Business Intelligence Solution Providers
Specialized in creating talent resource pool

Change Data Capture in IBM DataStage 8.5 - Lab#13

Description:

BISP is committed to provide BEST learning material to the beginners and advance learners. In the same series, we have prepared a complete end-to end Hands-on Guide for building financial data model in DataStage. The document focuses on how the real world requirement should be interpreted. The mapping document template with very simplified steps and screen shots makes the complete learning so easy. This is one of our documents from DataStage learning series. This document describes in step by step manner how to work on Change Capture stage in DataStage. With the help of this stage we want know which type of operation (updating or insertion) is applied on our source table and then we can see our output in dataset file. . Join our professional training program and learn from experts.

History:

Version	Description Change	Author	Publish Date

Mapping Sheet

Change_Capture Transformation.
Source- Oracle,Flat File
Target- DataSet

Transformation	Change Capture						
Source Name	Flat File, Oracle						
Target Name	DataSet File						

Flat File Source Details				Oracle Details				Flat Files	
Entity Name	Data Types	Is Null	Destination Entity Nan	Destination Entity Name	DataTypes	Is Null	DestinationField Name	DataTypes
ACCOUNT_OFFICER_CD	NUMBER(5)	N	src_creditcard	ACCOUNT_OFFICER_CD	NUMBER(5)	N	ACCOUNT_OFFICER_CI	NUMBER(5)
CREATED_BY	VARCHAR2(30)	y	src_creditcard	CREATED_BY	VARCHAR2(30)	y	CREATED_BY	VARCHAR2(30)
CREATION_DATE	TIMESTAMP(6)	y	src_creditcard	CREATION_DATE	TIMESTAMP(6)	y	CREATION_DATE	TIMESTAMP(6)
DEFINITION_LANGUAGE	VARCHAR2(4)	y	src_creditcard	DEFINITION_LANGUAGE	VARCHAR2(4)	y	DEFINITION_LANGUAG	VARCHAR2(4)
ENABLED_FLAG	VARCHAR2(1)	y	src_creditcard	ENABLED_FLAG	VARCHAR2(1)	y	ENABLED_FLAG	VARCHAR2(1)
LAST_MODIFIED_BY	VARCHAR2(30)	y	src_creditcard	LAST_MODIFIED_BY	VARCHAR2(30)	y	LAST_MODIFIED_BY	VARCHAR2(30)
LAST_MODIFIED_DATE	TIMESTAMP(6)	y	src_creditcard	LAST_MODIFIED_DATE	TIMESTAMP(6)	y	LAST_MODIFIED_DATE	TIMESTAMP(6)
LEAF_ONLY_FLAG	VARCHAR2(1)	y	src_creditcard	LEAF_ONLY_FLAG	VARCHAR2(1)	y	LEAF_ONLY_FLAG	VARCHAR2(1)
ACCOUNT_OFFICER_DISPLAY_CD	VARCHAR2(10)	y	src_creditcard	ACCOUNT_OFFICER_DISPLAY	VARCHAR2(10)	y	ACCOUNT_OFFICER_DI	VARCHAR2(10)
							CHANGE_CODE	TRYINT

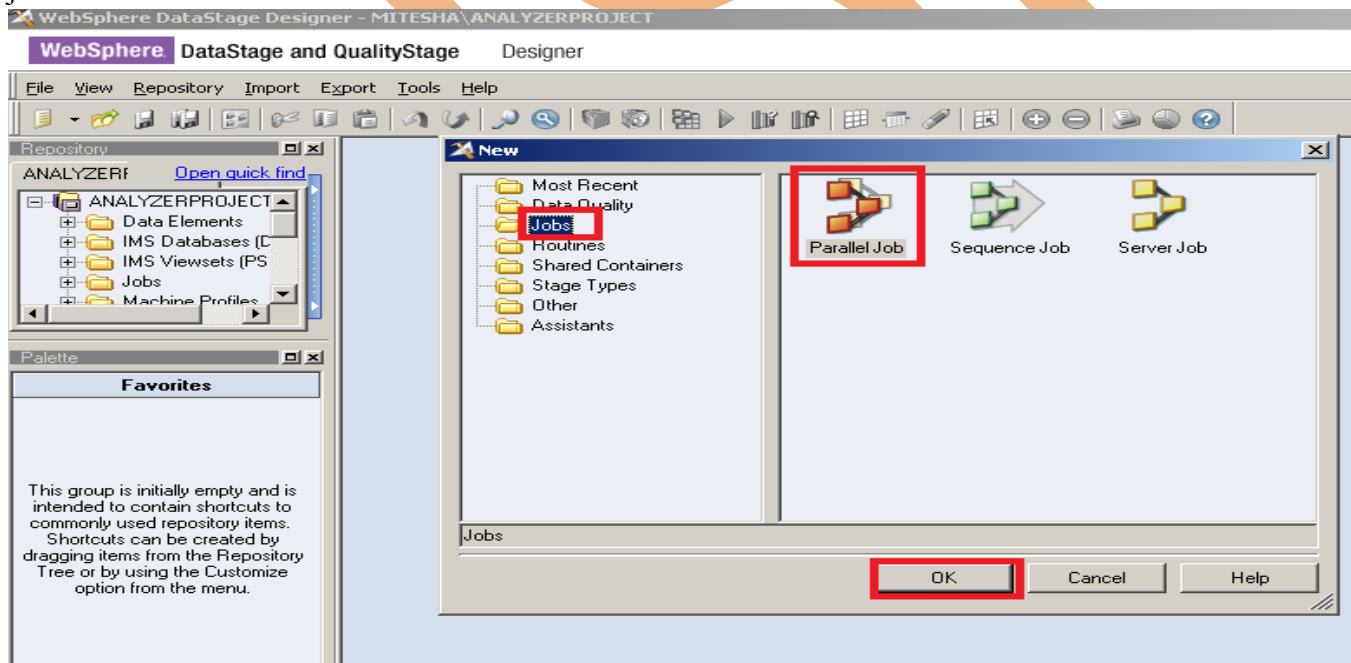## Loading Data In SQL Server, Oracle and Flat File with DataStage
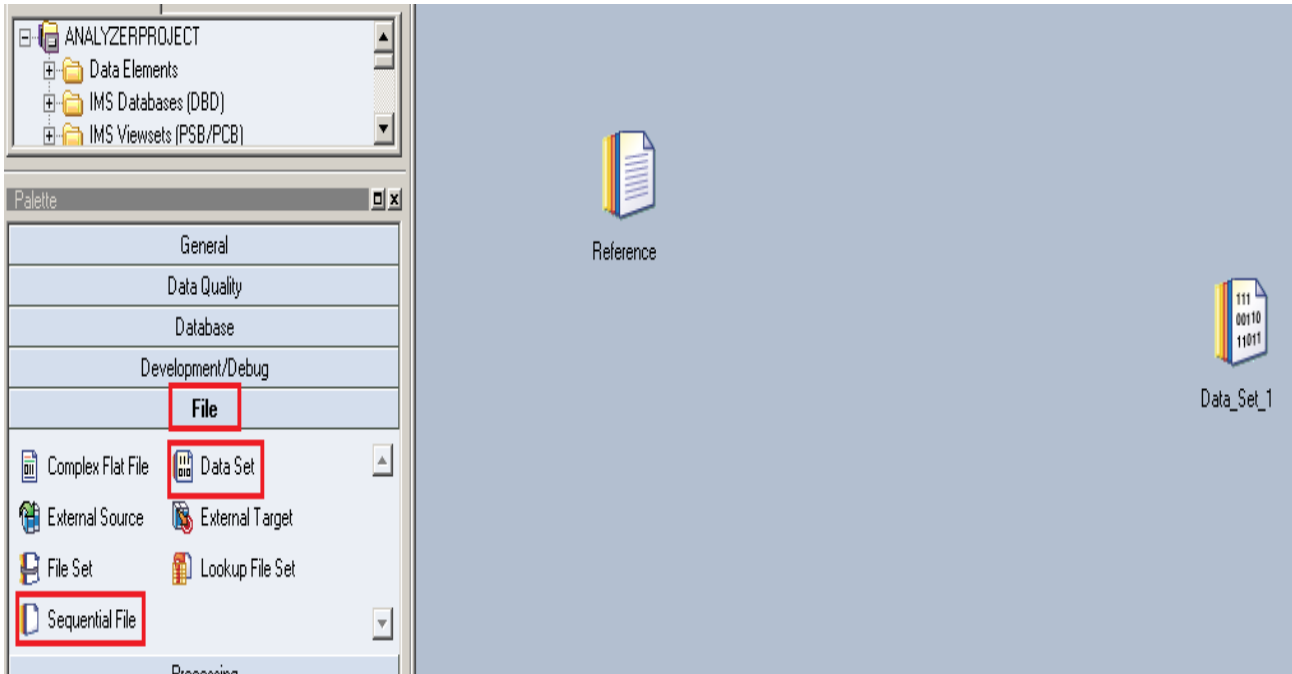**Step-1 :-** In Windows Click Designer Client of DataStage


Step-2 :- New Window opens, then click Jobs ,then click Parallel Job because we are using parallel jobs.


Step-3:- First go on File Palette then scroll the cursor and choose Sequential File and Dataset file. Drag them to Parallel pane and rename. These files are used for reference data and output data.

Step-4:- Choose Oracle Enterprise from Database Palette and drag it to parallel pane. This is used for taking input data from oracle.
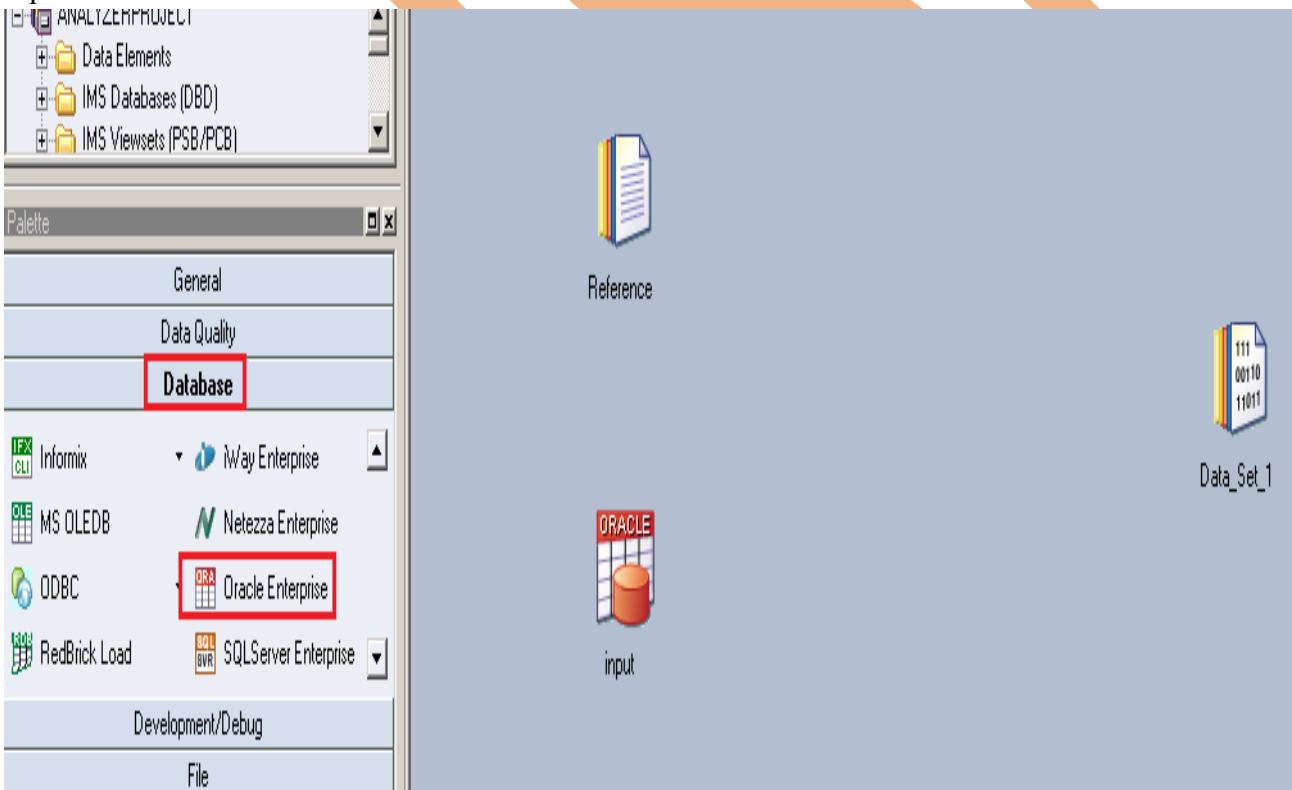
Step-5:- Now go to processing palette and choose ChangeCapture stage as we know that this stage is useful for capturing change on your input file with the help of reference file.
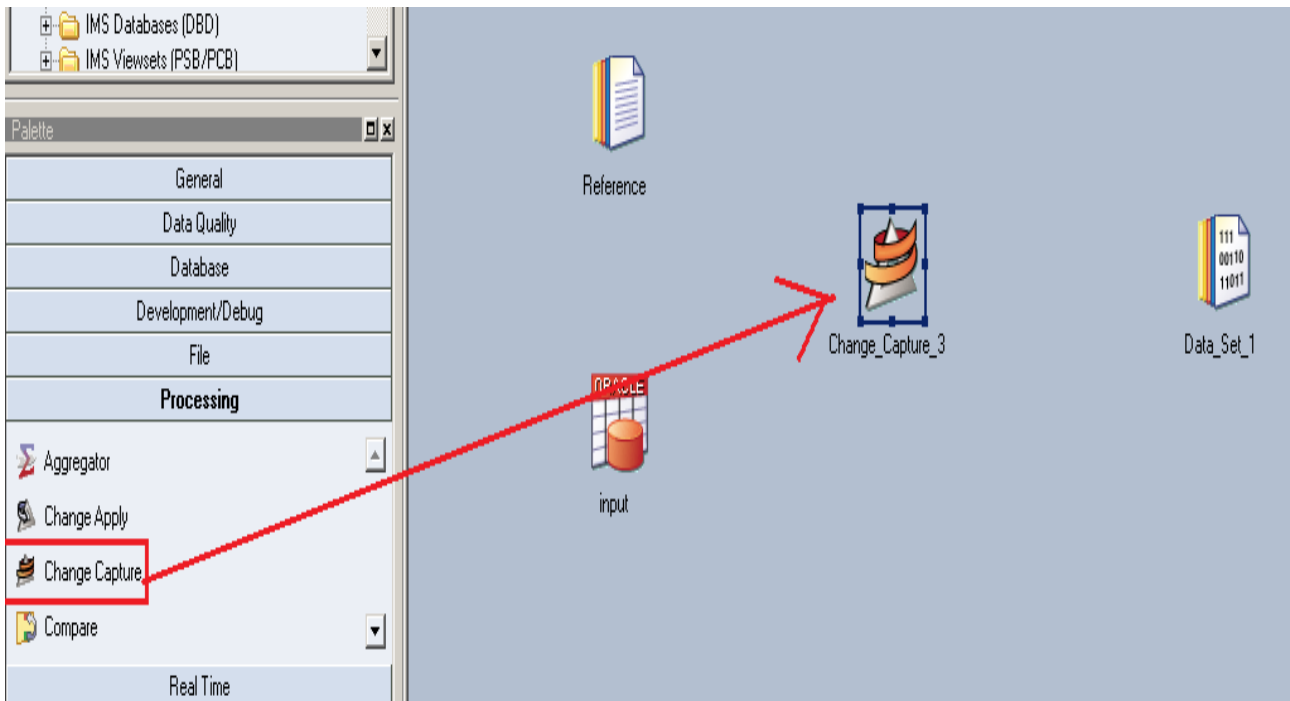
Step-6:- Now make connection with these stages. So we have to use this link stage for connecting these stages.
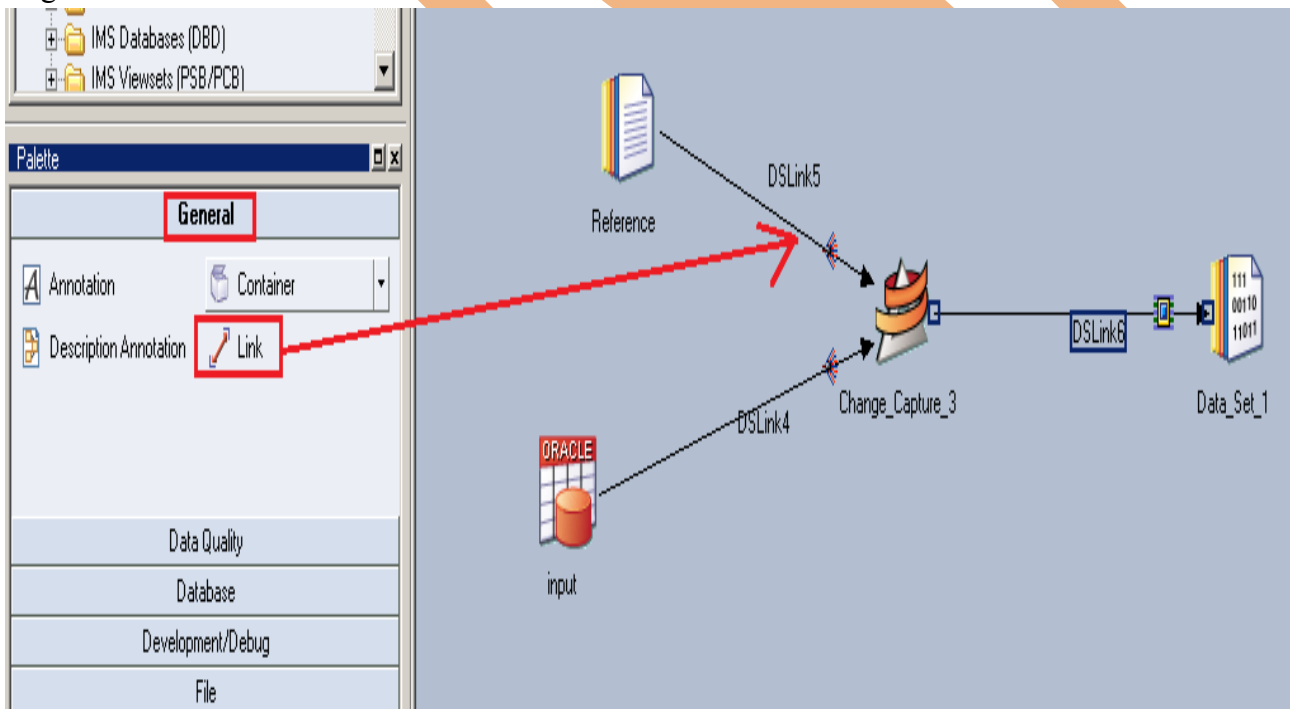

Step-7:- Now double click on oracle stage for fetching oracle table into datastage. There are multiple methods available but we choose table because we already know the table and schema name.
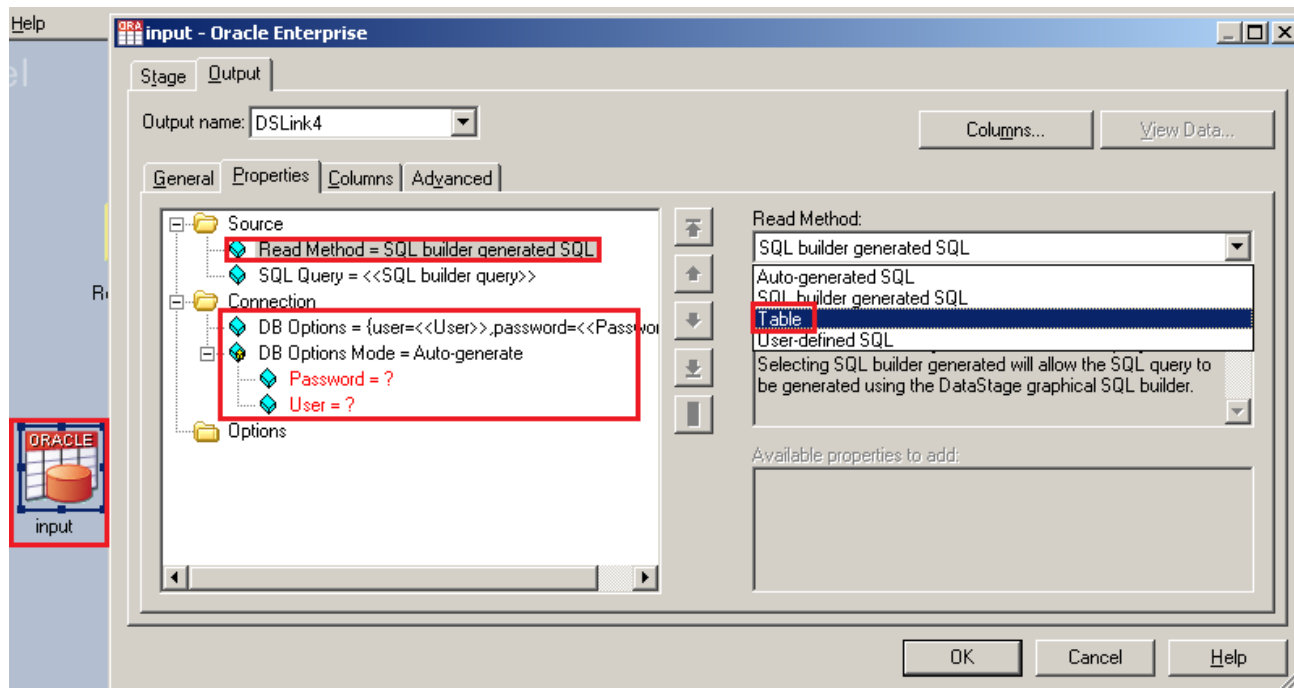
Step-8:- There's one more option. Once you click on connection tab then remote server options popup. Here we have to specify our server name of oracle.
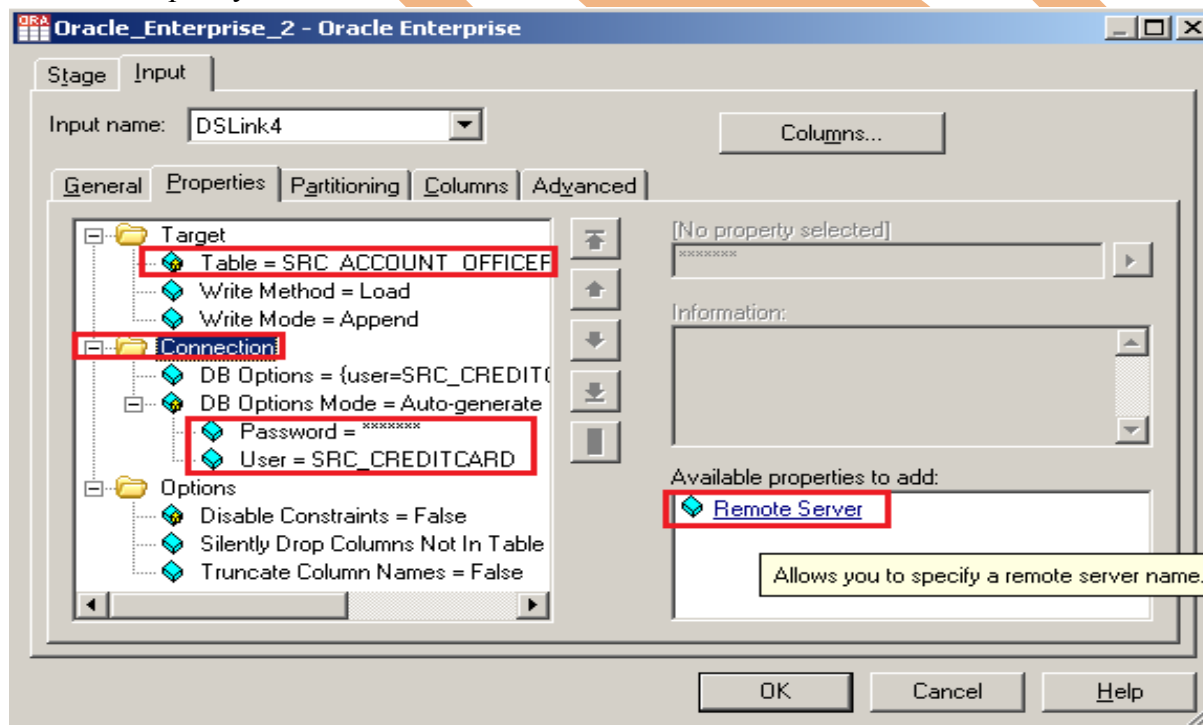

Step-9:- See below, my server name is ORCL and enter username/password and table name. Ensure that you put correct information then only you will be able to load the data into table.
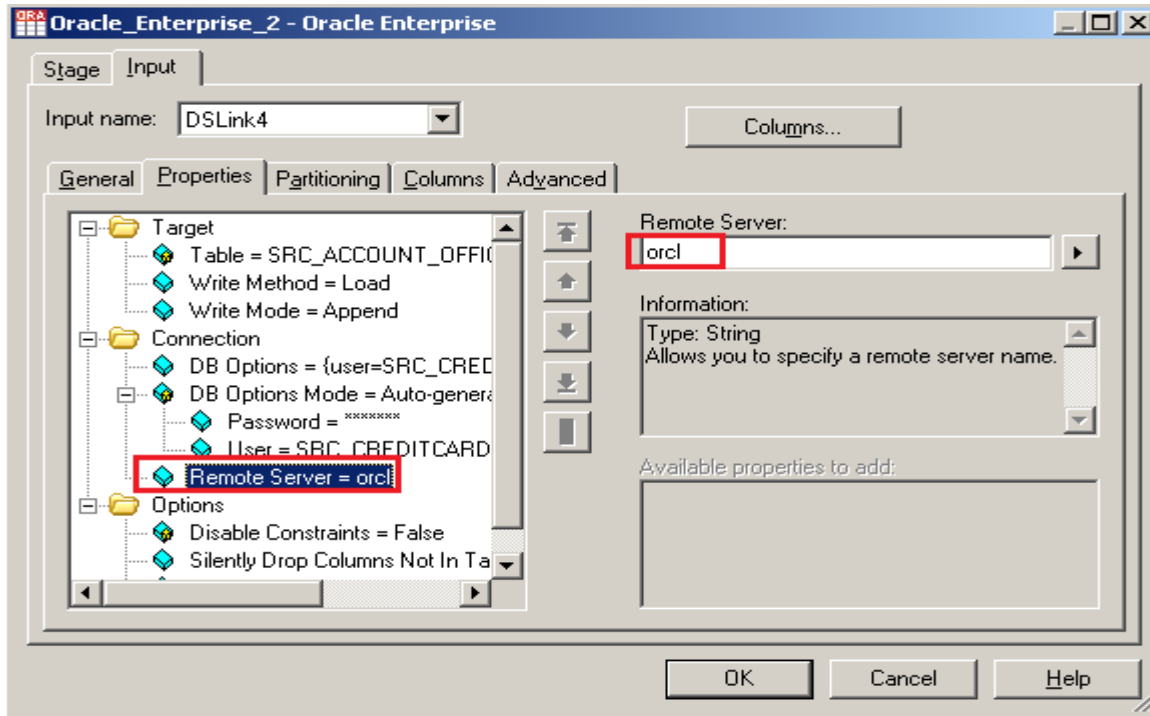
Step-10:- Now go to column tab and put all the columns name with data type and it should match with your oracle table columns.
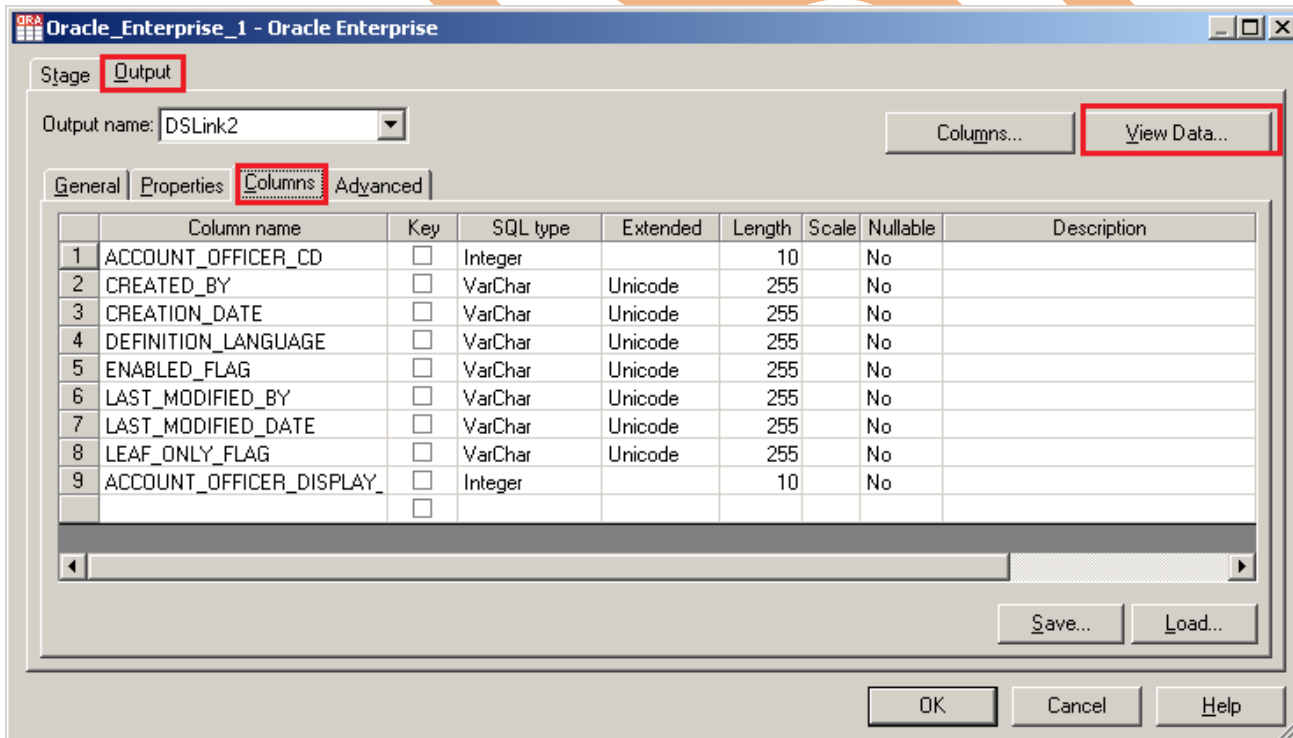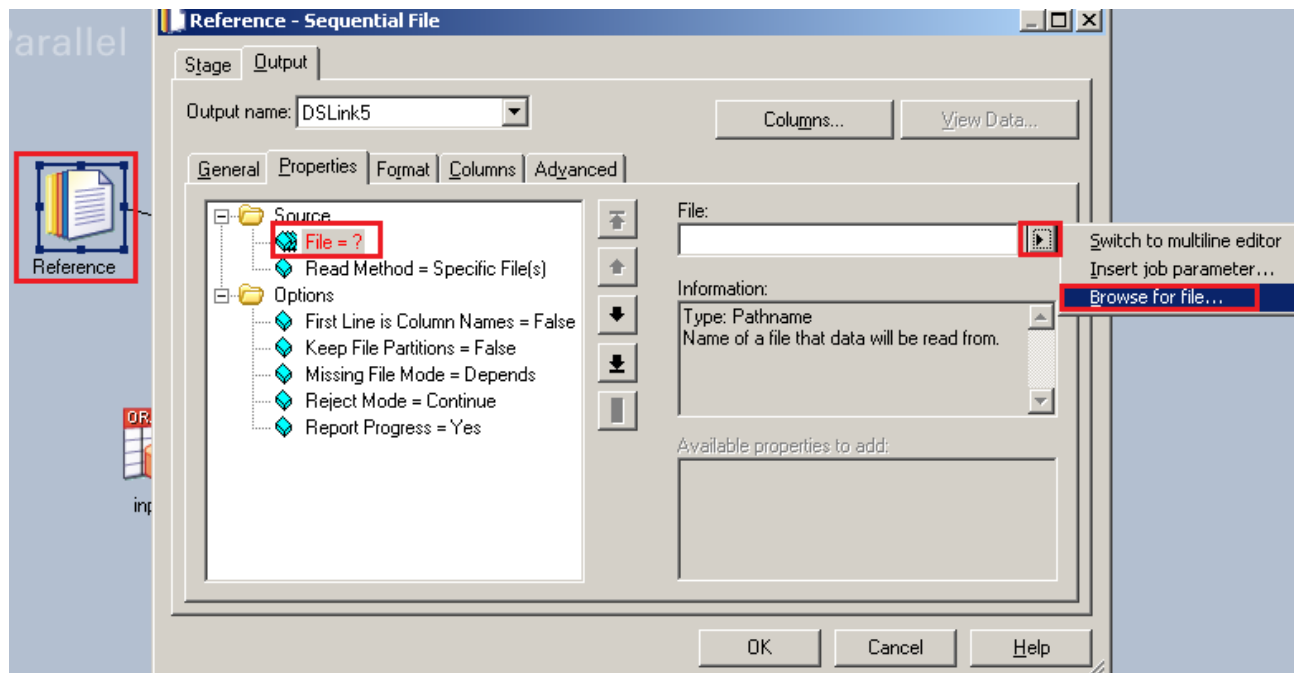

Step-11:- Now click on View Data and Ok. If your data is shown like this that means you have made successful connection between input file and Data Stage otherwise it will give some error like column mismatch error.
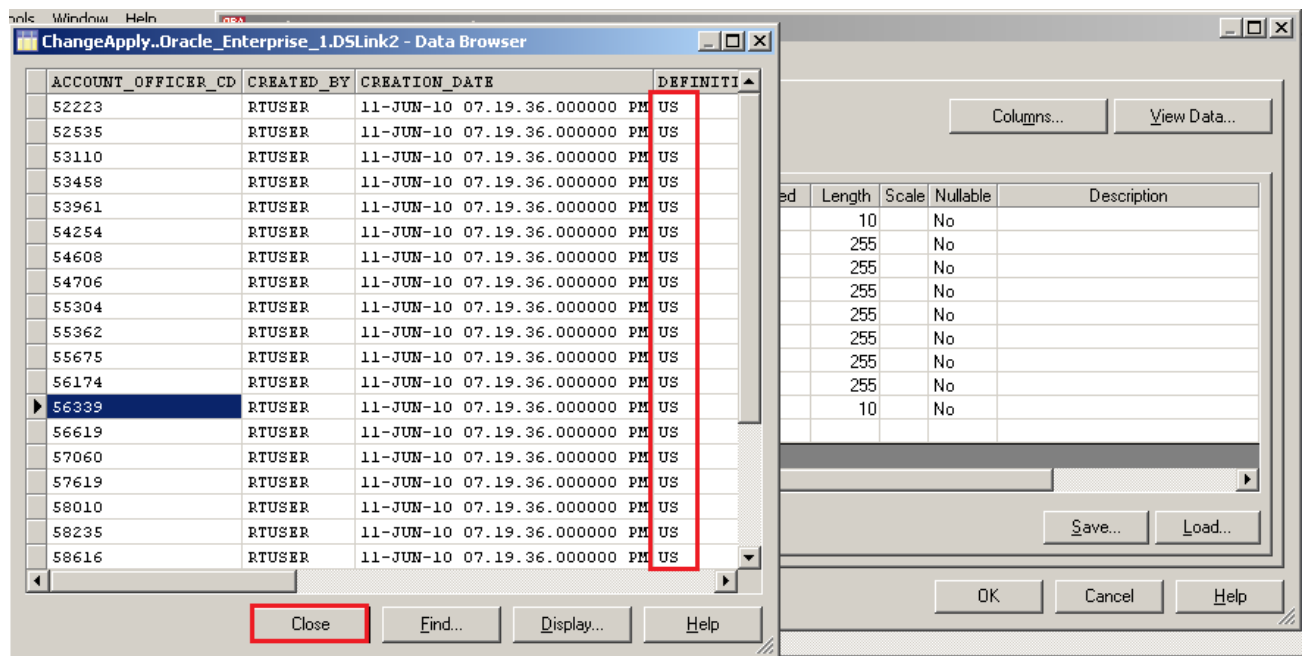
Step-12 : Now first before loading data into output, we have to create a table. So, we write a create table command and make sure to give same column name as you have in flat file. Go to SQL tab then write query and remember the username and password.

```
SQL> CREATE TABLE SRC_ACCOUNT_OFFICER_CD_FILT
  2      ( ACCOUNT_OFFICER_CD NUMBER(5,0) PRIMARY KEY,
  3    CREATED_BY VARCHAR2(30 BYTE),
  4    CREATION_DATE VARCHAR2 (100),
  5    DEFINITION_LANGUAGE VARCHAR2(4 BYTE),
  6    ENABLED_FLAG VARCHAR2(1 BYTE),
  7    LAST_MODIFIED_BY VARCHAR2(30 BYTE),
  8    LAST_MODIFIED_DATE VARCHAR2 (100),
  9    LEAF_ONLY_FLAG VARCHAR2(1 BYTE),
 10    ACCOUNT_OFFICER_DISPLAY_CD VARCHAR2(10 BYTE) DEFAULT NULL
 11      );
```
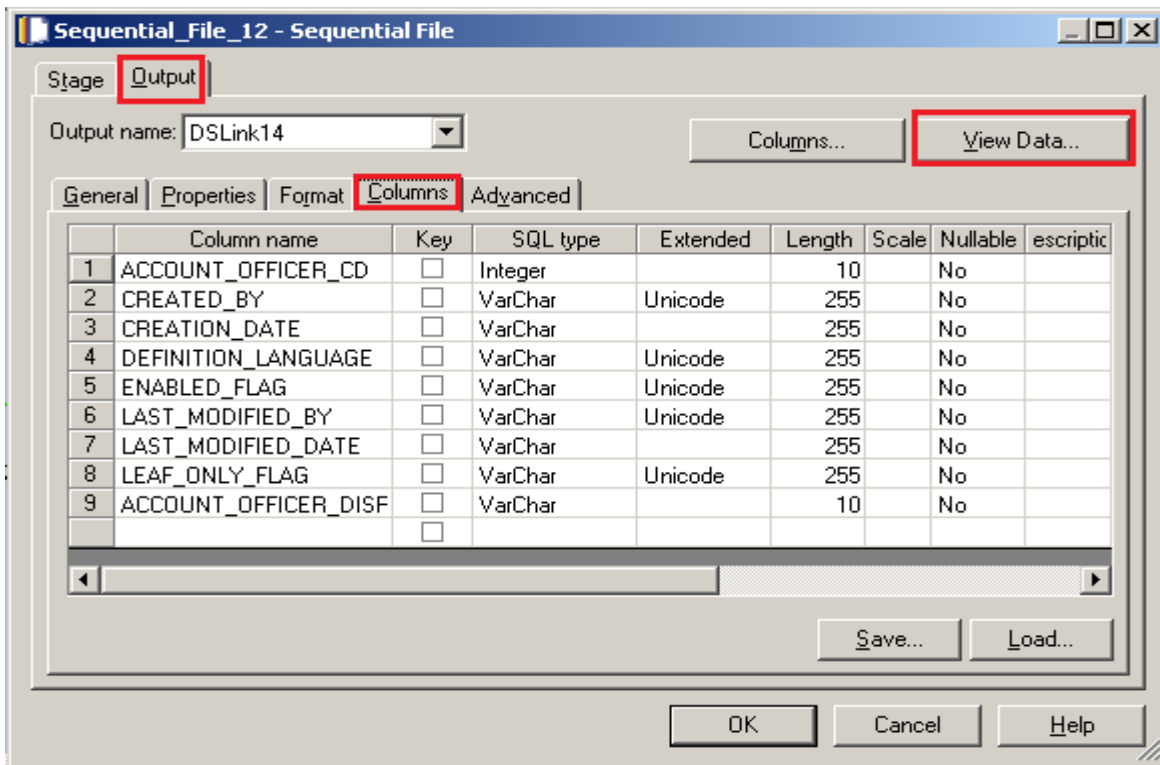
Step-13:- Double click on Reference file and browse an input file that can be .txt, .csv or any test file.

Step-14:- Go to Format tab and Remove double Quote because we don't need them then Click on OK.

Note:- If you use date as input then specify here which type of format you used.
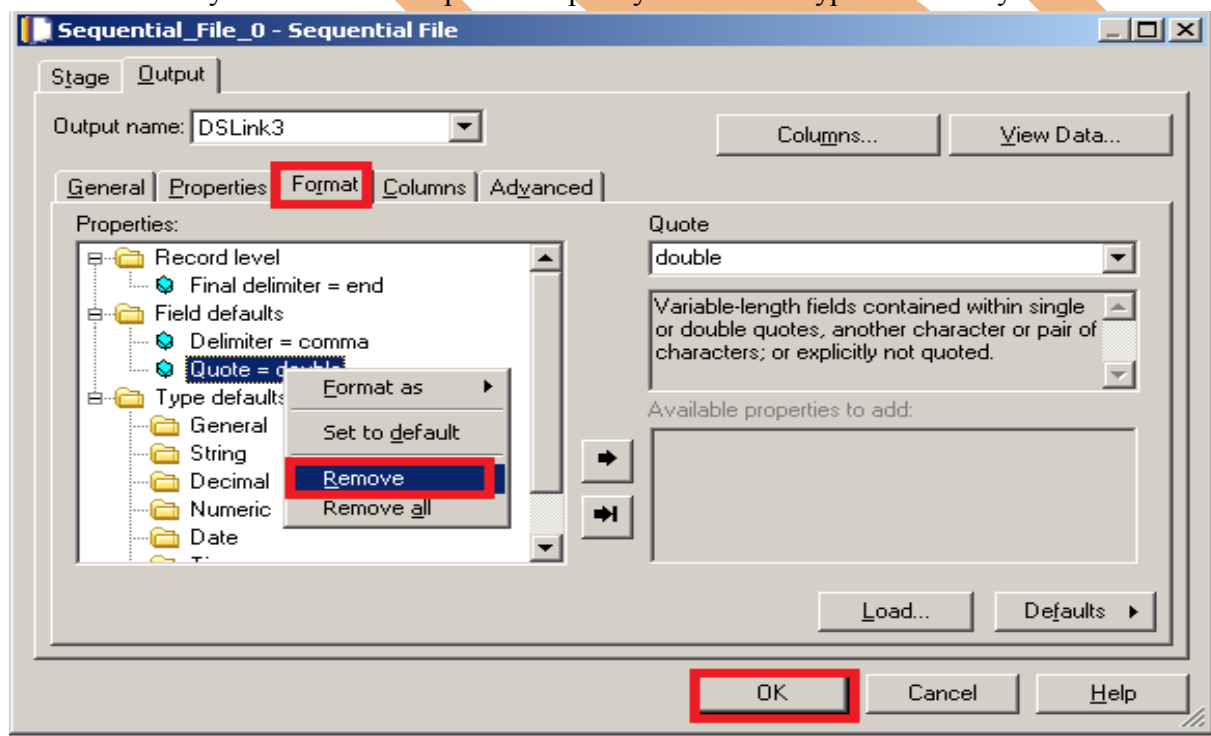

Step-15:- Go to Columns tab and enter same column names as declared in Input file and put datatype, length.

Step-16:- Now click on View Data and Ok. If your data is shown like this that means you have made successful connection between reference file and DataStage otherwise it will give some error like column mismatch error.
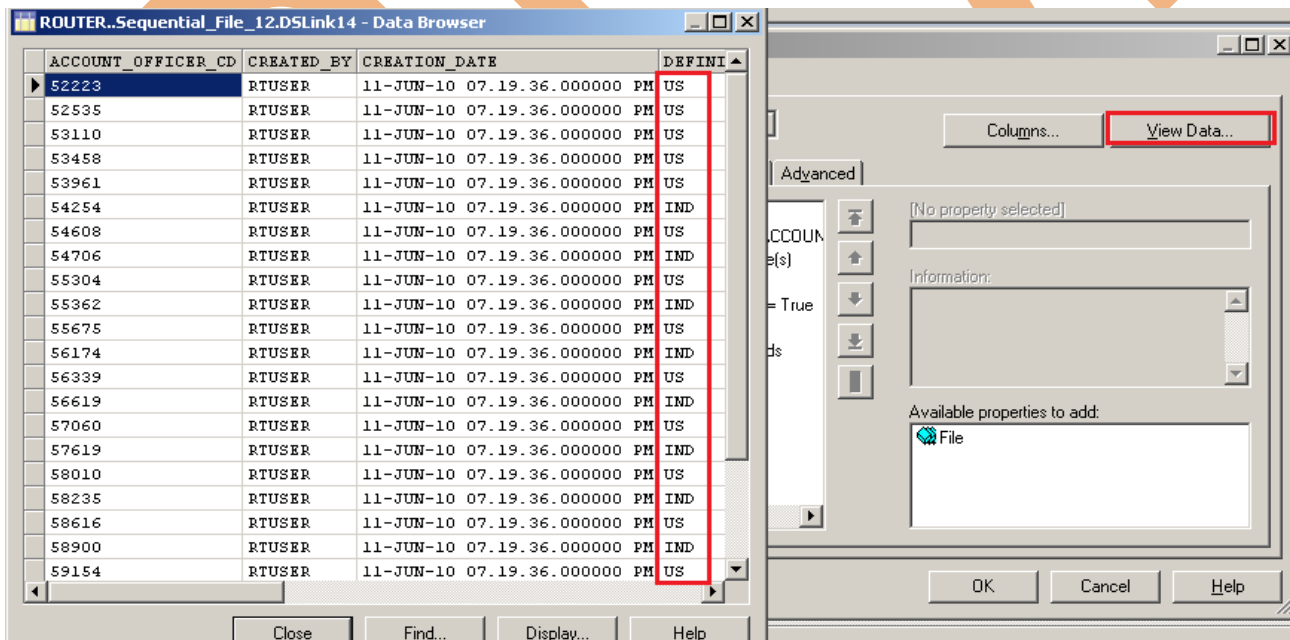

Step-17:- After choosing file, select 'True' from First line in Column Names dropdown for removing first line from Input file.
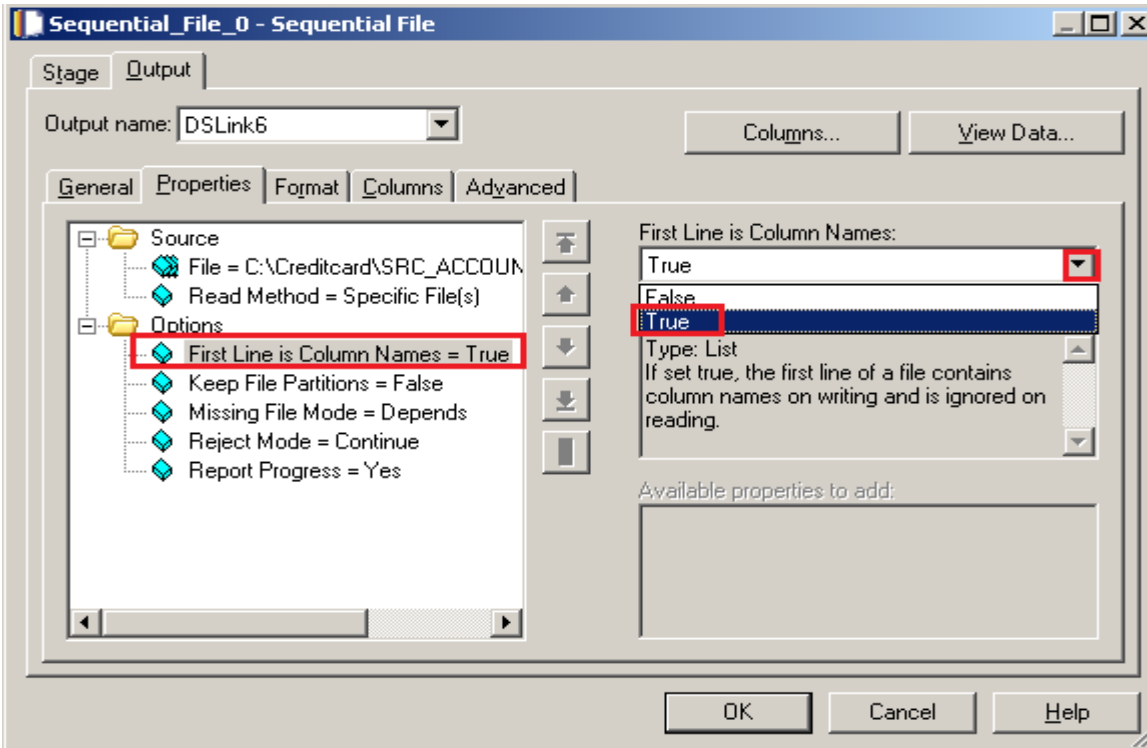
Step-18 :- Make input file and remember column names.

```
ACCOUNT_OFFICER_CD,CREATED_BY,CREATION_DATE,DEFINITION_LANGUAGE,ENABLED_FLAG,LAST_MODIFIED_BY,LAST_MODIFIED_DATE,LEAF_ONLY_FLAG,ACCOUNT_OFFICER_DISPLAY_CD,
52223,RTUSER,11-JUN-10 07.19.36.000000 PM,US,Y,RTUSER,11-JUN-10 07.19.36.000000 PM,Y,52223.36,
52535,RTUSER,11-JUN-10 07.19.36.000000 PM,US,Y,RTUSER,11-JUN-10 07.19.36.000000 PM,Y,52534.89,
53110,RTUSER,11-JUN-10 07.19.36.000000 PM,US,Y,RTUSER,11-JUN-10 07.19.36.000000 PM,Y,53110.24,
53458,RTUSER,11-JUN-10 07.19.36.000000 PM,US,Y,RTUSER,11-JUN-10 07.19.36.000000 PM,Y,53457.6,,
53961,RTUSER,11-JUN-10 07.19.36.000000 PM,US,Y,RTUSER,11-JUN-10 07.19.36.000000 PM,Y,53960.7,,
54254,RTUSER,11-JUN-10 07.19.36.000000 PM,IND,Y,RTUSER,11-JUN-10 07.19.36.000000 PM,Y,54253.88,
54608,RTUSER,11-JUN-10 07.19.36.000000 PM,US,Y,RTUSER,11-JUN-10 07.19.36.000000 PM,Y,54607.99,
54706,RTUSER,11-JUN-10 07.19.36.000000 PM,IND,Y,RTUSER,11-JUN-10 07.19.36.000000 PM,Y,54705.91,
55304,RTUSER,11-JUN-10 07.19.36.000000 PM,US,Y,RTUSER,11-JUN-10 07.19.36.000000 PM,Y,55303.81,
55362,RTUSER,11-JUN-10 07.19.36.000000 PM,IND,Y,RTUSER,11-JUN-10 07.19.36.000000 PM,Y,55362.01,
55675,RTUSER,11-JUN-10 07.19.36.000000 PM,US,Y,RTUSER,11-JUN-10 07.19.36.000000 PM,Y,55674.97,
56174,RTUSER,11-JUN-10 07.19.36.000000 PM,IND,Y,RTUSER,11-JUN-10 07.19.36.000000 PM,Y,56173.79,
56339,RTUSER,11-JUN-10 07.19.36.000000 PM,US,Y,RTUSER,11-JUN-10 07.19.36.000000 PM,Y,56338.61,
56619,RTUSER,11-JUN-10 07.19.36.000000 PM,IND,Y,RTUSER,11-JUN-10 07.19.36.000000 PM,Y,56618.92,
57060,RTUSER,11-JUN-10 07.19.36.000000 PM,US,Y,RTUSER,11-JUN-10 07.19.36.000000 PM,Y,57059.61,
57619,RTUSER,11-JUN-10 07.19.36.000000 PM,IND,Y,RTUSER,11-JUN-10 07.19.36.000000 PM,Y,57619.17,
58010,RTUSER,11-JUN-10 07.19.36.000000 PM,US,Y,RTUSER,11-JUN-10 07.19.36.000000 PM,Y,58010.18,
58235,RTUSER,11-JUN-10 07.19.36.000000 PM,IND,Y,RTUSER,11-JUN-10 07.19.36.000000 PM,Y,58235.35,
58616,RTUSER,11-JUN-10 07.19.36.000000 PM,US,Y,RTUSER,11-JUN-10 07.19.36.000000 PM,Y,58616.48,
58900,RTUSER,11-JUN-10 07.19.36.000000 PM,IND,Y,RTUSER,11-JUN-10 07.19.36.000000 PM,Y,58899.61,
59154,RTUSER,11-JUN-10 07.19.36.000000 PM,US,Y,RTUSER,11-JUN-10 07.19.36.000000 PM,Y,59153.7,,
59567,RTUSER,11-JUN-10 07.19.36.000000 PM,IND,Y,RTUSER,11-JUN-10 07.19.36.000000 PM,Y,59566.99,
60254,RTUSER,11-JUN-10 07.19.36.000000 PM,US,Y,RTUSER,11-JUN-10 07.19.36.000000 PM,Y,60254.29,
60740,RTUSER,11-JUN-10 07.19.36.000000 PM,US,Y,RTUSER,11-JUN-10 07.19.36.000000 PM,Y,60740.15,
60951,RTUSER,11-JUN-10 07.19.36.000000 PM,US,Y,RTUSER,11-JUN-10 07.19.36.000000 PM,Y,60951.47,
61134,RTUSER,11-JUN-10 07.19.36.000000 PM,US,Y,RTUSER,11-JUN-10 07.19.36.000000 PM,Y,61133.62,
```

Step-19:- Now double click on ChangeCapture stage and see their Change keys and Change Values are available. We have to put our Change key here and with this ChangeCapture analyses your input table and reference table. Change Values are used for check operation that means whatever values we specify ChangeCapture looks for any change in this.
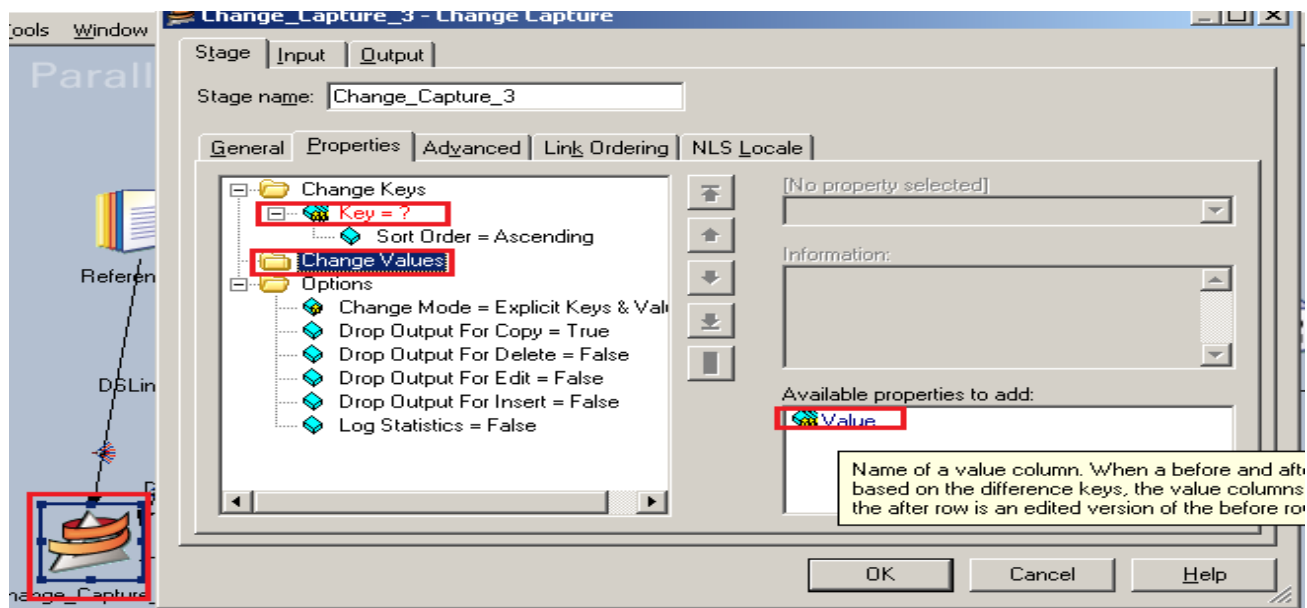
Step-20:- Now see here we choose ACCOUNT_OFFICER_CD key because this key is available in both the tables and is unique also .We choose DEFINITION_LANGUAGE as Change Value because we want to capture language based data changes i.e. when language changes or deletes then store the record.
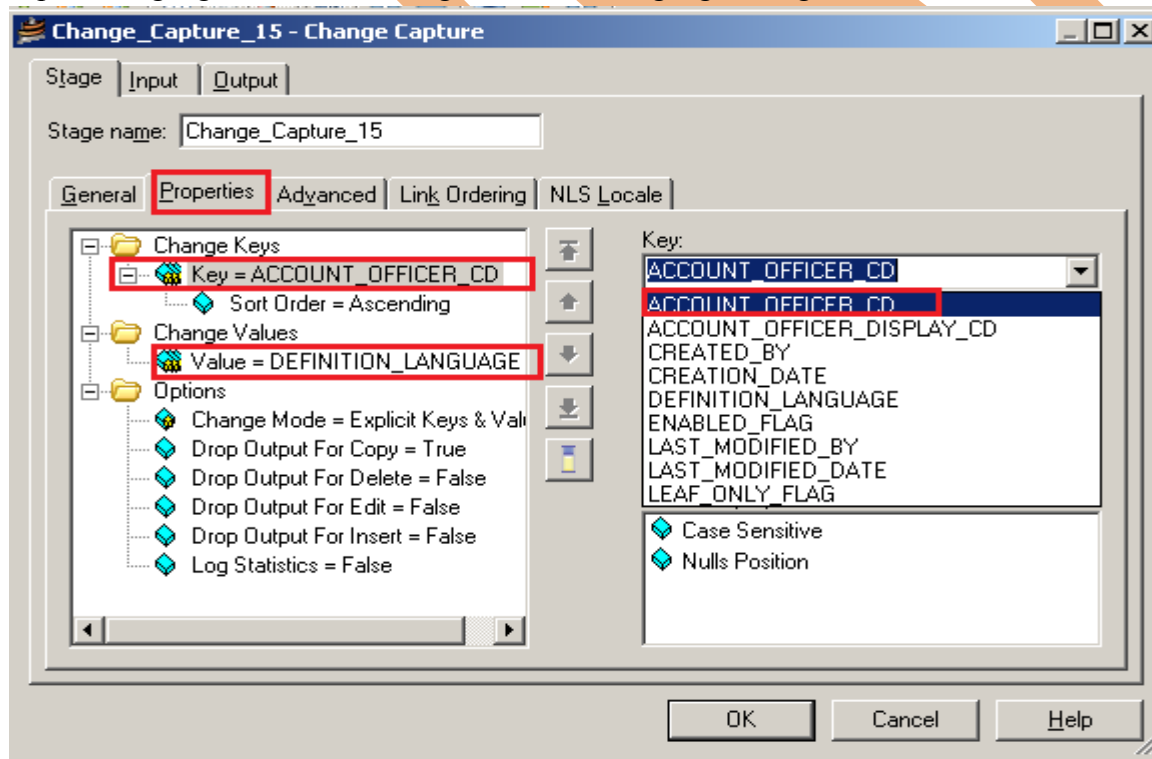

Step-21:- In Options section select the below properties to add as want to show the type of operation performed in Code Column Name and these codes have default values associated that will show in output file.
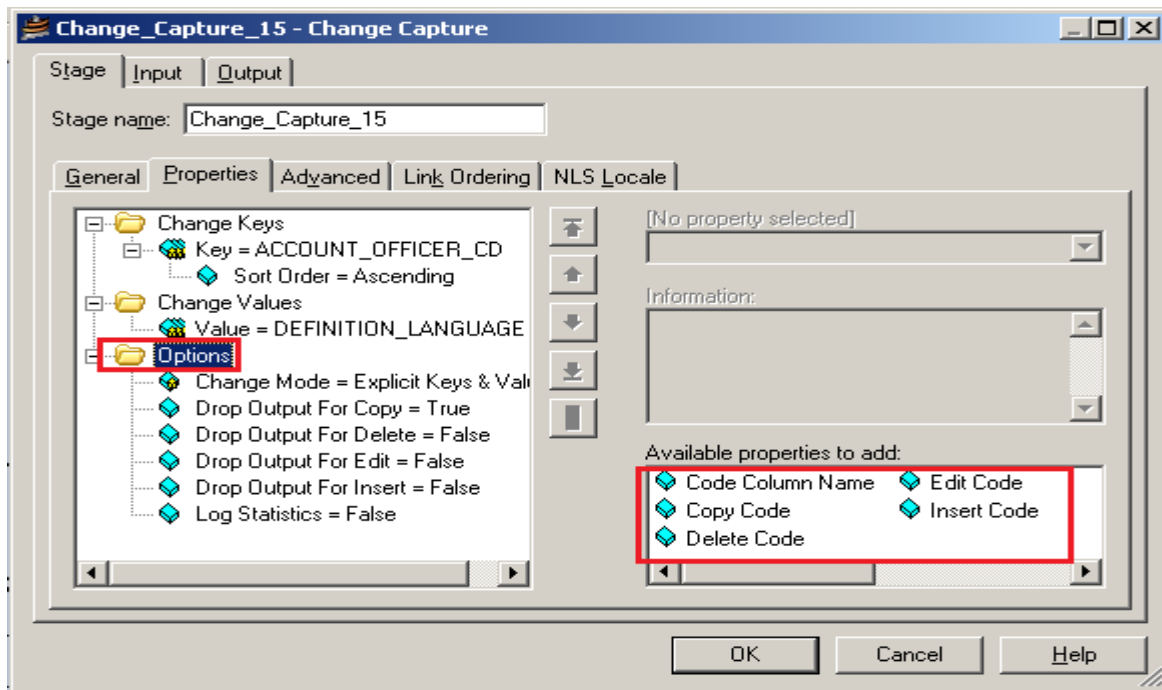
Step-22:- Now click on option and click Change Mode and select Explicit keys & values as Change Mode.
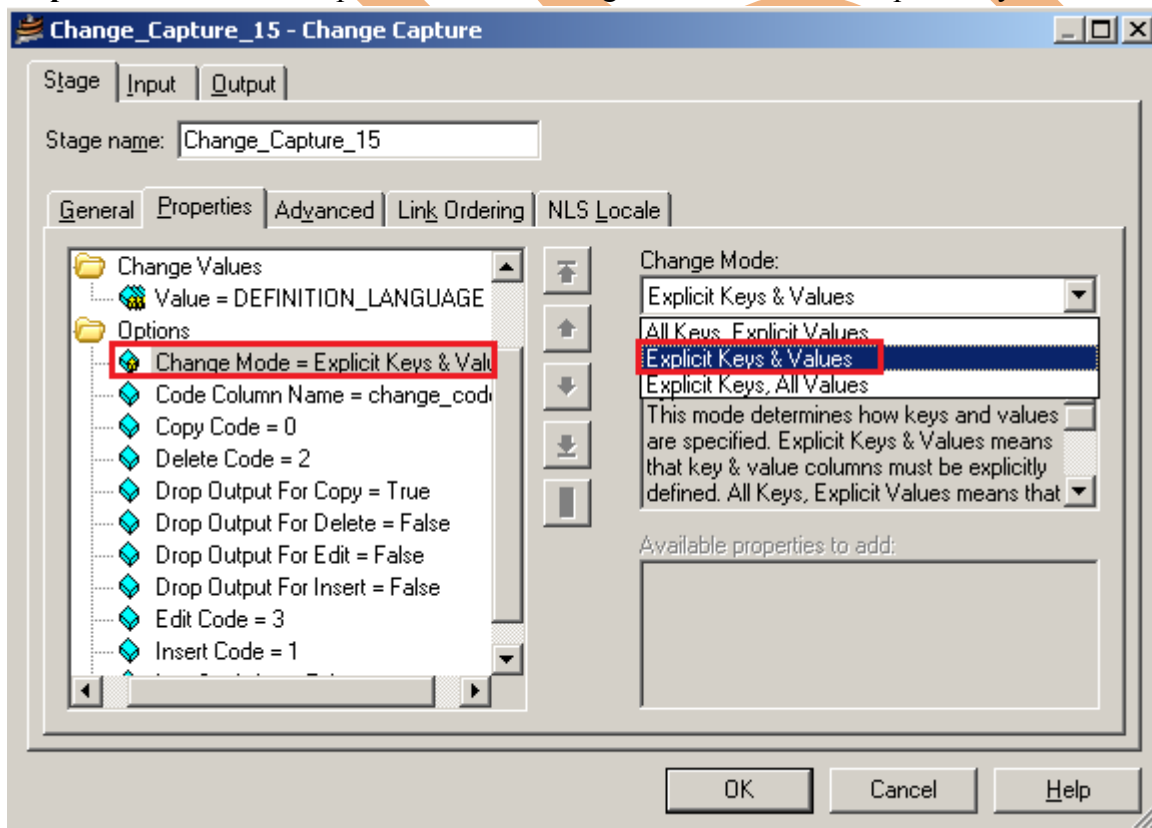
Step-23:- These are default values which we can also change. These values in output file shows that which type of operation you have perform**.**
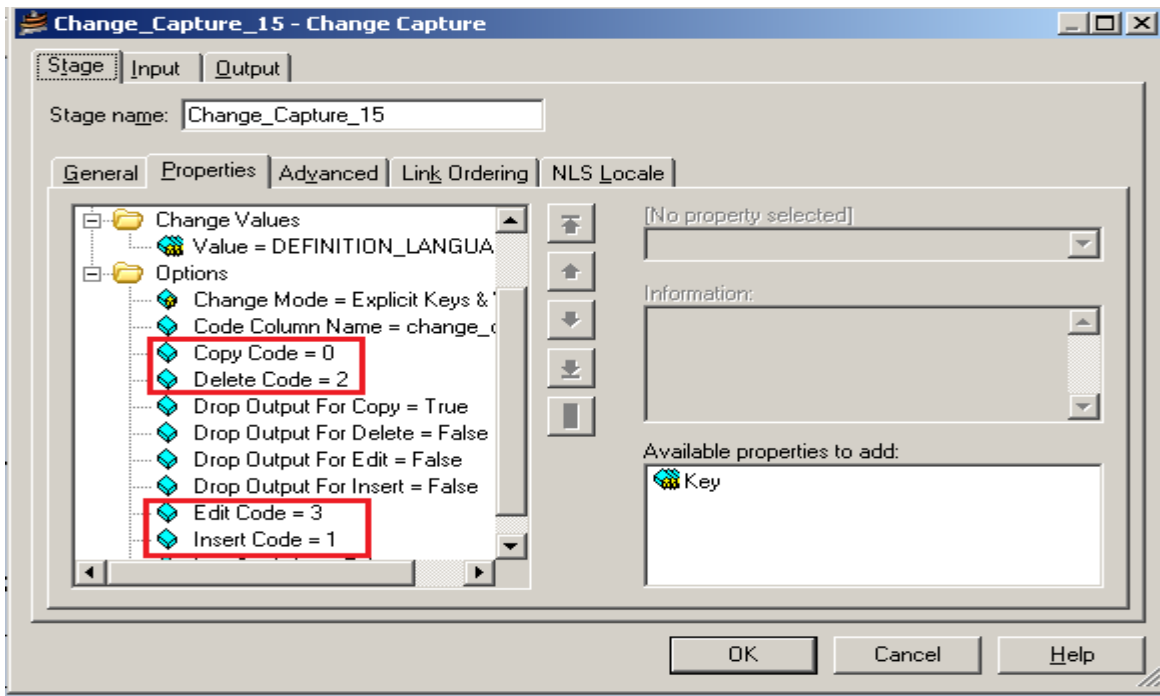
Step-24:- Go to Input tab and then choose Hash as Partitioning type. Remember always choose Hash if you use Key based transformation. When we select hash key then the sorting options will be visible.
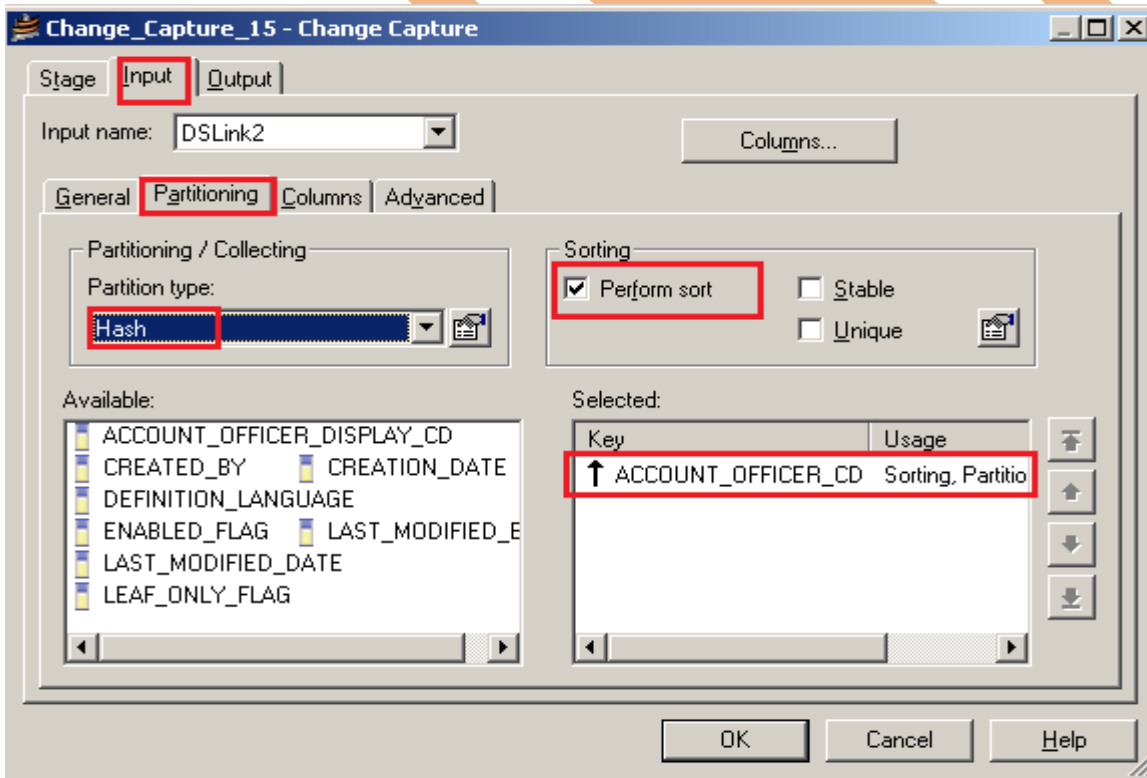
Step-25:- Apply same process as previous one because we have two input files.
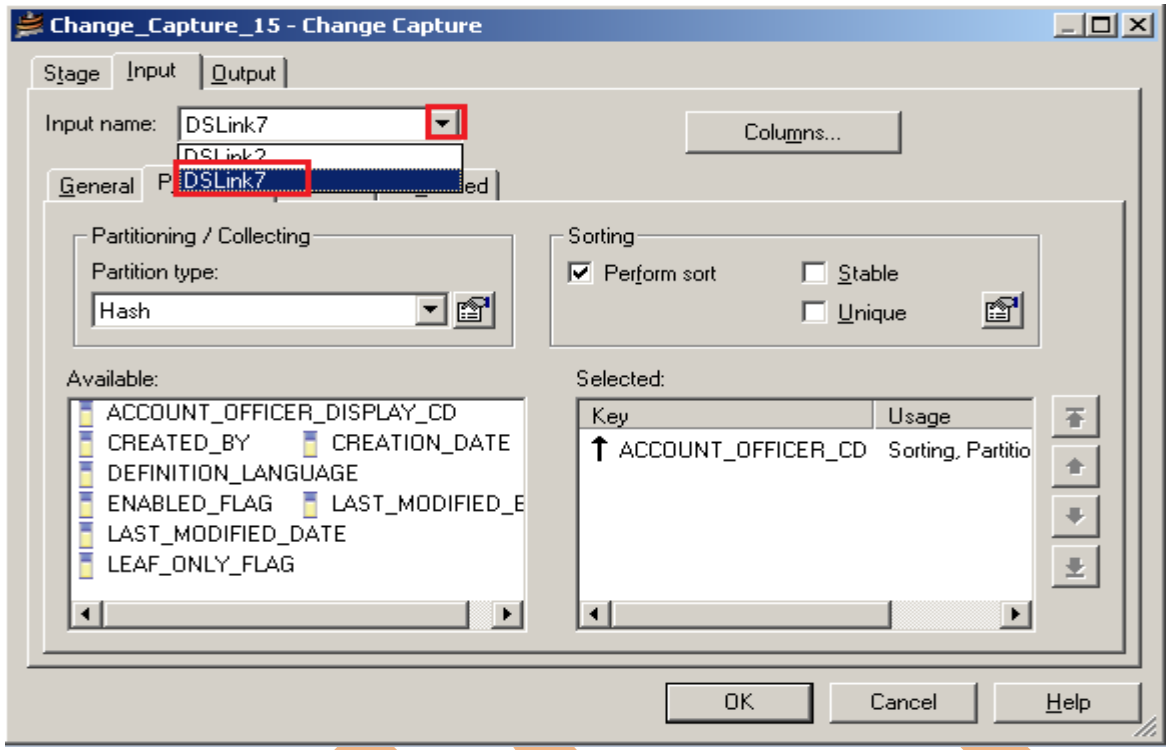
Step-26:- Now go to output tab then select all links and drag them into Out Link. Notice that there is one more column that is ChangeCode where our changing code value is stored.
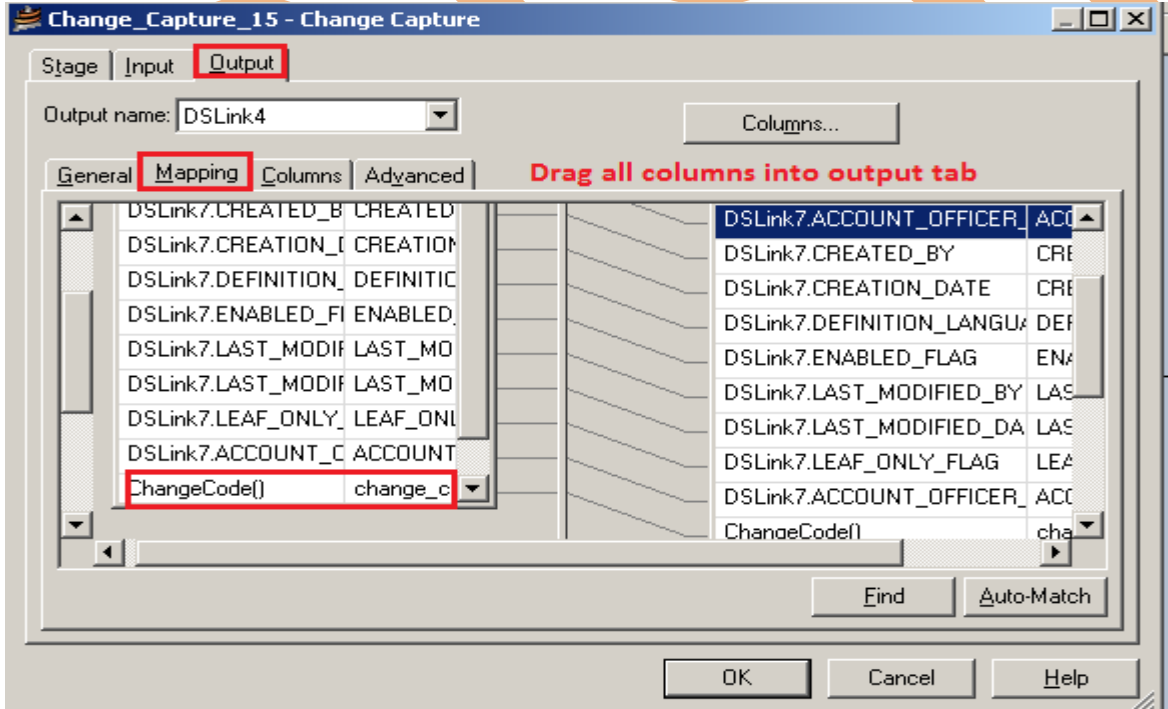
Step-27:- See here in output column there is one more column available and this column is useful to see the values of changing operation.
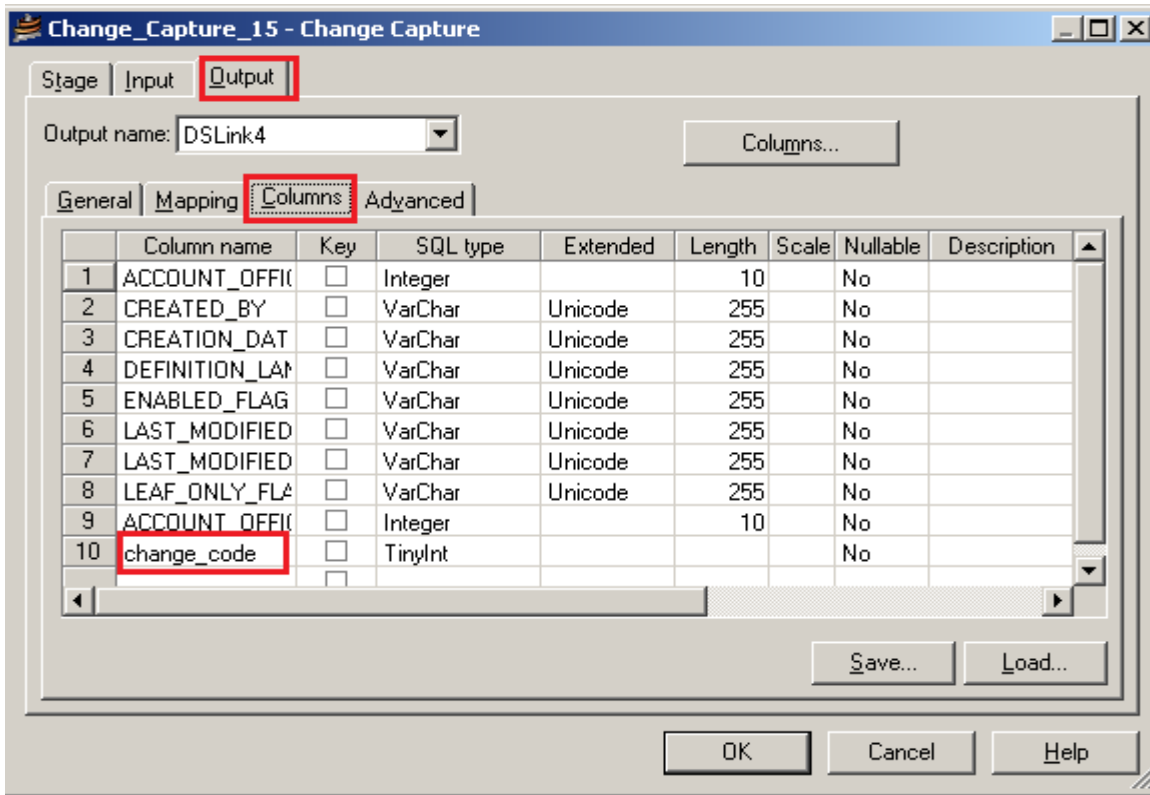
Step-28:- Click on compile button for compiling our job.
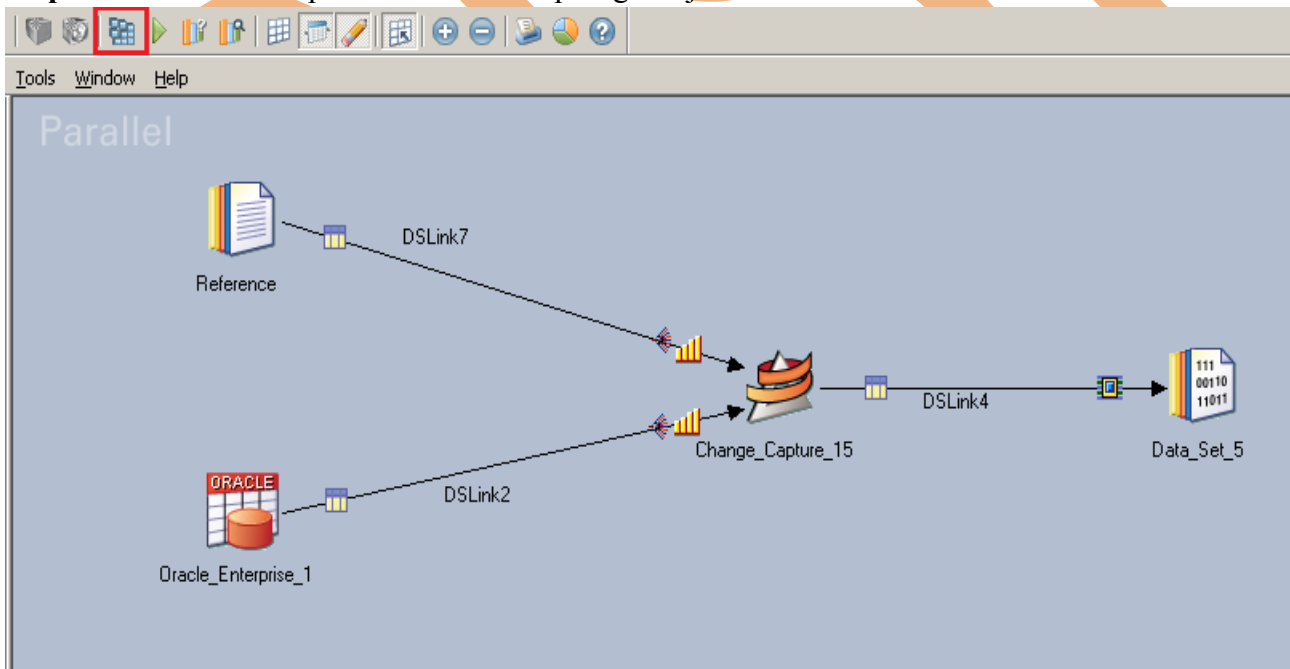
Step-29:- If this shows that means your compilation is done otherwise it shows error and now click on run button or (CTRL+F5).
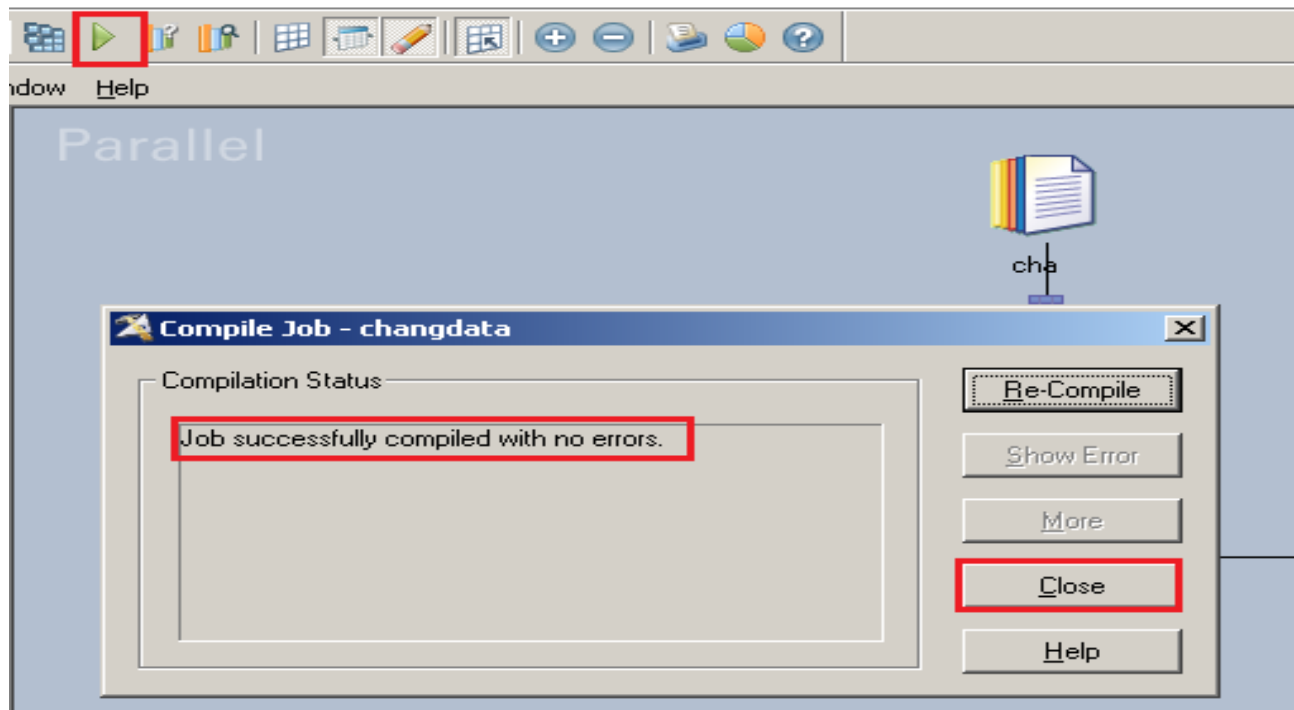
Step-30:- After click on run wait for a while.If it shows GREEN line that means your tranformation is successfully done otherwise if it shows RED Line that means not Done and BLUE Line means Under Process.
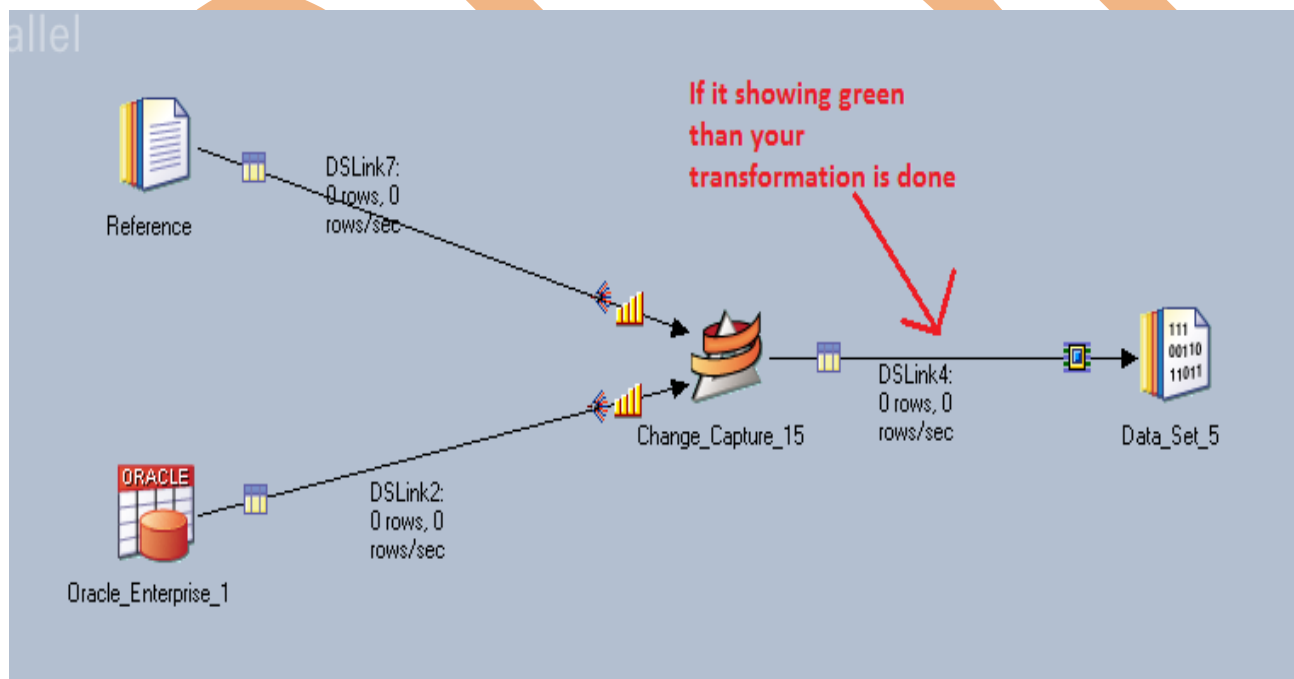
Step-31:- Now we want to open our dataset output file but it shows like this because dataset file doesn't open on notepad because extension is .DS . So we have to run this file in DataStage platform.

Step-32:- In DataStage designer go to Tools menu and choose Data Set Management and wait for a while. After some time a dialog box opens then specify your DS file and open.
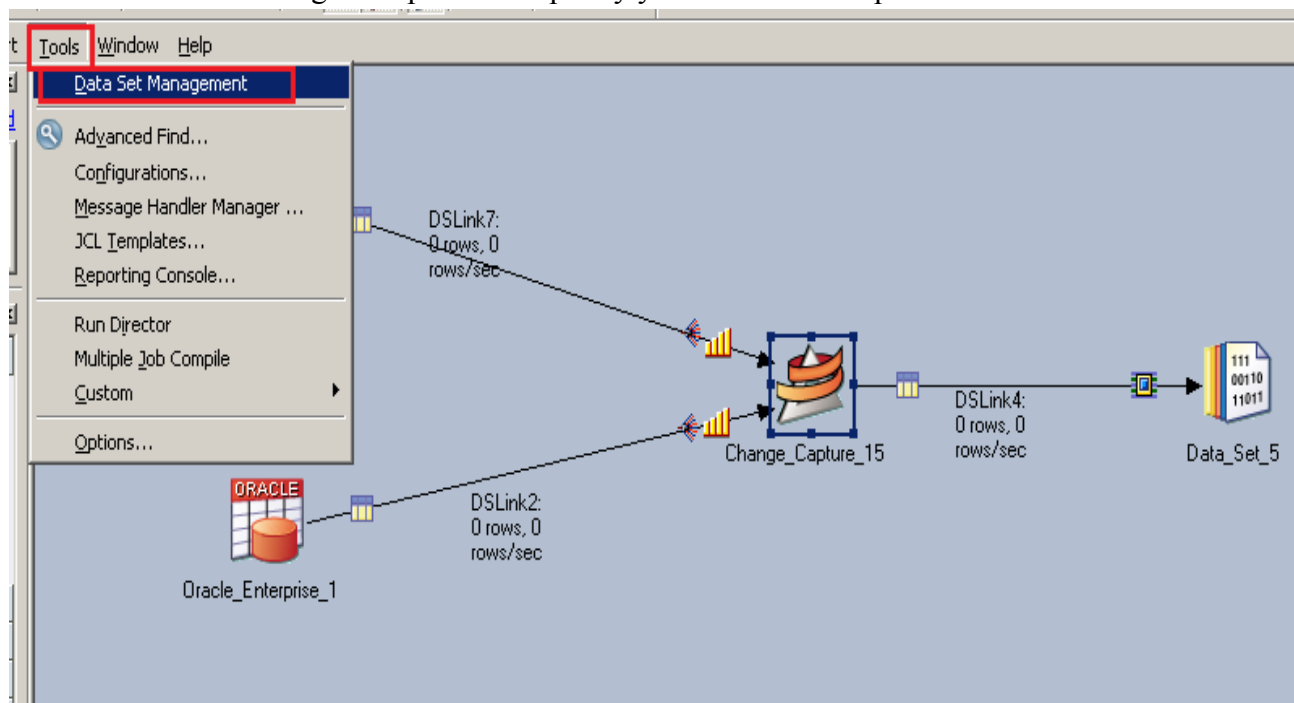
Step-33:- This is dialog box which opens when we click on DataSetManagement and choose ds file.
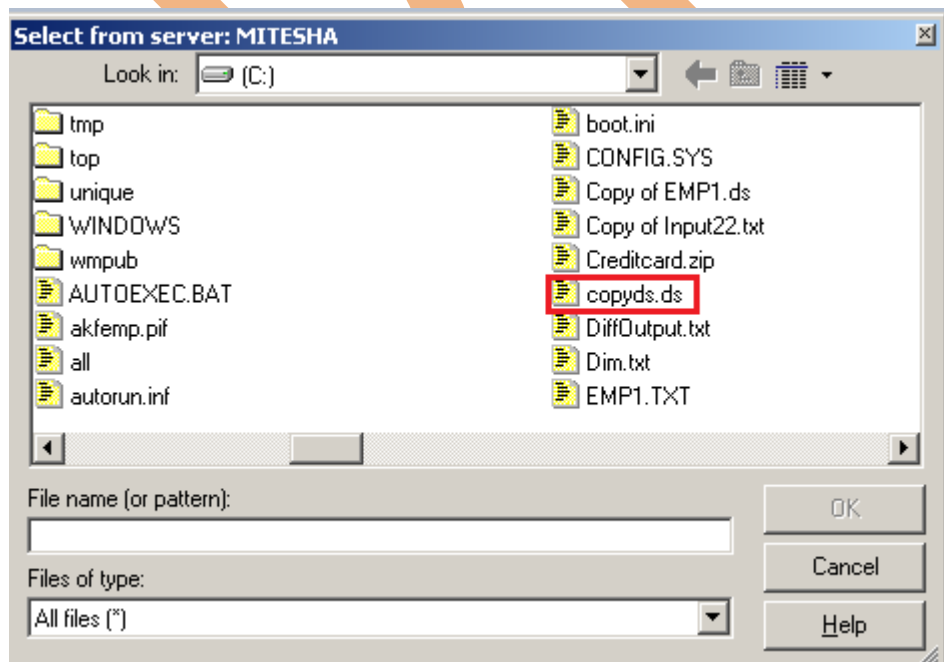
Step-34:- Click on Show button this will show the result and shows which type of operation you perform.