## Experiment No. : 6

Title: Demonstration of slowly changing dimension using SQL Server Integration Services

Objectives: 1. To study Slowly changing dimension and types of slowly changing dimensions

2. To demonstrate Type 0, Type 1 and Type 2 SCD in SQL Server

Key concepts: SCD

Theory:

SSIS Slowly Changing Dimension Type 2 (SCD 2): If you want to maintain the historical data of a column, then mark them as Historical attributes. If your Dimension table members (or Columns) marked as Historical attributes, then it will maintain the current record, and on top of that, it will create a new record with changing details. In real-time, the type of change provided by this SCD 2 or SSIS Slowly Changing Dimension Type 2 is beneficial. For example, if we want to maintain the Employee’s previous and current Country then we can mark these columns as Historical attributes.

In Type 2 SCD, a new record is added to the Destination table to represent the new information. Therefore, both the original and the new record will be present. The new record gets its own primary key.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Emp\_ID | Emp\_Name | Country | Start\_Date | End\_Date |
| 1 | Roopesh | India | 01-01-11 | NULL |

Now, when the Country is changed to the Employee “Roopesh” from India to USA then  the data will be stored in the following way –

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Emp\_ID | Emp\_Name | Country | Start\_Date | End\_Date |
| 1 | Roopesh | India | 01-01-11 | 13-03-12 |
| 2 | Roopesh | USA | 13-03-12 | NULL |

The null End\_Date in row two indicates the current tuple version. In some cases, a standardized surrogate high date (e.g. 9999-12-31) may be used as an end date, so that the field can be included in an index, and so that null-value substitution is not required when querying. So, in this case we can say the records with NULL as End\_Date is the current record for an employee.

The major disadvantage of using this is the table SIZE grows faster as records will get added whenever there is change in data to maintain history and the advantage of using this Type is that we can track historical information accurately.  I hope you got some useful info regarding SCD type 2 and now let’s jump into SCD Transformation.

I have created two tables one Test\_Stage(Which I am using as Staging table) and another one is Test(Which I am using as Main table) with the following syntax. You can clearly see that Main table is having TWO Extra columns to track history.

Create table Test\_Stage  
(  
Emp\_Id int,  
Emp\_Name Varchar(100),  
Country Varchar(100)  
)

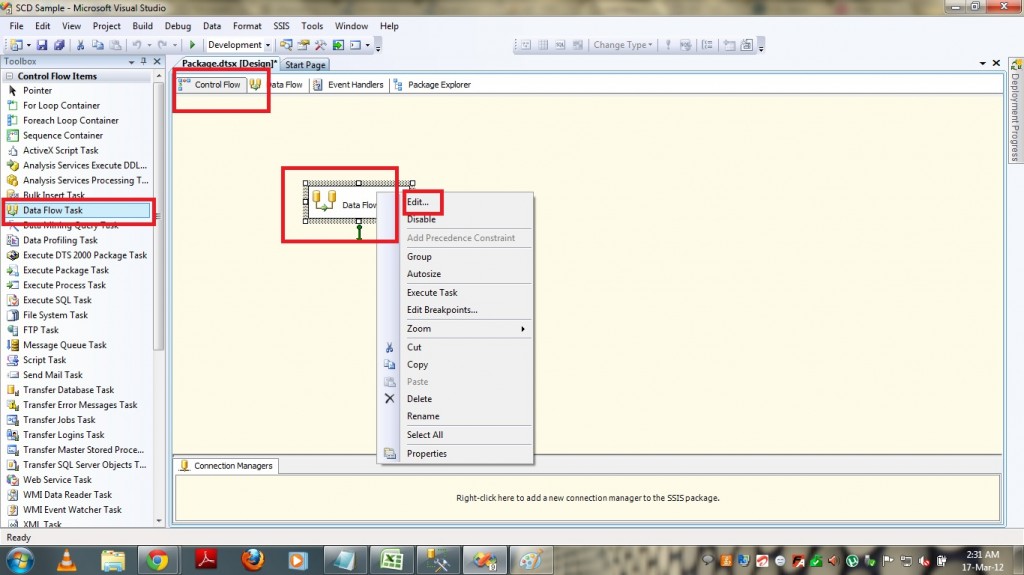
Create table Test  
(  
Emp\_Id int,  
Emp\_Name Varchar(100),  
Country Varchar(100),  
Start\_Date Datetime,  
End\_Date DateTime  
)

Now I inserted some records into my staging table i.e Test\_Stage and PFB the scripts used –

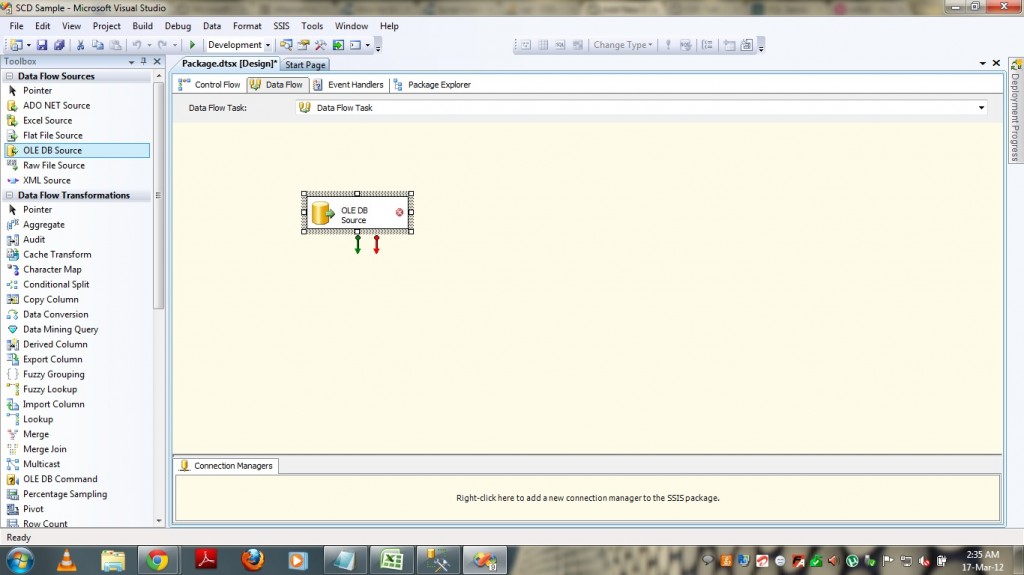
Insert into test\_stage Values(1,’Roopesh’,’India’);  
Insert into test\_stage Values(2,’Lokesh’,’UK’);  
Insert into test\_stage Values(3,’Vinay’,’USA’);  
Insert into test\_stage Values(4,’Rakesh’,’China’);  
Insert into test\_stage Values(5,’Venki’,’Japan’);

Now Source data is ready and PFB the steps you have to follow to use Slowly Changing Dimension Transformation to implement Type 2 SCD.

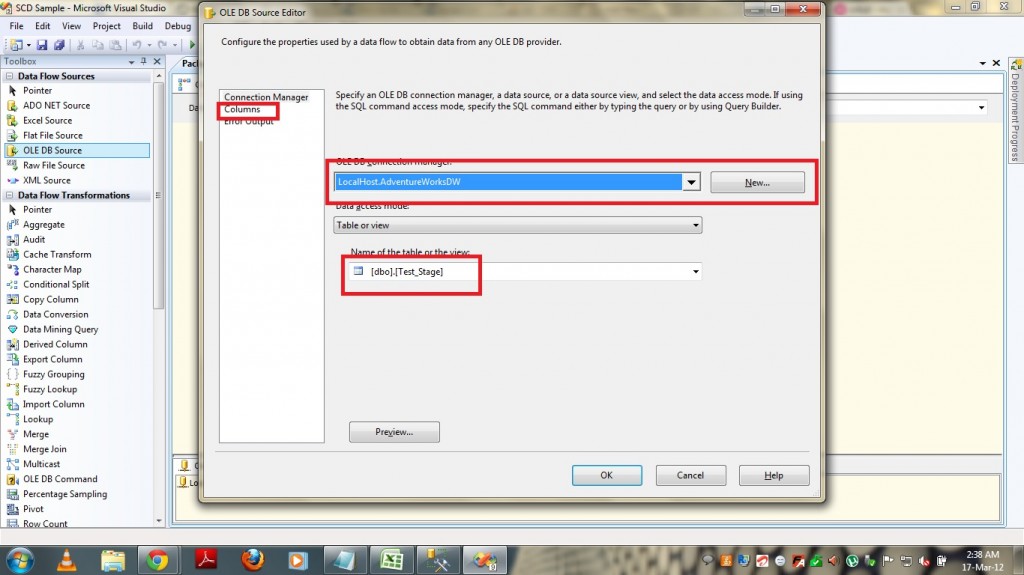
* Open SSIS Package and drag a dataFlow Task from toolbox to control Flow Pane as shown below.
* Either double click or Right click on Data Flow Task and select EDIT as shown below.

[](http://www.msbiguide.com/wp-content/uploads/2012/03/SCD-1.jpg)

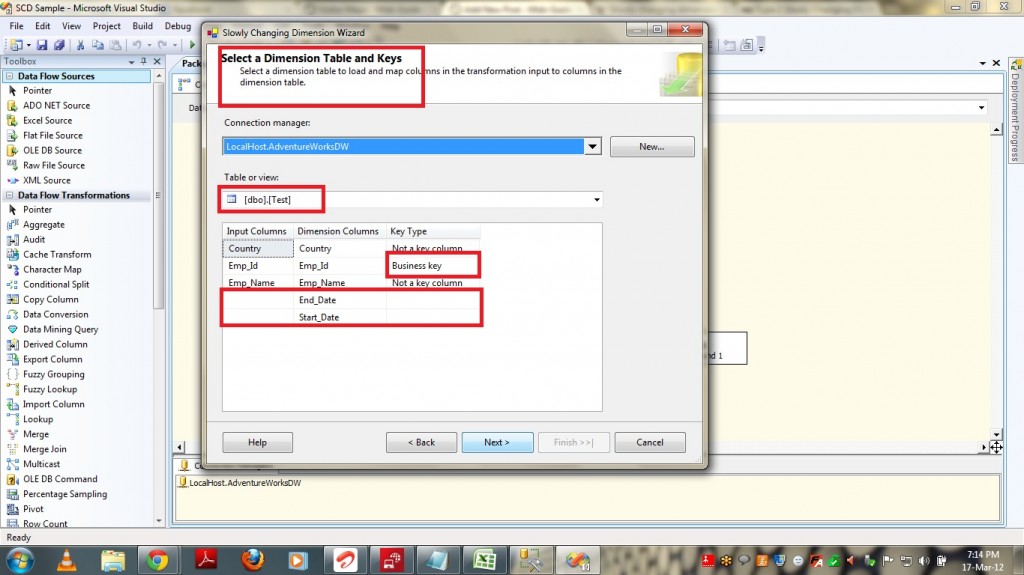
* One the Data Flow pane is opened then drag and drop OLEDB SOURCE from “Data Flow Sources” as shown below.

[](http://www.msbiguide.com/wp-content/uploads/2012/03/SCD-2.jpg)

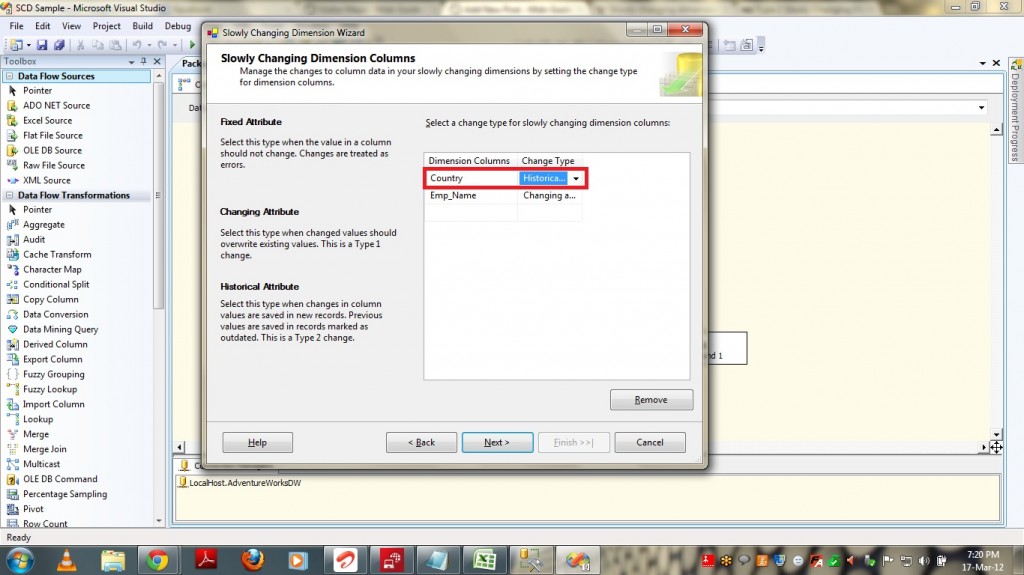
* Now select EDIT by Right clicking on OLE DB Source and provide the Source details. remember that you have to provide Staging table connection details here.
* Select columns tab and check whether columns are coming correctly or not.

[](http://www.msbiguide.com/wp-content/uploads/2012/03/SCD-3.jpg)

* Now drag and drop Slowly Changing Dimension Transformation in to the Data Flow Pane and provide the connection between OLE DB Source and SCD Transformation.
* Now right click on the SCD Transformation and click on EDIT menu.
* Select the dimension table(in our case it is “Test”) and Keys under “Select Dimension Table and Keys” page as shown below.

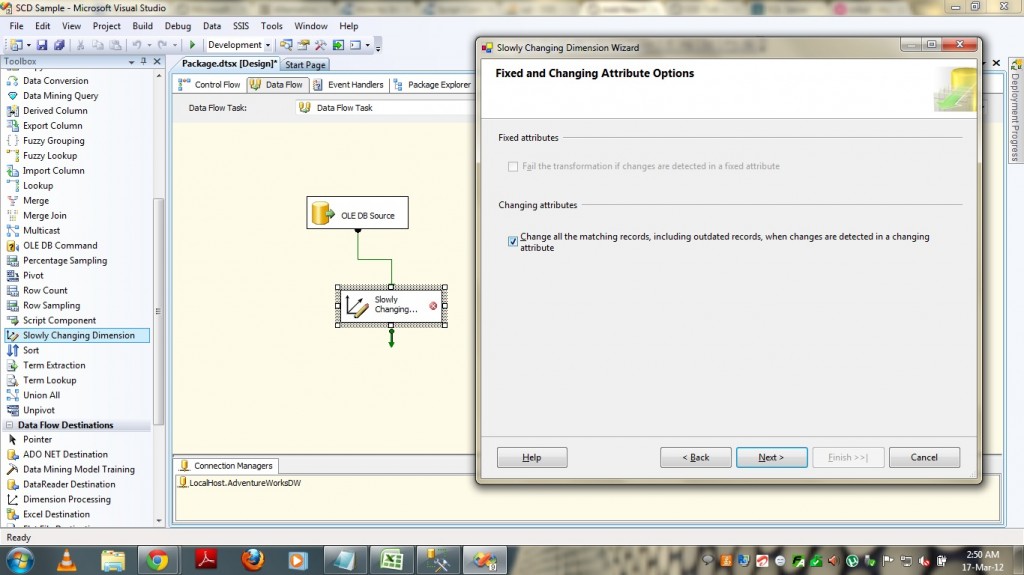
[](http://www.msbiguide.com/wp-content/uploads/2012/03/SCD-13.jpg)

* In the above screenshot, the Business Key is the column based on which we will filter the Newly added records from the existing one. In simple this is the column which we use to look up.
* After selecting Next, Under “Slowly Changing Dimension Columns” select the columns that MAY change and select Change Type as “Changing Attribute” for those which we DOESN’T required to maintain HISTORY and select “Historical Attribute” for those whose HISTORY needs to be tracked. This is very important part of SCD and in other words this is the ONLY change we have when compared to SCD Type 1 and SCD Type 2 implementation in SSIS. When we select “Changing Attribute” for any attribute then it WON’T CREATE a new record when there is a change in this value and if you select “Historical Attribute” then if there is any change in this attribute value then a NEW RECORD will be created to track history.

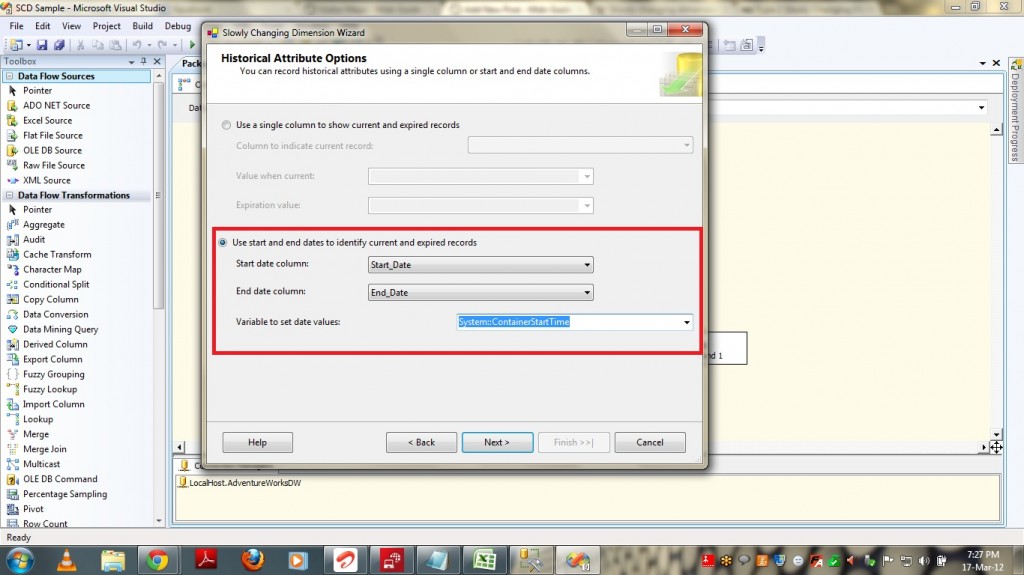
[](http://www.msbiguide.com/wp-content/uploads/2012/03/SCD-14.jpg)

* In the above screenshot we can see that for Country attribute I selected “Historical Attribute” as I wish to track the historical changes for that attribute and hence when ever there is a change in that particular attribute column value then only it tracks history by creating a new record and for other attribute Emp\_Name, it will directly apply UPDATE on the table.

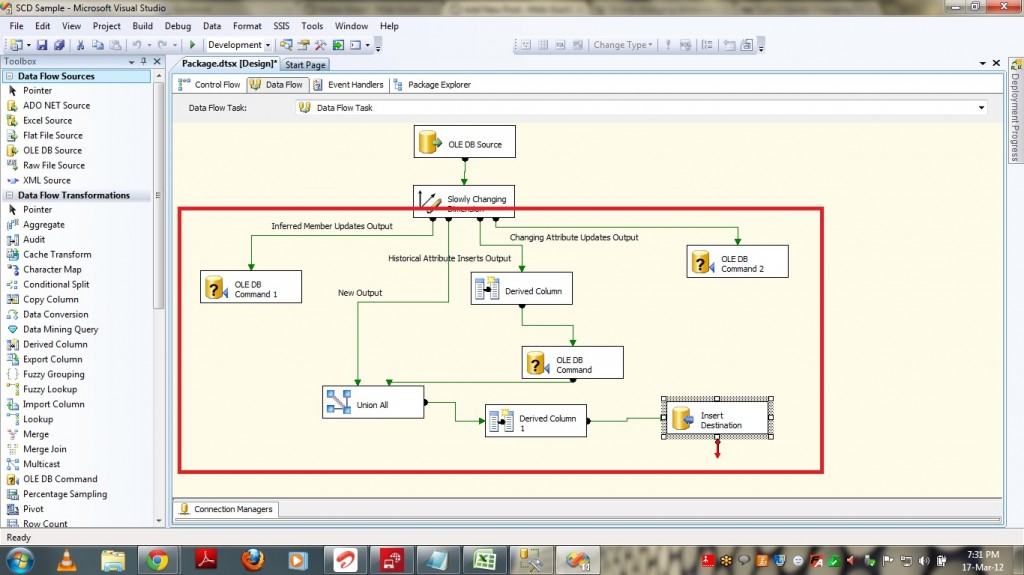
* Check the option “Change all the matching records including the outdated records…….” if you wish to have that functionality and click Next.

[](http://www.msbiguide.com/wp-content/uploads/2012/03/SCD-6.jpg)

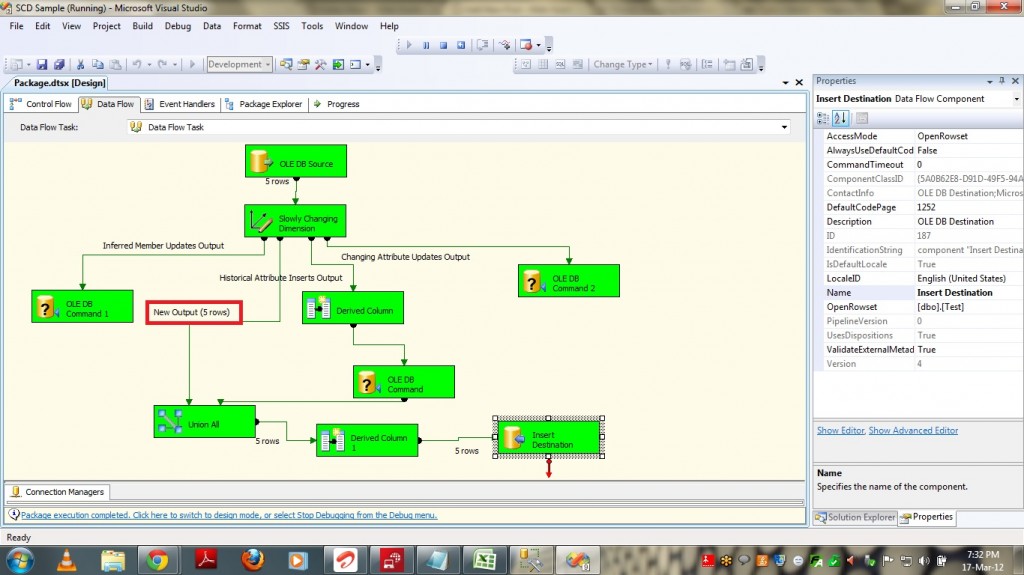
* Under “Historical Attributes Option” page select “Use Start and End date to identify current and expired records” option and by selecting start date column and End date column click on NEXT.

[](http://www.msbiguide.com/wp-content/uploads/2012/03/SCD-15.jpg)

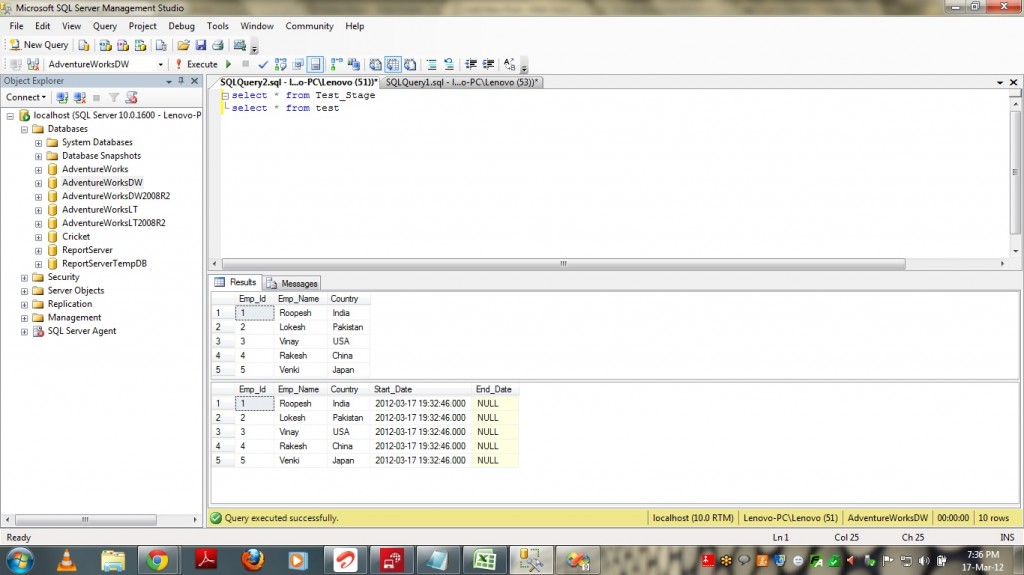
* Click Next and finally click on Finish by leaving the default options as is.
* Once you click on Finish, SSIS server creates all the functionality that is required to implement SCD Type 1 based on the information provided by us. It will add Data flow tasks for Inserting new records and updating existing records as shown below.

[](http://www.msbiguide.com/wp-content/uploads/2012/03/SCD-16.jpg)

* Now if I run the package then it will check for newly added records and inserts those records and updates the already existing records if there is any change in data. As there is NO RECORDS in TEST table in this run it will load all the records under New Insert as shown below.

[](file:///F:\tib\covid-lockdown\DWBI\Practical\BWBI-Lab\imp-How%20to%20Define_Implement%20Type%202%20SCD%20in%20SSIS%20using%20Slowly%20Changing%20Dimension%20Transformation%20_%20Msbi%20Guide_files\SCD-17.jpg)

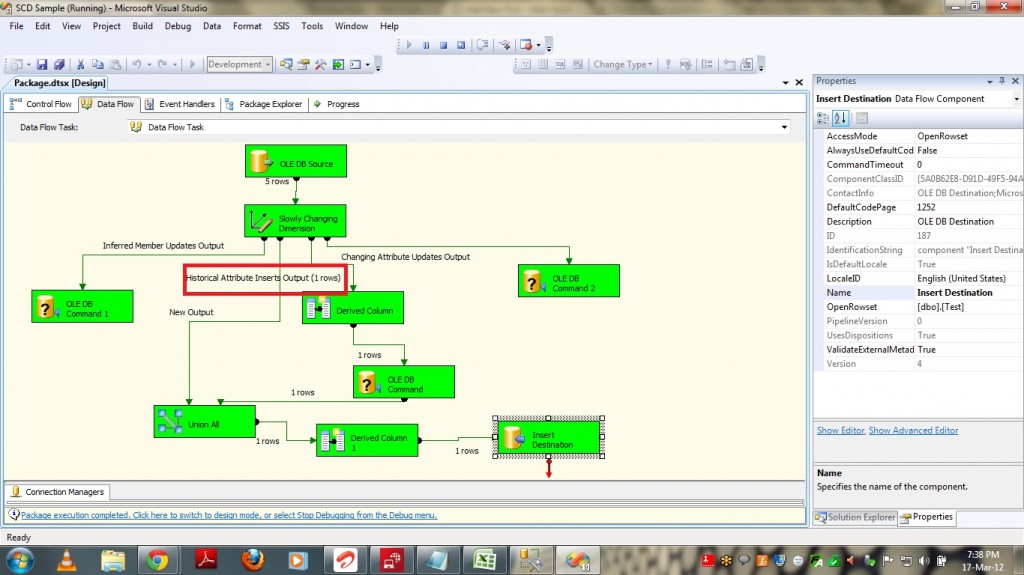
* PFB the Screenshot of the data tables output after running the package. You can see data got loaded in to our table “TEST” with Start\_Date as current Date.

[](http://www.msbiguide.com/wp-content/uploads/2012/03/SCD-18.jpg)

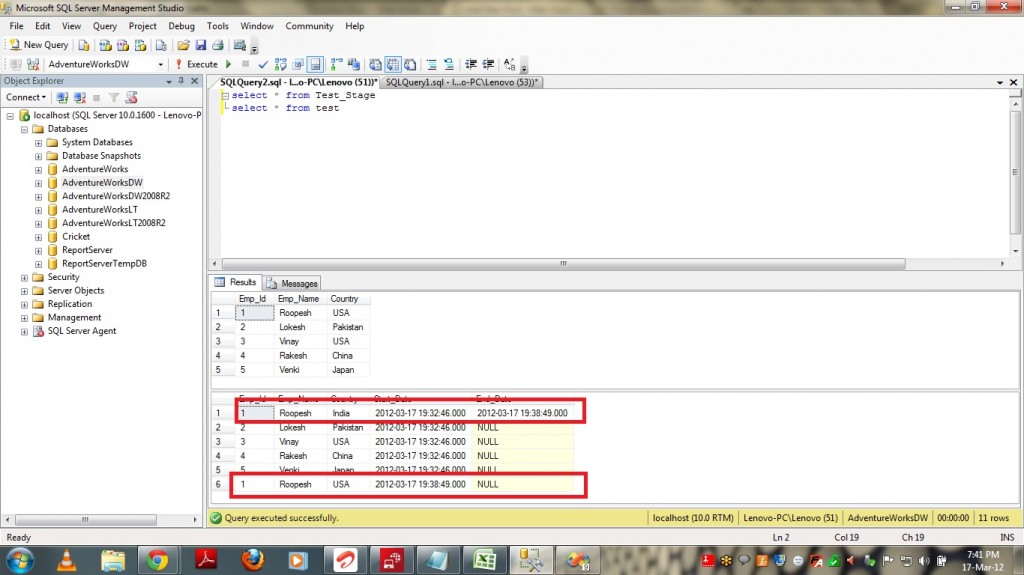
* Now I updated the existing records in staging table using the following queries and once I run the package again we can see data getting updated into our main table.

              update test\_stage set Country = ‘USA’ where Emp\_Id = 1

* Once I run the package you can see the package showing “1 rows” under “Historical Attributes Inserts Output” section as shown below.

[](http://www.msbiguide.com/wp-content/uploads/2012/03/SCD-19.jpg)

* Now if you run the queries you can see the data updated in TEST table.

[](http://www.msbiguide.com/wp-content/uploads/2012/03/SCD-20.jpg)

Now we implemented SCD Type 2 using Slowly Changing Dimension Transformation.

Bottom of Form