Business Intelligence Solution Providers

Specialized in creating talent resource pool

Lookup Transformation in IBM DataStage 8.5 - Lab#12

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 contains step by step process for storing data into SQL Server when we have two flat files and want the data merge based on account number and store in target. Join our professional training program and learn from experts.

History:

Version	Description Change	Author	Publish Date
0.1	Initial Draft	Upendra Upadhyay	12th Aug 2011
0.1	Review#1	Amit Sharma	18th Aug 2011

Mapping Sheet

Look Up Records.
Source- Flat File, Flat File
Target- SQL SERVER

Select * from SRC_ACCOUNT_OFFICER_CD A join
SRC_ACCOUNT_OFFICER_CD_LOOK B
**on A. ACCOUNT_OFFICER_CD =B. ACCOUNT_OFFICER_CD;**

Transformation	LookUp Transformation							
Source Name	Flat File, Flat File							
Target Name	SQL SERVER							
Flat File Source Details			Flat File Source Details			SQL SERVER Details		
Entity Name	Data Types	Is Null	Entity Name	Data Types	Aggregate Expression	Destination Entity Name	DestinationField Name	Is Null
	NUMBER(5)	N	ACCOUNT_OFFICER_CD	NUMBER(10)		SA	ACCOUNT_OFFICER_CD	N
CREATED_BY	VARCHAR2(30)	y	ADDRESS	VARCHAR2(10)		SA	CREATED_BY	y
CREATION_DATE	TIMESTAMP(6)	y			Merge two files based on account number and goes to sql server	SA	CREATION_DATE	y
DEFINITION_LANGUAGE	VARCHAR2(4)	y				SA	DEFINITION_LANGUAGE	y
ENABLED_FLAG	VARCHAR2(1)	y				SA	ENABLED_FLAG	y
LAST_MODIFIED_BY	VARCHAR2(30)	y				SA	LAST_MODIFIED_BY	y
LAST_MODIFIED_DATE	TIMESTAMP(6)	y				SA	LAST_MODIFIED_DATE	y
LEAF_ONLY_FLAG	VARCHAR2(1)	y				SA	LEAF_ONLY_FLAG	y
ACCOUNT_OFFICER_DISPLAY_CD	VARCHAR2(10)	y				SA	ACCOUNT_OFFICER_DISPLAY	y
						SA	ADDRESS	y

Steps to perform LookUp Transformation.
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.
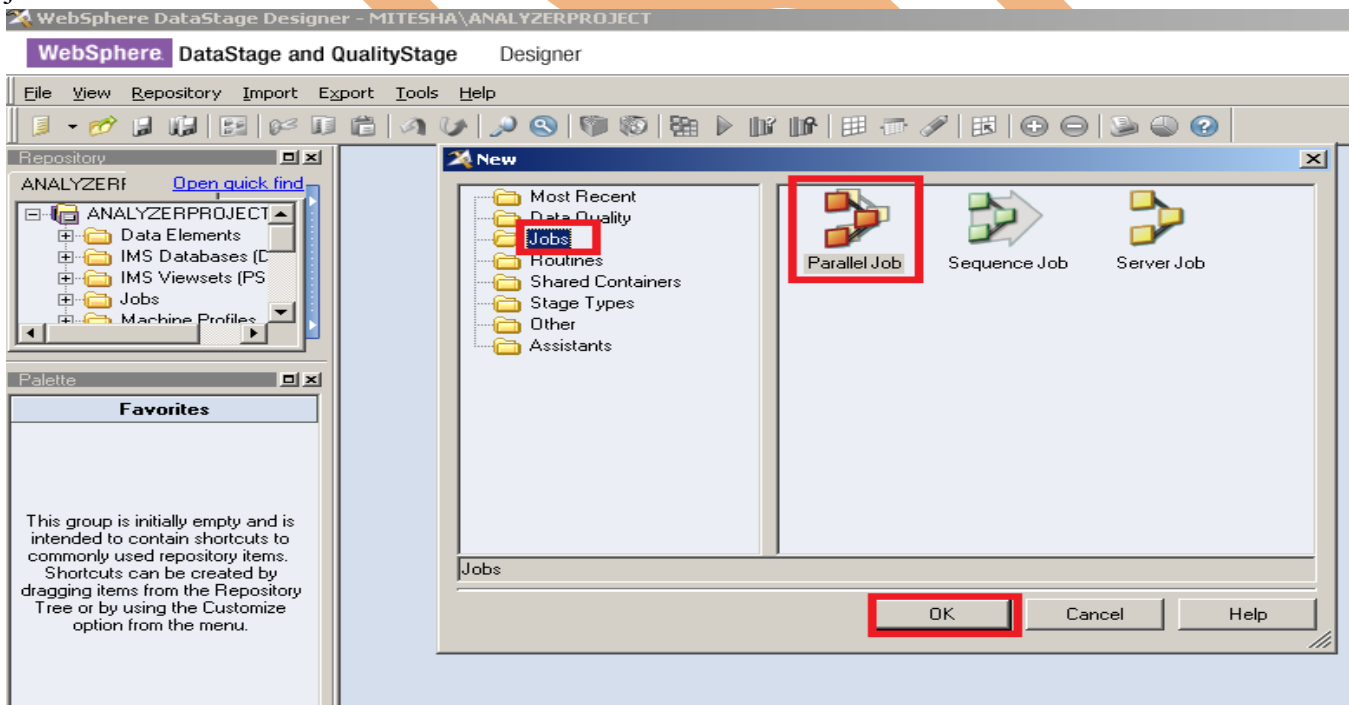

Step-3 :- Go to File Palette then scroll the cursor and select Sequential File and drag it to Parallel Job and see these are 3 files, one for input data and another for reference data for lookup and third one is for rejected data when lookup reject data then goes to reject file.
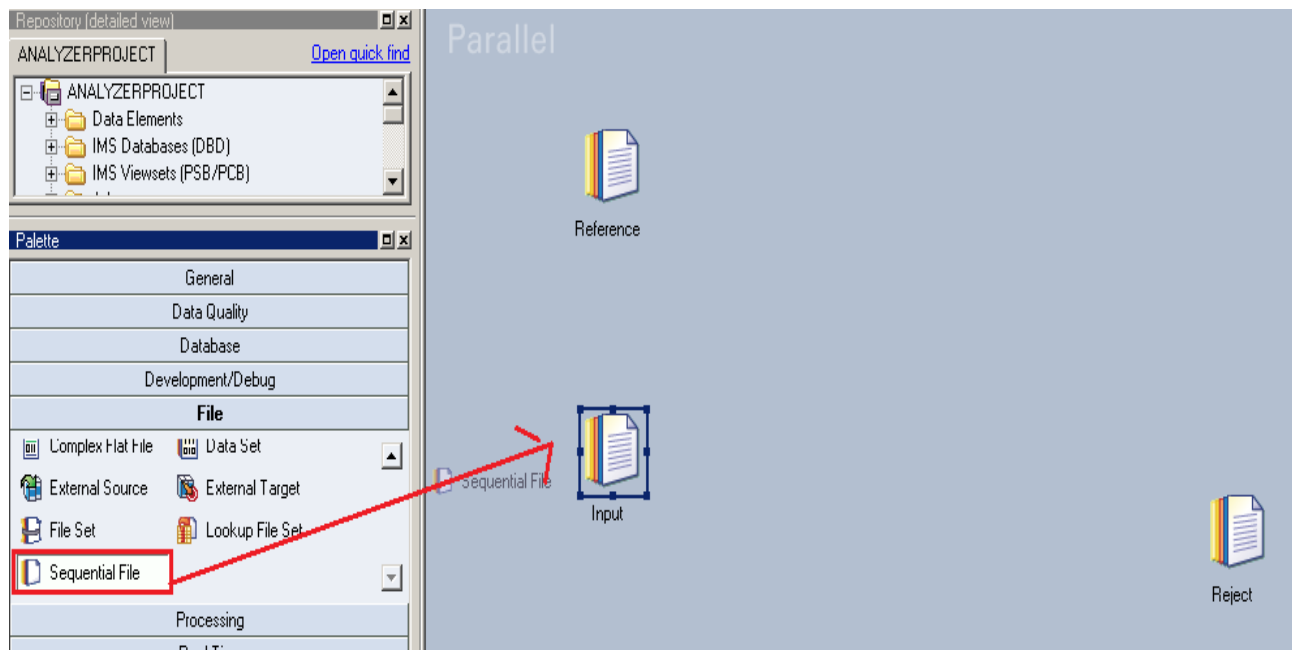
Step-4:- Choose SQL Server Enterprise processing method from Database Palette and drag it to parallel job, this is used for loading data into Sql server from flat files.
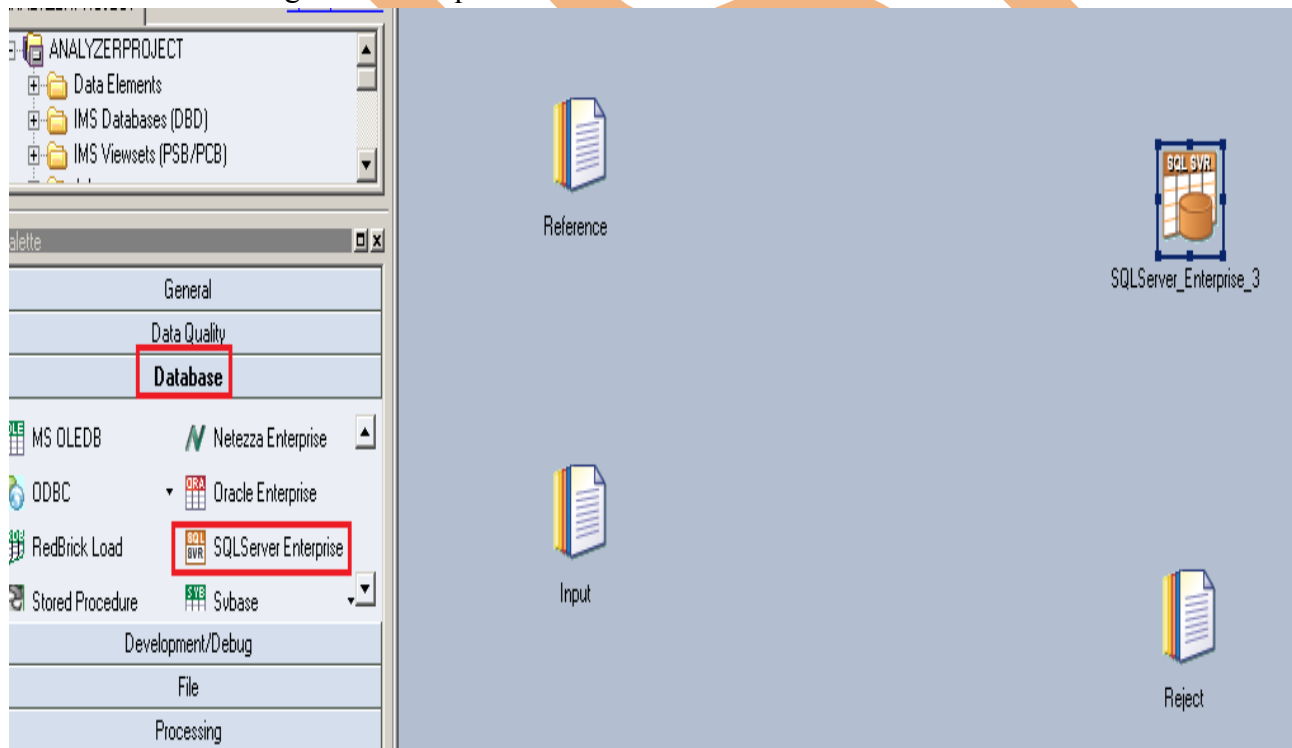
Step-5 :- Choose Lookup from Processing Palette and drag and drop it to parallel window. This is used to merge data from two source based on any common column, the rejected data will go to separate file.
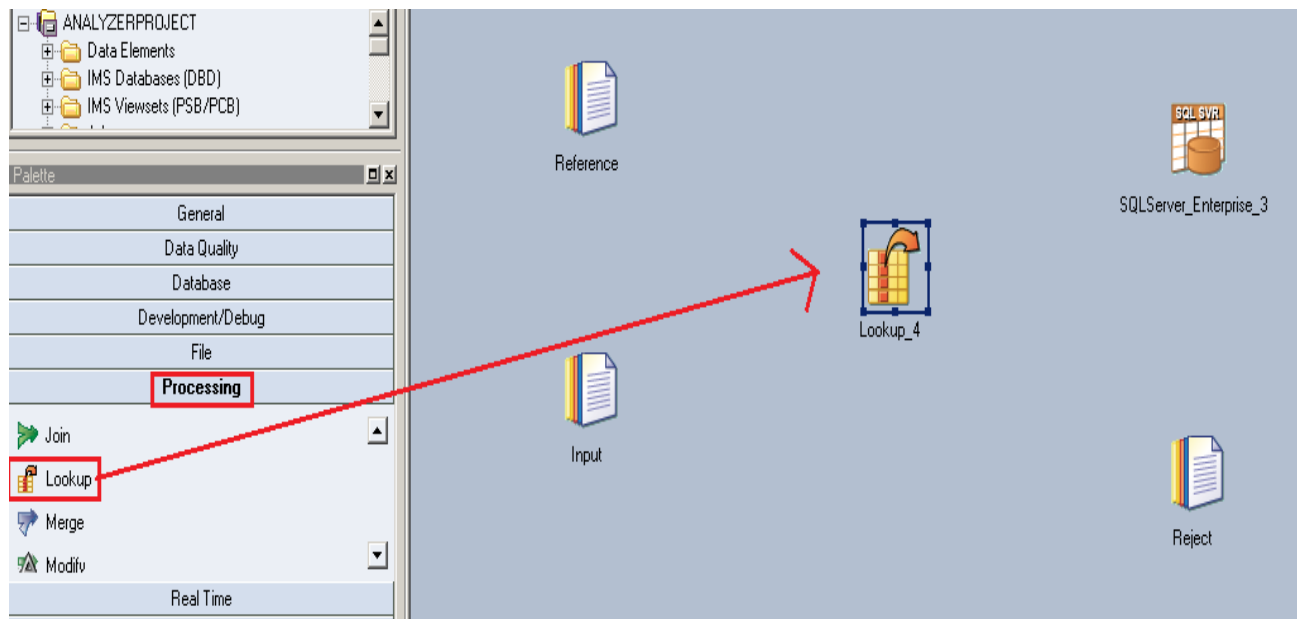
Step-6 : This shows the simple link and Reference link. Second sequential file is called reference file used for lookup every time and dashed line is called reject link and the rejected data goes to Rejected file.
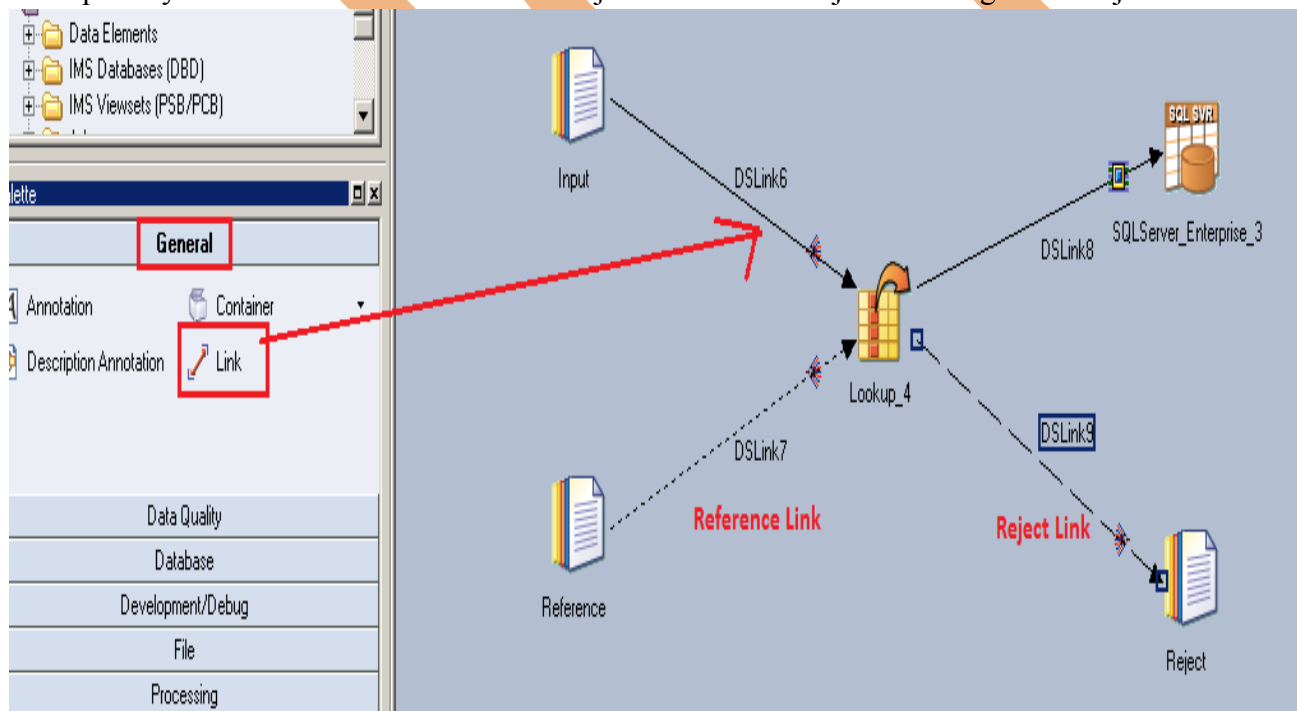
Step-7:- Now first before loading data into output, we have to create a table in sql server. So, we write a create table command and make sure to give same column name as you have in flat file. Goto SQL tab than write query and remember the username and password.

```
MITESHA.mast...QLQuery1.sql*    Table - dbo.Table_1    Summary

CREATE TABLE SRC_ACCOUNT_OFFICER_CD_LOOK
    (      ACCOUNT_OFFICER_CD NUMERIC(5,0) PRIMARY KEY,
    CREATED_BY VARCHAR(30),
    CREATION_DATE VARCHAR(100),
    DEFINITION_LANGUAGE VARCHAR2(4),
    ENABLED_FLAG VARCHAR(1),
    LAST_MODIFIED_BY VARCHAR(30),
    LAST_MODIFIED_DATE VARCHAR(100),
    LEAF_ONLY_FLAG VARCHAR(1),
    ACCOUNT_OFFICER_DISPLAY_CD VARCHAR2(10 BYTE) DEFAULT NULL,
                    ADDRESS VARCHAR(10)
    );
```

Step-8:- Double click on Input file and browse an input file that can be .txt,.csv any test file.
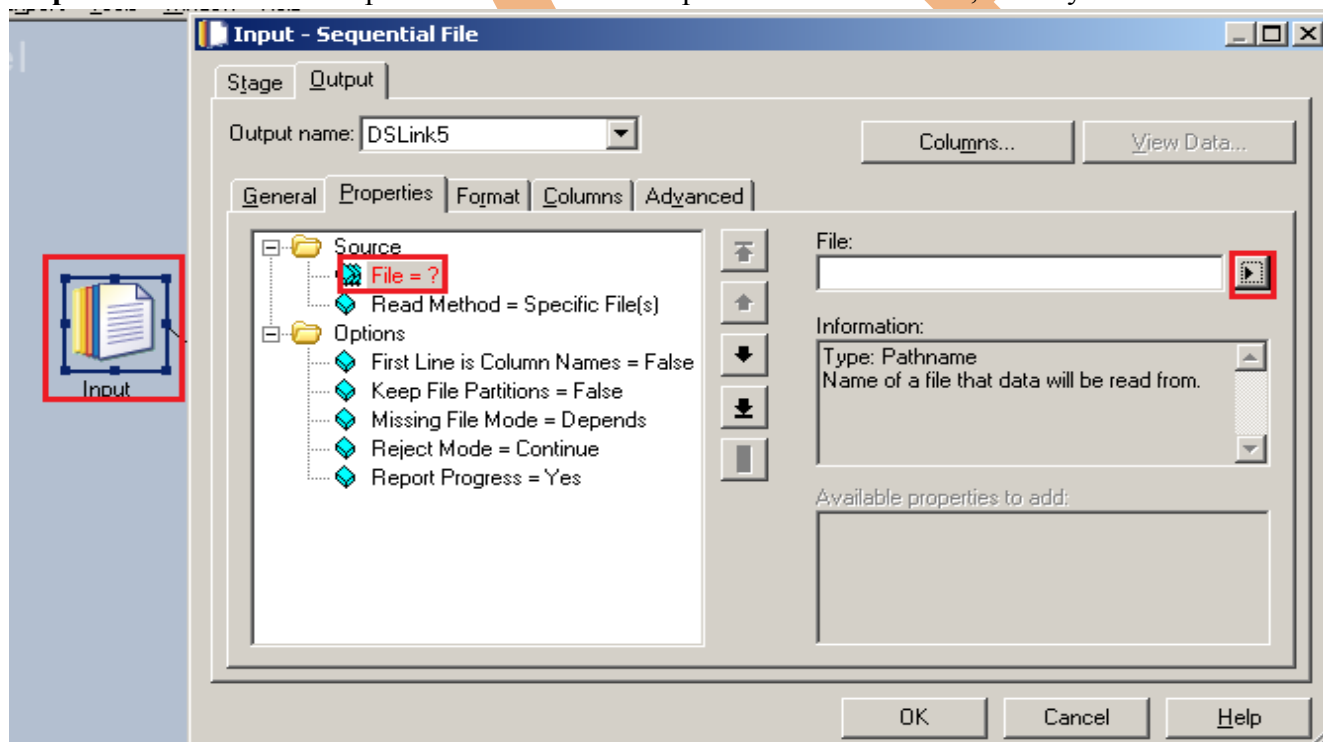
Step-9 :- Make input file and remember column names.

```
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,US,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,US,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,US,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,US,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,US,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,US,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,US,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,US,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,US,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-10 :- After choosing file, select 'True' from First line in column Names dropdown for removing first line from Input file. The first line is a header row.
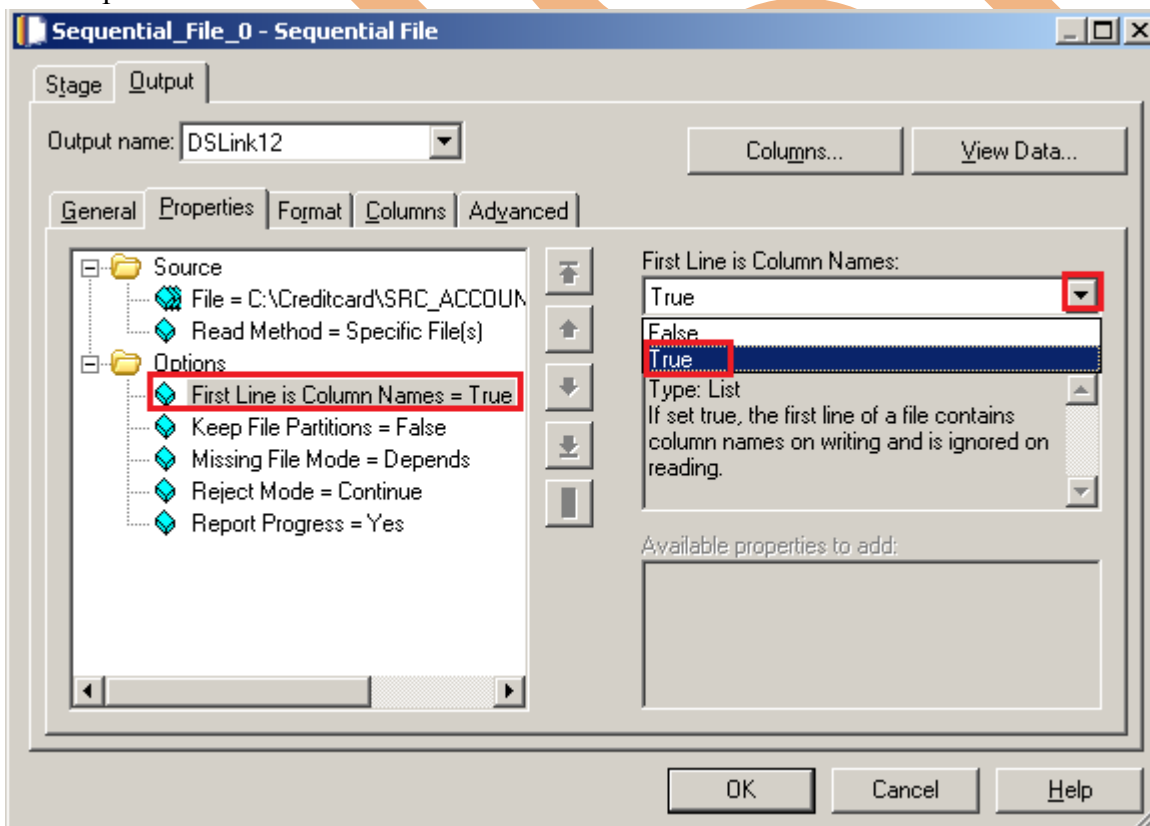
Step-11:- Go to Columns tab and enter same column names as declared in Input file and put datatype, length then click on View Data. This will show your Input Data.
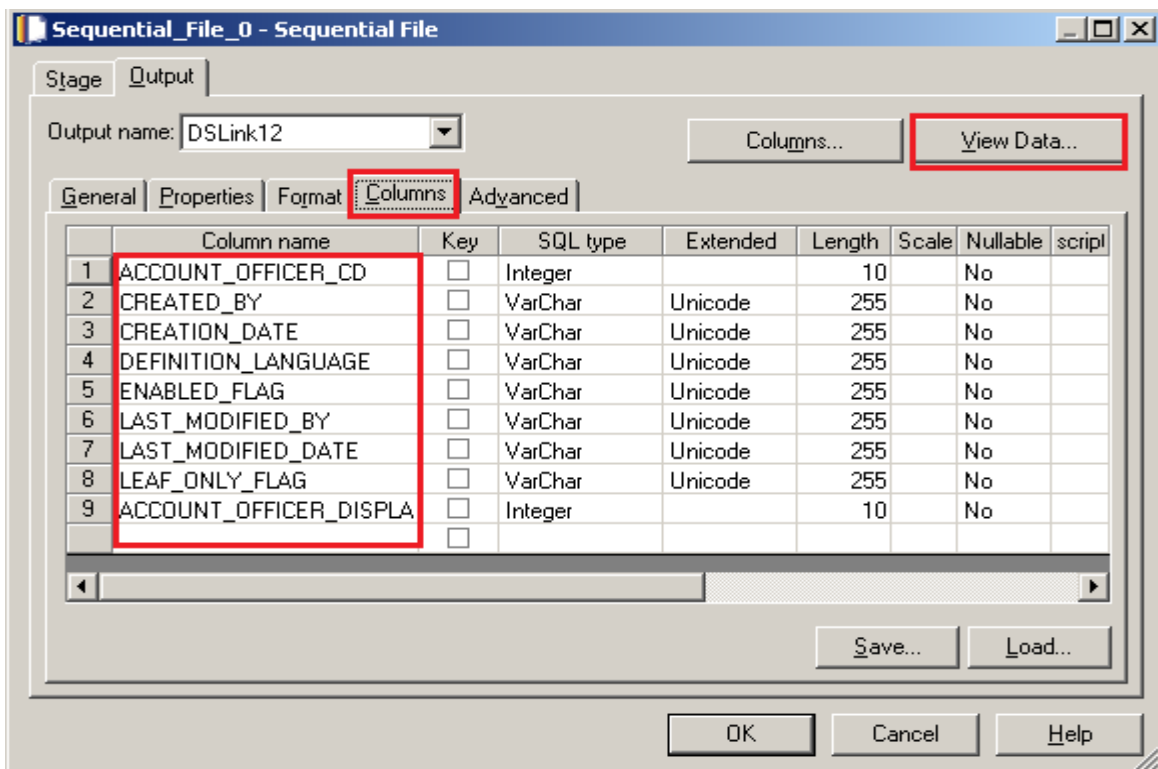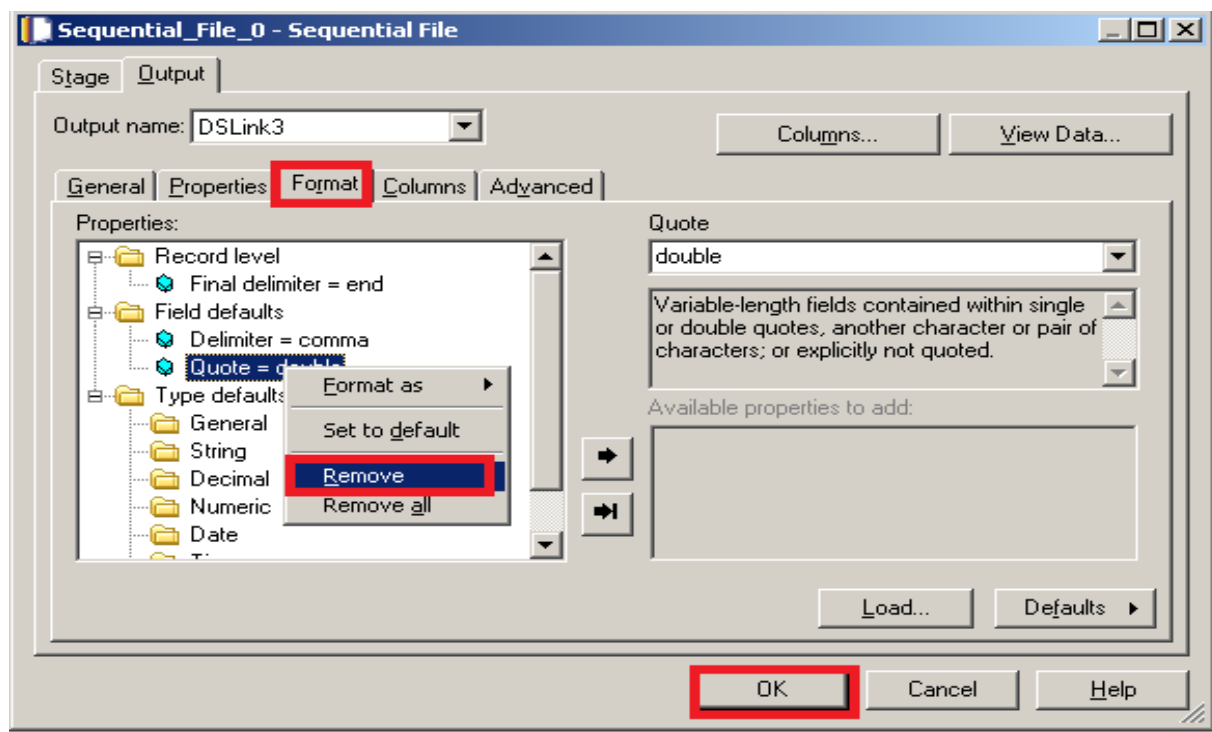
Step-12:- After Clicking on View Data if your data shows like this then our Data is valid and if some error comes then we have to rectify them and check our column names.
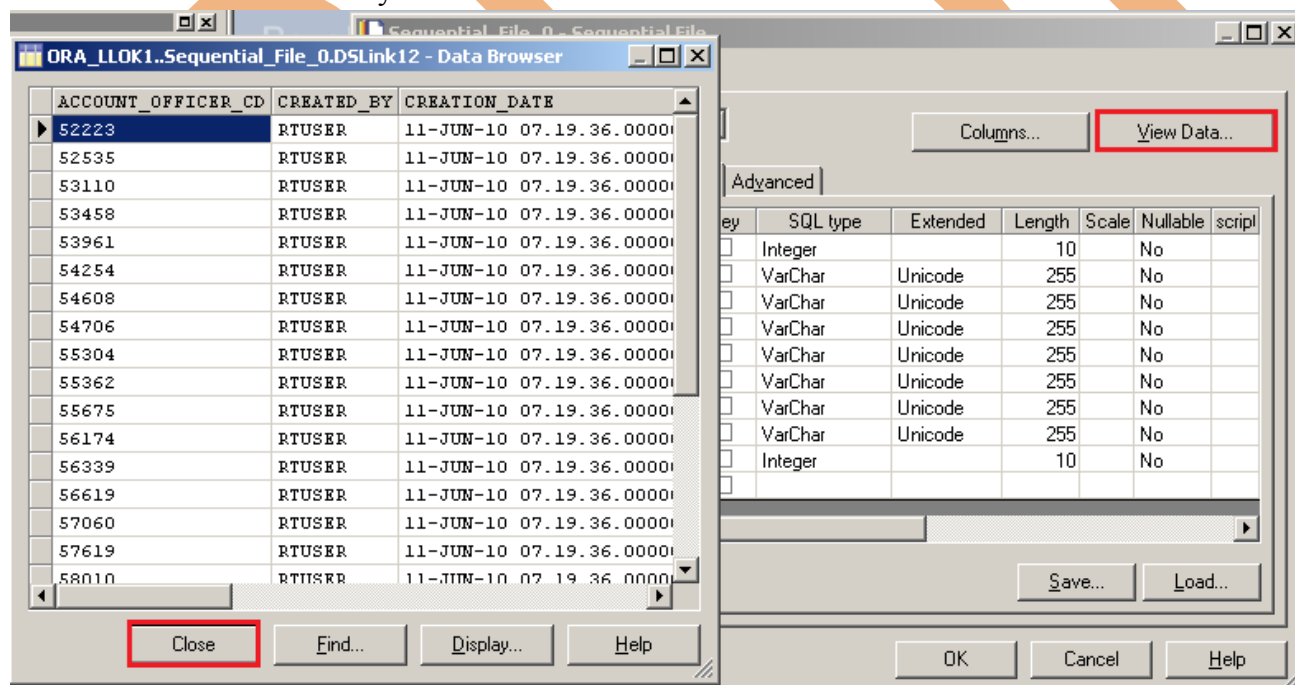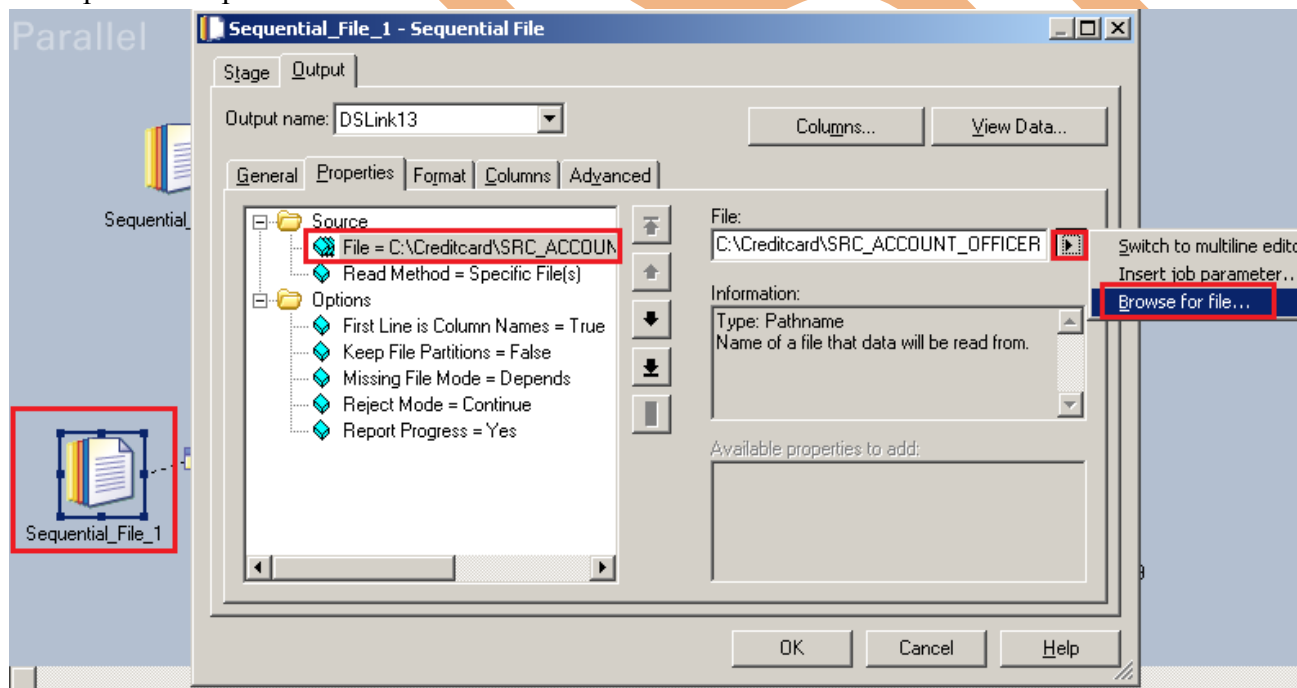

Step-13:- Go to Format tab and Remove double Quote because we don't need them, Click on OK.

Note:- If you use date as a input than specify here which type of format you used
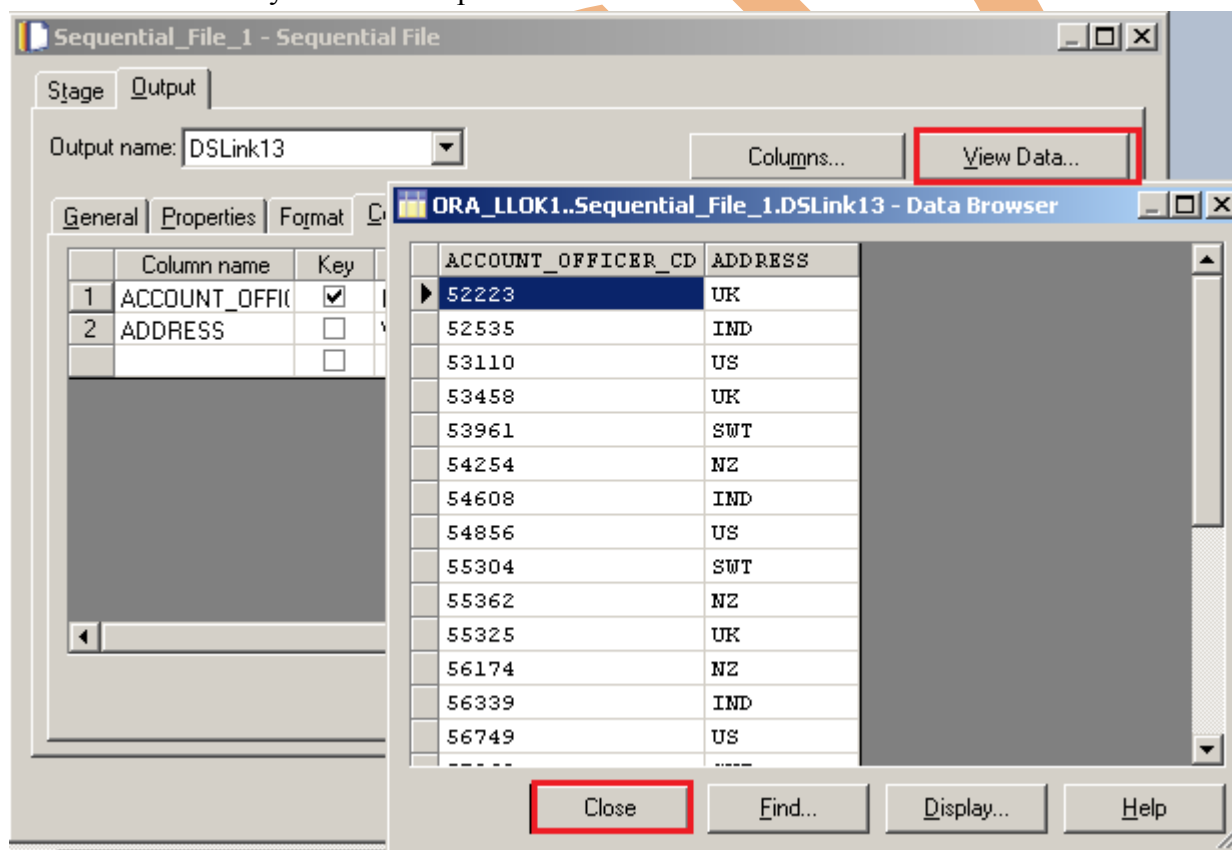
**Step-14** Now Click on Reference file same as Input sequential file and put the referential input file and apply same process as previous one.

Step-15:- This is the second sequential file. Input file and this file have same Account_Officer_cd for matching. If matched then it goes to output file otherwise goes into Rejected File.

```
ACCOUNT_OFFICER_CD,ADDRESS
52223,UK
52535,IND
53110,US
53458,UK
53961,SWT
54254,NZ
54608,IND
54856,US
55304,SWT
55362,NZ
55325,UK
56174,NZ
56339,IND
56749,US
57060,SWT
57619,NZ
58010,CANADA
58145,IND
58616,US
58900,SWT
50004,IND
59567,US
60254,IND
68960,CANADA
60951,UK
61134,US
```

Step-16:- Same as Input file after putting the data see view data, if it shows like this that means your data is correct and this is my referential input file.

Step-17:- Now click on Lookup, then click on constraints and drag them all to output link for making link this as Input file data.
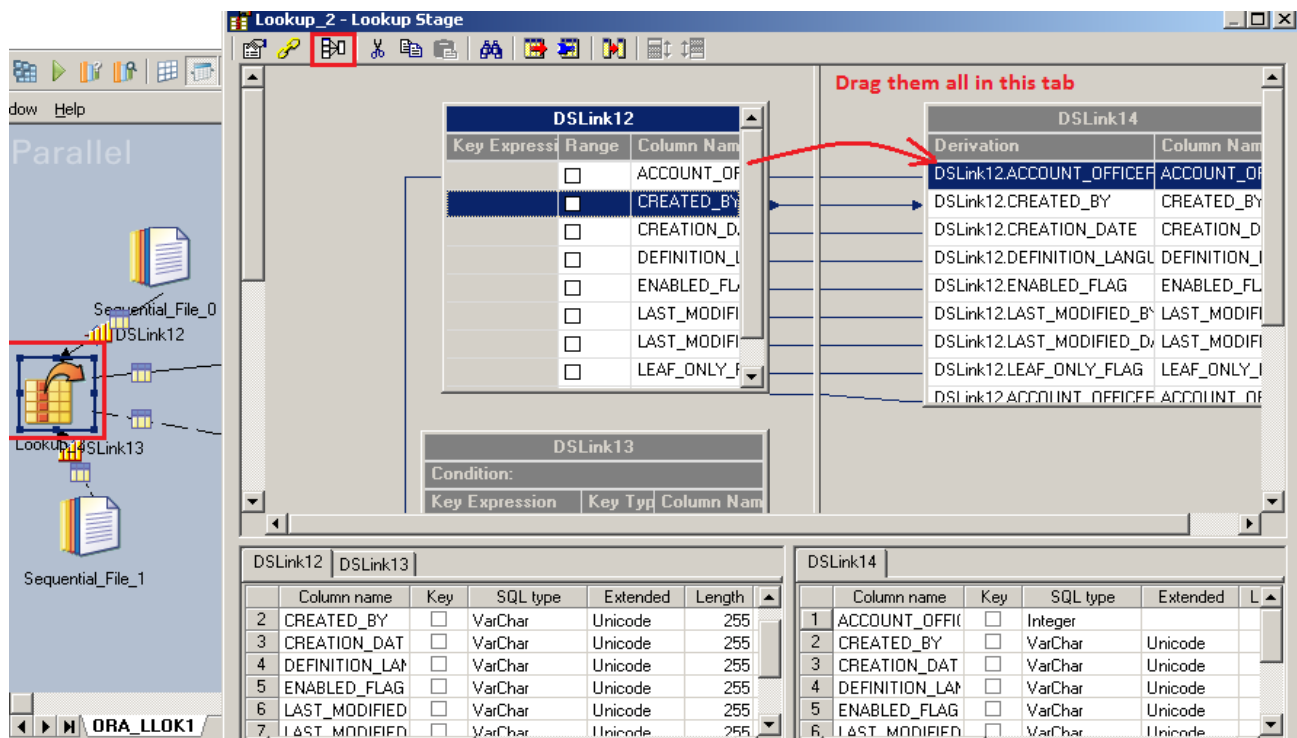
Step-18:- Now make connection between Input file and reference file. Drag Account_Officer_cd to reference Account_Officer_cd to make an Referential Integrity constraints and remaining columns drag to the output tab.
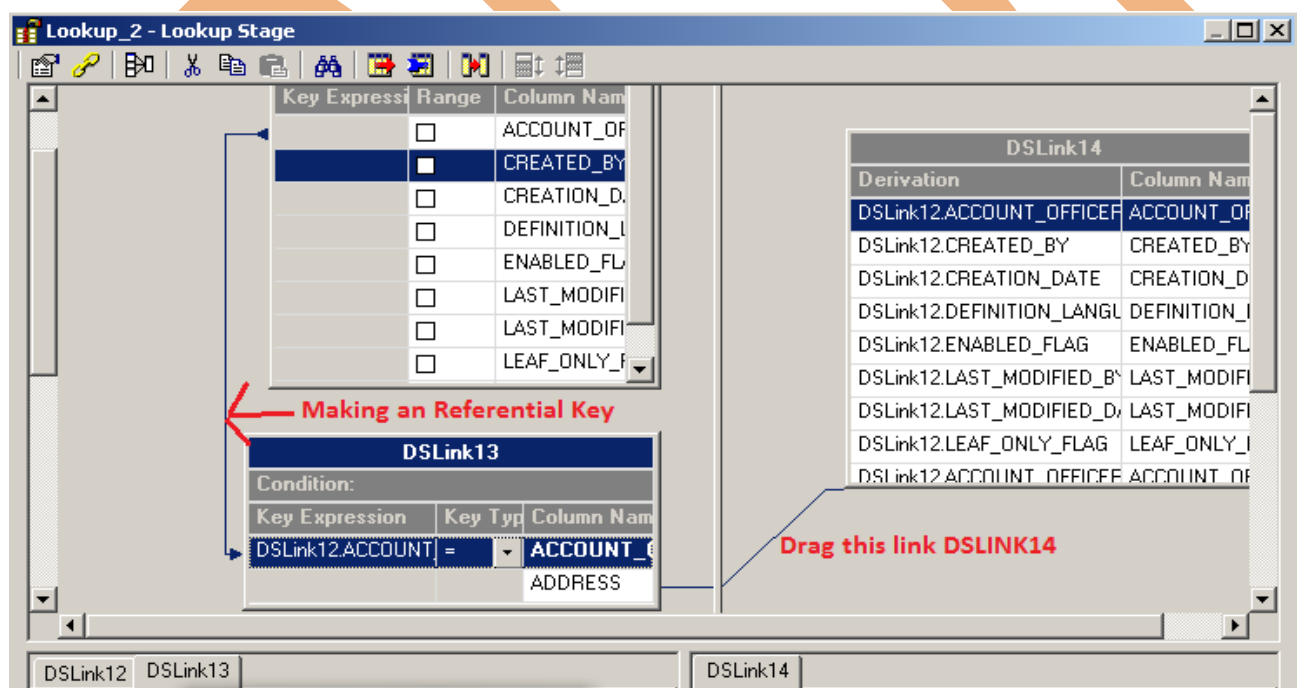
**Step-19:-** Select Reject from Lookup Failure then unmatched records will go to the rejected file
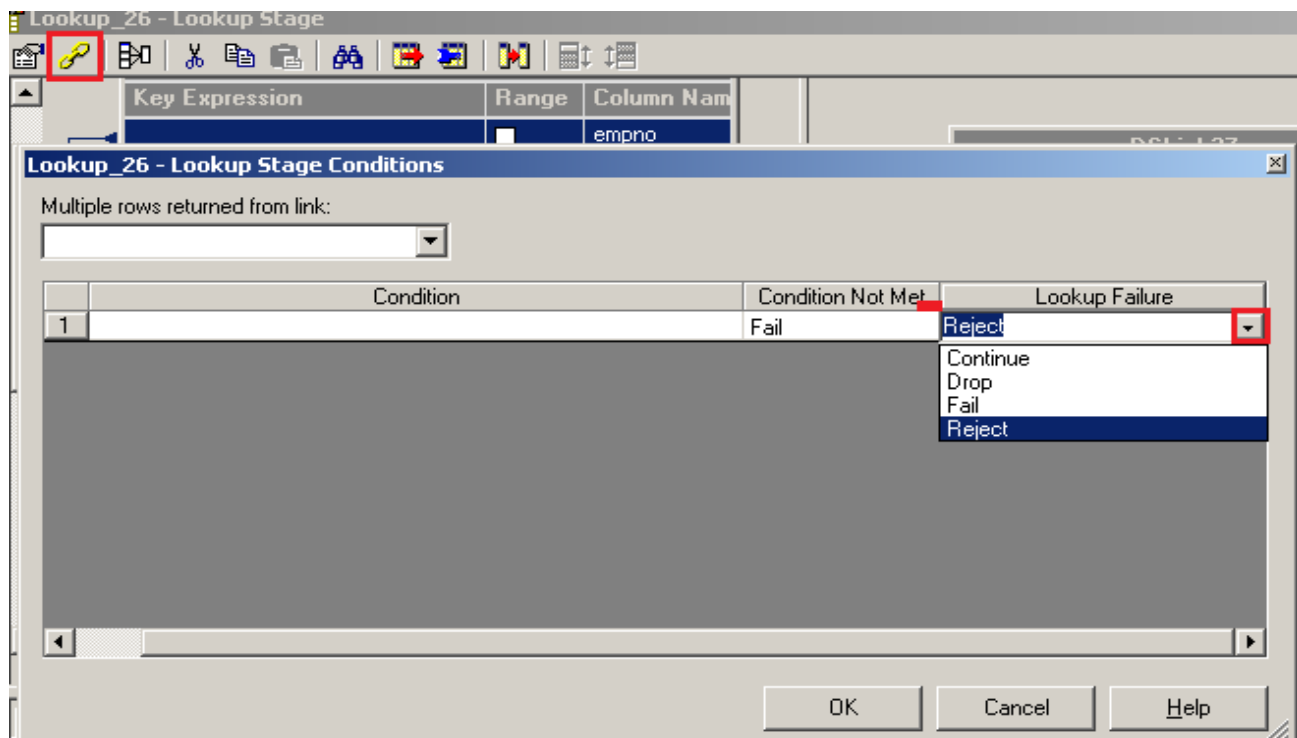
Step-20:-Now click to input and choose hash key partition because these are key based stage so we have to choose hash and make an unique key..
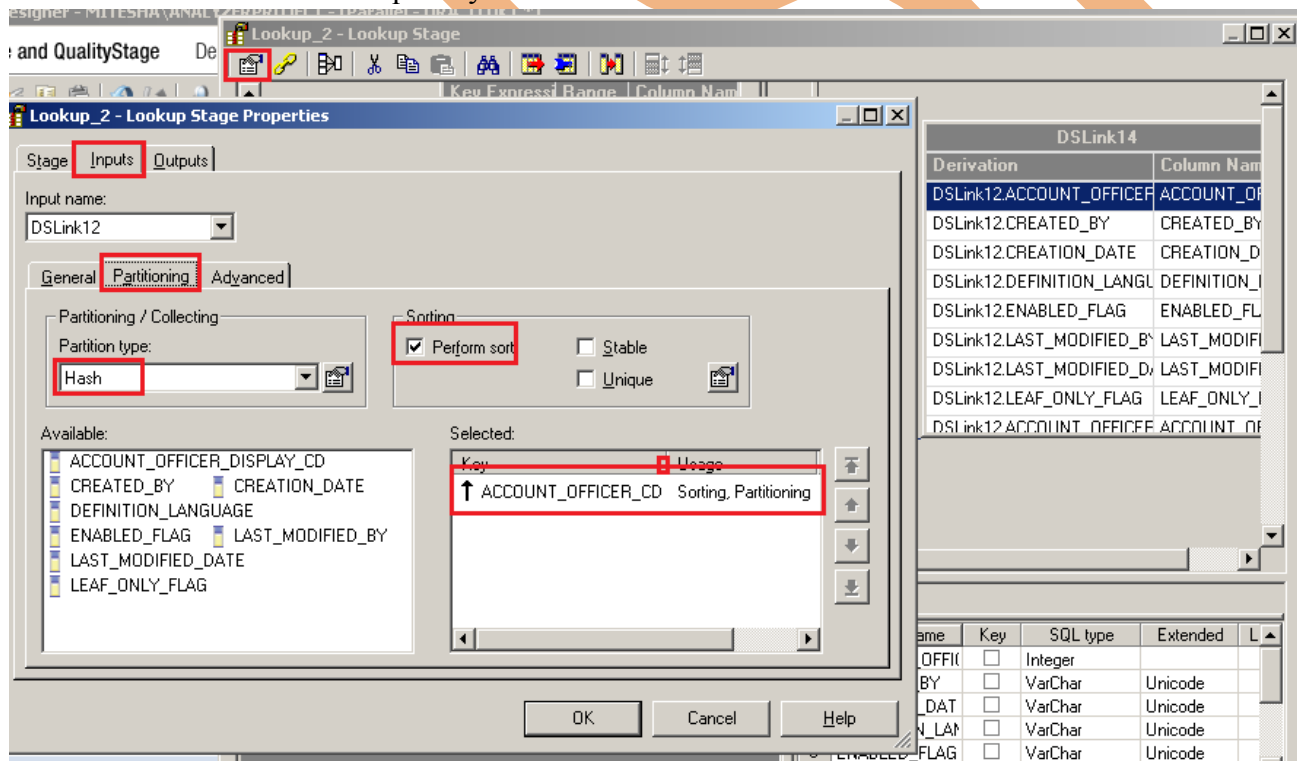

Step-21:- Again apply same process in second link as previous and choose same key and perform sorting.
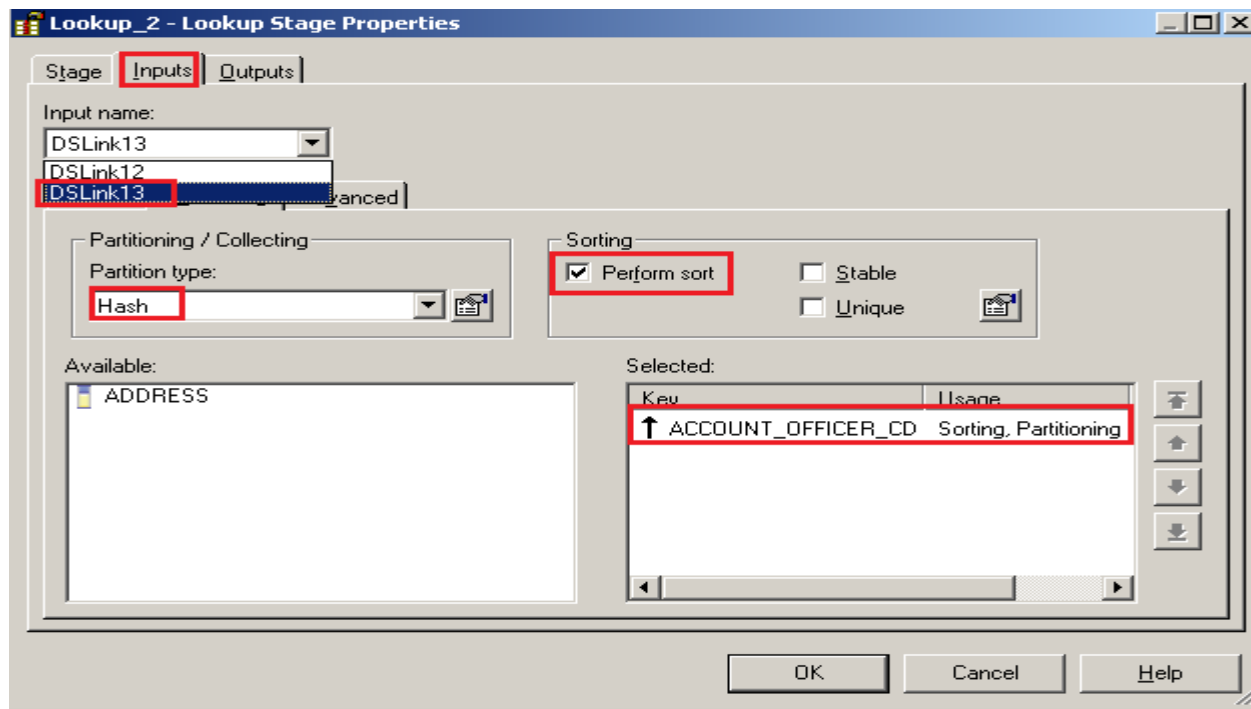
Step-22 :- Now double click on Sql Server enterprise stage then it looks like below screenshot. Here we have to enter our table name where we want to insert our data. Multiple options are there for write mode dropdown. If we directly want to load data then we simple used write method as load otherwise we manually put queries on it. Also we have to specify username and password and this password should match with oracle username and password. Below we specify Replace as write Mode that means if already data is present in table than replace the data with new data.
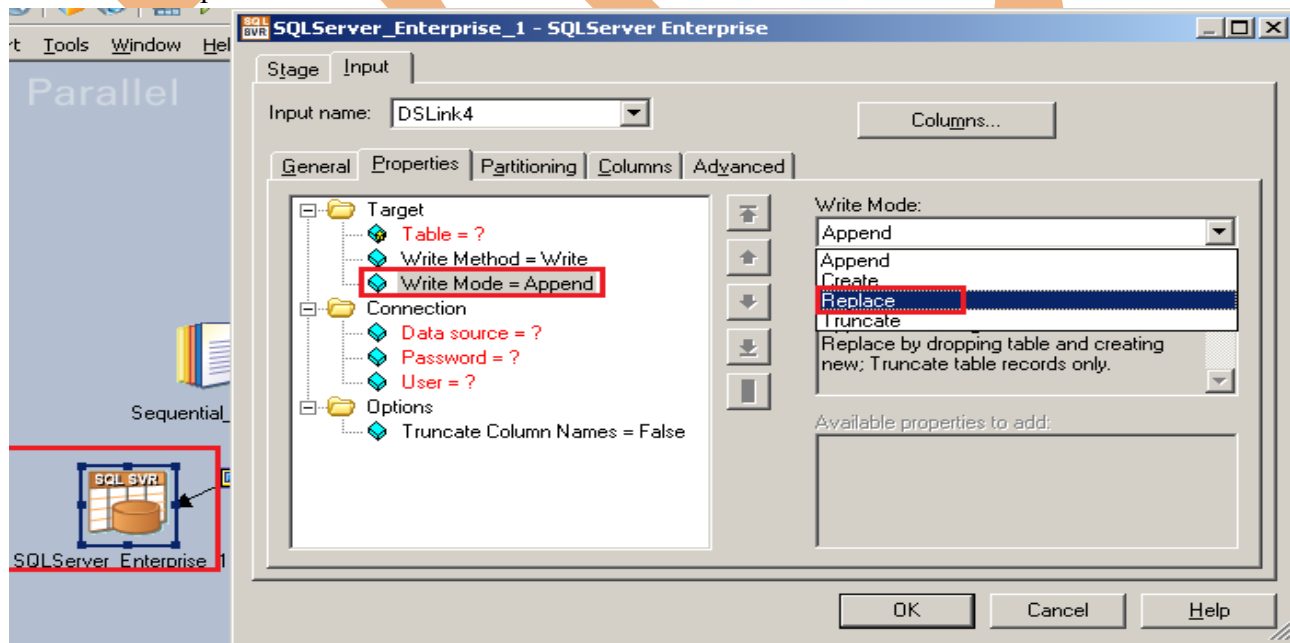
Step-23:- Click on column tab and see all the columns are showing or not.
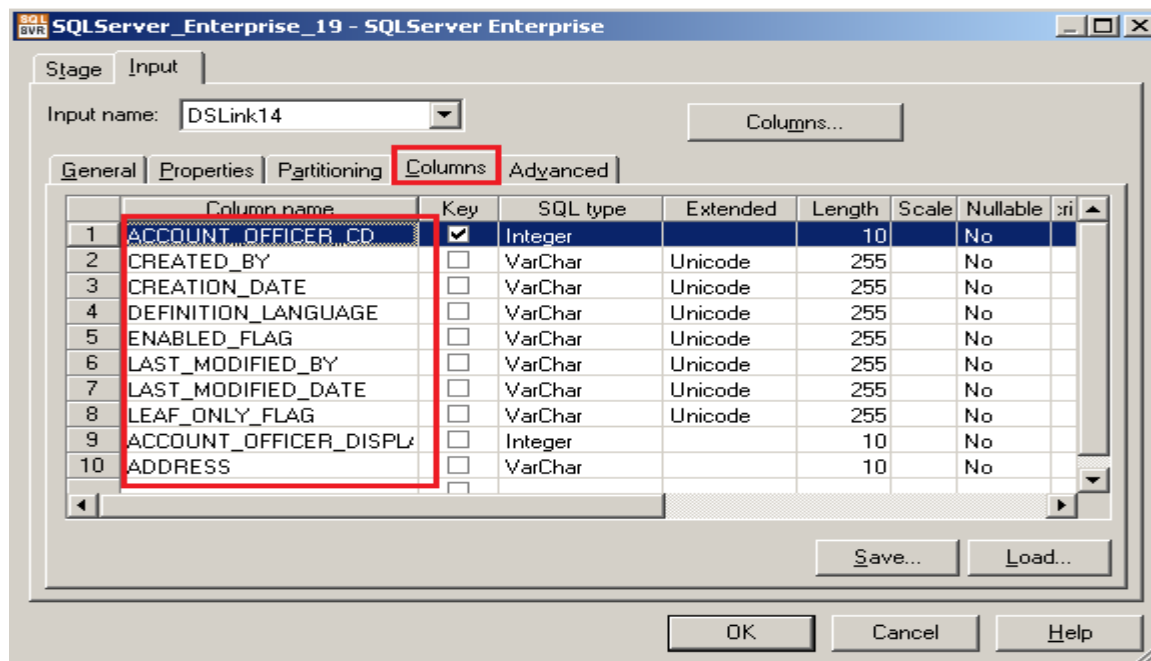
Step-24:- Now here we have to specify the table name and schema name and make an ODBC connection. System DSN and all the other entries should be same as where your table is situated and whatever you specify in System DSN.
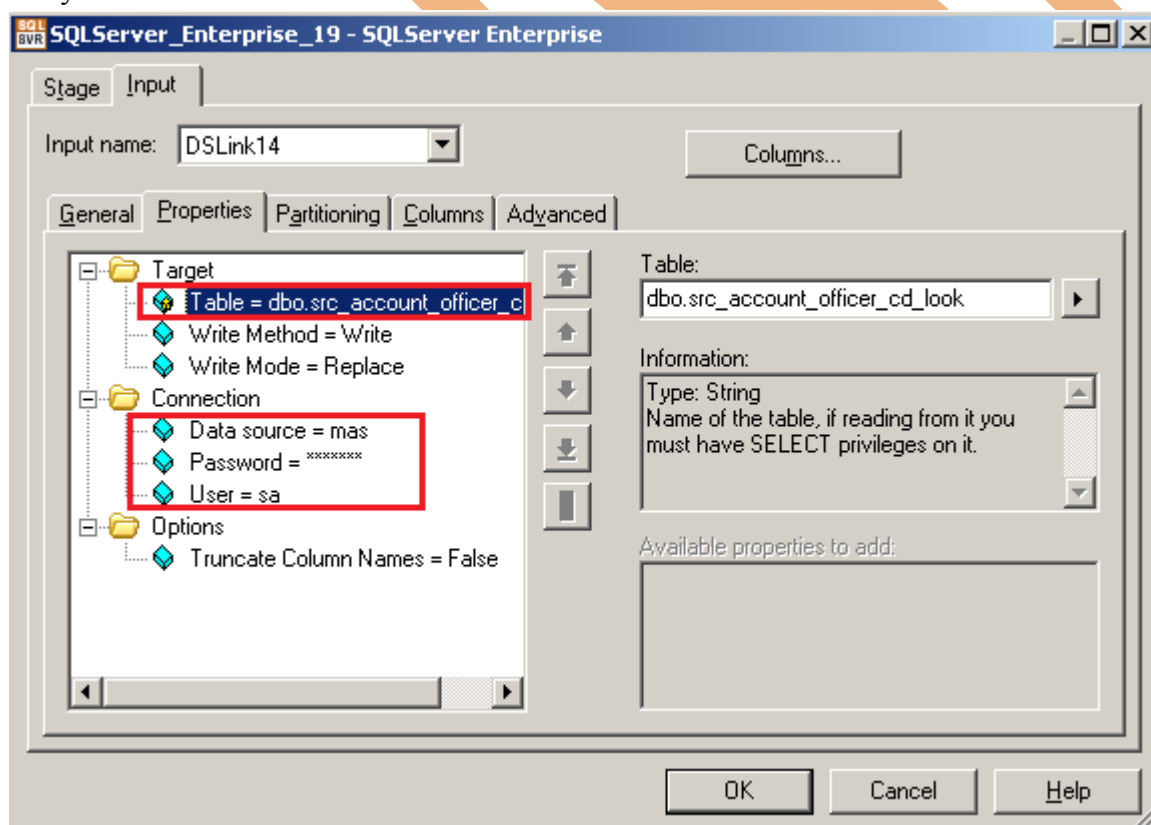
Step-25:- This shows the simple link and Reference link that means second sequential file is called reference file for lookup every time and dashed line is called reject link. When it fails to find any matching records then it goes to Rejected file.
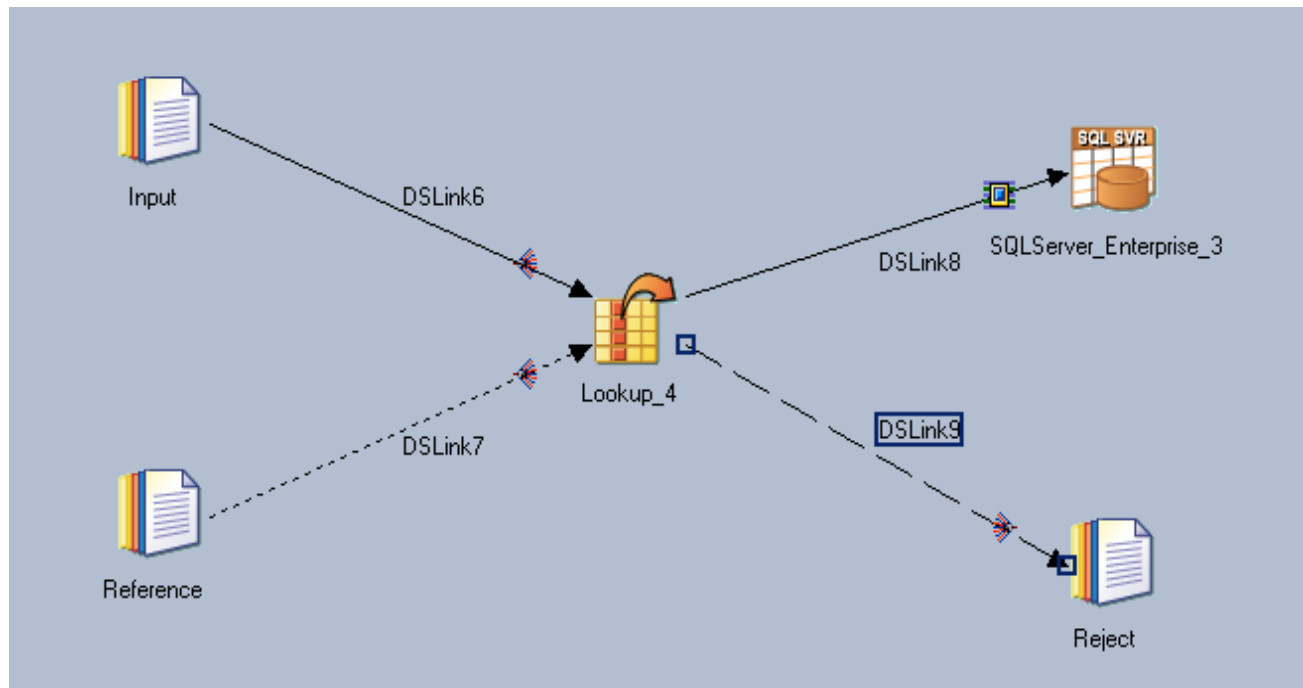
Step-26:- This screenshot shows that your compilation is done otherwise it will throw error. Now click on run button or (CTRL+F5).
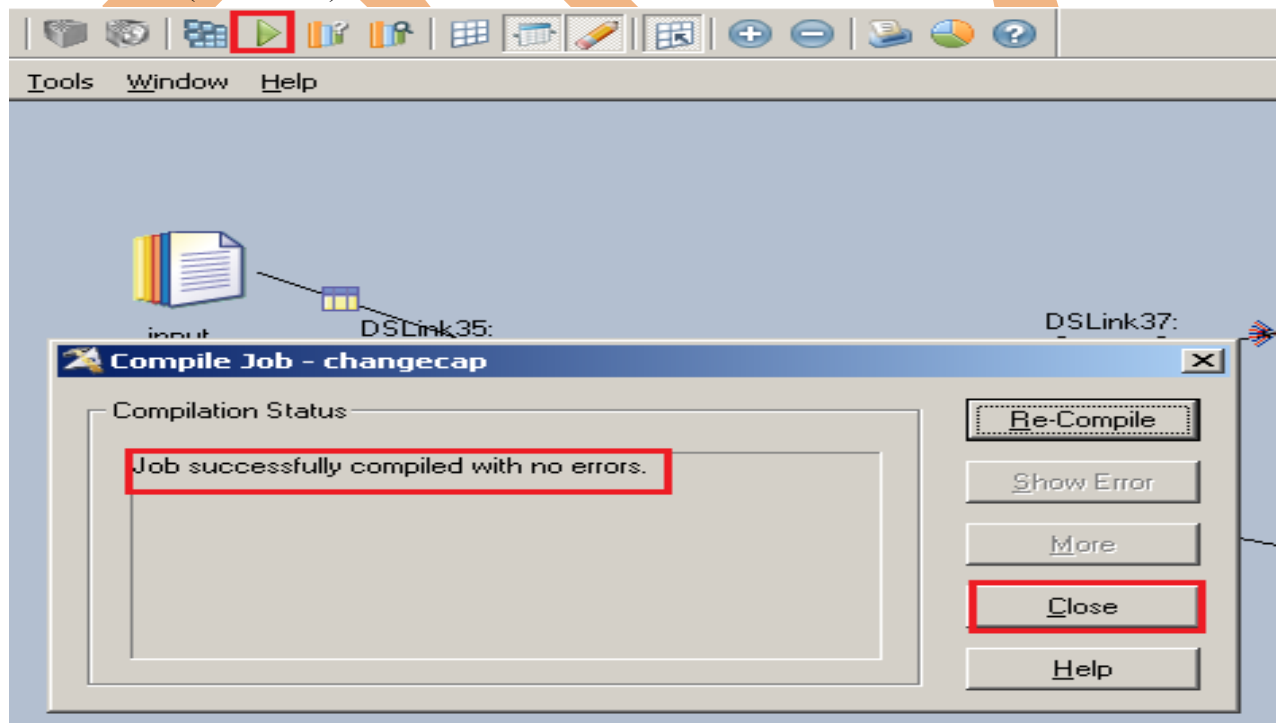
Step-27:- After click on run wait for a while than it shows GREEN line that means your transformation is successfully done otherwise if it shows RED Line that means not done and BLUE Line means Under Process.

Step-28:- Now Goto SQL Console and connect with the same login credentials as I mentioned in sql server enterprise stage sa/password123 so you can choose your own username and password. Before loading, remember to check structure of this file is available and then you should be able to load the data. then simple query:

<SELECT * FROM SRC_ACCOUNT_OFFICER_CD;>

Step-29:- This is Rejected data,Go to the location where you saved this flat file and open it to see the rejected data.

```
54706,"RTUSER","11-JUN-10 07.19.36.000000 PM","US","Y","RTUSER","11-JUN-10 07.19.36.000000 PM","Y",54705
55675,"RTUSER","11-JUN-10 07.19.36.000000 PM","US","Y","RTUSER","11-JUN-10 07.19.36.000000 PM","Y",55674
56619,"RTUSER","11-JUN-10 07.19.36.000000 PM","US","Y","RTUSER","11-JUN-10 07.19.36.000000 PM","Y",56618
58235,"RTUSER","11-JUN-10 07.19.36.000000 PM","US","Y","RTUSER","11-JUN-10 07.19.36.000000 PM","Y",58235
59154,"RTUSER","11-JUN-10 07.19.36.000000 PM","US","Y","RTUSER","11-JUN-10 07.19.36.000000 PM","Y",59153
60740,"RTUSER","11-JUN-10 07.19.36.000000 PM","US","Y","RTUSER","11-JUN-10 07.19.36.000000 PM","Y",60740
```

This is Rejected Data