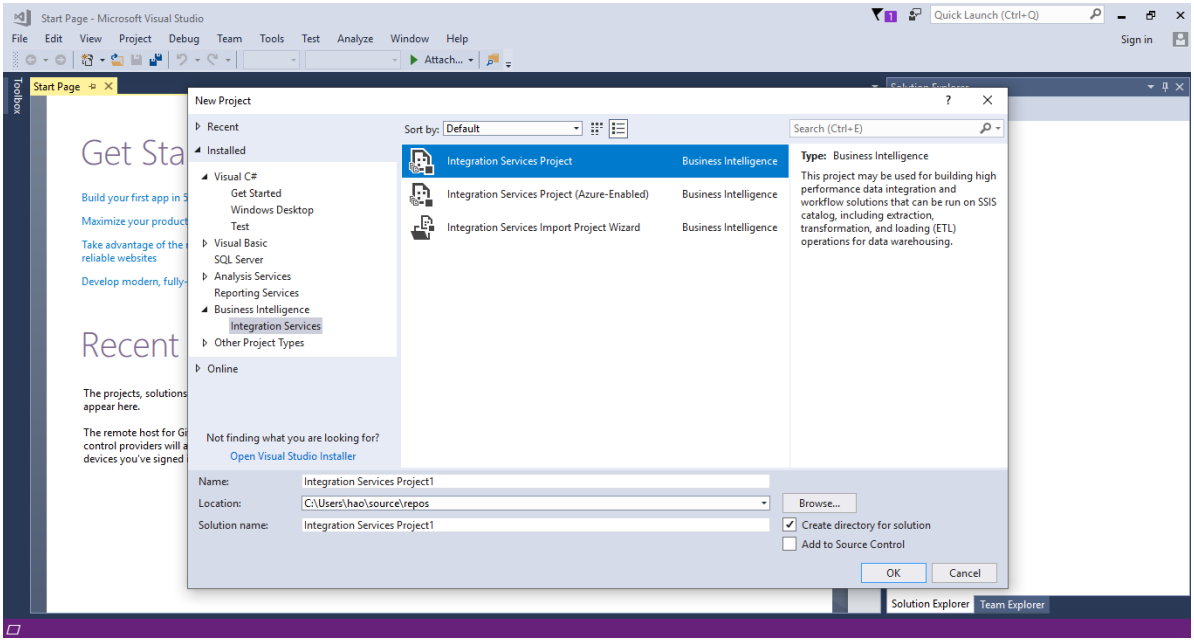


Now, Microsoft Visual Studio will be used to load the data to the northwind data warehouse from the Northwind database, after doing certain transformations.

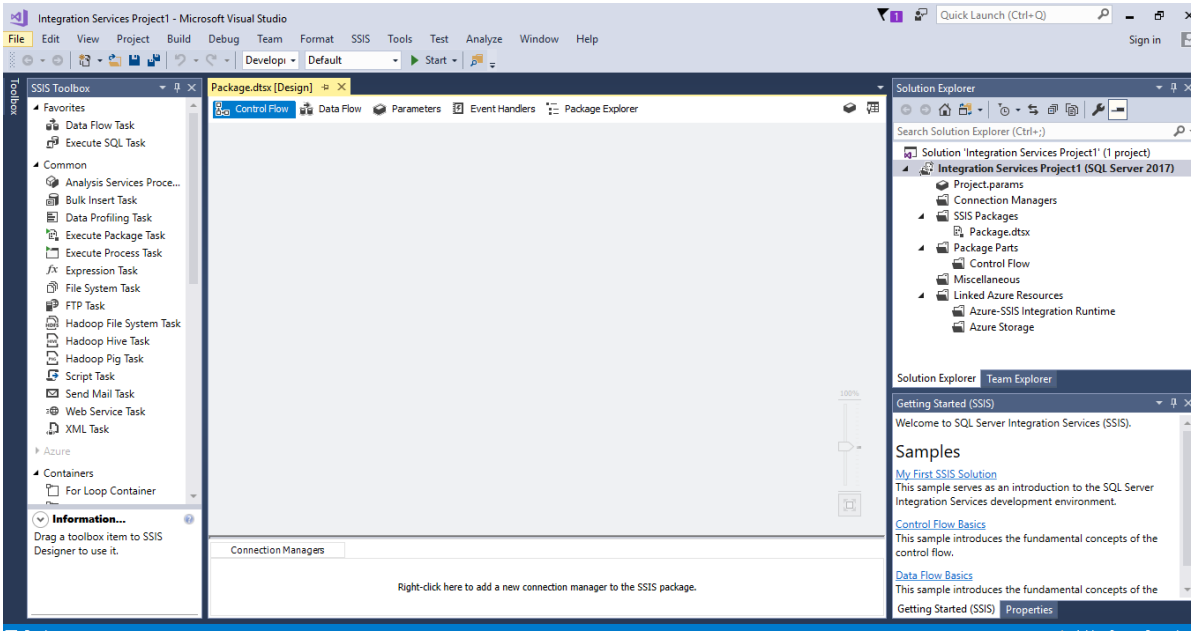
4. Open Microsoft Visual Studio. Choose `File-> New->Project`



5. A window `New Project` pops up. Choose `Business Intelligence->Integration Services Project`

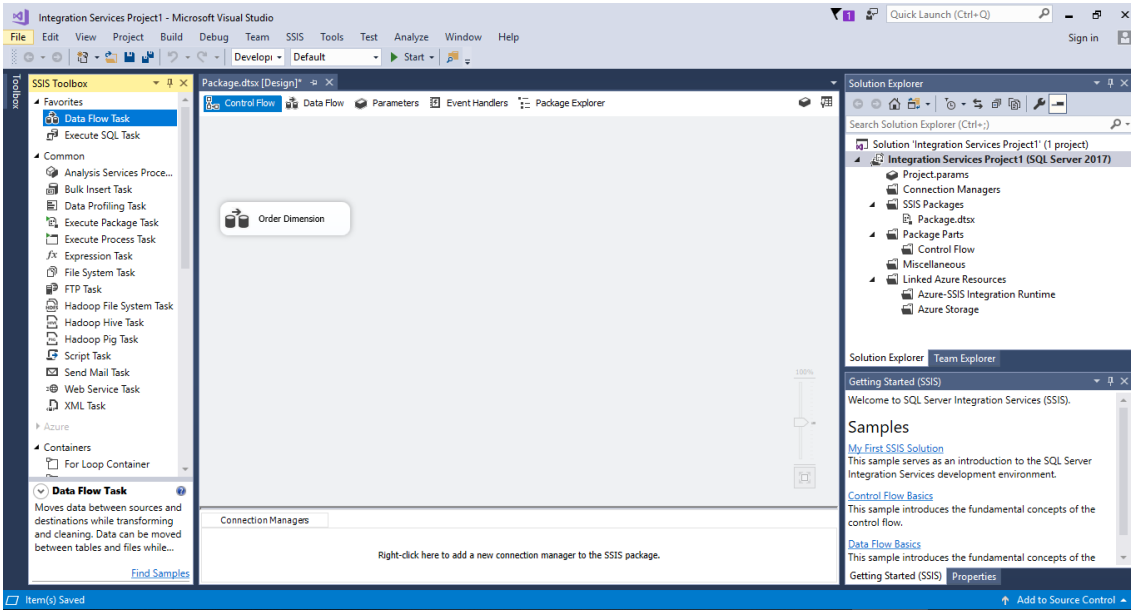


6. You should get a blank screen as below:

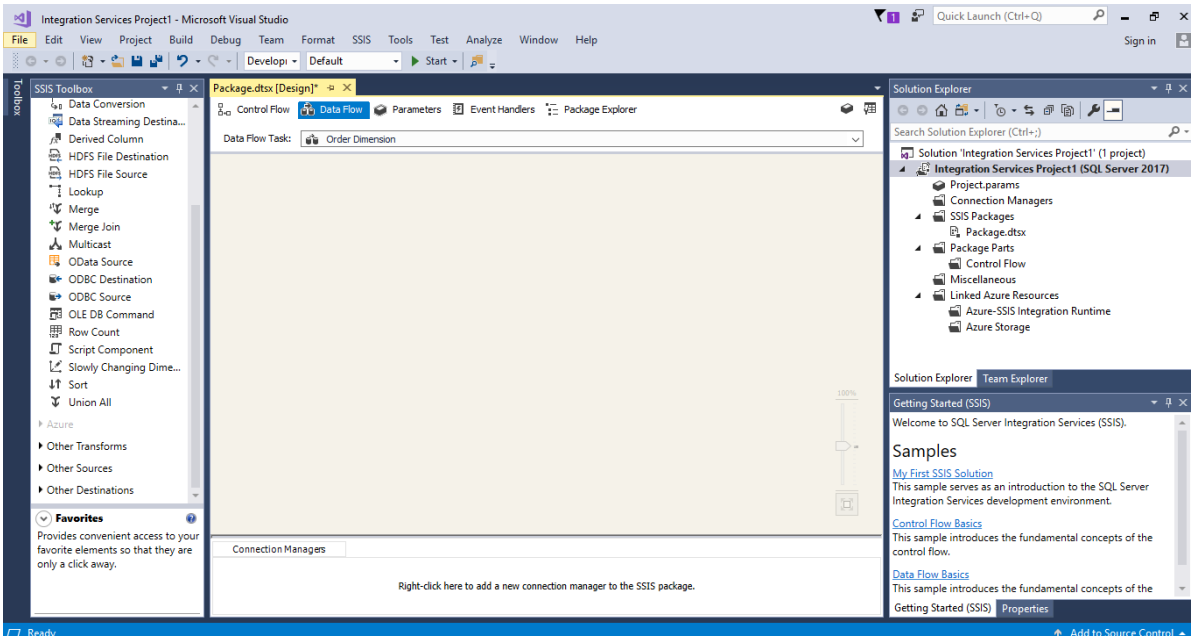


Using Microsoft Visual Studio, Microsoft SQL server is accessed for the data integration of the Northwind Data warehouse.

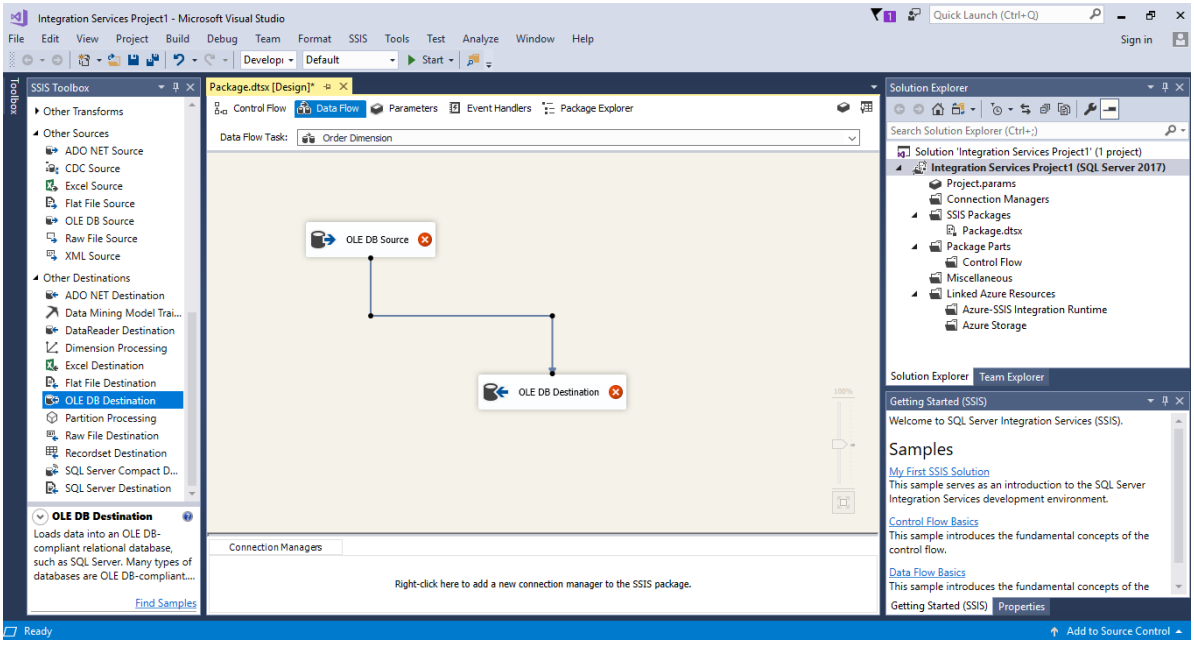
7. From left sidebar, Drag and drop `Data Flow task` onto the workspace. Rename it as `Order Dimension`



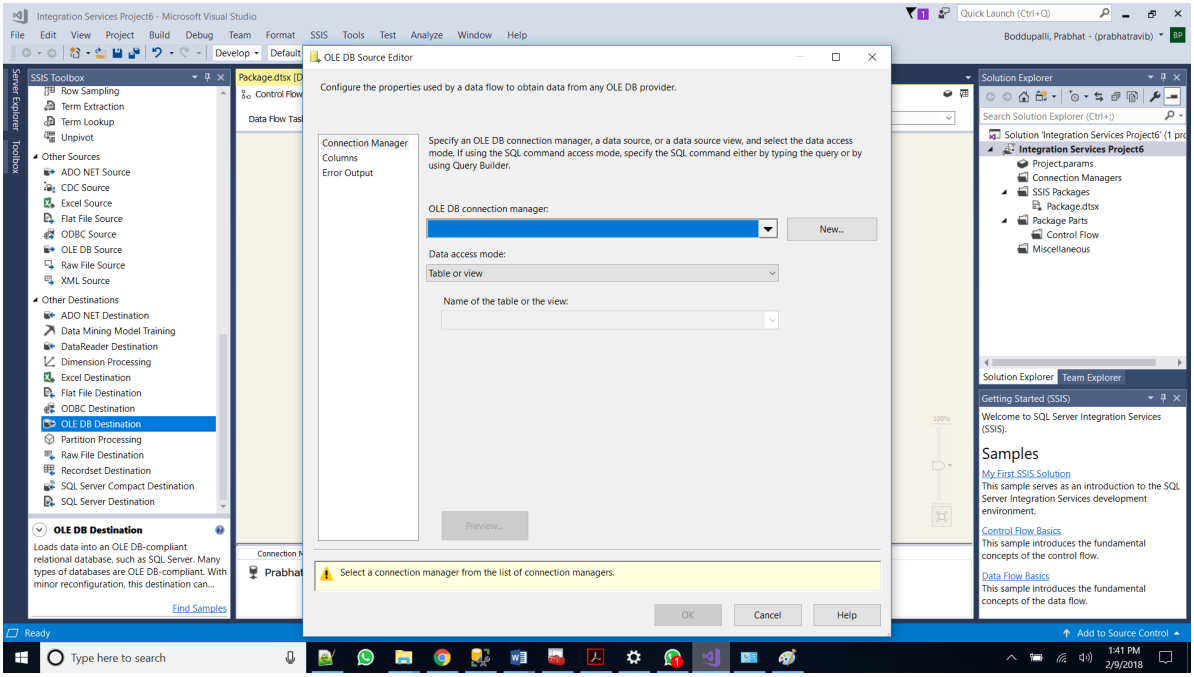
8. Double Click on the `Order Dimension` data flow. You should see a page like below open.



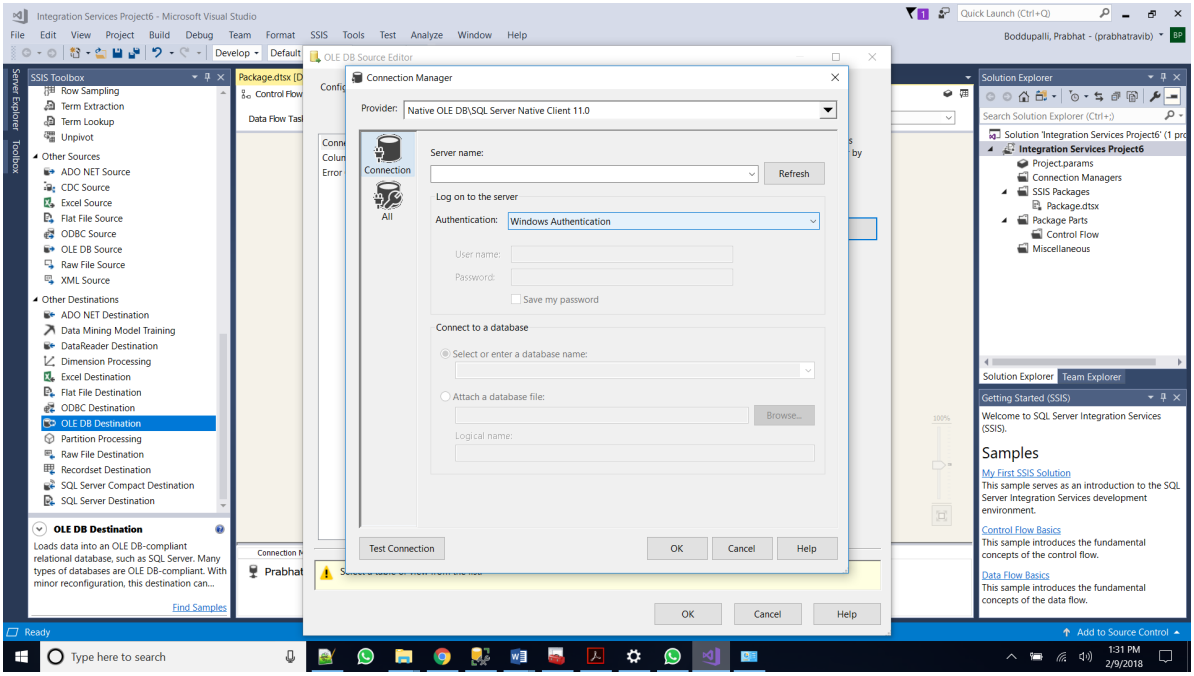
9. On the left-hand bar, Drag `OLE DB Source` from `Other Sources` on to the workspace and `OLE DB Destinations` under `Other Destinations` onto the workspace. Join both using the blue arrow.



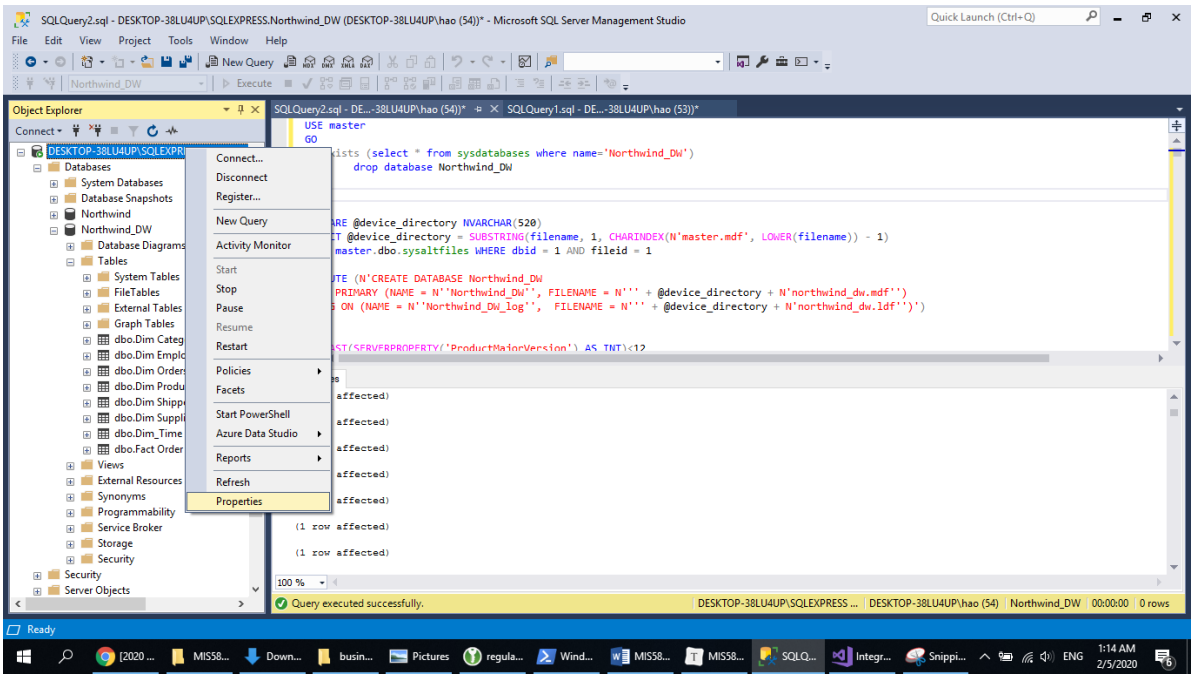
10. Double Click on the `OLE DB Source`.

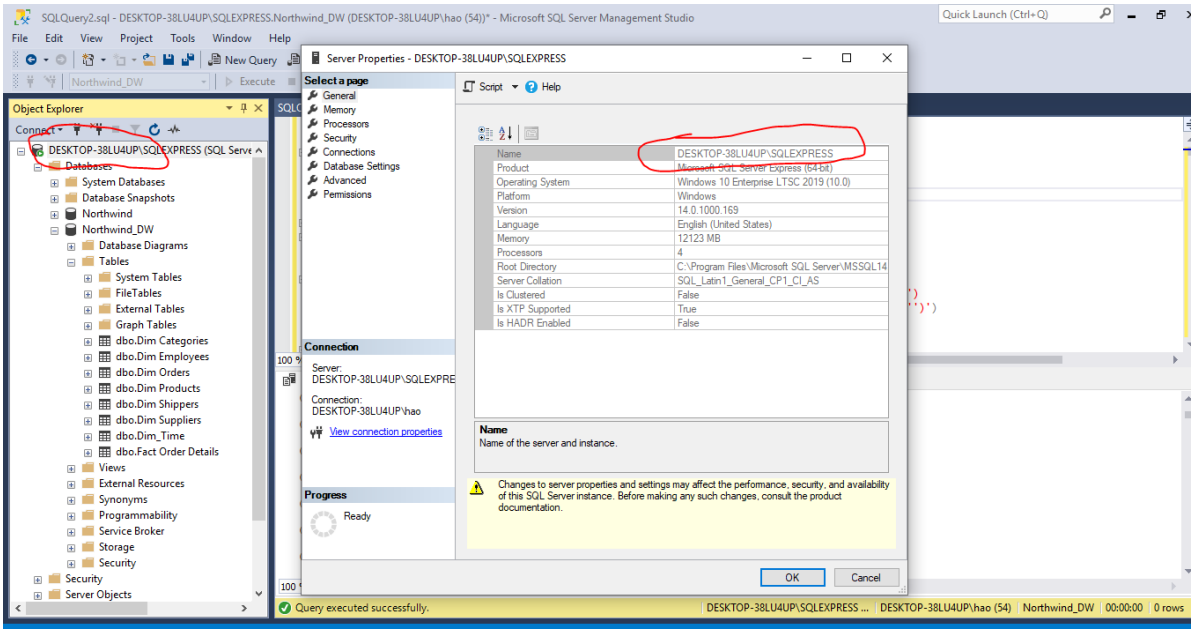


11. Click `New` Button to create a new connection manager. This creates a new connection to a database installed on Microsoft SQL server.

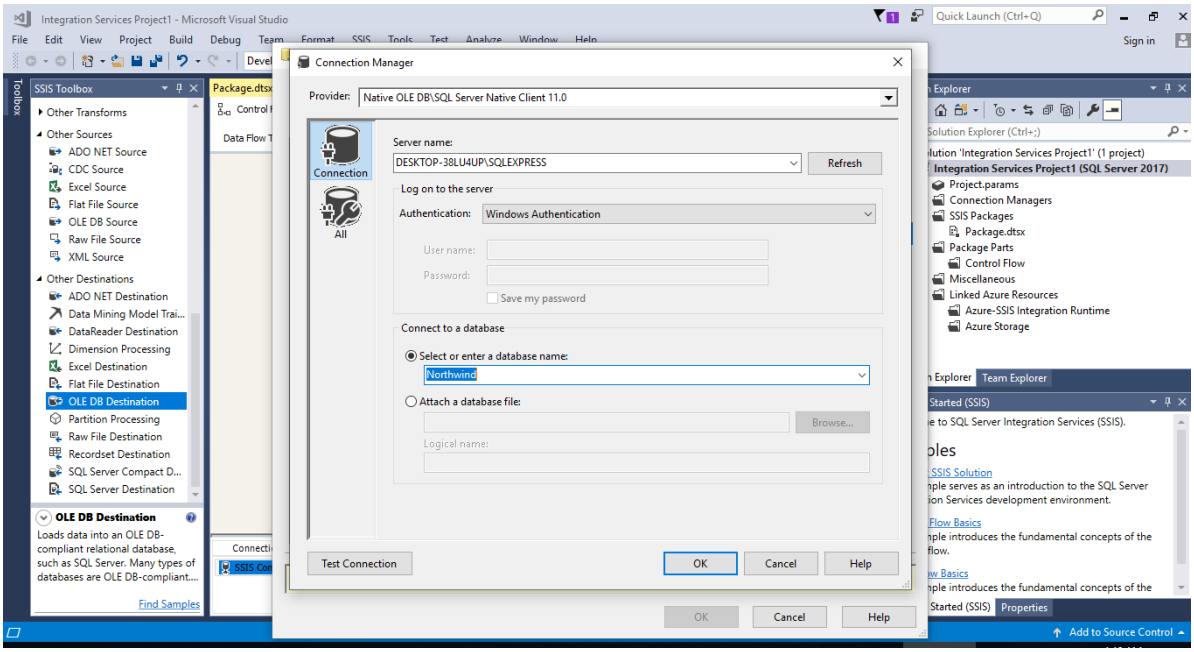


12. In the field `Server Name`, Type in the name of the Microsoft SQL Server installed on your computer. In SQL Sever Management Studio, the name to choose is shown by right click on your server name, choose `Properties`.

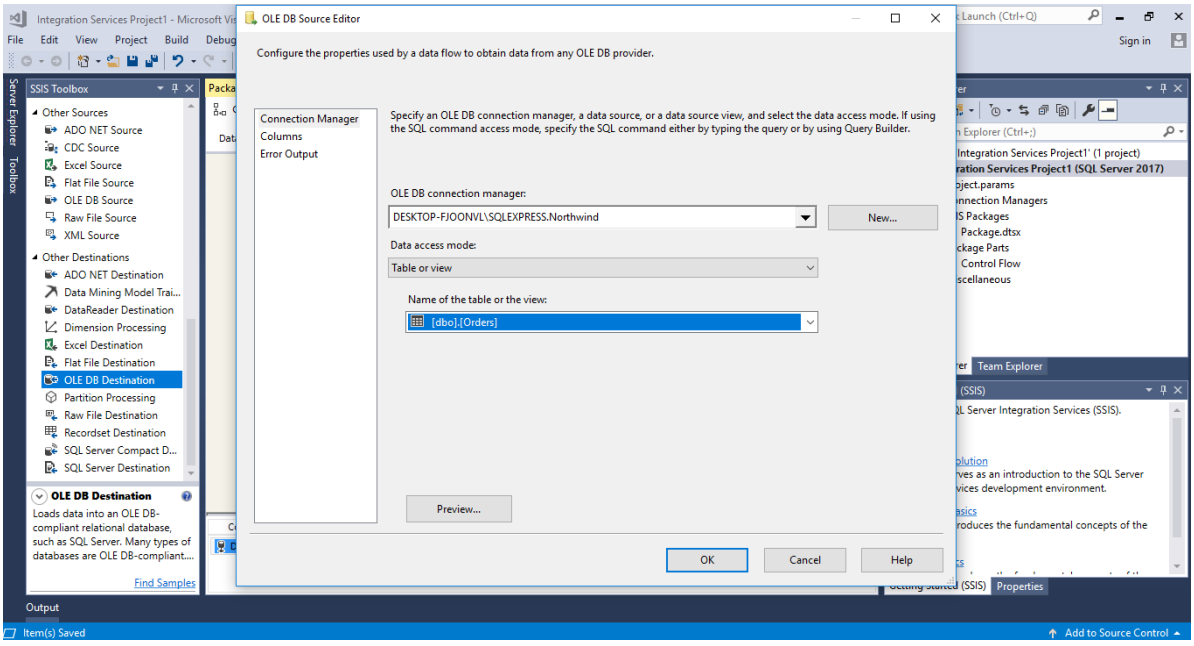




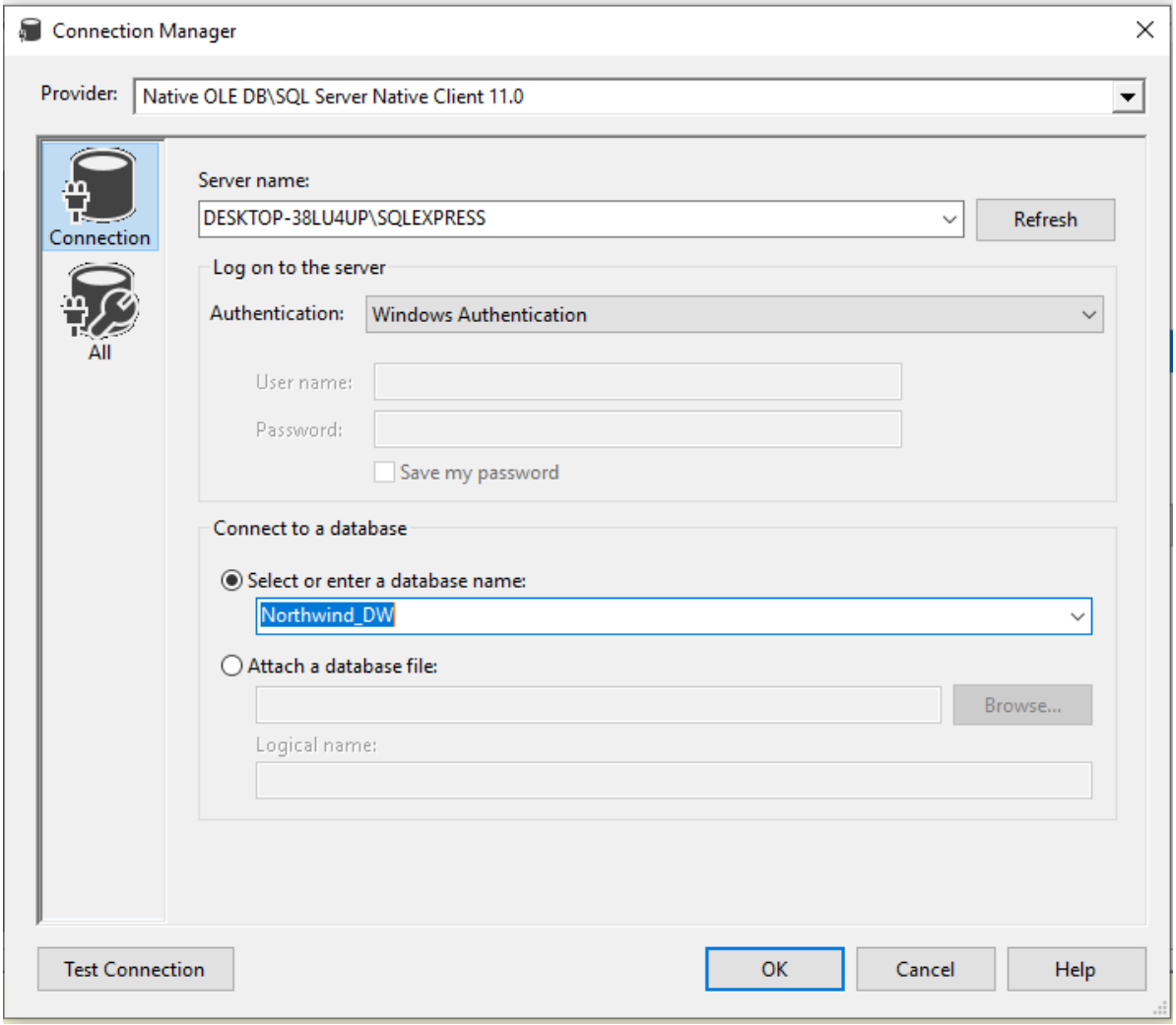
13. Next, under Select or enter a database name, choose `Northwind` Database. Click OK.



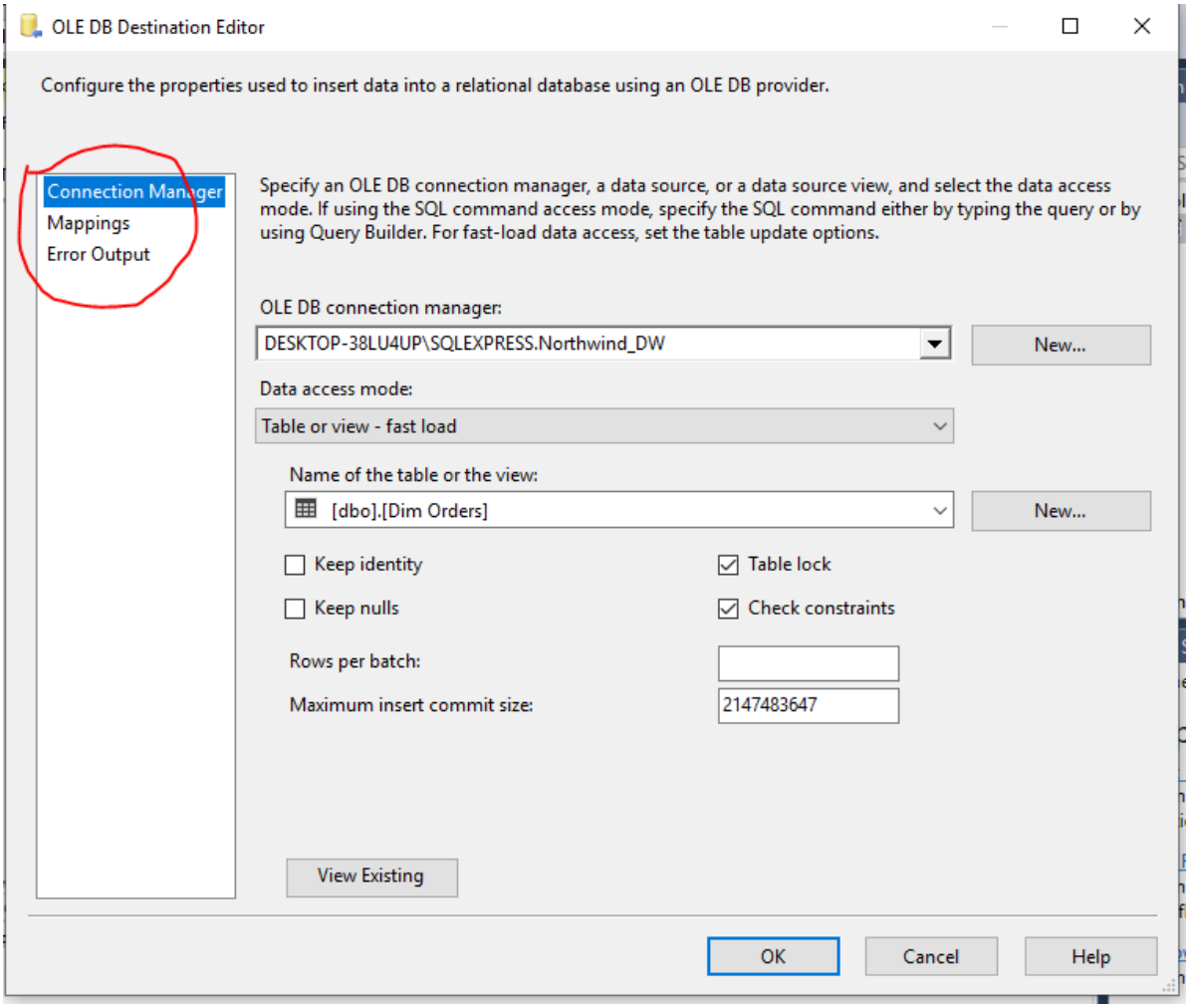
14. Now choose `Orders` table under `Name of the table or view`. Click OK.

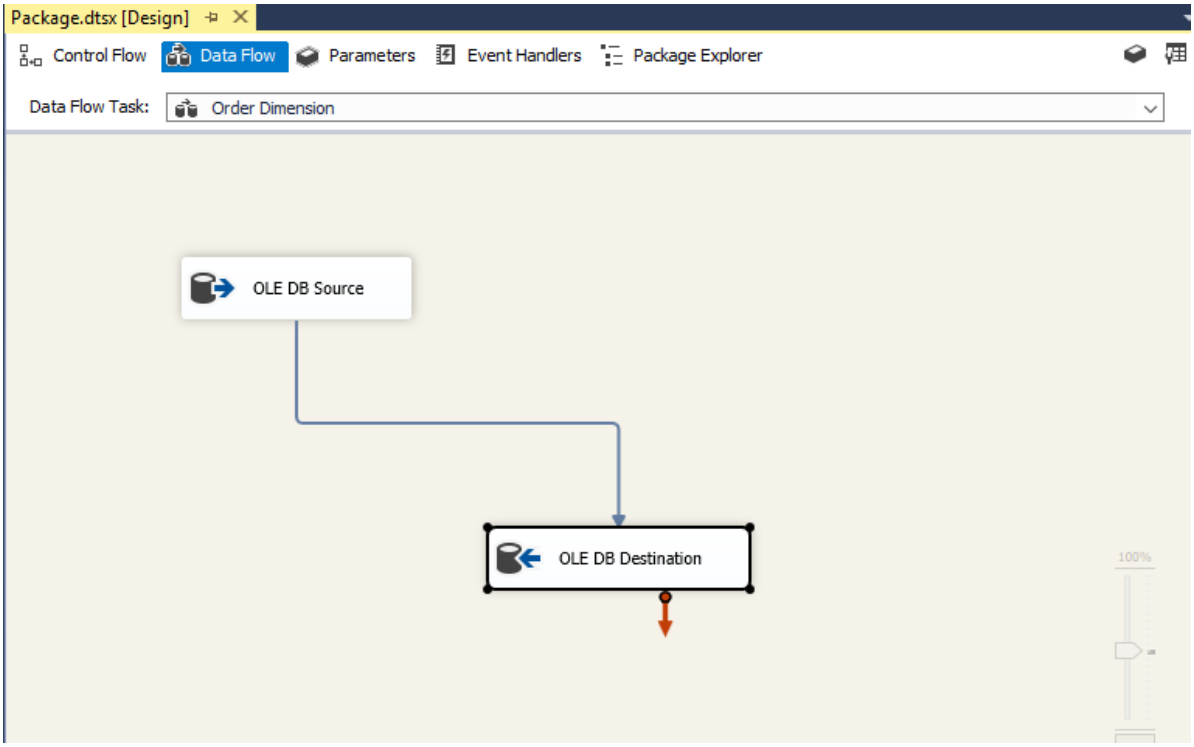


15. In the Data Flow workspace, double click on the `OLE DB destination` and Create a connection manager to the Northwind_DW Database. Repeat Steps 11-13 to create a new connection manager similar to what was done for the northwind database.

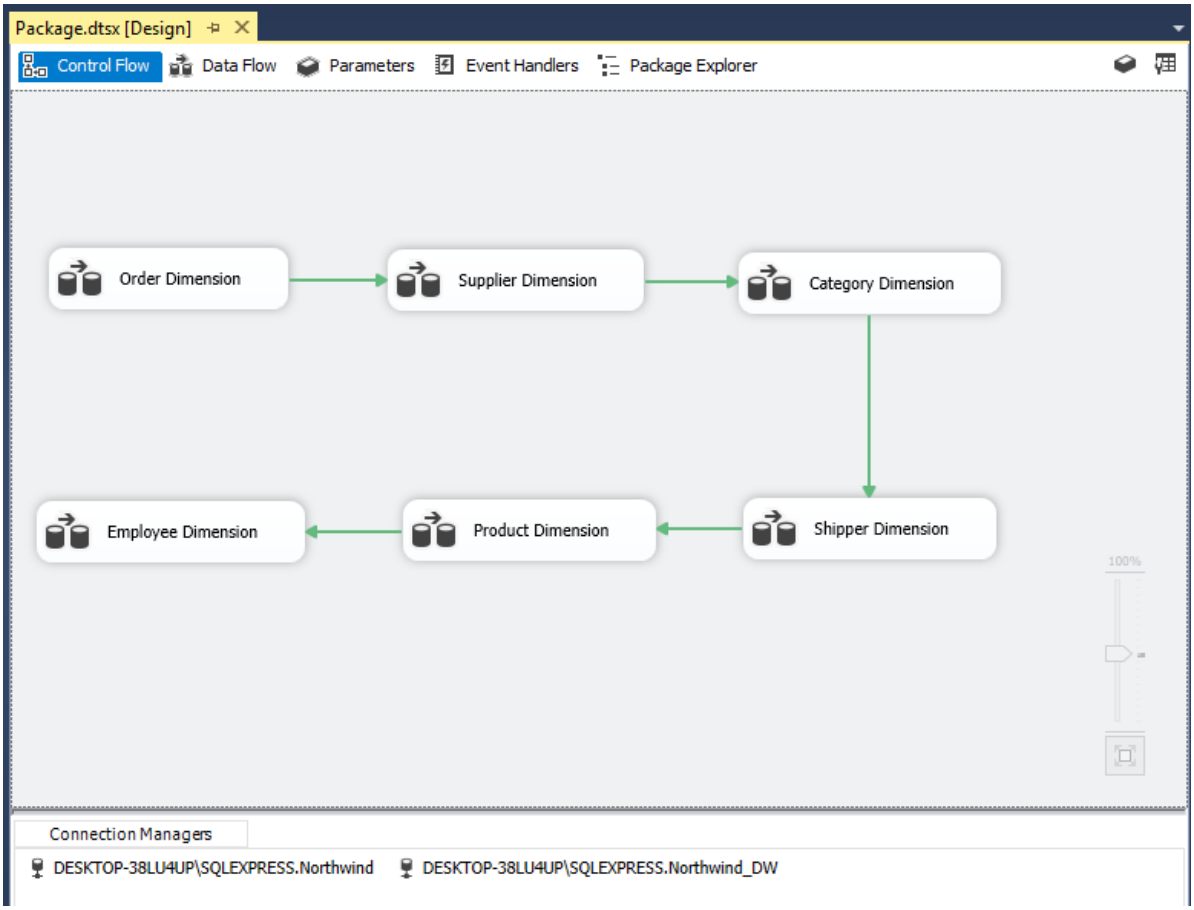


16. After the new connection manager has been set up, Choose `Dim Orders`. This is a dimension created to hold basic qualitative attributes about orders. Click `Mappings`, then Choose OK. `Errors` of `OLE DB Source` and `OLE DB Destination` disappear. (`Keep identity`: link1, link2)



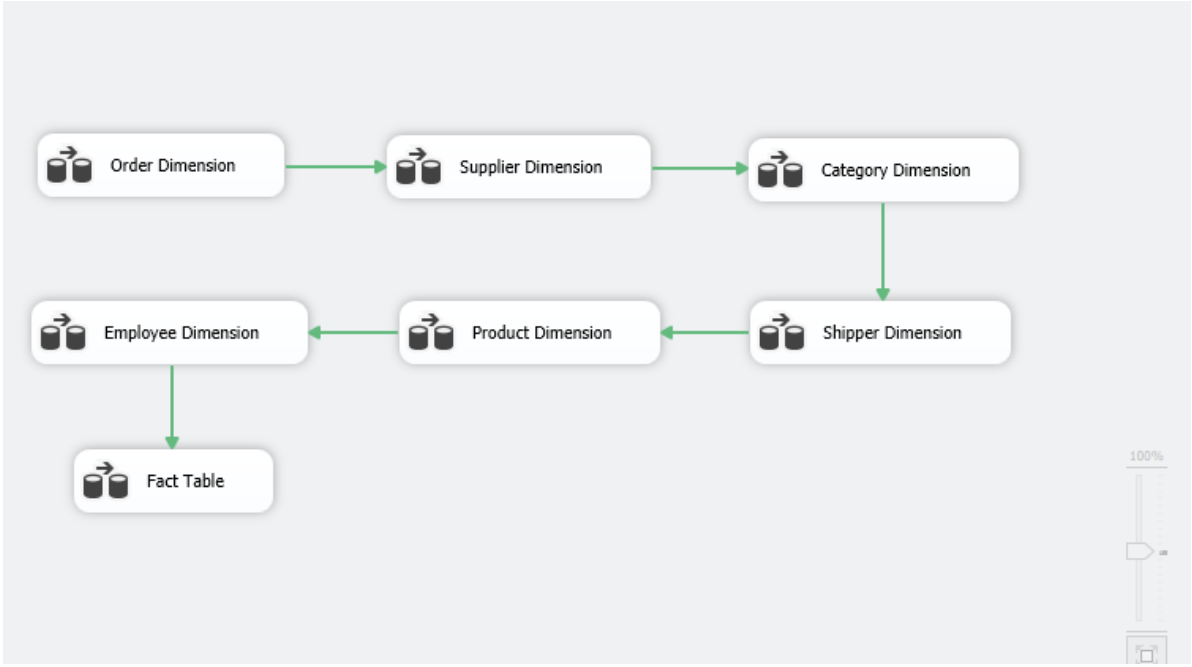


17. Repeat this procedure for `Supplier Dimension`, `Category Dimension`, `Shipper Dimension`, `Product Dimension`, `Employee Dimension`. Then join them as shown below:

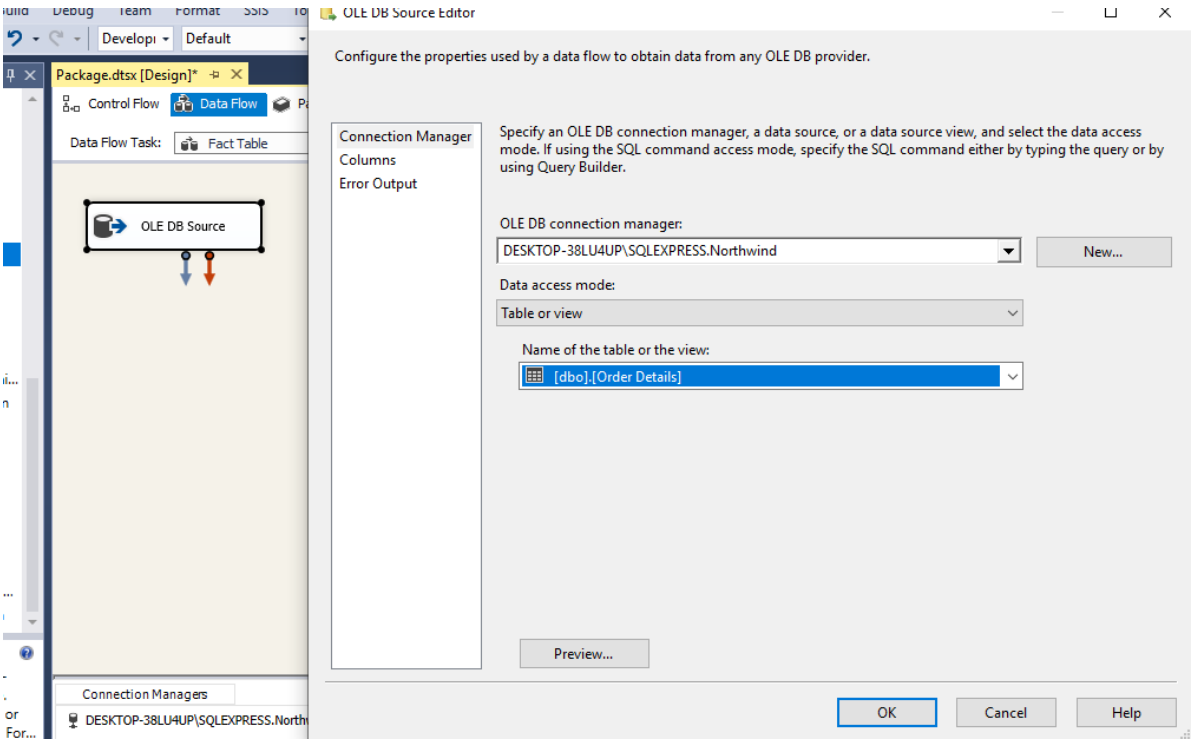


18. The Inventory Fact table is loaded from the Inventory `order details` table. We already create an empty Fact table in `Northwind_DW`, foreign key placeholders are created, which will be populated later.

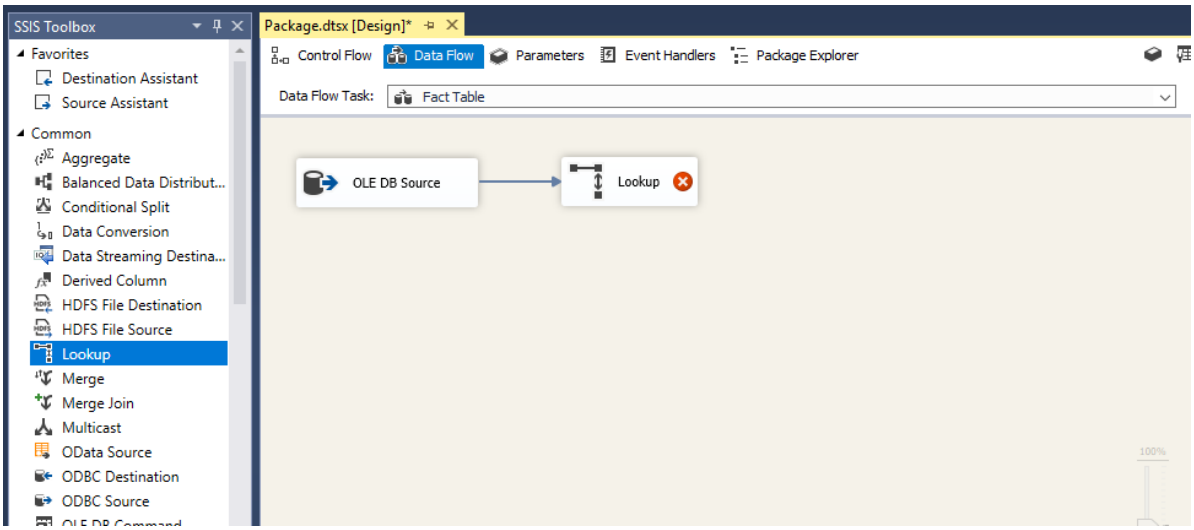
    Similarly, drag `Data Flow Task`, rename it as `Fact Table`, double click, Like before:



Drag the DB source from the left pane, select `Order Details` from Northwind database.

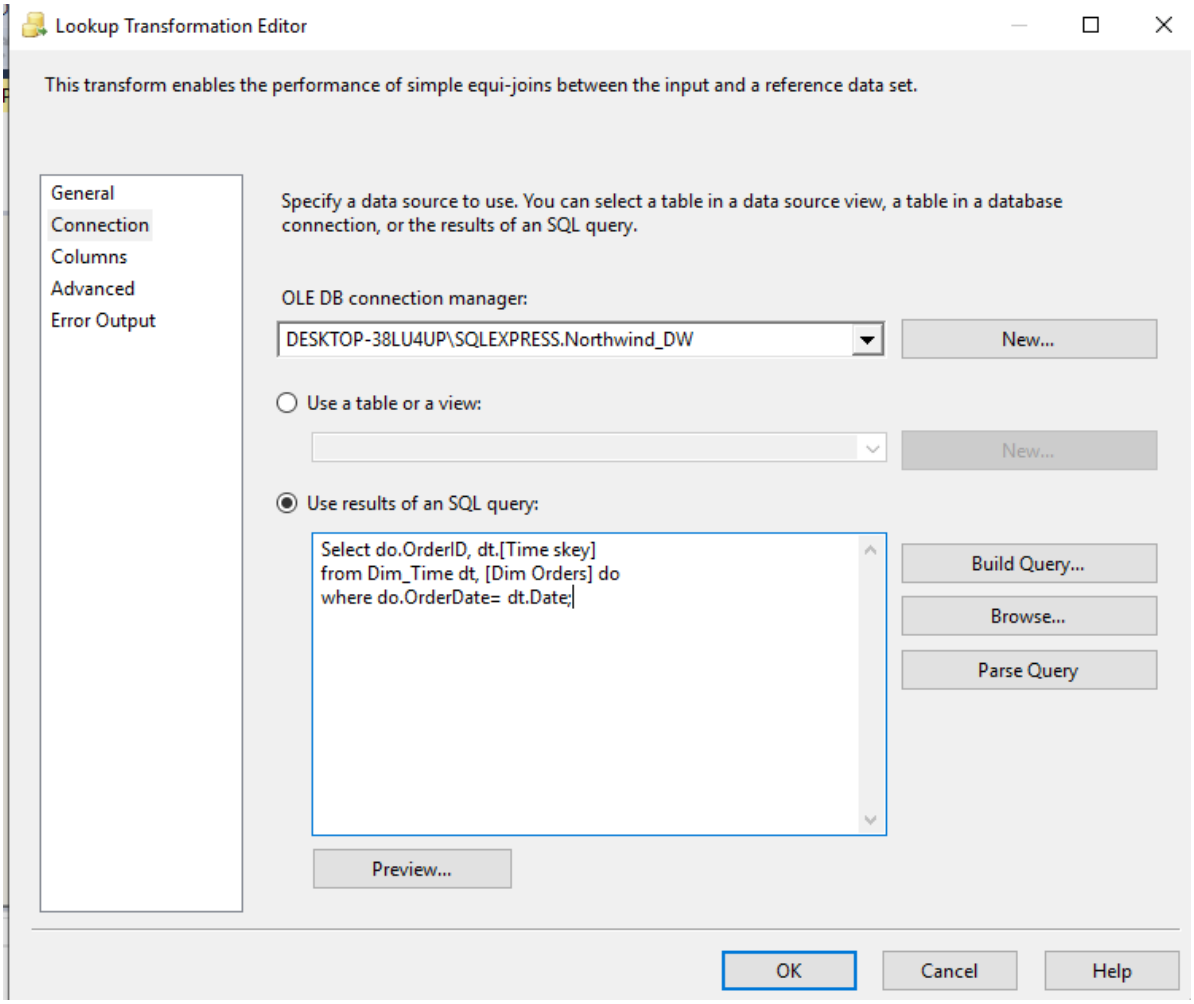


19. Next choose the Lookup task from the left pane under `common` and drag it into the workspace. Join using a blue arrow.

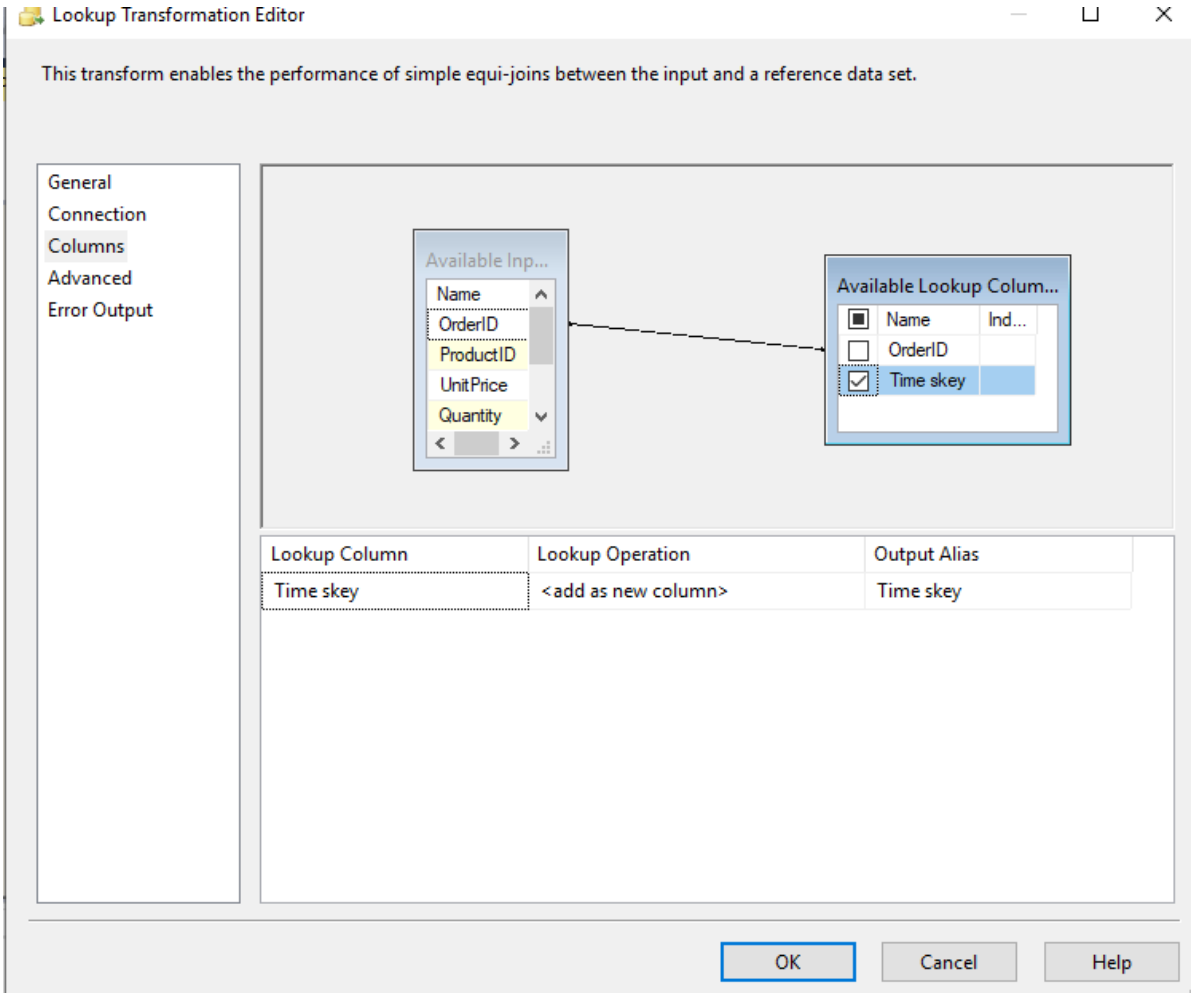


20. Click on the lookup icon. Click `Connection`, Choose the `Northwind_DW` connection manager. Choose `Use results of a SQL query`. And paste the query below.

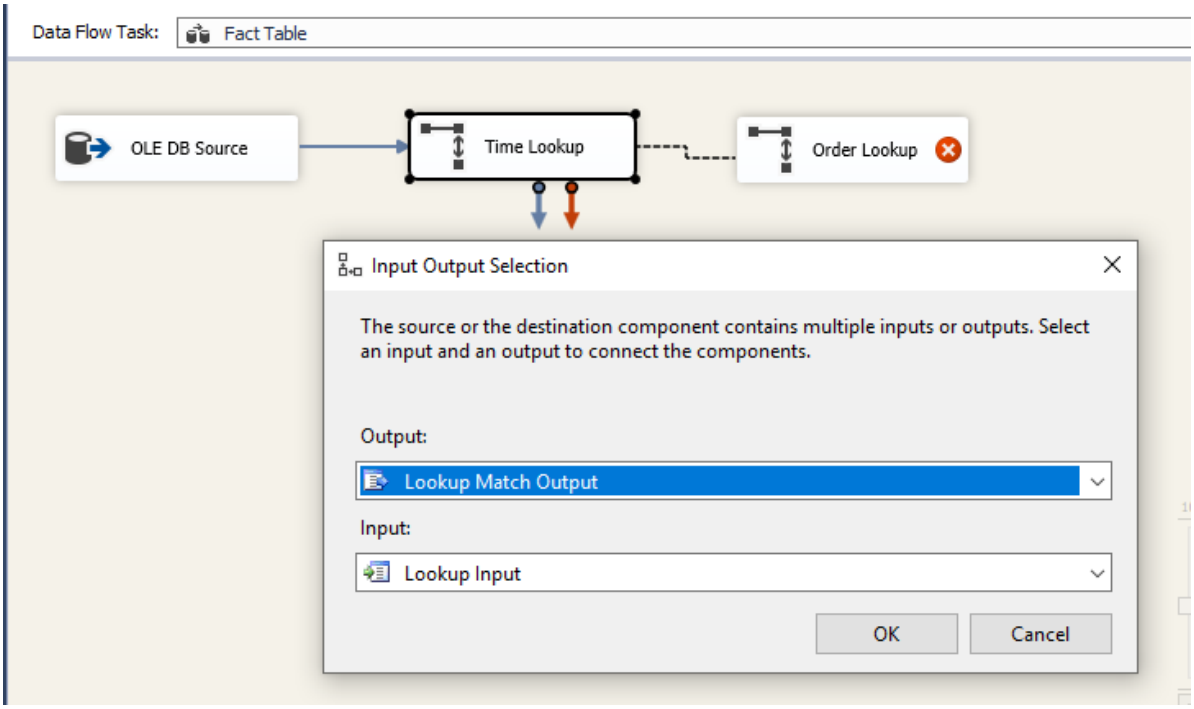
```
Select do.OrderID, dt.[Time skey]
from Dim_Time dt, [Dim Orders] do
where do.OrderDate= dt.Date;
```



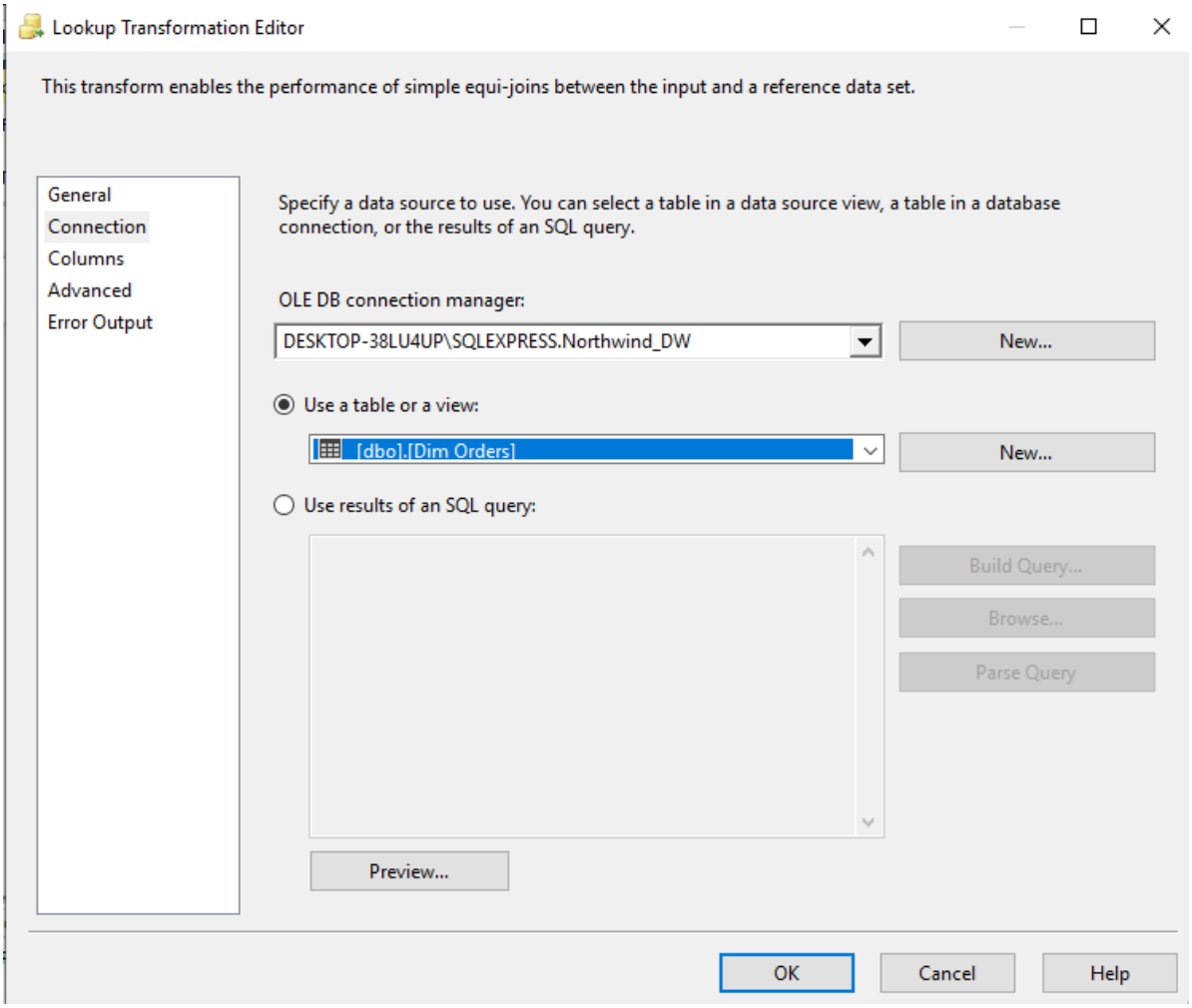
21. In the `columns` section of the editor, link `OrderID`. Choose `Time Skey`. Click OK.



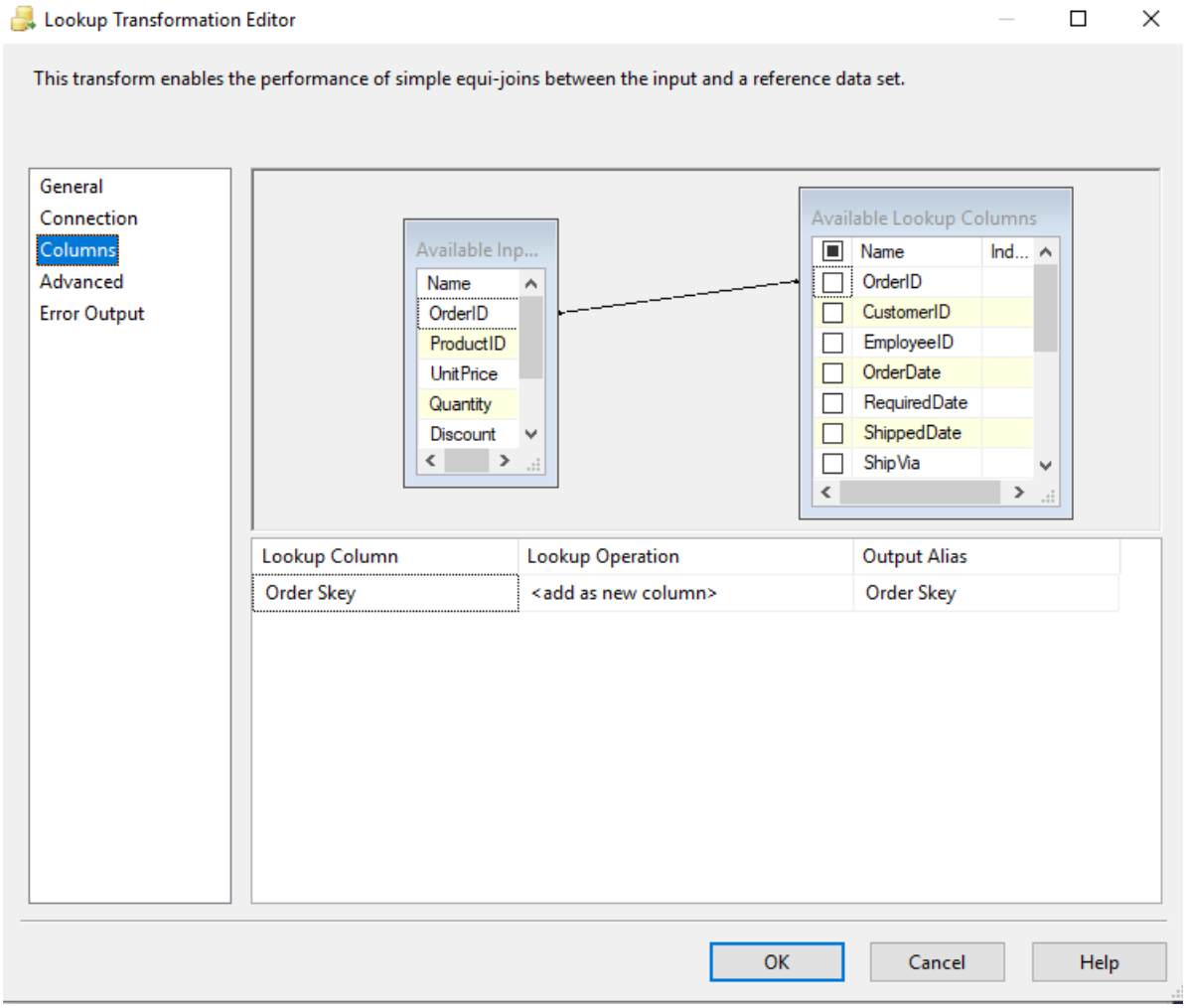
22. Rename the lookup as `Time Lookup`. Join with `OLE DB source` using blue line.
23. Choose another lookup and drag it onto the workspace, rename it as `Order Lookup`. Connect time lookup blue to the order lookup, choose `lookup match output`



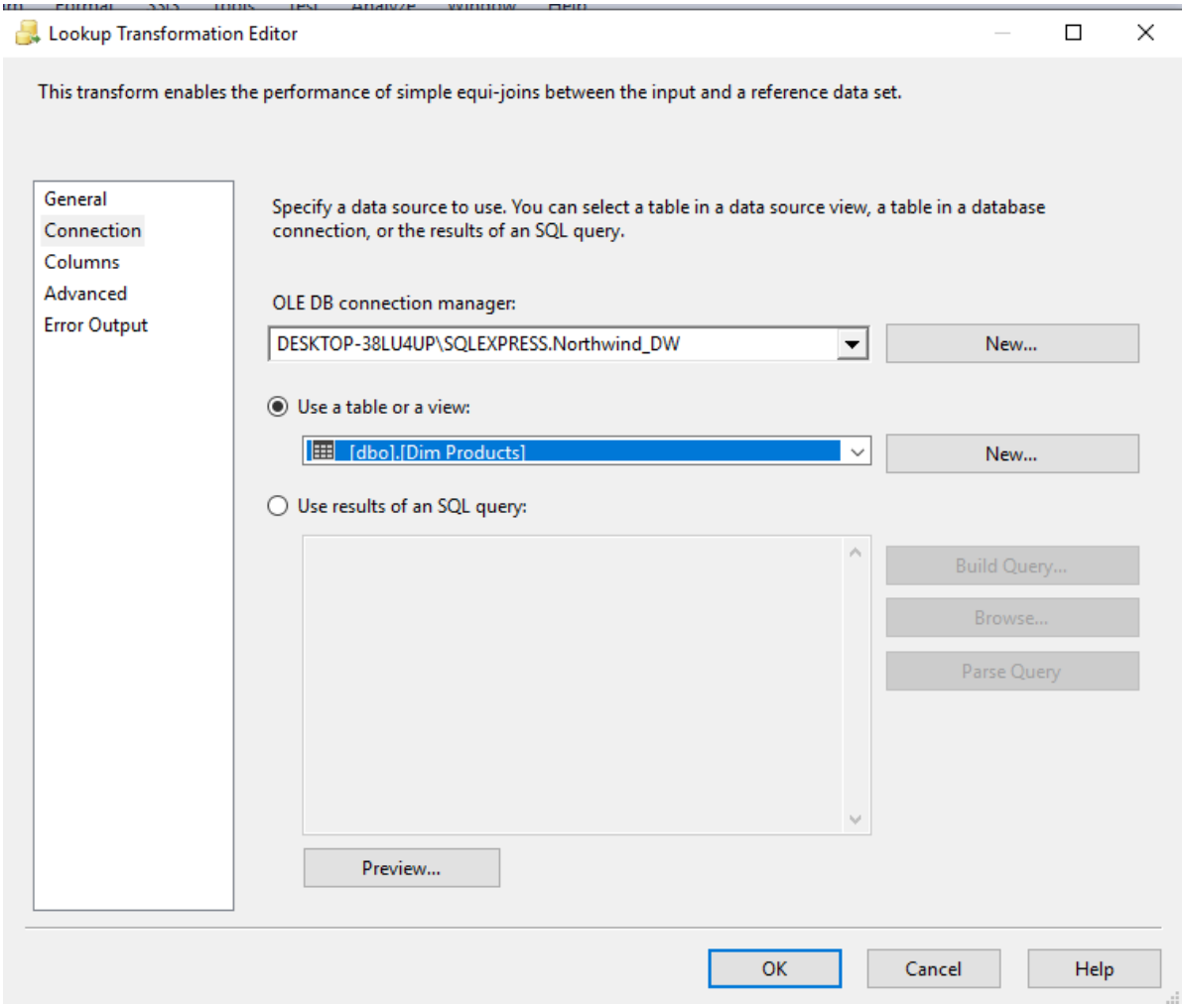
Double click on `Order Lookup` to open the editor. Choose `Northwind_DW` connection manager. Choose `Dim Orders` dimension.

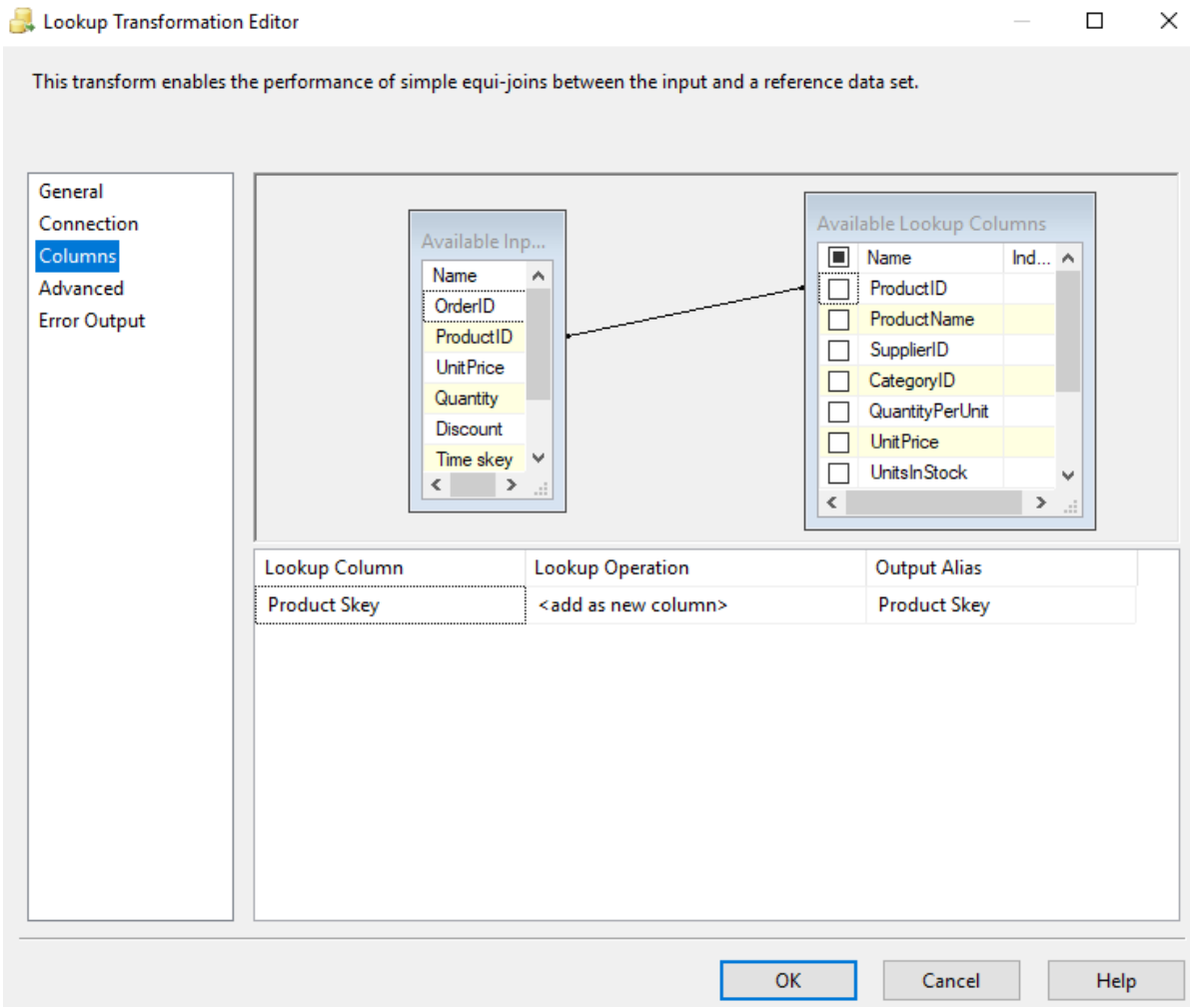


In `Columns` section of editor, link `OrderID`, choose `Order Skey`.



24. Choose another lookup and drag it onto the workspace, rename it as `Product Lookup`. Connect `Order Lookup` to `Product Lookup`. Double click on it to open the editor. Choose `Northwind_DW` connection manager. Choose `Dim Products` dimension.

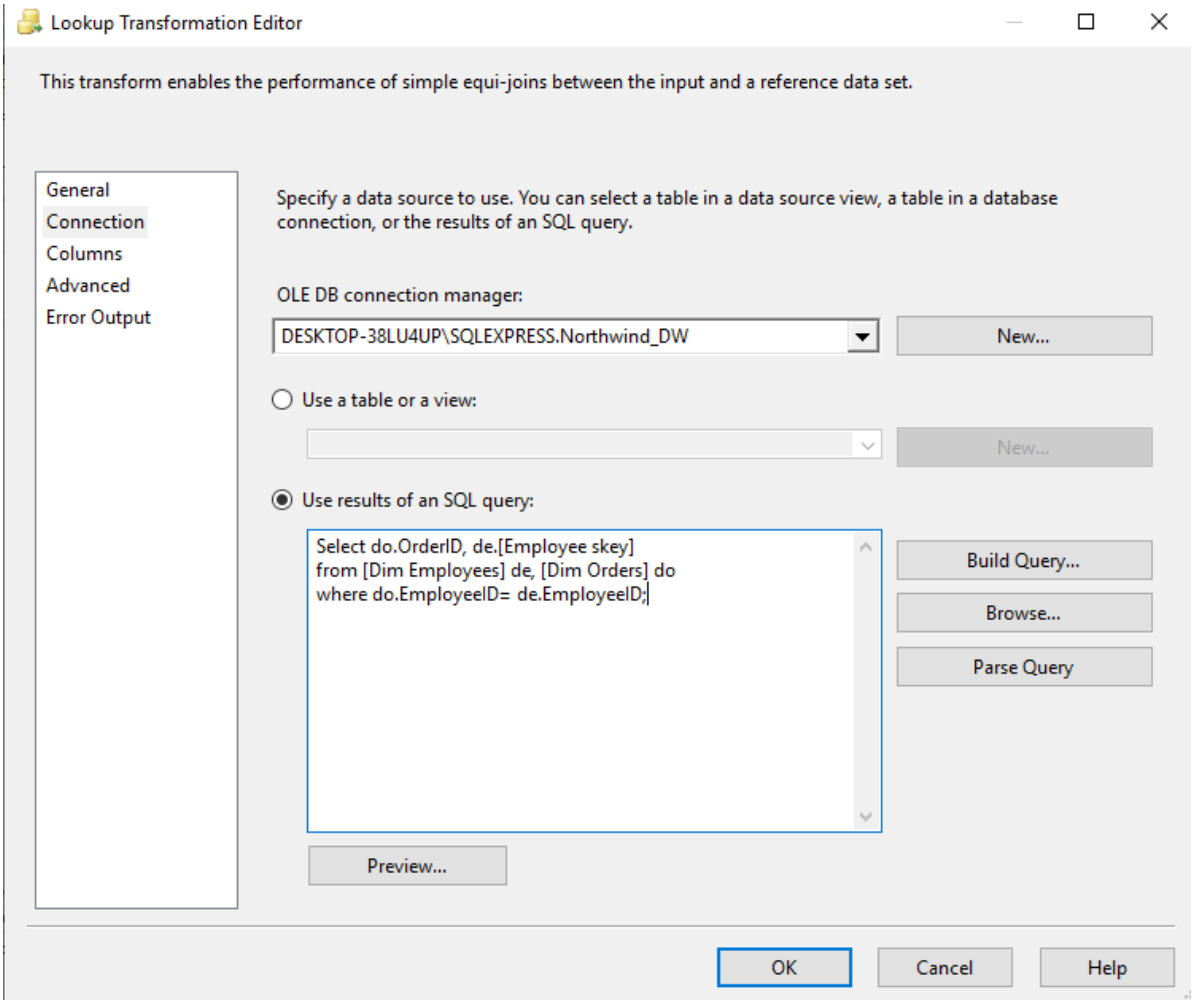

Lookup Transformation Editor
This transform enables the performance of simple equi-joins between the input and a reference data set.
General
Connection
Columns
Advanced
Error Output
Specify a data source to use. You can select a table in a data source view, a table in a database connection, or the results of an SQL query.
OLE DB connection manager:
DESKTOP-38LU4UP\SQLEXPRESS.Northwind_DW
New...
Use a table or a view:
[dbo].[Dim Products]
New...
Use results of an SQL query:
Build Query...
Browse...
Parse Query
Preview...
OK
Cancel
Help

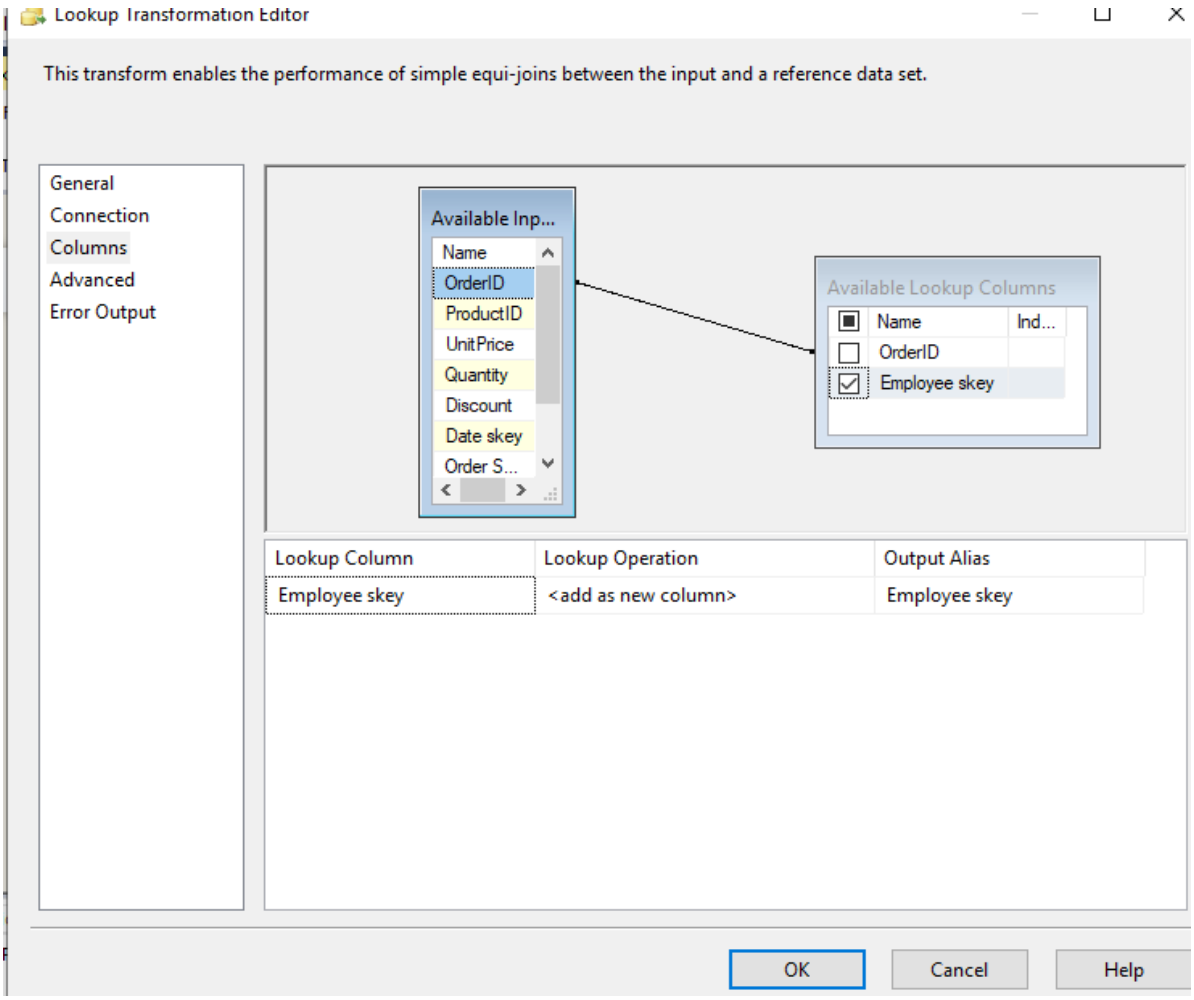

Lookup Transformation Editor
This transform enables the performance of simple equi-joins between the input and a reference data set.
General
Connection
Columns
Advanced
Error Output
Available Inp...
Name
OrderID
ProductID
UnitPrice
Quantity
Discount
Time skey
Available Lookup Columns
Name
Ind...
ProductID
ProductName
SupplierID
CategoryID
QuantityPerUnit
UnitPrice
UnitsInStock
Lookup Column
Lookup Operation
Output Alias
Product Skey
<add as new column>
Product Skey
OK
Cancel
Help

25. Similar to `Time Lookup`, use sql codes below to create lookups for employee, category, supplier, shipper.

`Employee Lookup`

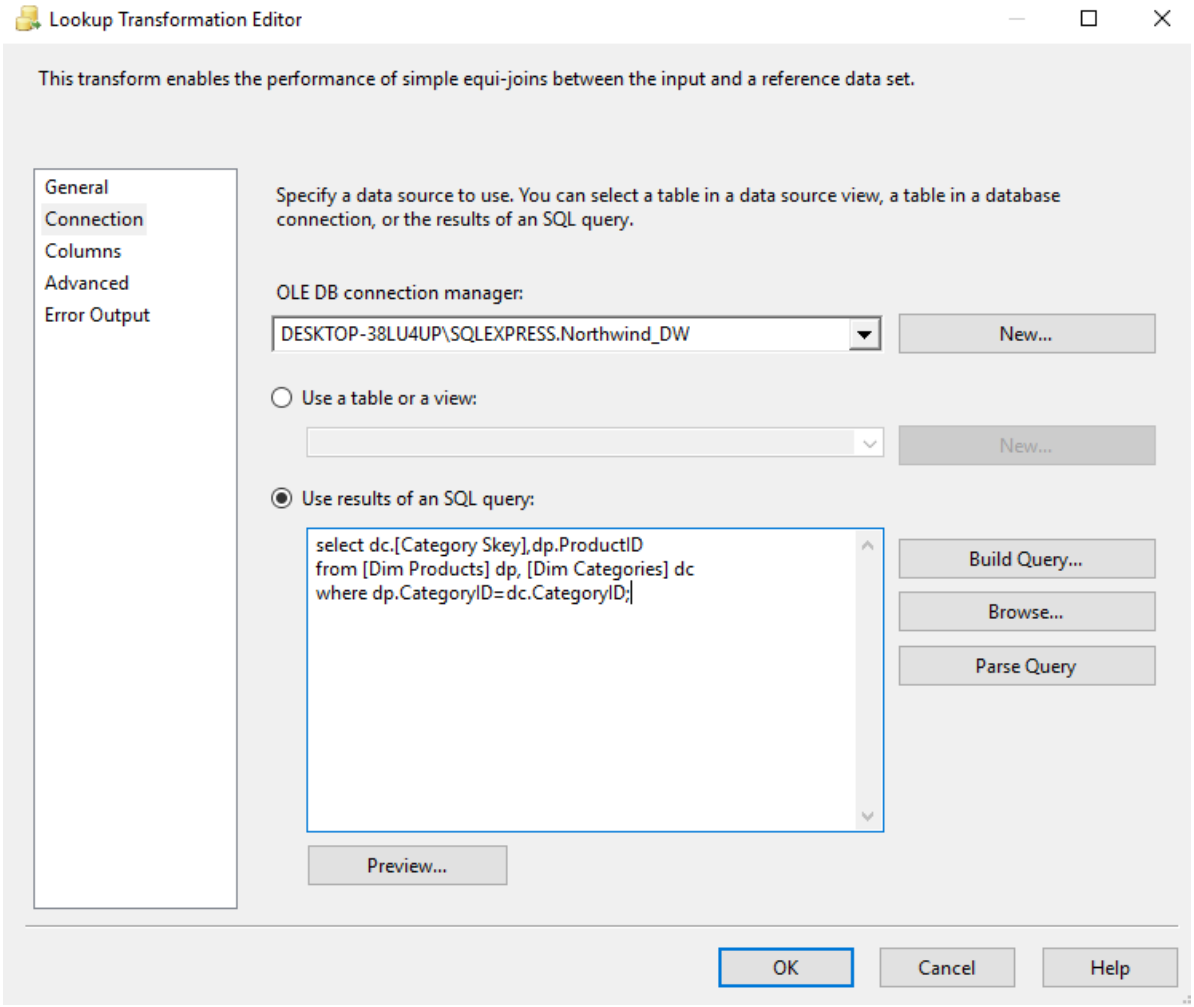
```
Select do.OrderID, de.[Employee skey]
from [Dim Employees] de, [Dim Orders] do
where do.EmployeeID= de.EmployeeID;
```

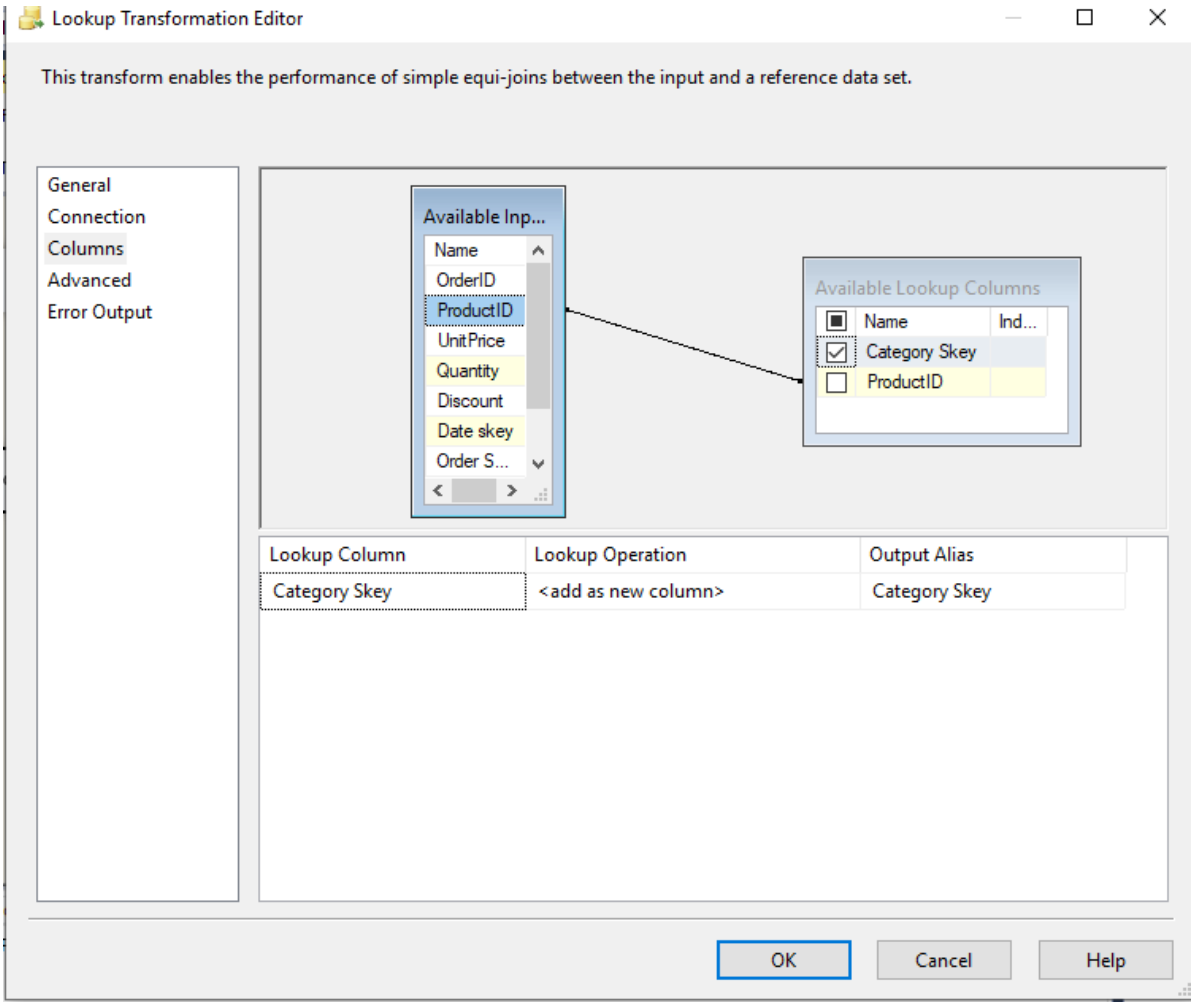




`Category Lookup`

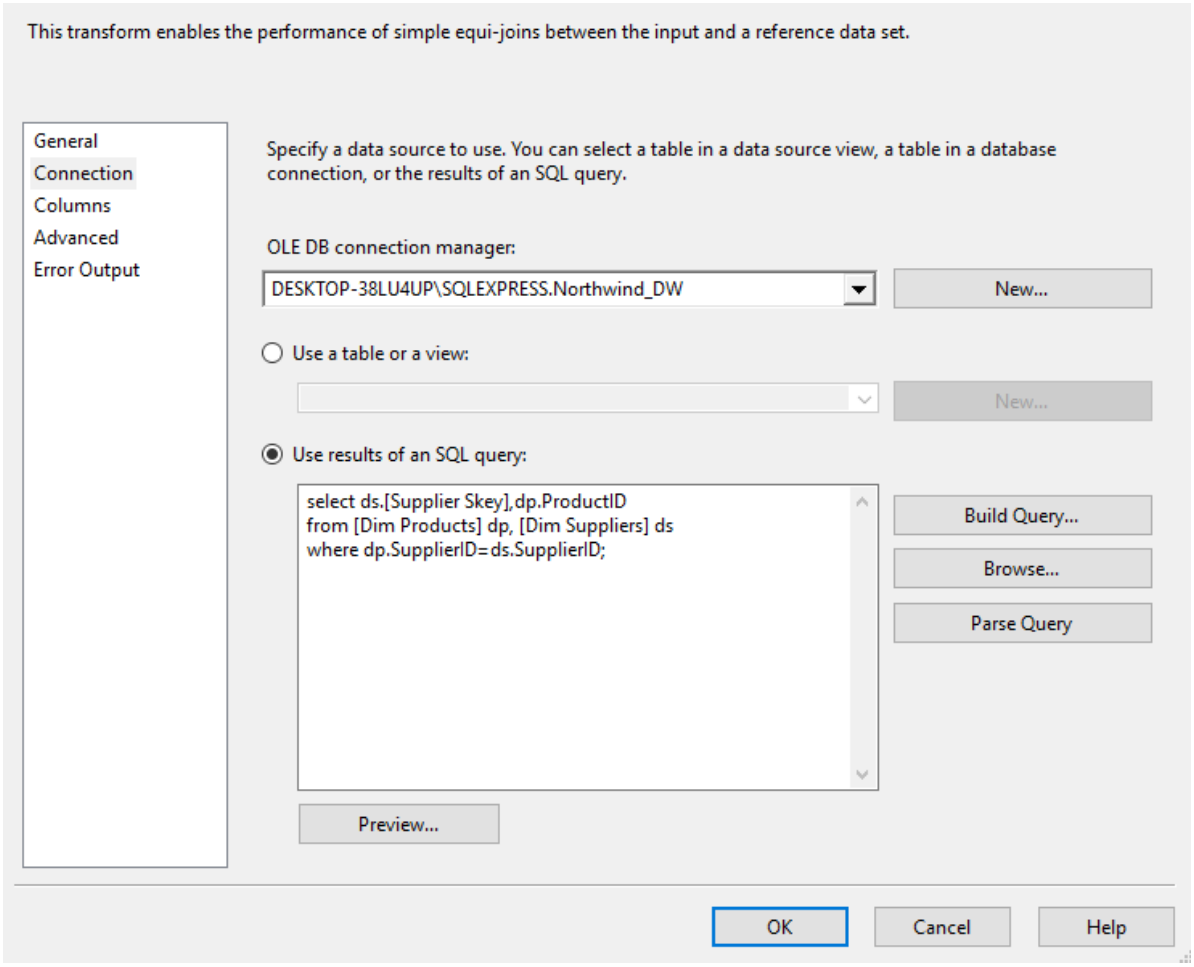
```
select dc.[Category Skey],dp.ProductID
from [Dim Products] dp, [Dim Categories] dc
where dp.CategoryID=dc.CategoryID;
```

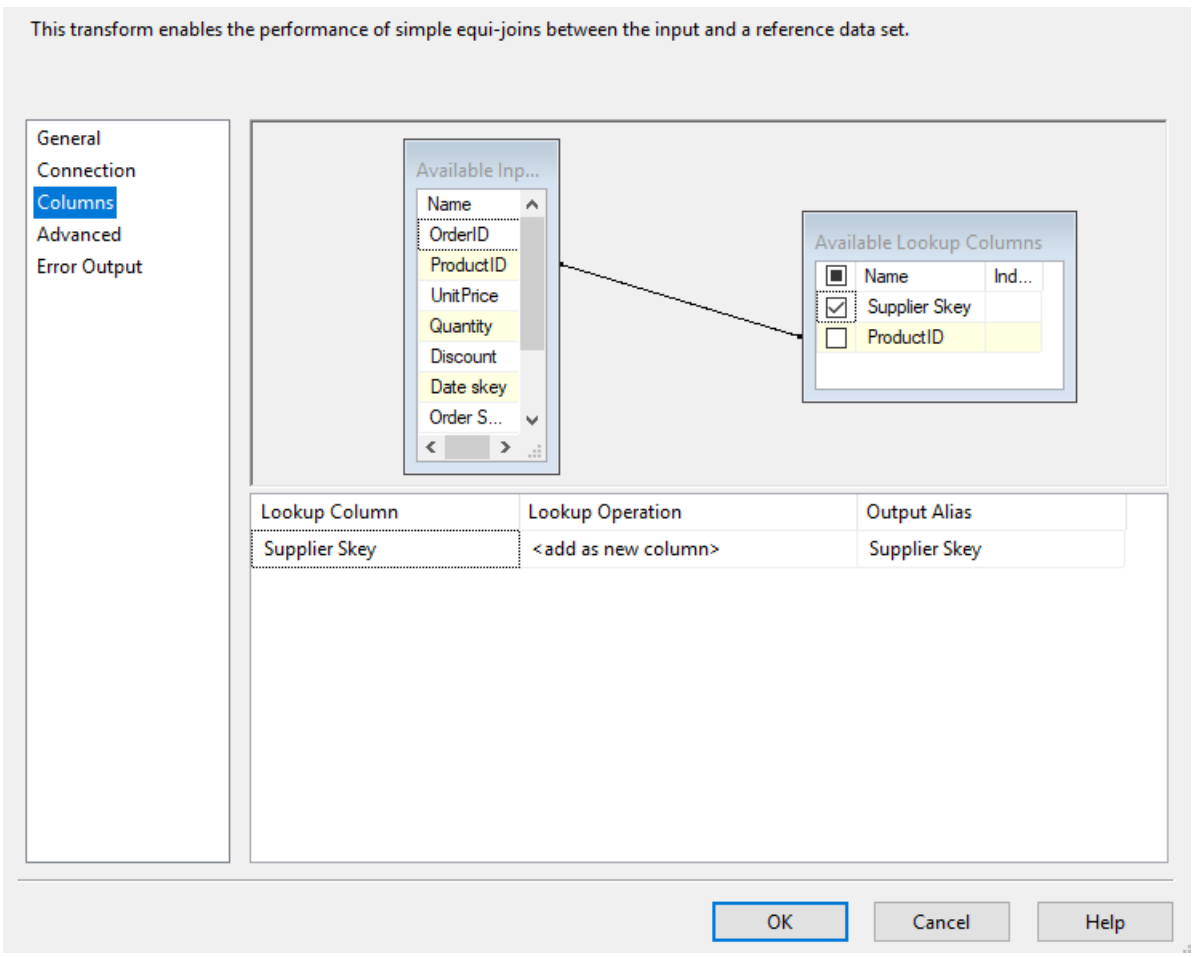




`Supplier Lookup`

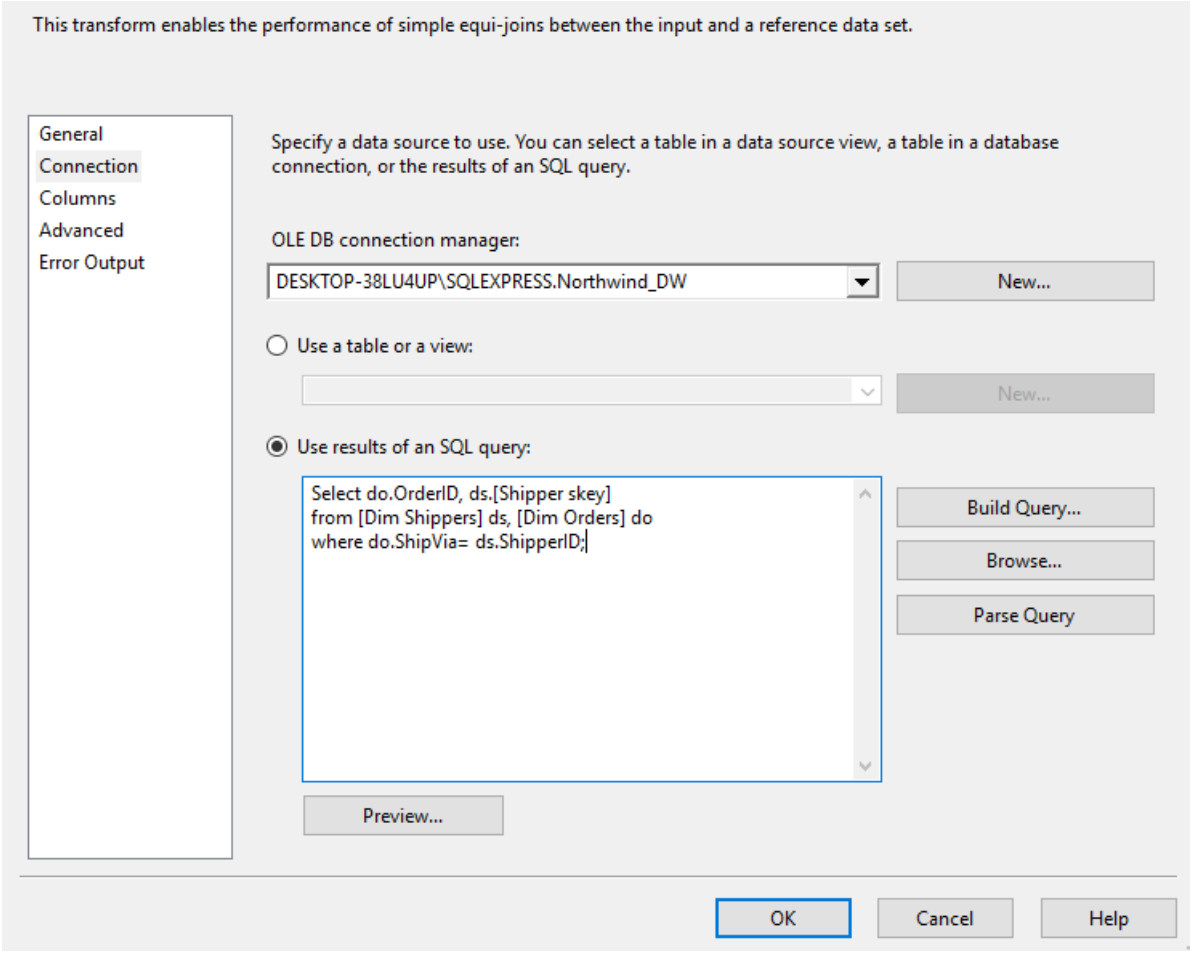
```sql
select ds.[Supplier Skey],dp.ProductID
from [Dim Products] dp, [Dim Suppliers] ds
where dp.SupplierID=ds.SupplierID;
```

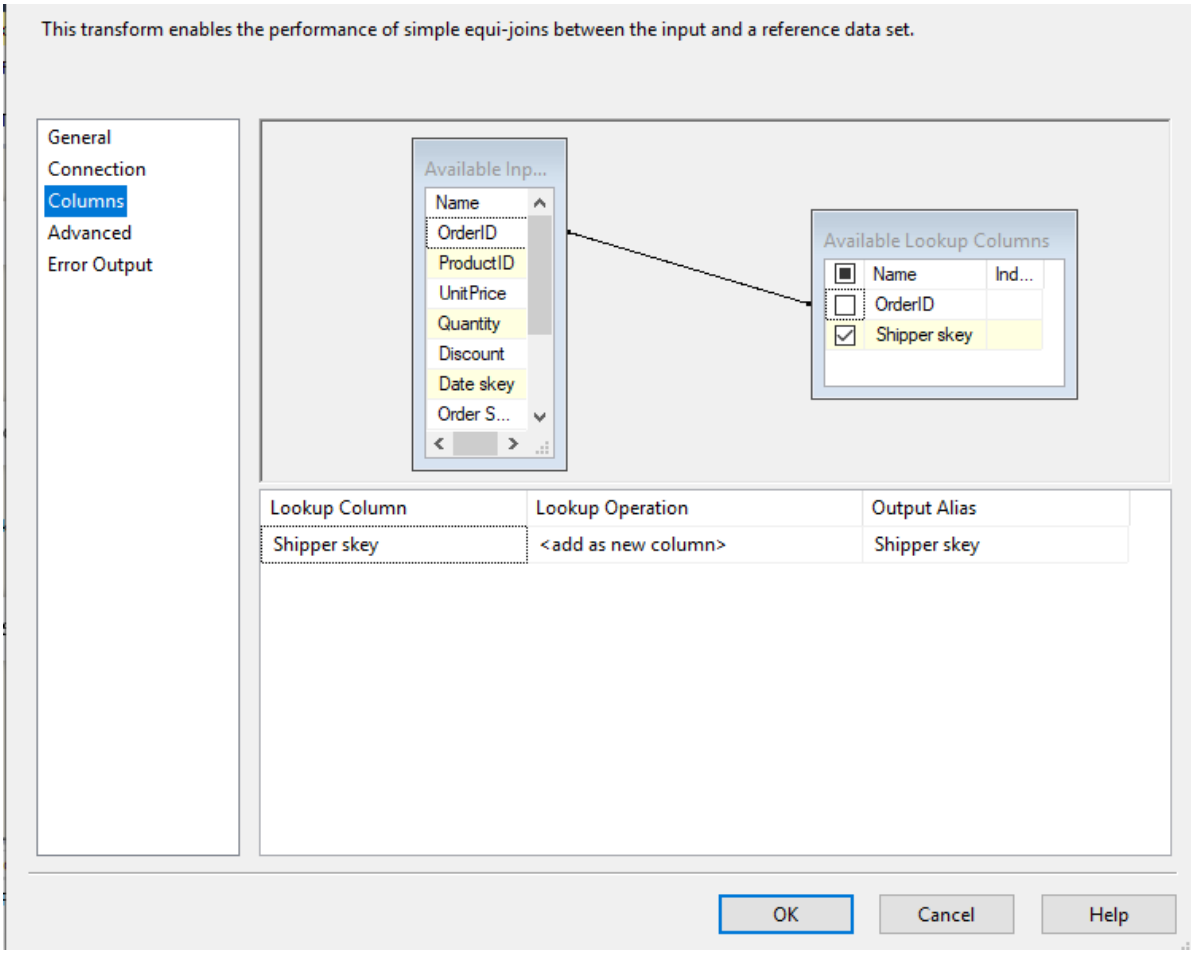




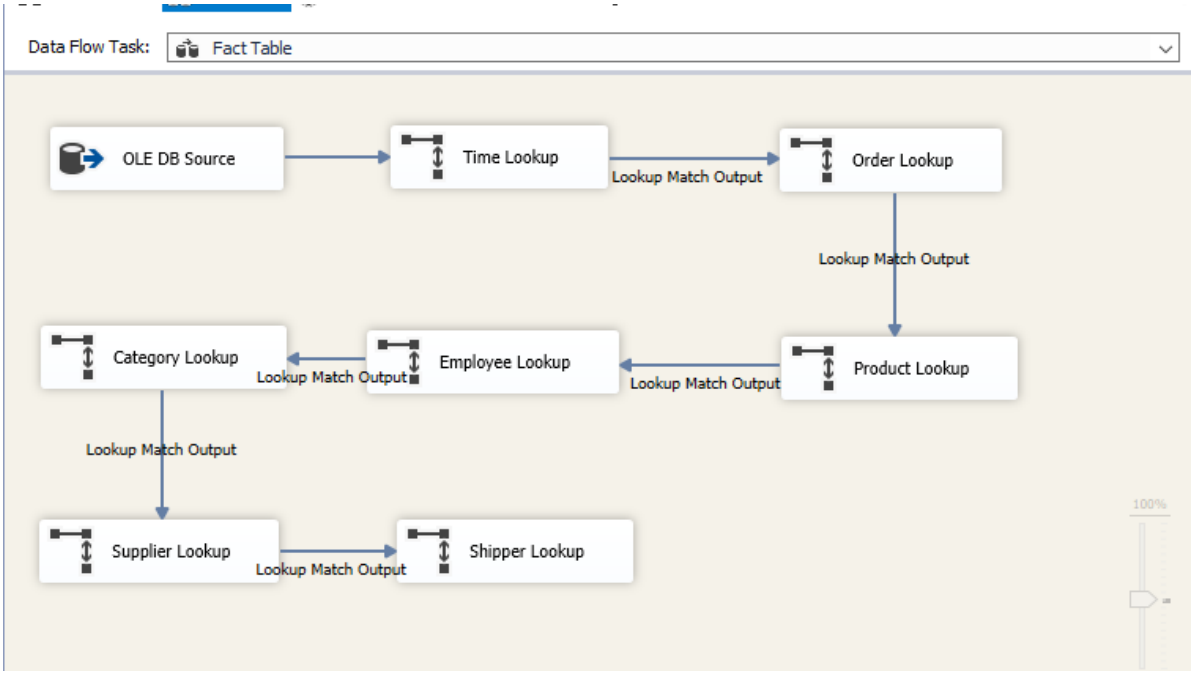
`Shipper Lookup`

```
Select do.OrderID, ds.[Shipper skey]
from [Dim Shippers] ds, [Dim Orders] do
where do.ShipVia= ds.ShipperID;
```

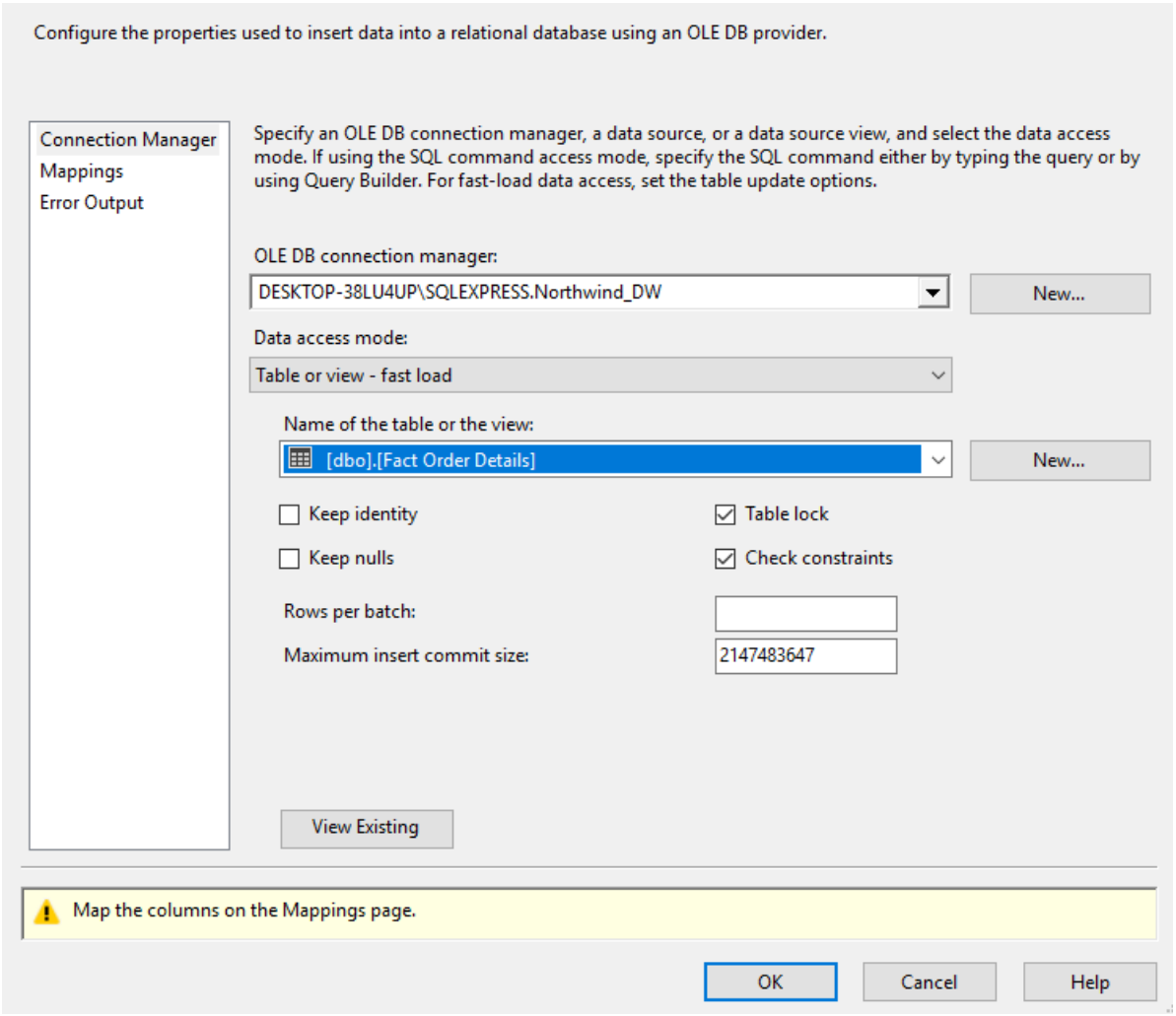




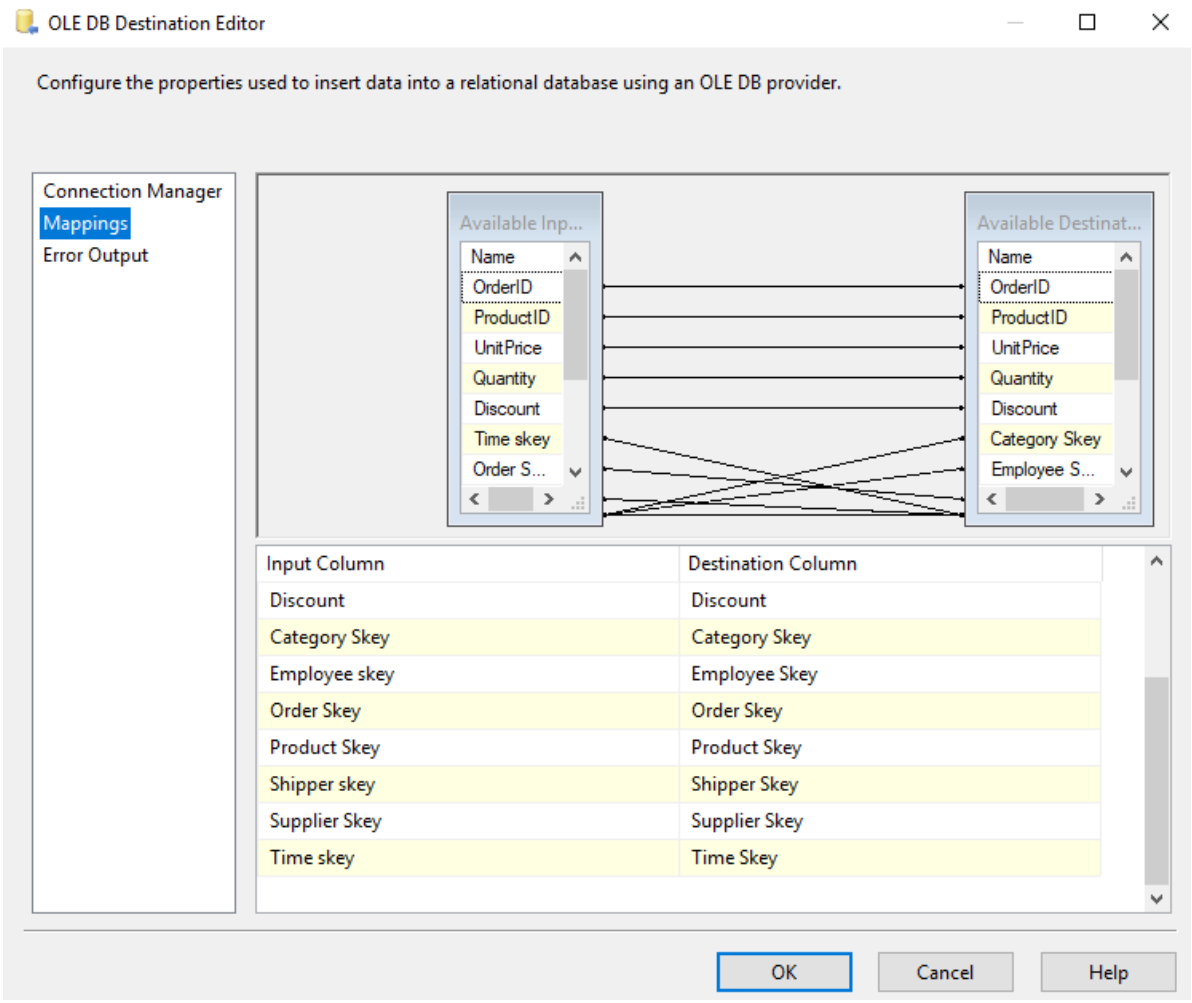
So we have lookups for fact table:

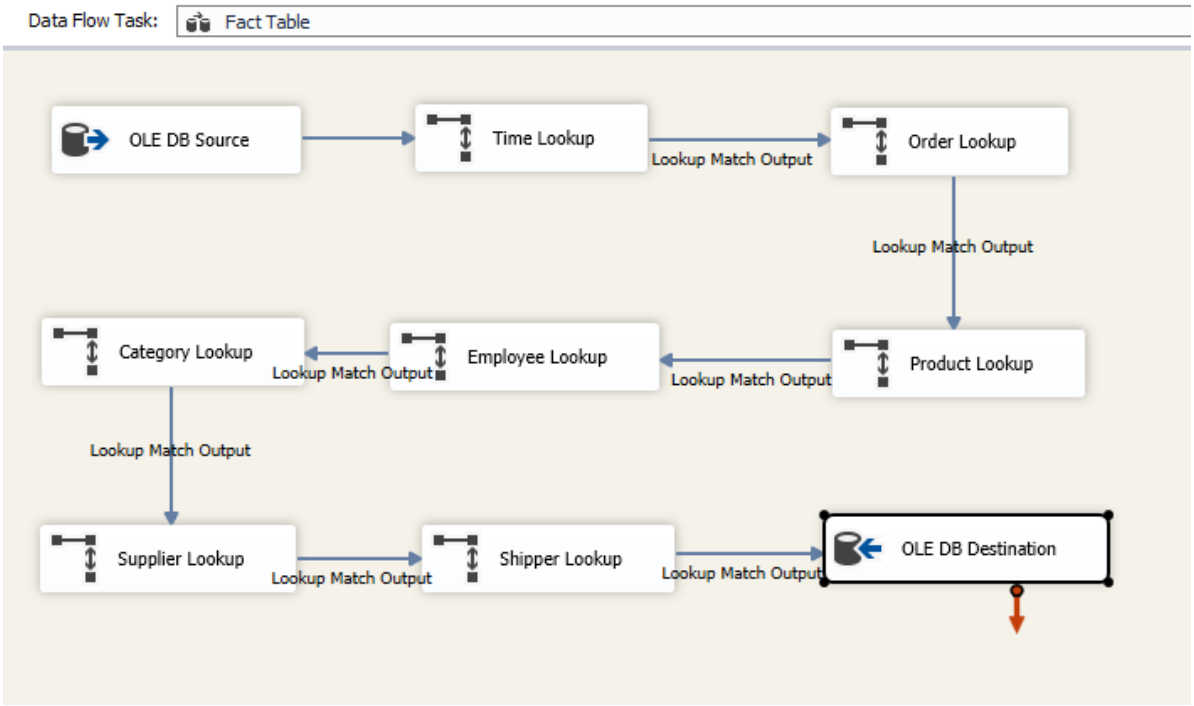


26. Drag `OLE DB destination`, double click, choose `Fact Order Details` from Northwind_DW:



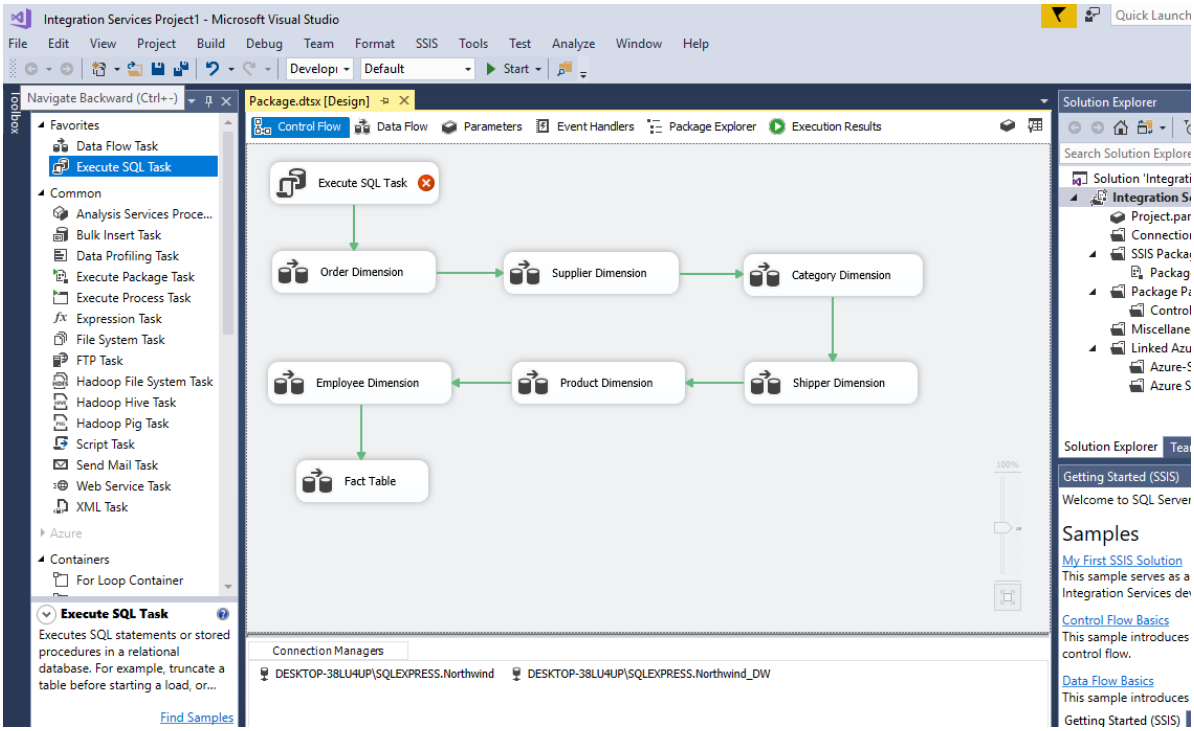
The Foreign keys in the inventory fact table are populated from the dimension tables by matching the primary keys from their respective dimension tables.





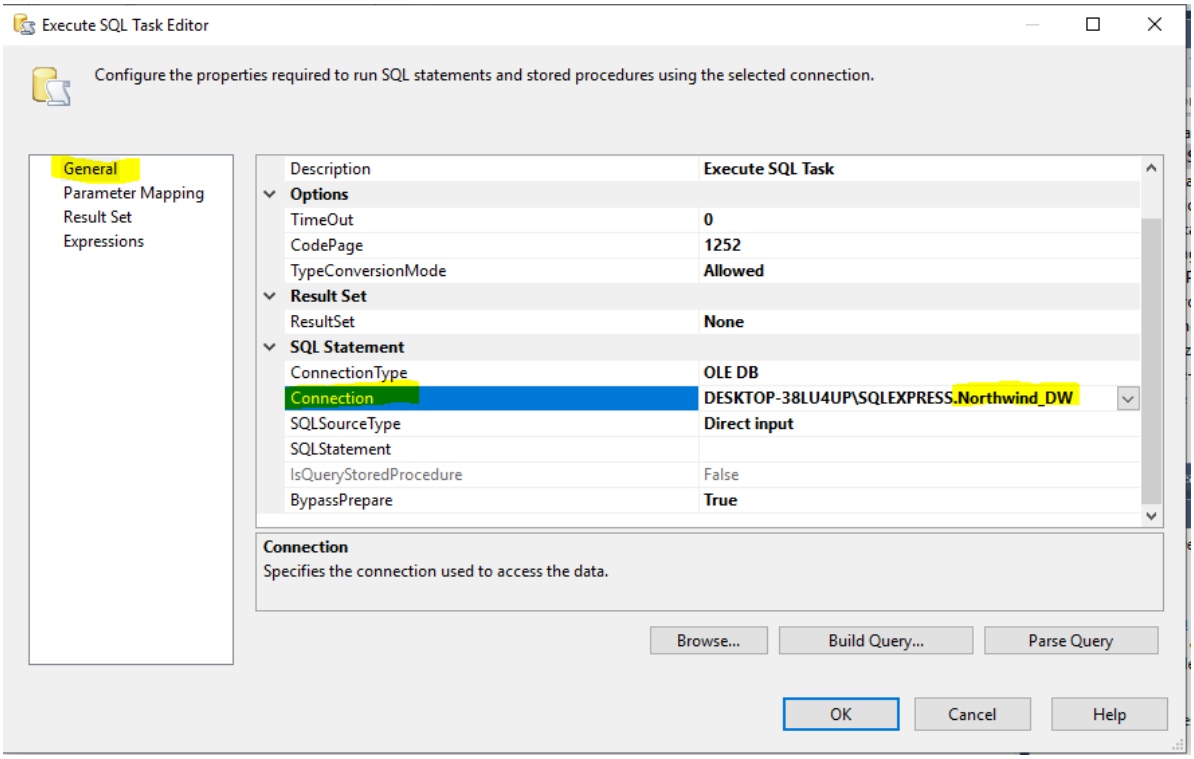
27. Truncate the data warehouse database tables:

Click the `Control Flow` tab. Drag `Execute SQL Task` task onto the design surface of the Data Flow tab. connect `Execute SQL task` to `Order Dimension`



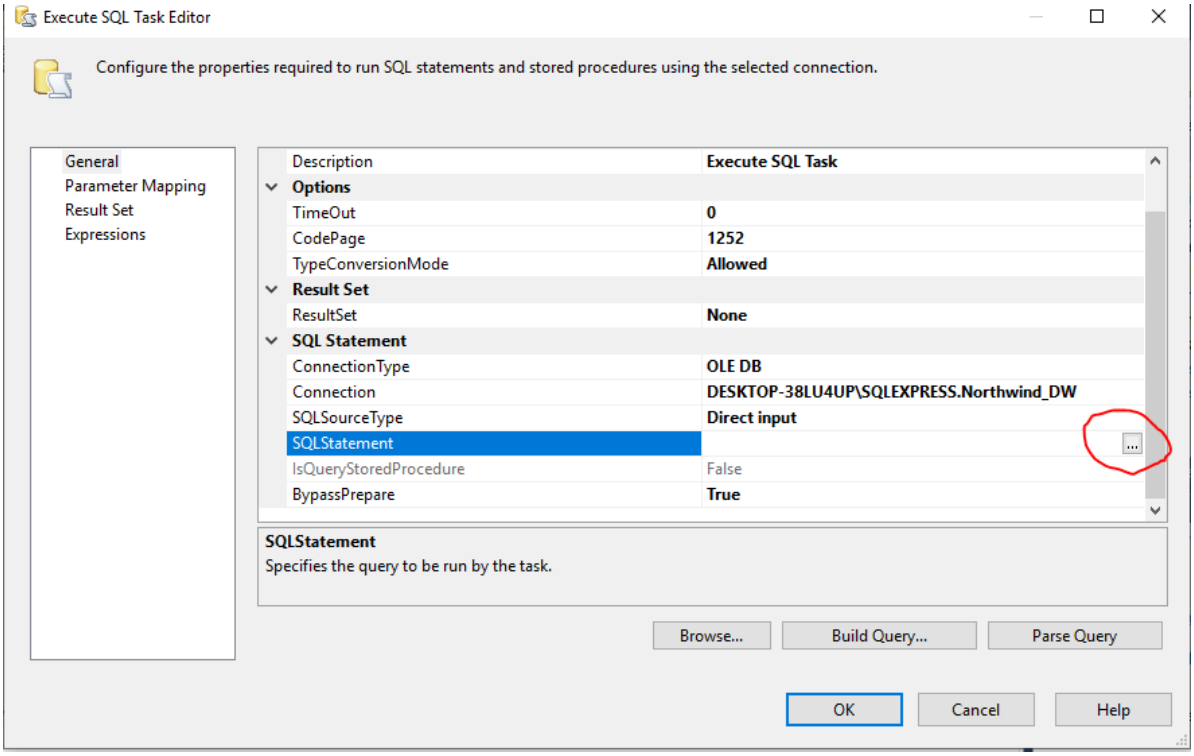
Right-click the task component and when click Edit.

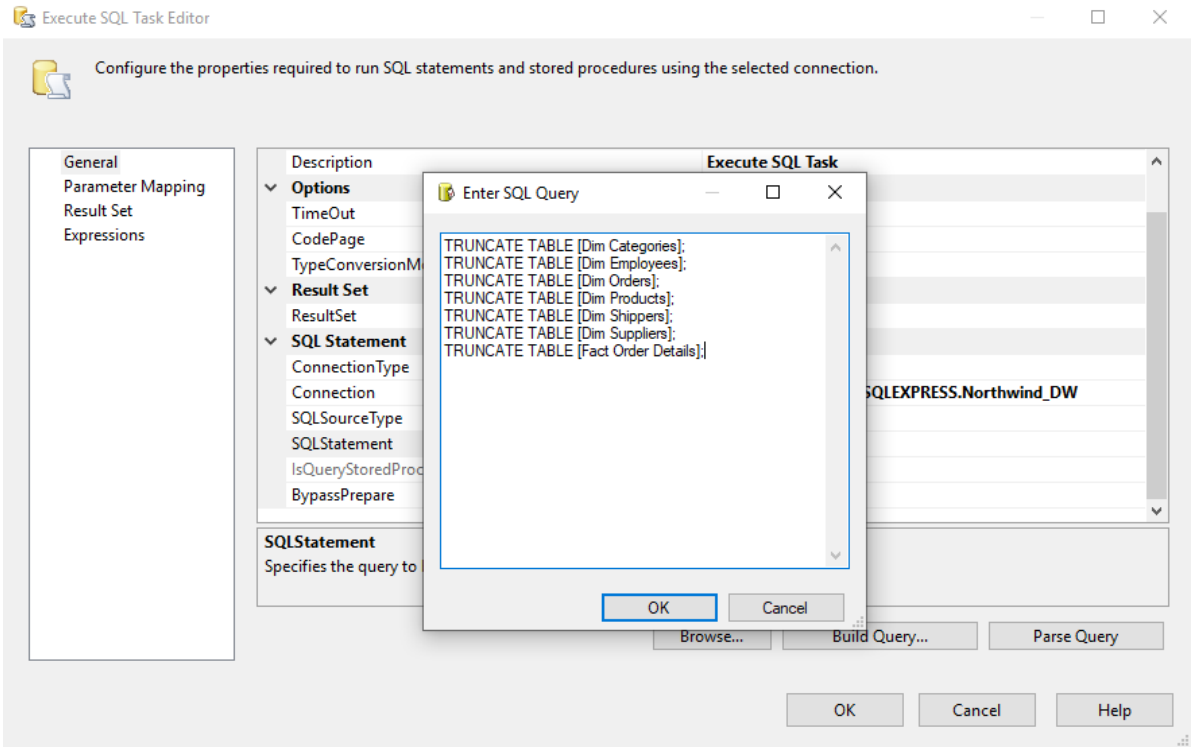
In General tab, Configure your SQL Connection to Northwind_DW



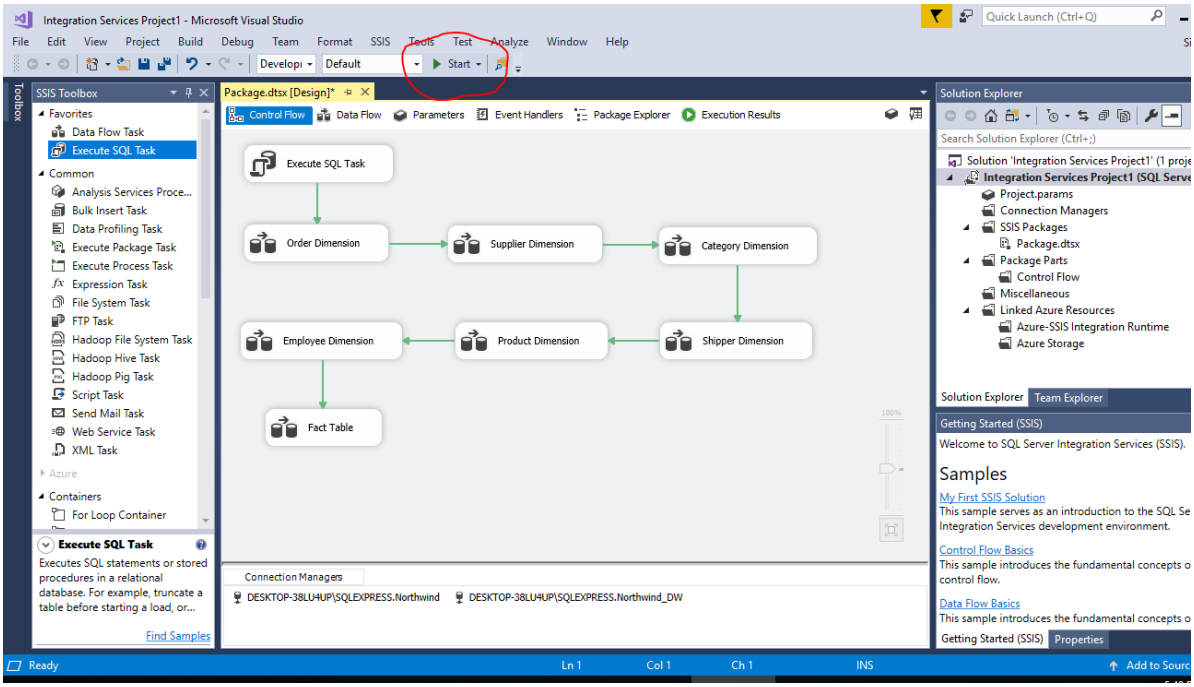
Add the attached SQL query to truncate all tables next to `SQLStatement`. Click OKs.

```
TRUNCATE TABLE [Dim Categories];
TRUNCATE TABLE [Dim Employees];
TRUNCATE TABLE [Dim Orders];
TRUNCATE TABLE [Dim Products];
TRUNCATE TABLE [Dim Shippers];
TRUNCATE TABLE [Dim Suppliers];
TRUNCATE TABLE [Fact Order Details];
```

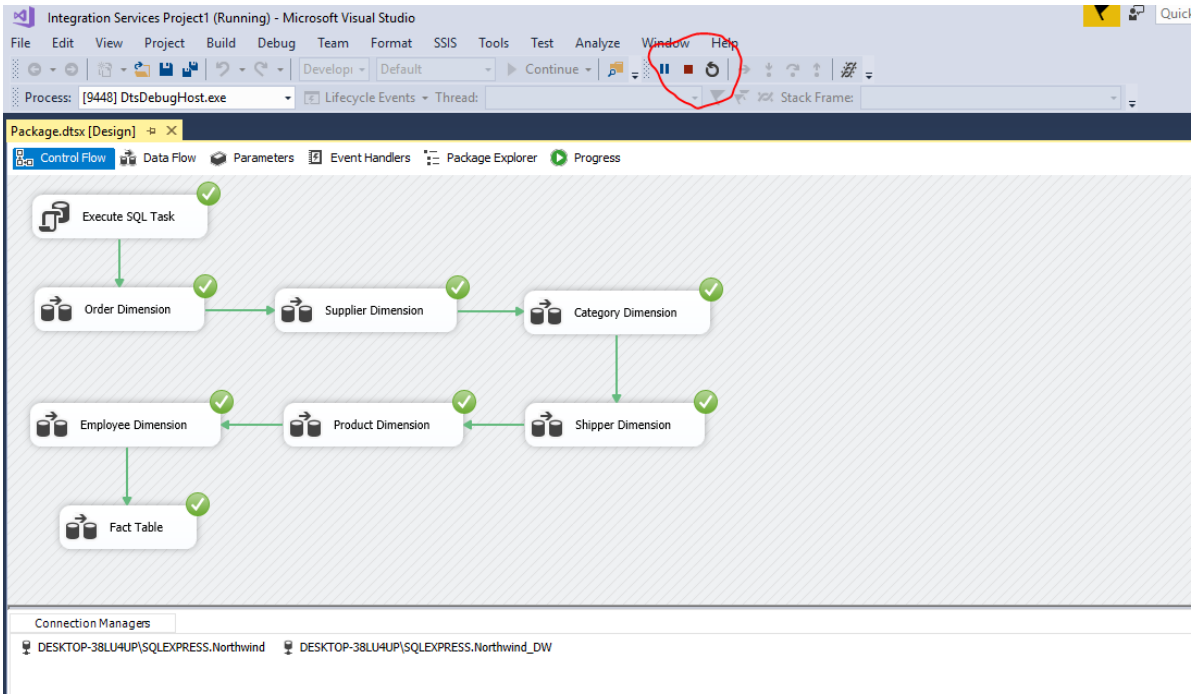




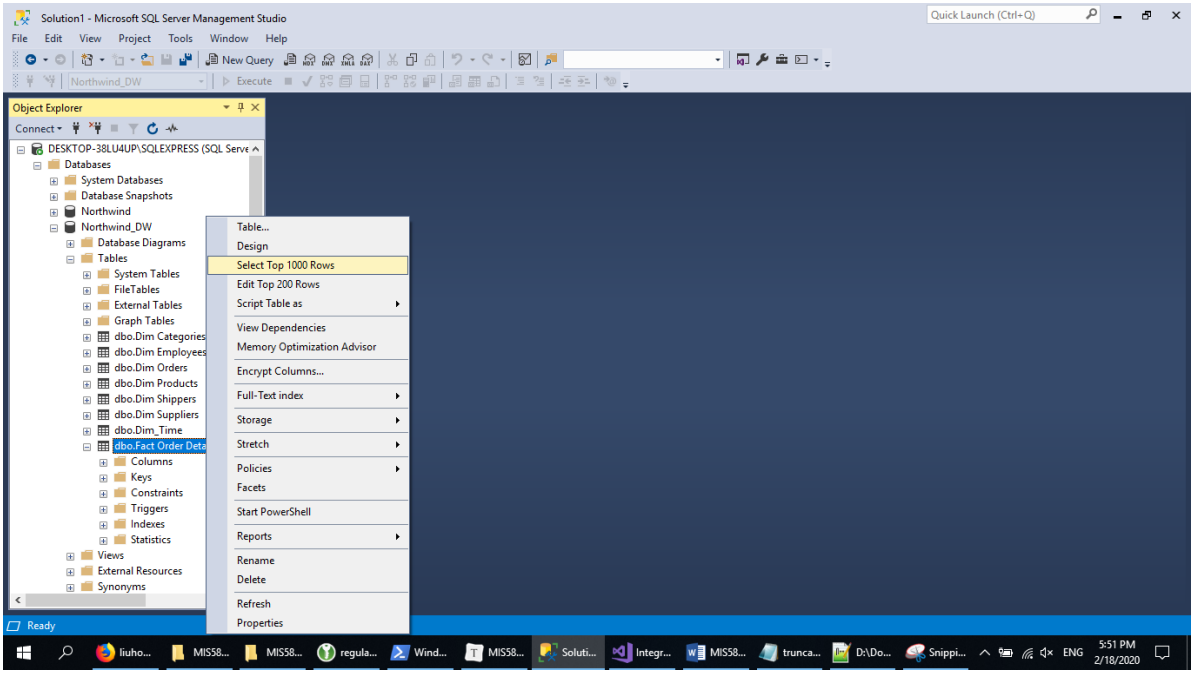
28. Go to control flow and click `start`. You should get a page like below:

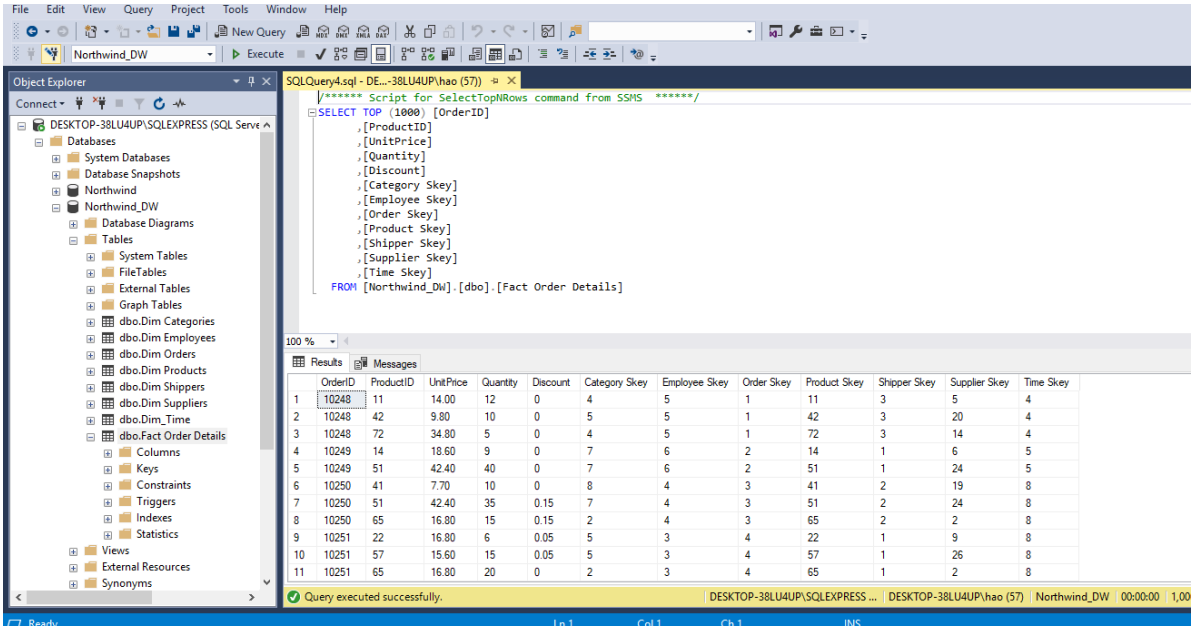


(you can stop it by clicking red square button)



29. Go to Microsoft SQL server. The tables in the Northwind_DW database must have values now.



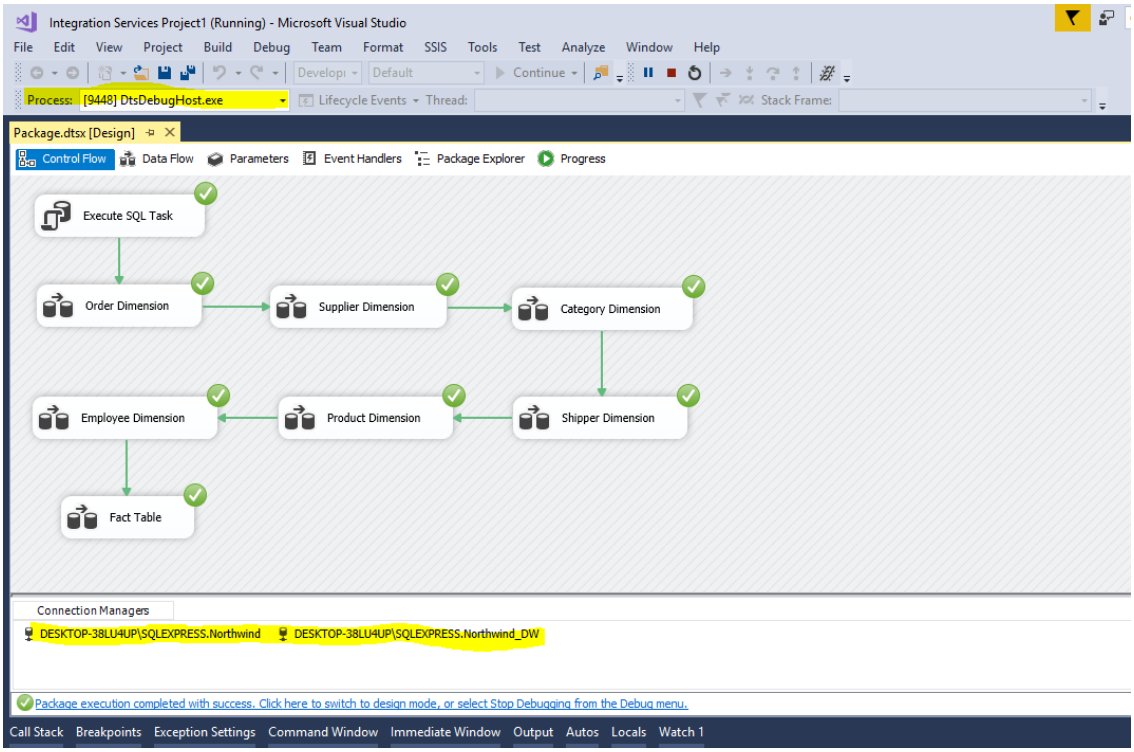


Done!

Screenshots for submission

Note: Put these screenshots in MS word and save it as pdf for submission.

1. Results of successful execution in Visual Studio in Step 28 like the following (areas labeled in yellow must be included in your screenshot):



2. Top rows of `Fact table` in `Northwind_DW` in step 29 (areas labeled in yellow must be included in your screenshot):