Experiment No. :4

Title:Demonstration of simple transformation in ETL using SQL Server Integration Services

Objectives: 1. To study ETL
              2. To Design ETL task to perform Extract of records from Database Table, Transform Lower case characters to upper case and load into the Datawarehouse Table
Key concepts: ETL

Theory:
What is Data Warehouse?
As name implies Data warehouse, It is warehouse for database to store large aggregated data collected from wide range of sources within an organization. Source can be soft files, database files or some excel files.

For example: Baskin-Robbins (Famous for world's largest chain of ice cream specialty shops) has many shops in India as well as across the world. Let's say there is a Baskin-Robbins shop in our area and it has its own system of saving customer visit and product purchase history. So these data must be stored in a excel. Once in a week all these area-data is been collected and stored in a centralized city-data center which is nothing data-warehouse for all small-small areas. Same way all this city-data must be collected and stored in a state-data. A large data store which is accumulated from wide-range of souces is known as Data Warehouse
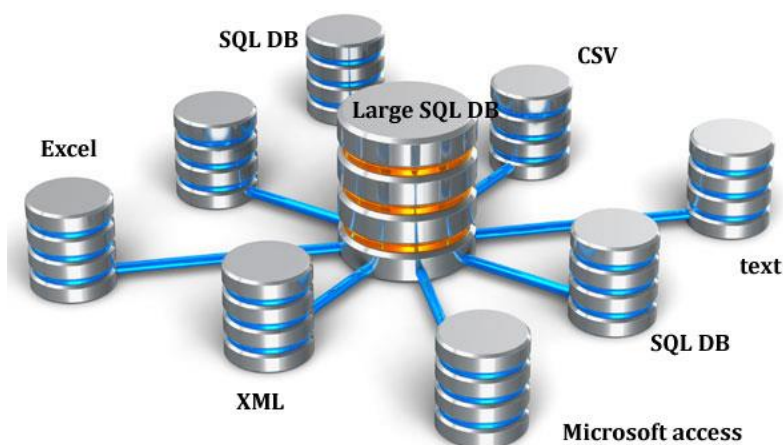
What is an ETL process ?
ETL stands for Extraction, Transformation and Loading. It is a process in data warehousing to extract data, transform data and load data to final source. ETL covers a process of how the data are loaded from the source system to the data warehouse. Let us briefly describe each step of the ETL process.

Extraction

Extraction is the first step of ETL process where data from different sources like txt file, XML file, Excel file or various sources collected.

Transformation

Transformation is the second step of ETL process where all collected data is been transformed into same format i.e. format can be anything as per our requirement before loading it to data-warehouse i.e. it may be data-type format, data merge format, splitting format, alphabet joining format, currency format etc.

Loading

Final step of ETL process, The big chunck of data which is collected from various sources and transformed then finally load to our data warehouse.
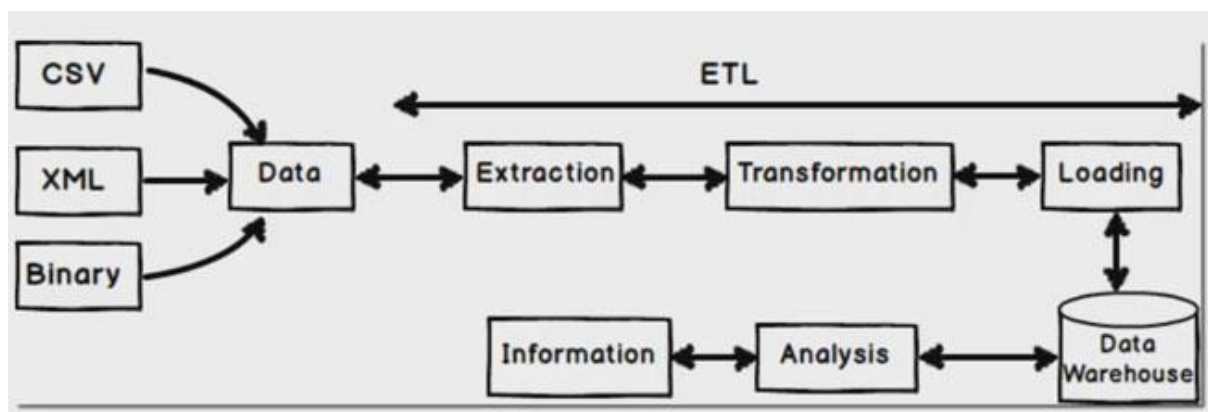
ETL process with SSIS Step by Step

We do this example by keeping baskinrobbins (India) company in mind i.e. customer data which is maintained by small small outlet in an excel file and finally sending that excel file to USA (main branch) as total sales per month. This data is necessary at head quaters (main branch) to track performance of each outlet.

So here also we will do same thing i.e. We will collect customer product purchase sales data from small-small outlet (In an Excel Format) - Extraction

Since baskinrobbins is located in USA we need to convert or transform product purchase amount to USD currency and we will also convert product name to uppercase for unique representation - Transformation

Finally loading this transformed data to database / datawarehouse (SQL Server Database) - Loading

Step 1:- Extract Data from Excel File

Create a simple excel file with a columns names as CustomerCode, CustomerName, ProductPurchase, Quantity, Amount, CustomerVisitedDate respectively

| CustomerCode	CustomerName	ProductPurchase	Quantity	Amount	CustomerVisitedDate
101	Shiv	Café ice cream	2	200	1/6/2016
102	Shaam	caramel icecream	2	250	2/6/2016
103	Allwyn	violette icecream	8	1523	1/6/2016
104	Khadak	cassis icecream	3	750	1/6/2016
105	Robin	chocolate icecrear	5	1000	2/6/2016
106	Ajay	pista icecream	4	800	1/6/2016
107	Sukesh	mango icecream	1	150	1/6/2016
108	Praful	Café ice cream	9	2250	2/6/2016

Create SSIS project in MS SQL server 2012

Now go to SSIS Toolbox and drag and drop "Data Flow Task" to control panel as show in below image.

Just double click on "Data Flow Task" to take you to "Data Flow".

Now go to SSIS Toolbox and from Other Souces tab just drap and drop Excel Souces. Why excel source because our inital data which we want to extract it is in excel format.

So just drag excel file as shown below image and right click and rename it so that if any developer reads it can easily able to understand.

Now right click on that excel source -> Edit -> Click on New button - Browse Excel file from your computer
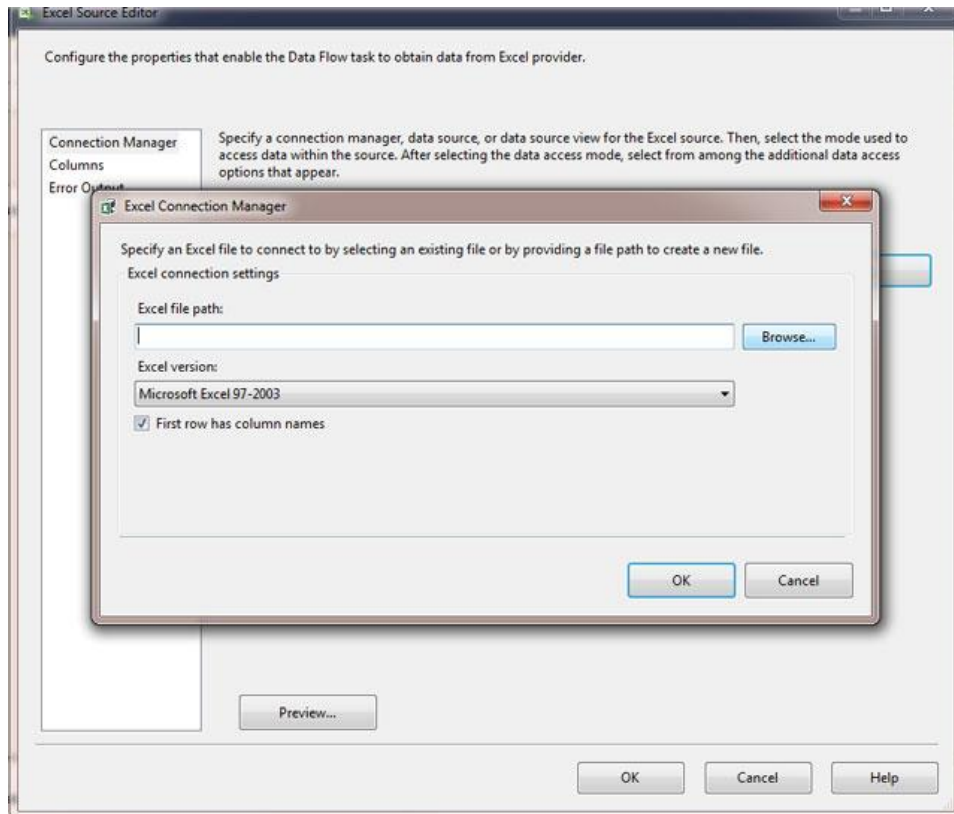
Since our first column of excel file is having column names so we need to check this below check box as you see in above image.

Now select Data access mode as "Table on view" then select Excel sheet name from drop down.
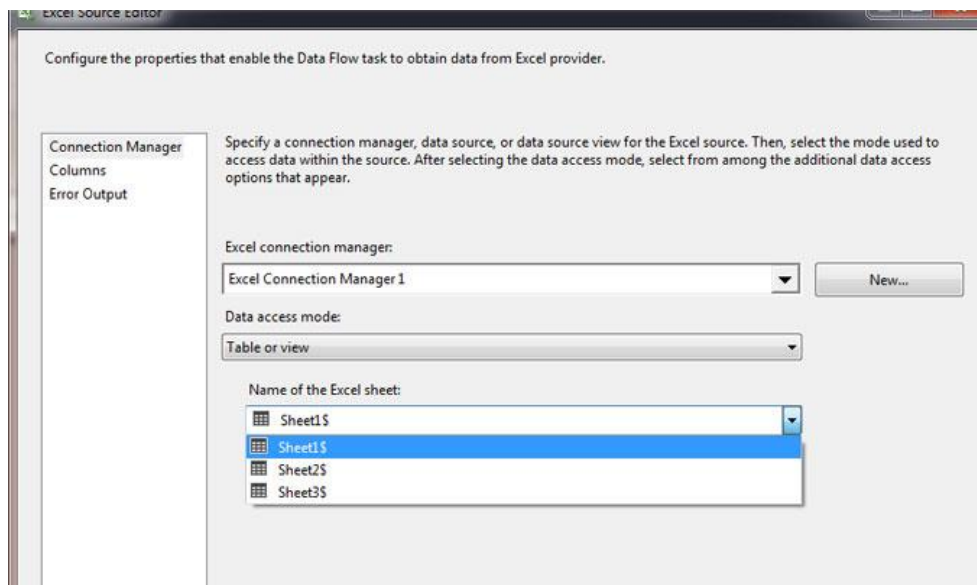
Step 2: Transform Data (Convert to US currency and Upper case)
In our excel file we have column name called "Amount" and that amount is in Indian currency. So we need to convert that Indian amount to USD amount so to do that we will drag and drop "Derived Column" from SSIS toolbox.

Now if you see on Excel Source file box there are two arrows "Red" and Blue". Just drag that "Blue" arrow and join it to "Derived column" as shown in below image and rename that "Derived column".
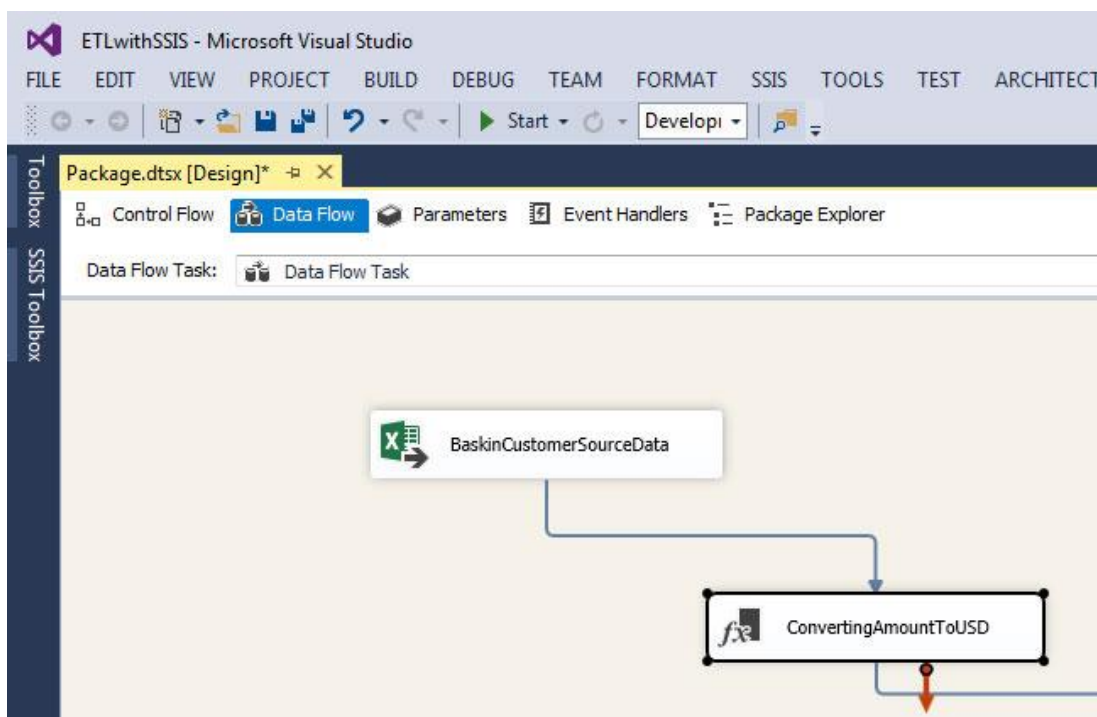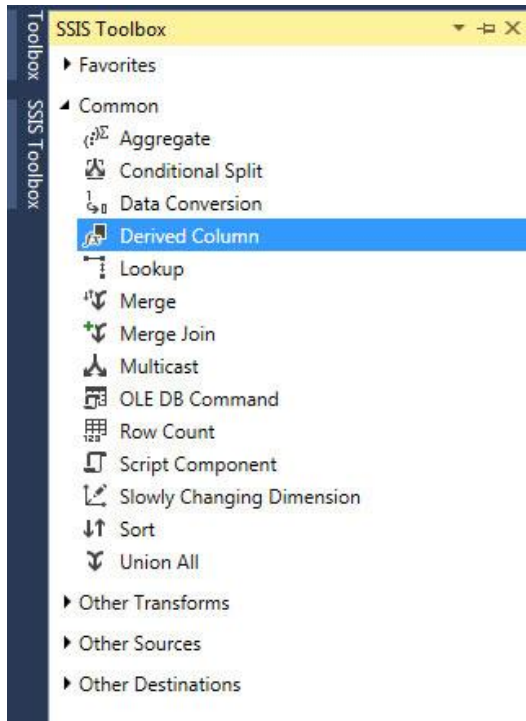
right click on "Derived column" i.eConvertingAmounttoUSD and click on Edit as
show below image.
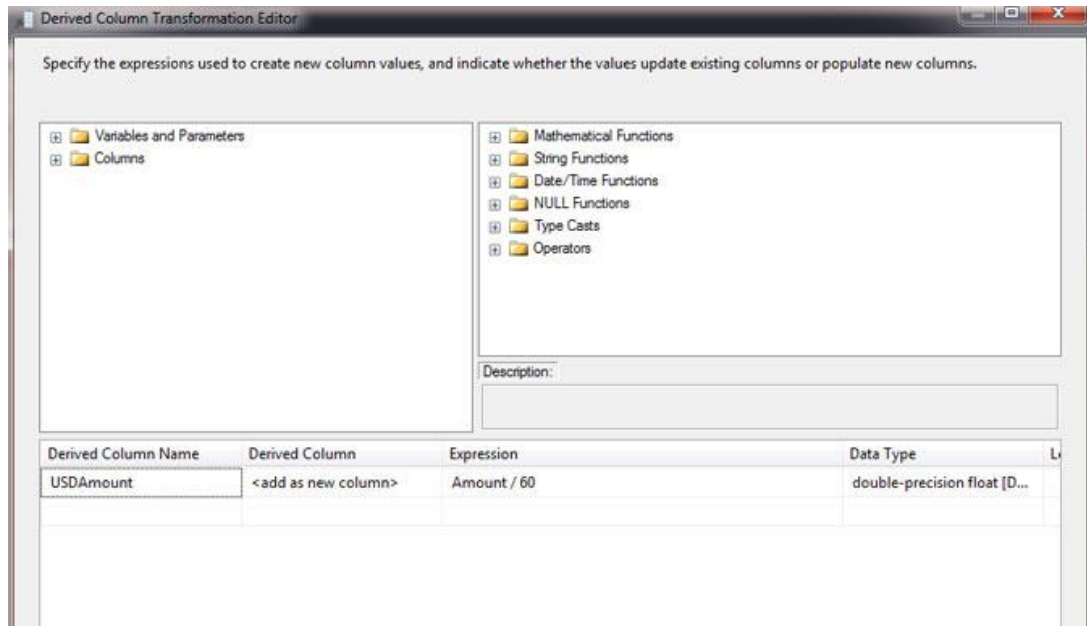
Now in the below Derived Column Name give a new column name "USDAmount" and
Derived column select "add a new column" and in the Expressing give the Formula
i.e. "Excel Column (Amount / 60)" and Data-type select as double precision float.

We have now USD amount, next step we need stand representation of product
name i.e. in UPPER CASE so for that we will add another "Derived Column" same
way. Now here we need to drag "Blue" arrow from "ConvertingAmounttoUSD" to
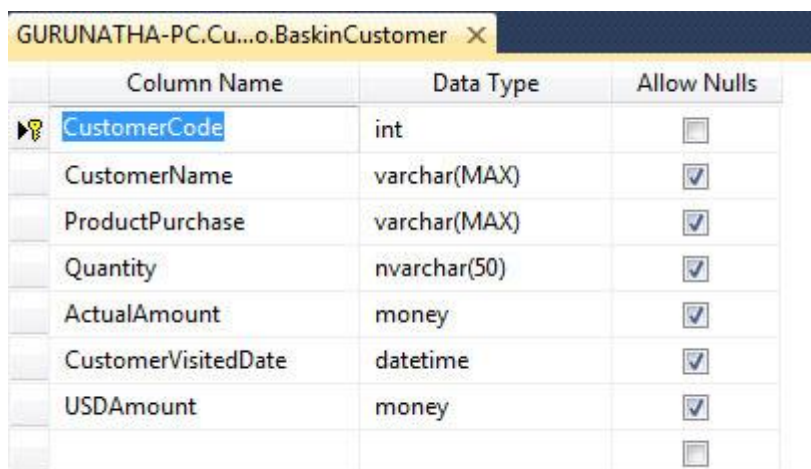new derived column ("CapitalProductName") as shown below image.
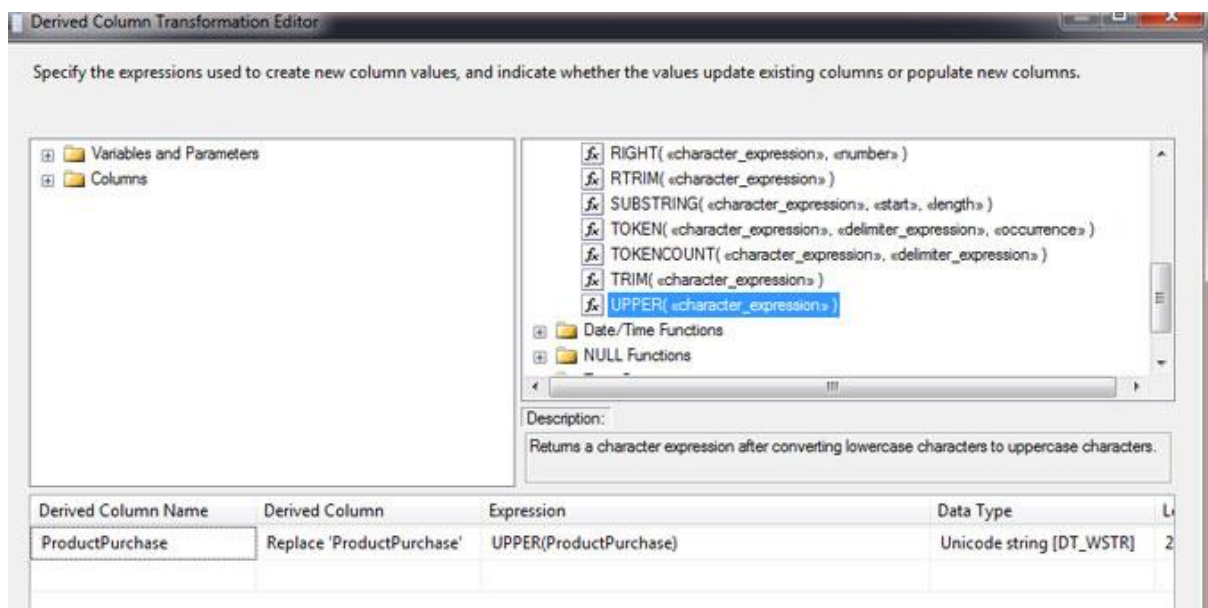
Same way right click on "CapitalProductName" go to EDIT choose column "ProductPurchase" from columns section and drag it to "Derive Column Name. In the Derived Column select "Replace 'ProductPurchase'.

Expression you can either choose from string function of you can type it i.e. UPPER (Column name). finally click on OK button.

So as you can see we have completed data transformation part i.e. Convert Amount to USD and Changed Product Names to UPPER case.

Step 3 : Loading Data To SQL Server

Before you start this step just open up your SQL server management studio and create a new database if need or just a new table with same excel column names as shown below image


Now go to your SSDT and from SSIS toolbox under Other Destination select "ADO.NET Destination"and drag it to "Data Flow" and drag arrow from "CapitalProductName" to "ADO.NET Destination".
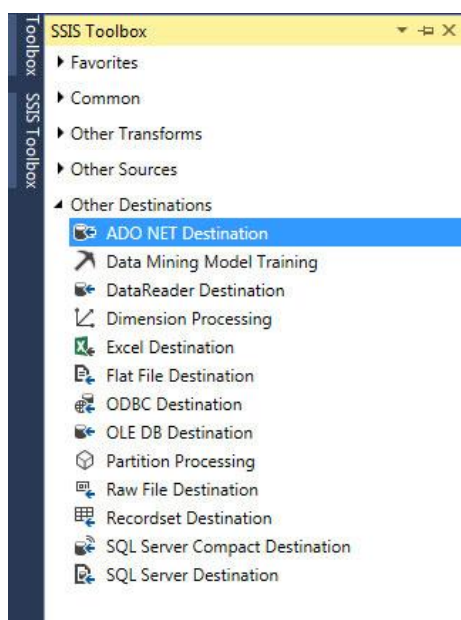
Right click and click on edit add your server name if you do not know your extact server name go to sql server management studio -> File -> click on connect and copy that server name. Now come back to SSDT and click New button -> Again New -> give here server name.Once you give server name automatically database dropdown will populate. Select database and click on OK button. Finally choose table as shown in below image and save it (OK button).
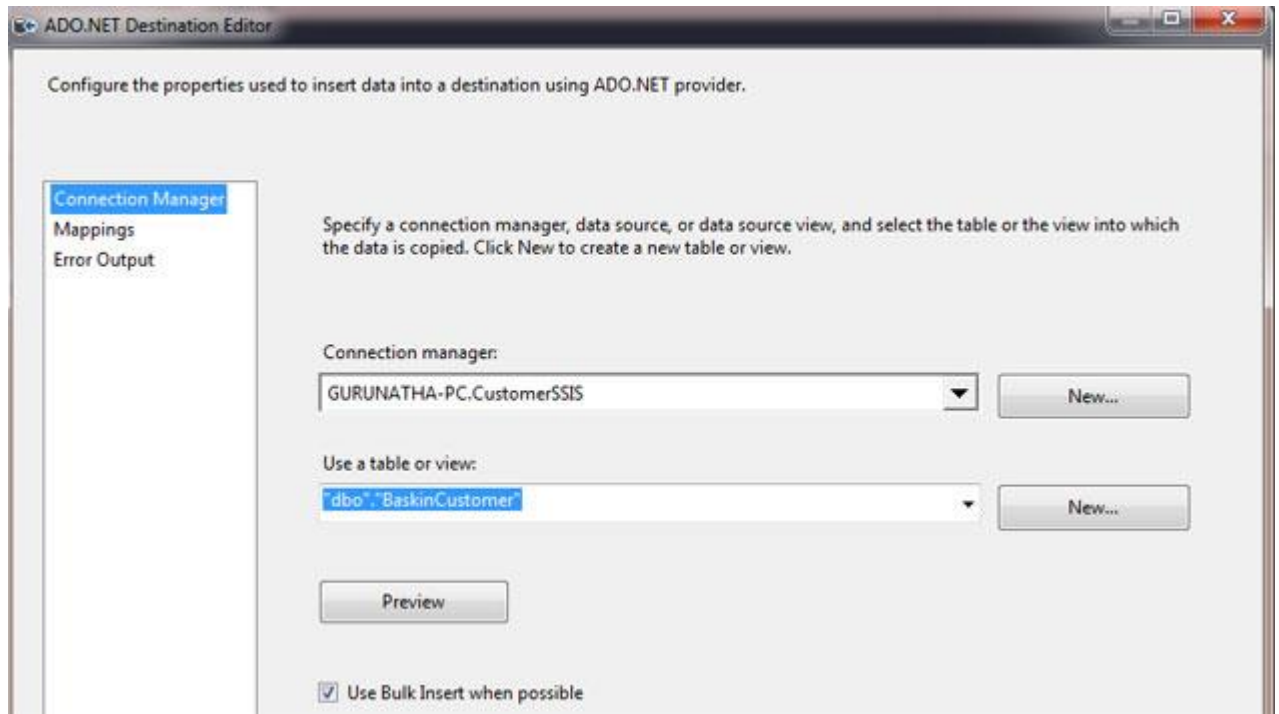
So what we did here it that we have created a path from where excel data flow from excel file -> Transform USD & UPPER CASE -> SQL Server Data Warehouse.

Final step just go to DEBUG and click on Start button from top menu of SSDT or just click on F5. Automatically data will flow from Excel Source -> SQL Server.

Self-Study:

All students are required to perform ETL experiment using Talend ETL tool