IBM Data Stage Lab Guide#16
Implement Change Data Capture

Description:

BISP is committed to provide BEST learning material to the beginners and advance learners. In the same series, we have prepared a list of beginner's guide and FAQs for IBM Data Stage. We have built complete financial Data Model and various data transformation techniques. Download many such learning documents, student guide, Lab Guide and Hands-on practice materials. This document guides you through implementation approach for Change Data Capture. ***Join our professional training to learn from Experts.***

History:

Version	Description Change	Author	Publish Date
0.1	Initial Draft	Varun Khare	12th Aug 2012
0.1	Review#1	Amit Sharma	18th Aug 2012

BISP

Mapping Sheet

Change Capture Job.

Source- Oracle, Flat File

Target- Dataset

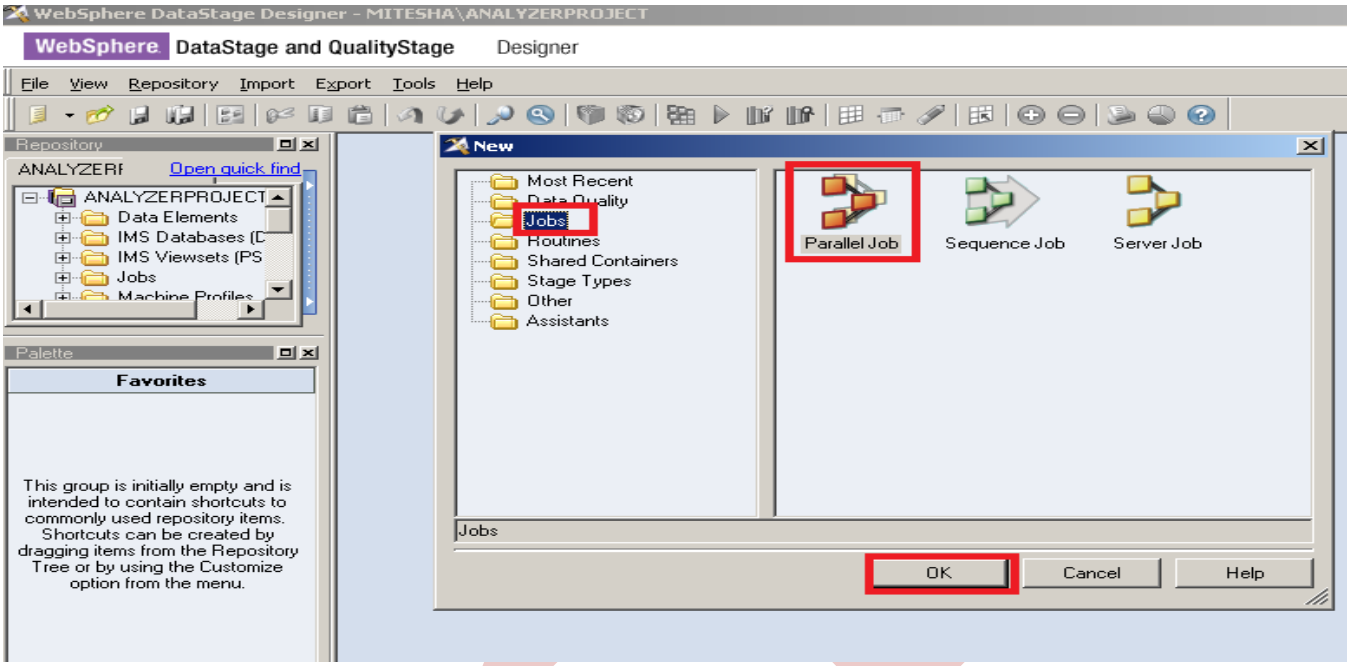
Transformation	Change Capture
Source Name	Flat File, Oracle
Target Name	DataSet File

Flat File Source Details			Oracle Details				Flat Files	
Entity Name	Data Types	Is Null	Destination Entity Nar	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

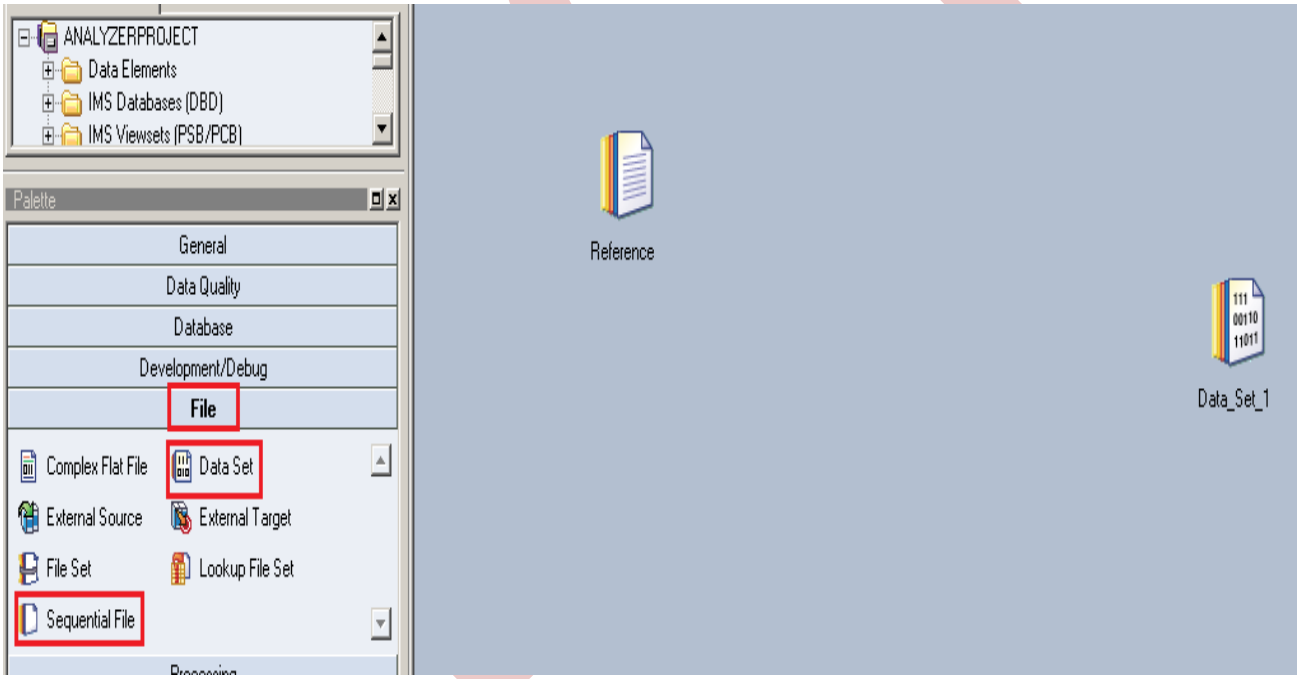
Steps to implement CDC.

Step-1 In Windows Click Designer Client of DataStage

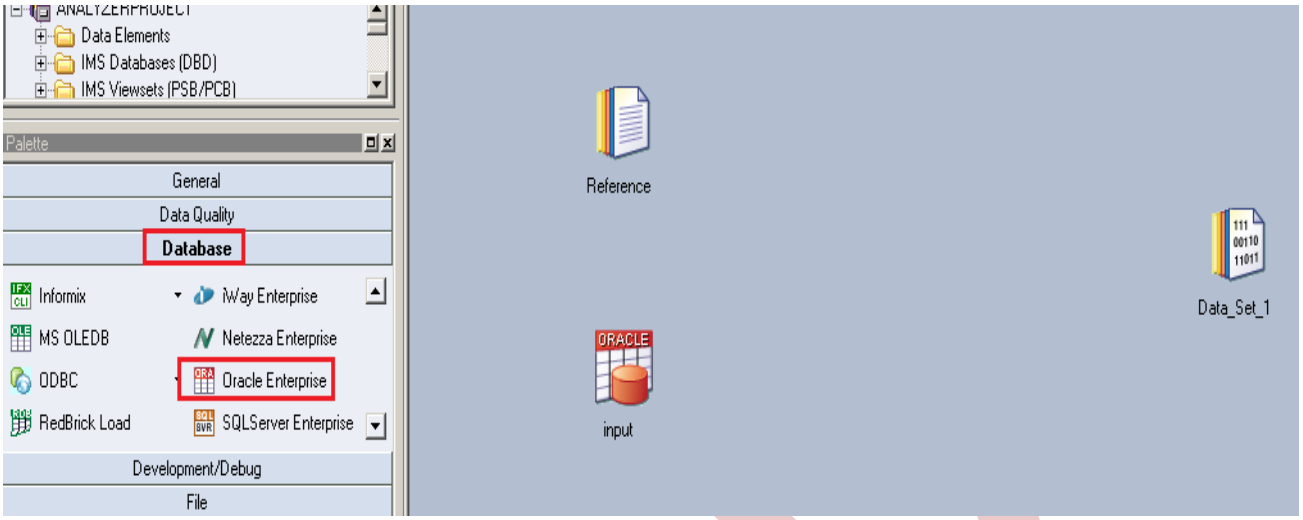
Step-2 :- New Window opens, then click Jobs ,then click Parallel Job because we are using parallel jobs.



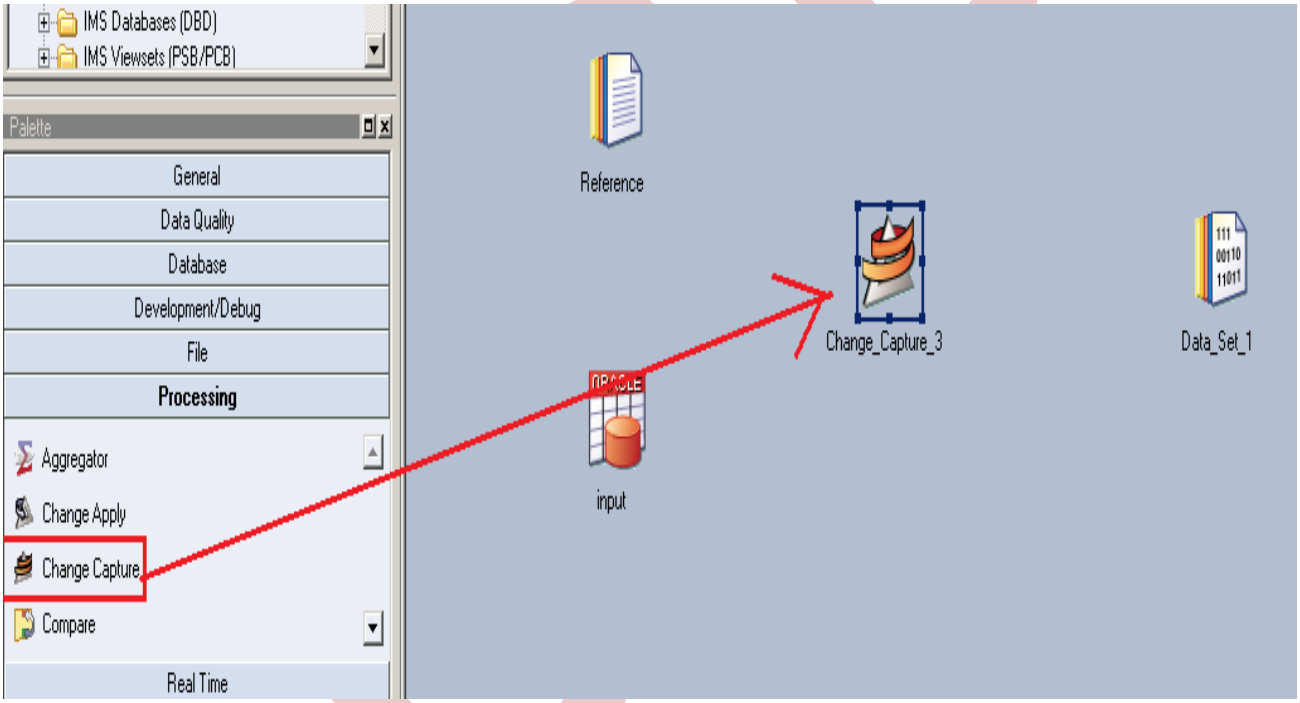
Step-3 :- First select File Palette, choose Sequential File and Dataset file drag it to Parallel pane and rename it. This file is used for reference data and output data.



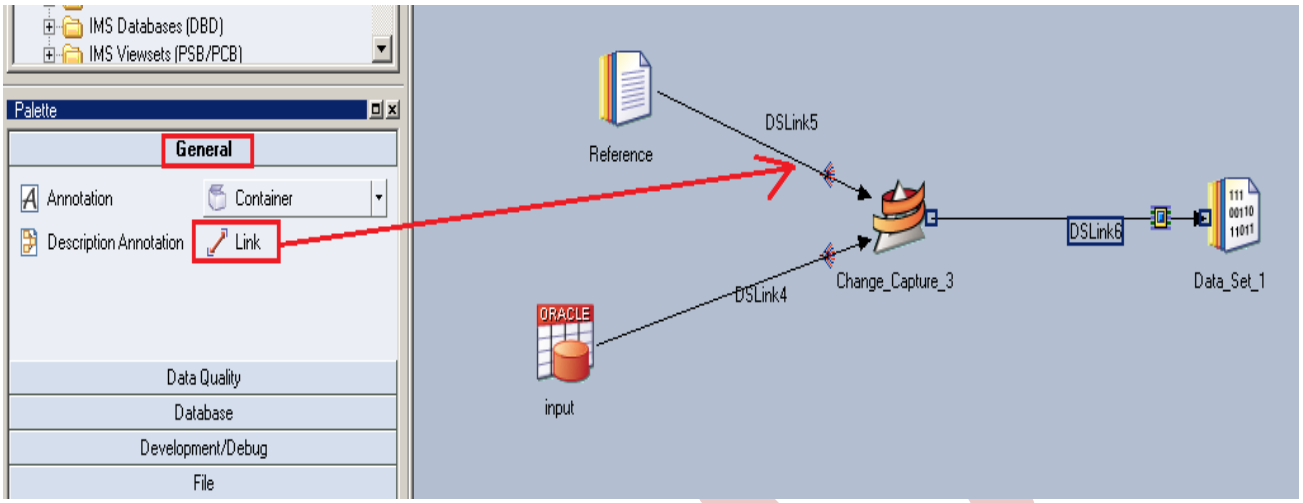
Step-4 :- Choose Oracle Enterprise from database Data base Palette and drag it to parallel pane. This is used for taking input data from oracle.



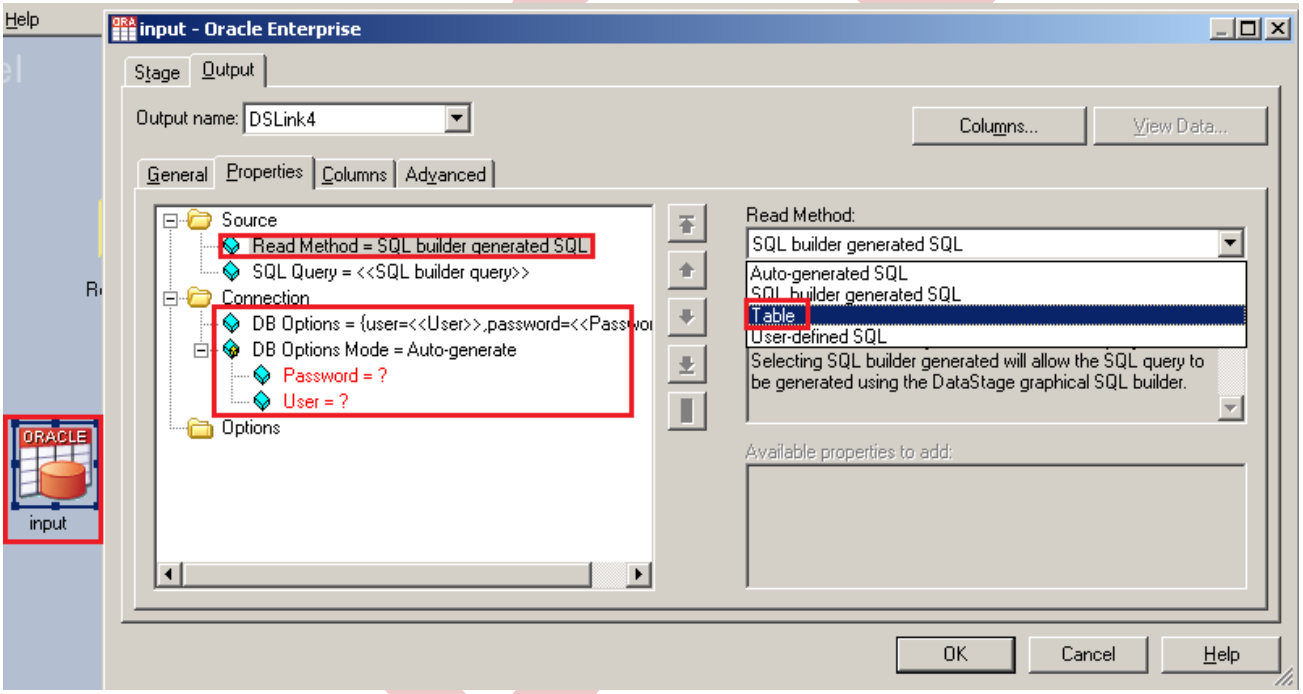
Step-5 :- Now go to processing palette and choose change capture stage as we know that this stage is useful for capture change on your input file with the help of reference file.



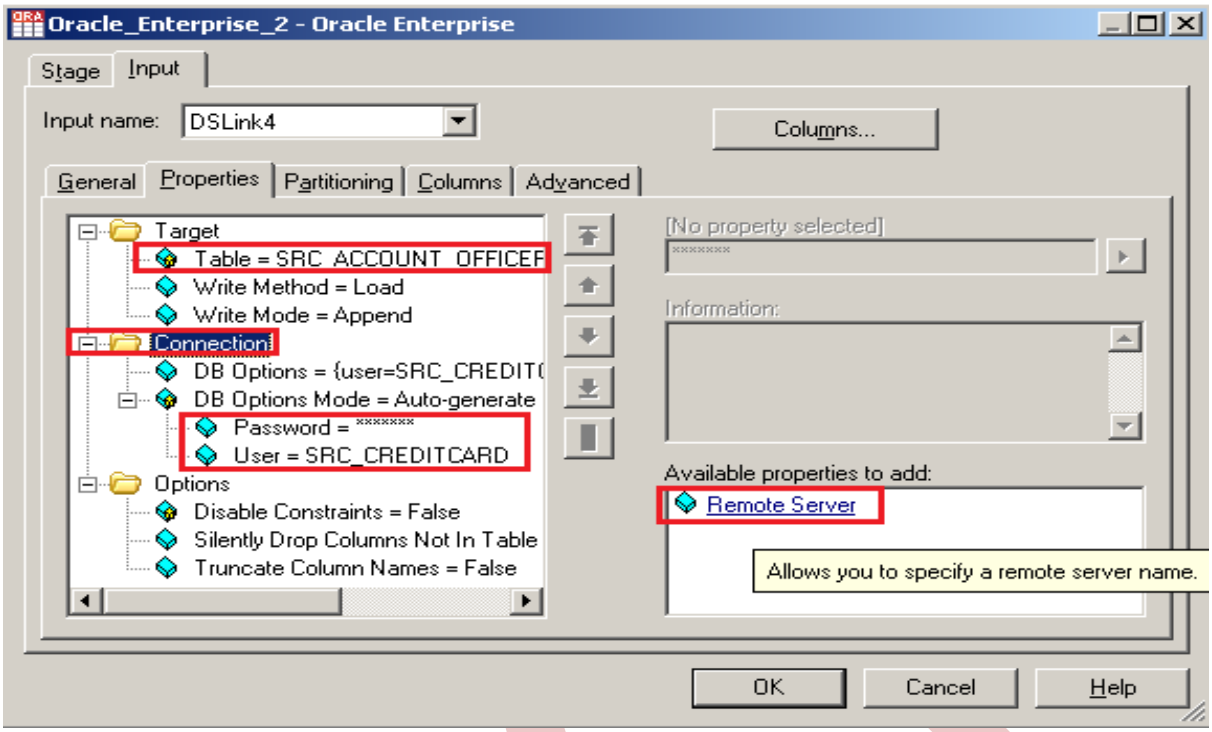
Step-6:- Now make an connection with these stages that's why we have to use this link stage for connecting these stages.



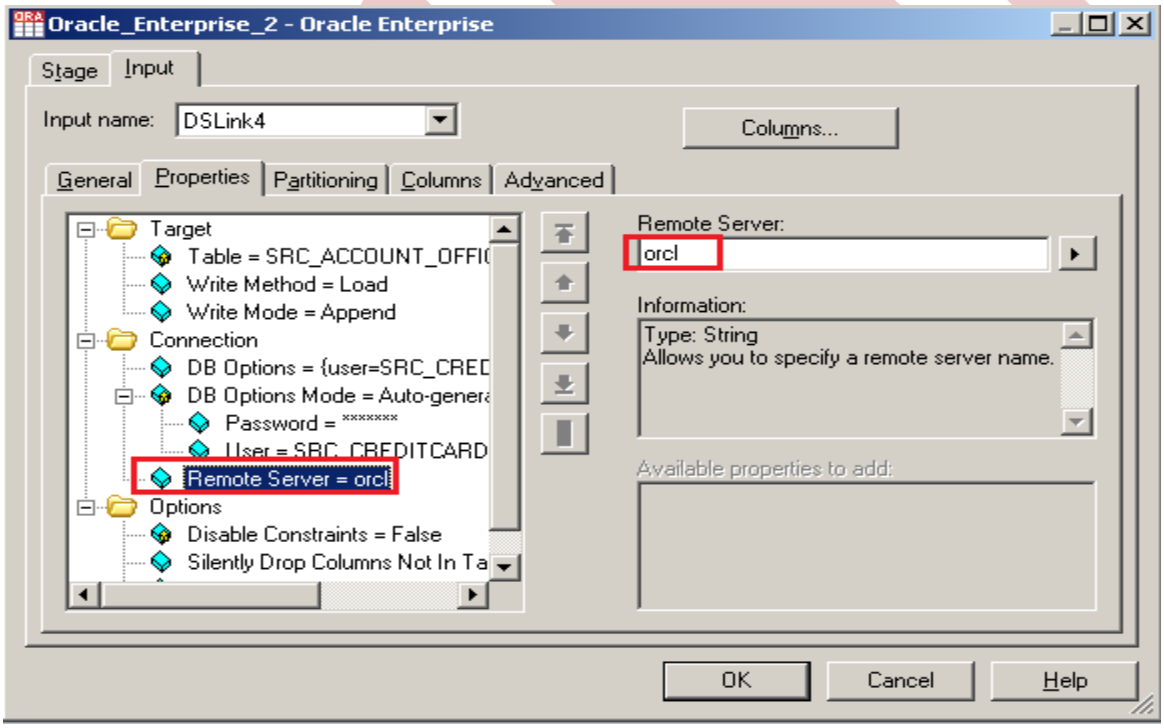
Step-7:- Now double click on oracle stage to fetch oracle table into DataStage. There are multiple methods available, We choose table as Read Method as we know the table and schema name so we directly call the table.



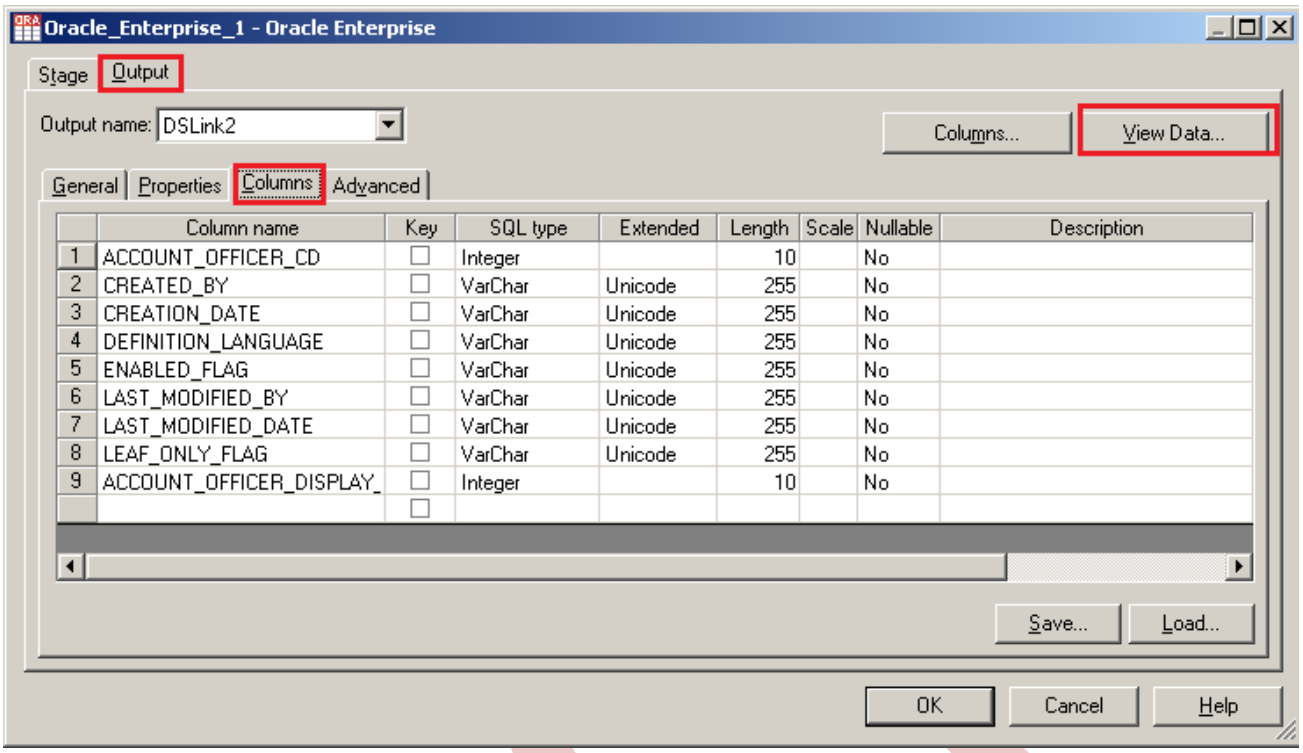
Step-8:- There's one more option here once you click on connection tab, remote server options popup, specify the server name of oracle.



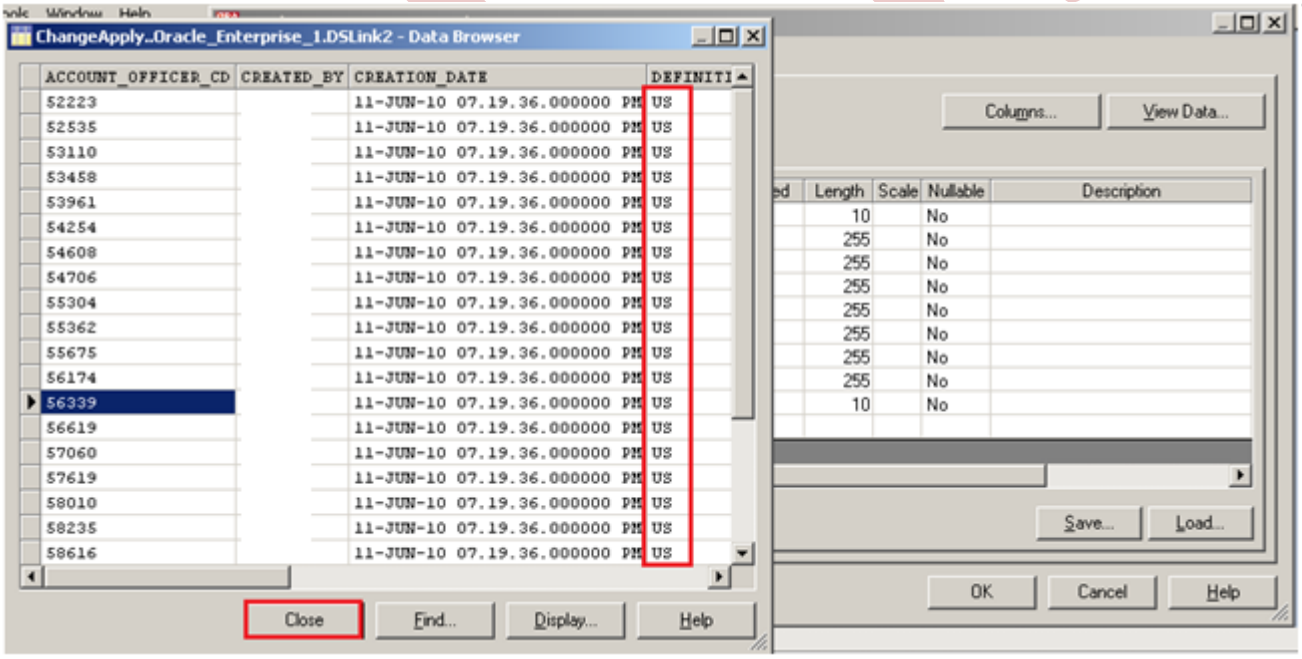
Step-9:- See below, my SID name is ORCL and username/password and table name. Ensure that you put correct information than only you will be able to load the data into table.



Step-10:- Now go to column tab and specify all columns name with data type. Make sure it should be matched with your oracle table columns.



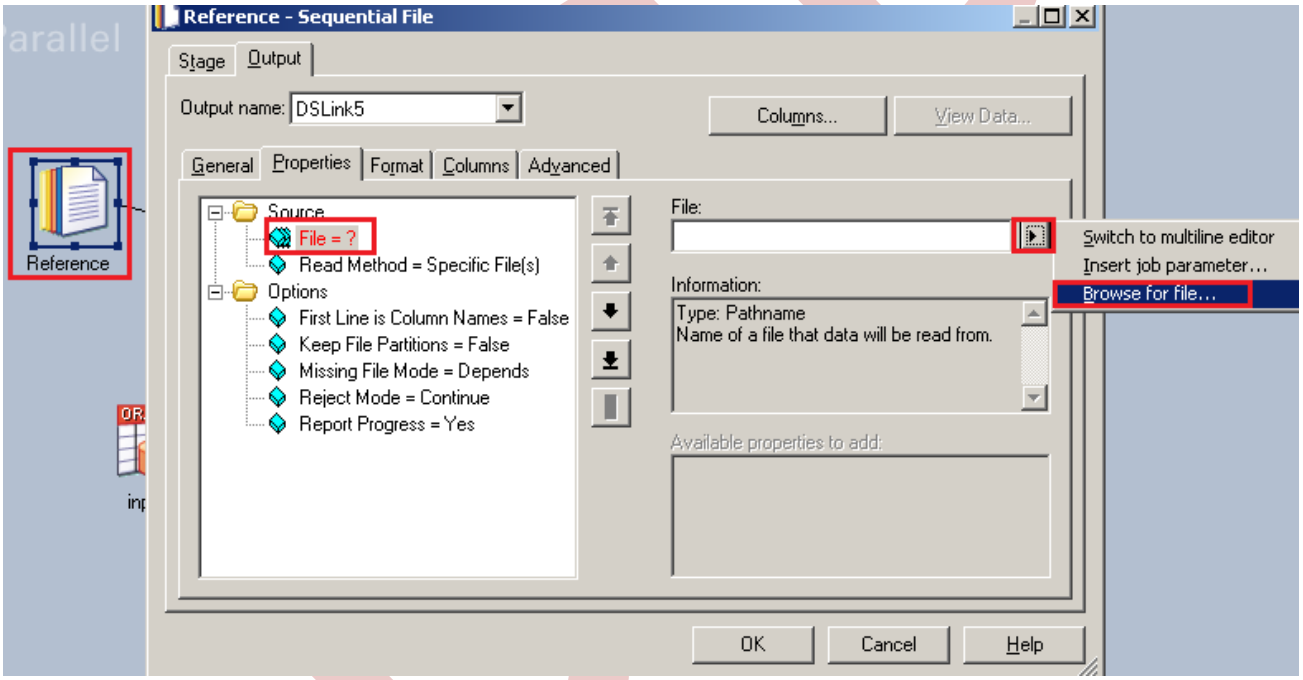
Step-11:- Now click on View Data and Ok. If your data is shown as given below it means, you have made successful connection between input file and Data Stage otherwise it may give some error like column mismatch error.



Step-12 Now, first before loading data, we have to create a table in oracle. 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 than write query and remember the username and password.

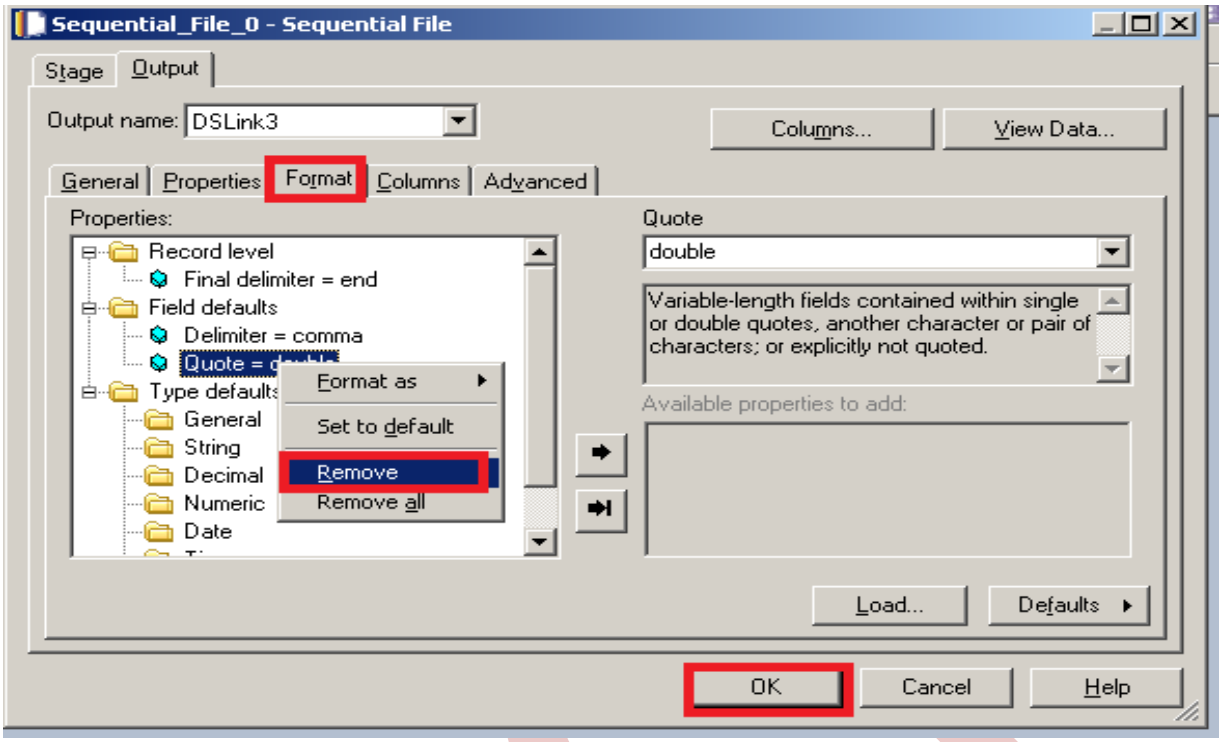
```
SQL> CREATE TABLE SRC_ACCOUNT_OFFICER_CD_FILT
  2     ( ACCOUNT_OFFICER_CD NUMBER(5,0) PRIMARY KEY,
  3   CREATED_BY VARCHAR2(30 BYTE),
  4   CREATION_DATE VARCHAR2 (100),
  5   DEFINITION_LANGUAGE VARCHAR2(4 BYTE),
  6   ENABLED_FLAG VARCHAR2(1 BYTE),
  7   LAST_MODIFIED_BY VARCHAR2(30 BYTE),
  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 any test file.

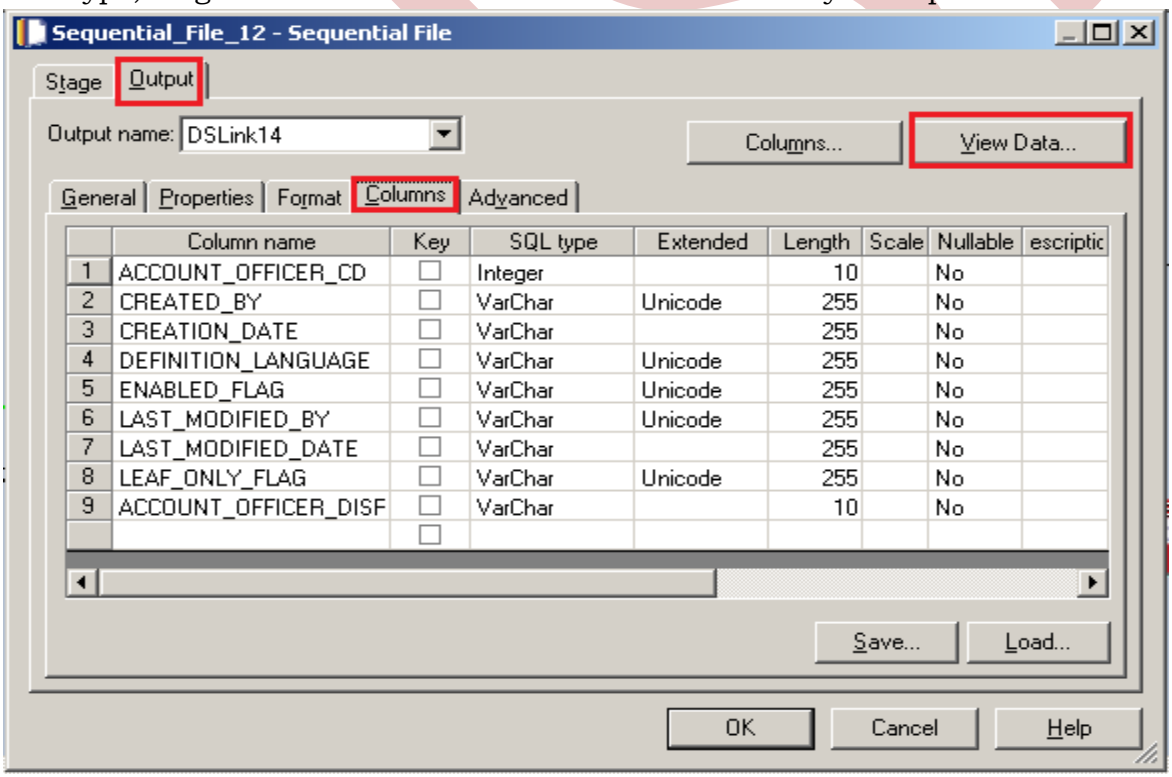


Step-14:- Go to Format tab and Remove double Quote because we don't need them, then Click OK.

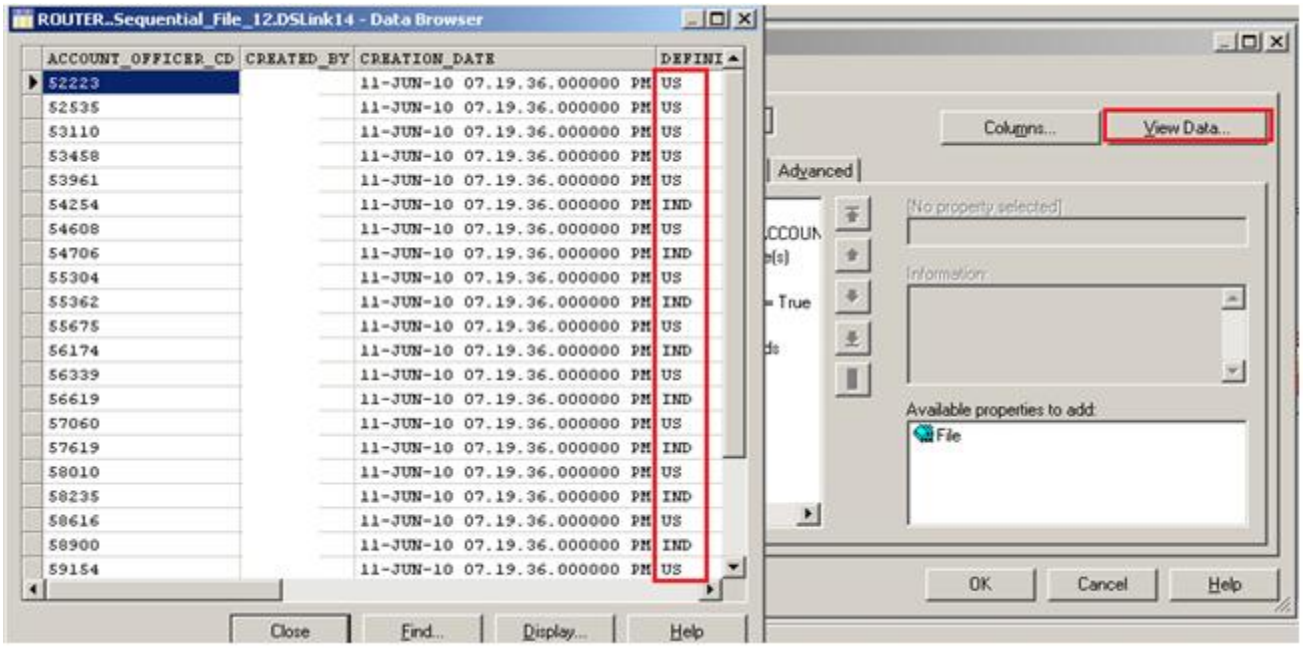
Note:- If you use date as a input than specify here which type of format you used



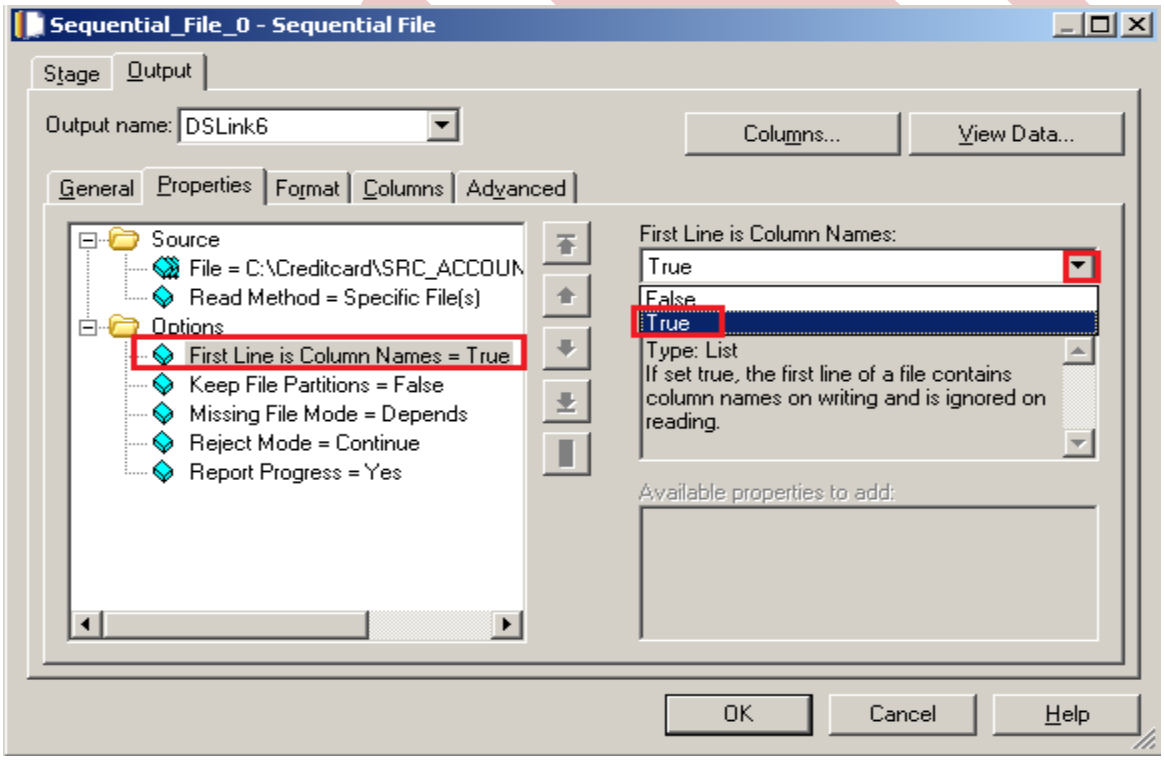
Step-15: Go to Columns tab and enter same column names as declared in Input file and put data type, length than click on View Data. This will show your Input Data.



Step-16:- 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 Datastage otherwise it will give some error like column mismatch error.



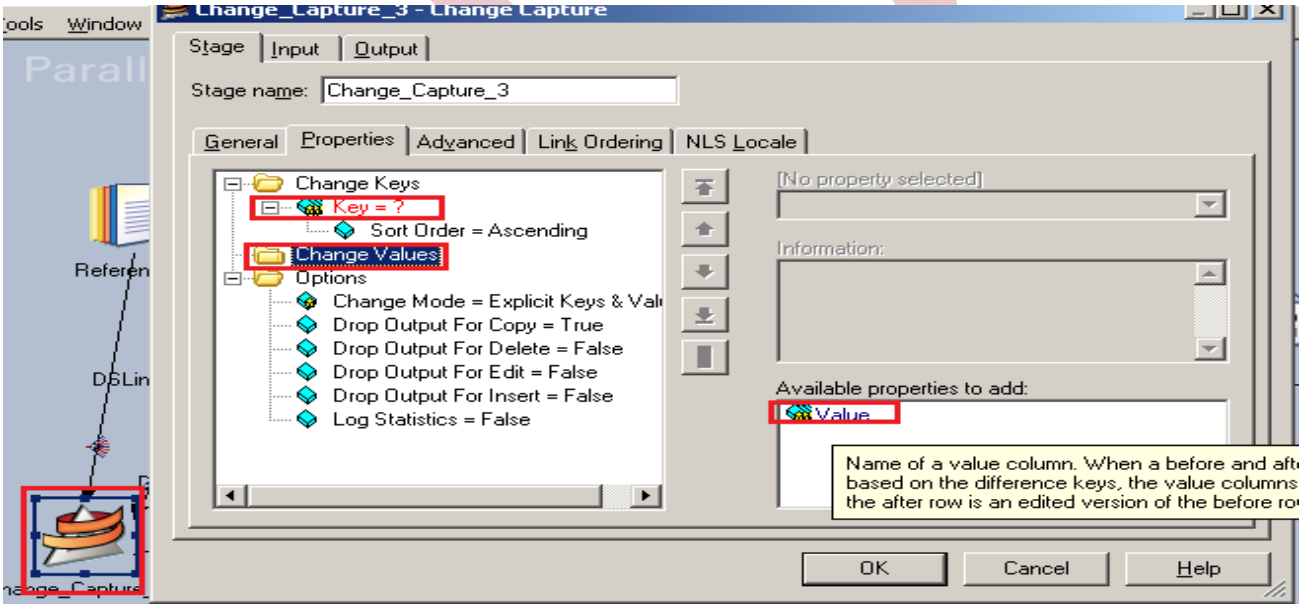
Step-17:- After choosing file, select 'True' from First line in column Names dropdown for removing first line from Input file, First row is a header row.



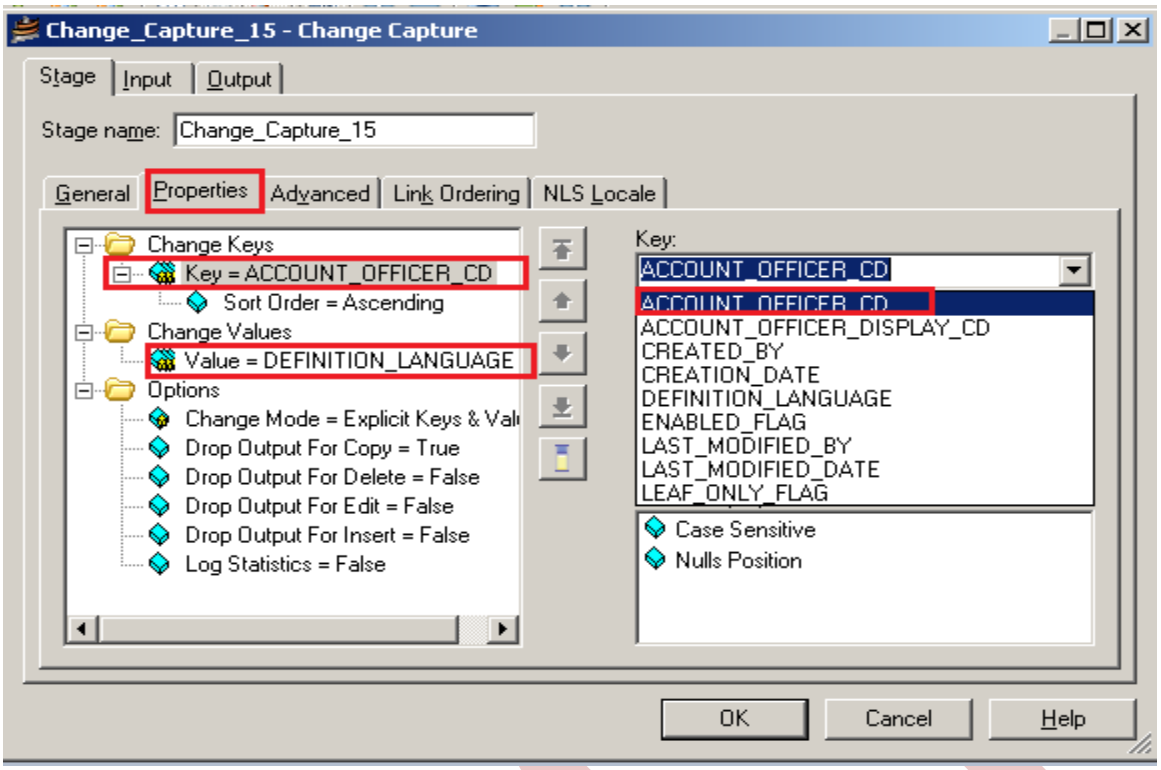
Step-18 :- Make input file and remember column names.

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

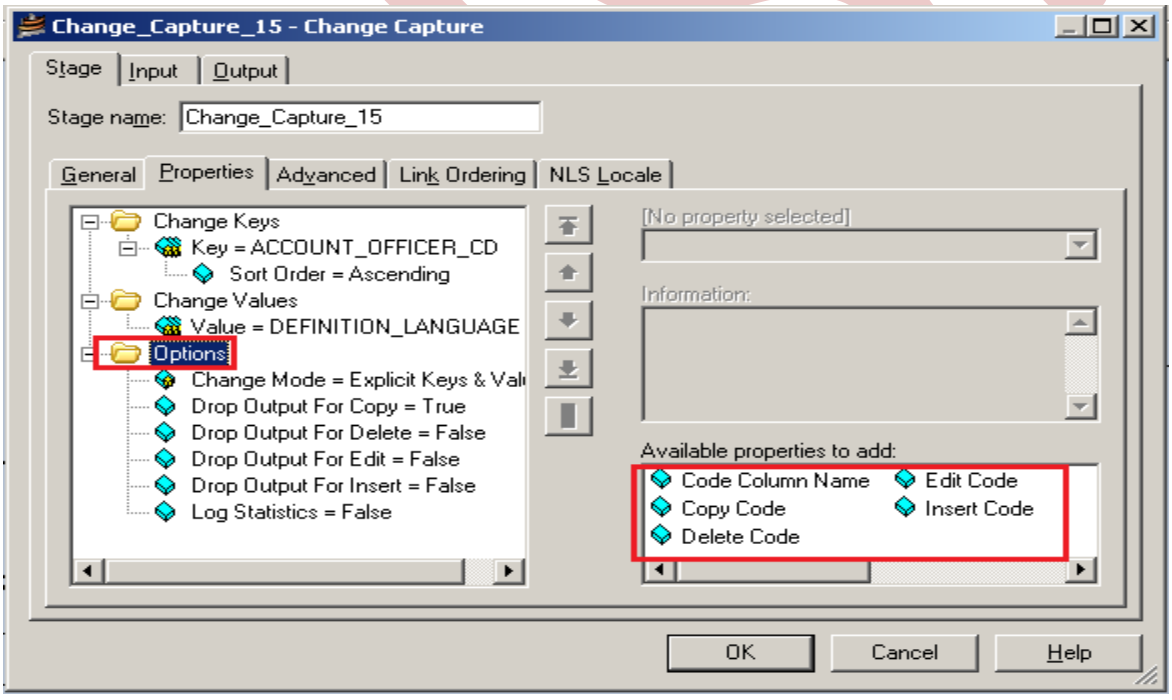
Step-19:- Now double click on change capture stage and there is a Change key and Change values are available so we have to put our key here with this change capture analysis your input table and reference table and values are used for check operation that means whatever values we specified only change capture see that operation.



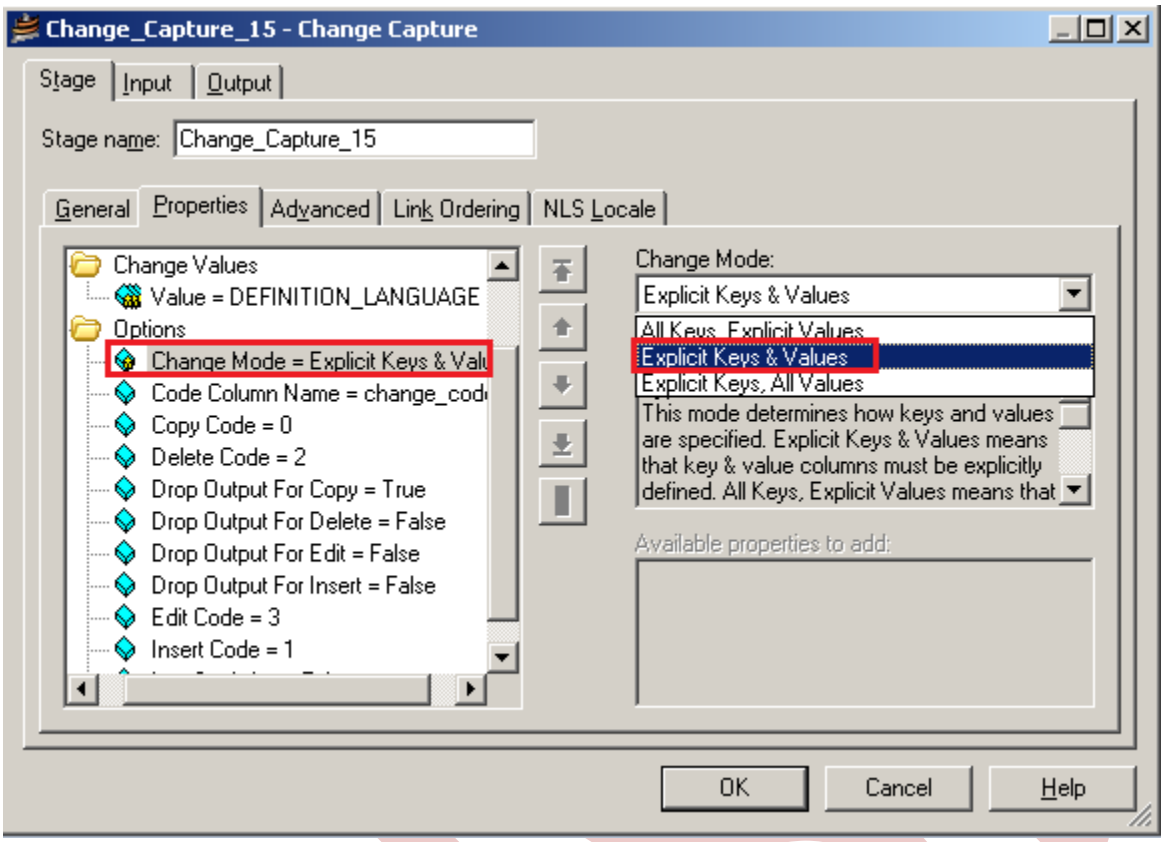
Step-20:- Now see here we choose this key because this key available both the table and unique also and we choose value this because we want to capture language based data means when language change or delete than store.



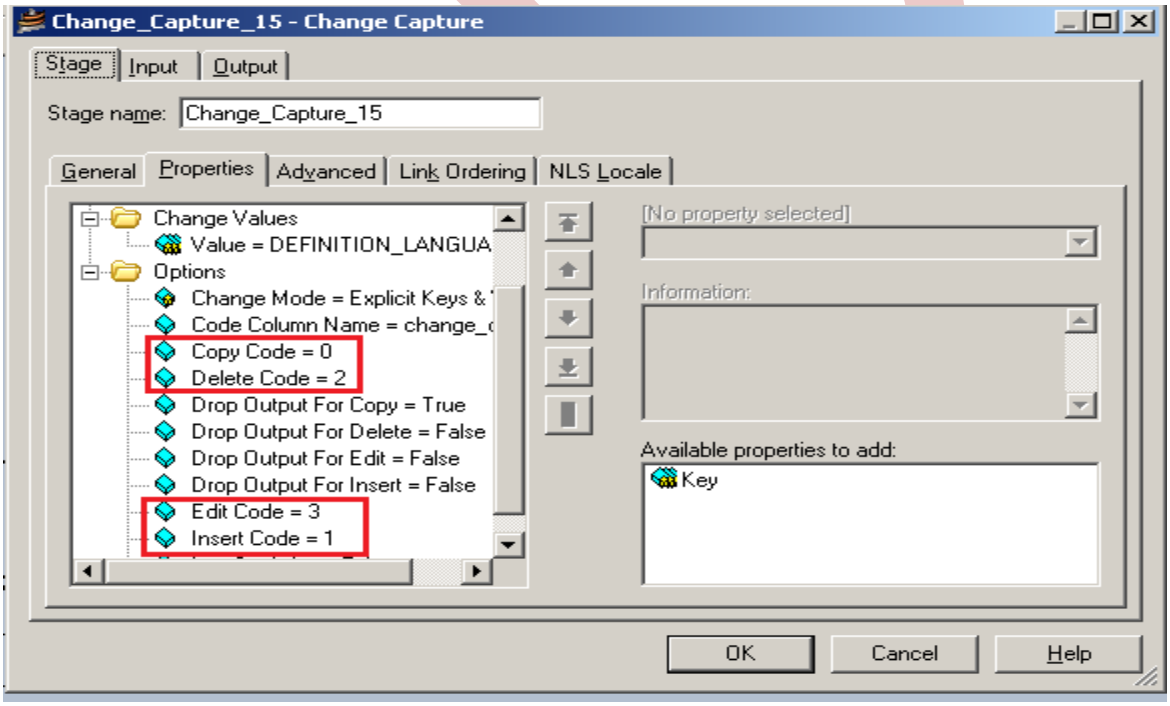
Step-21:- In options, choose option as given below.



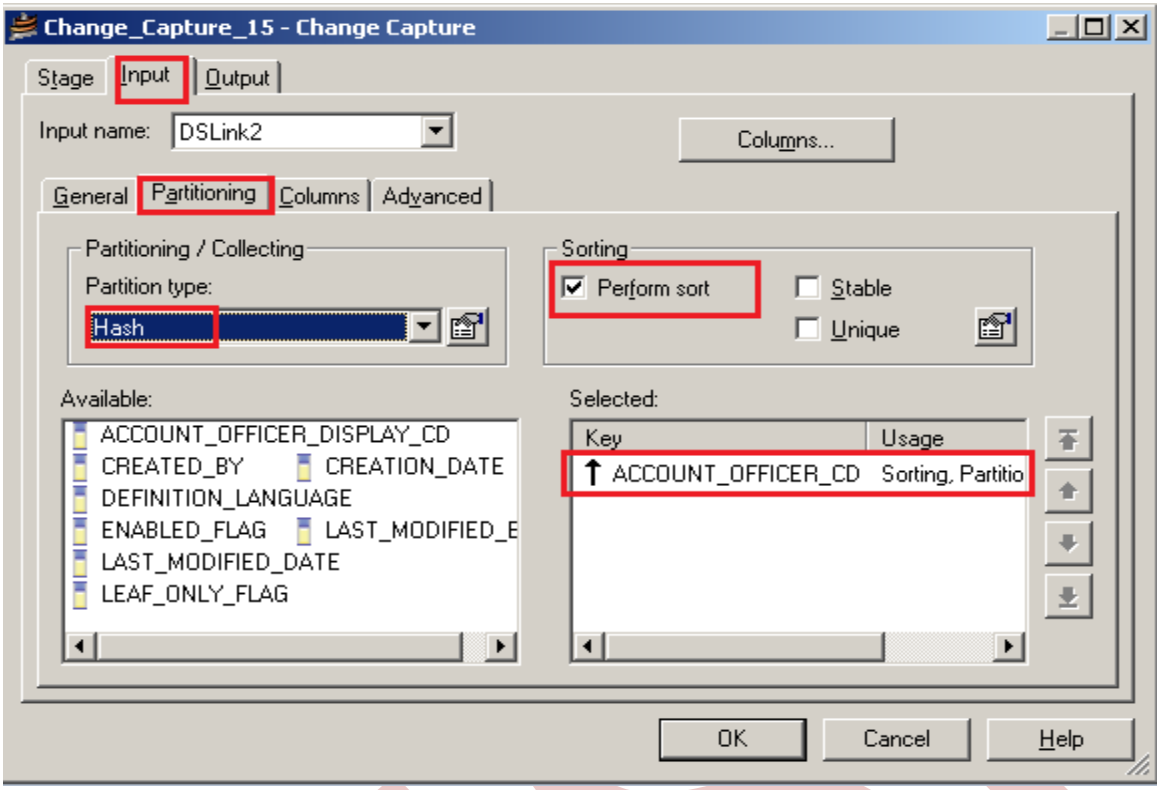
Step-22:- Now click on option and click change mode as we want make explicit keys & values.



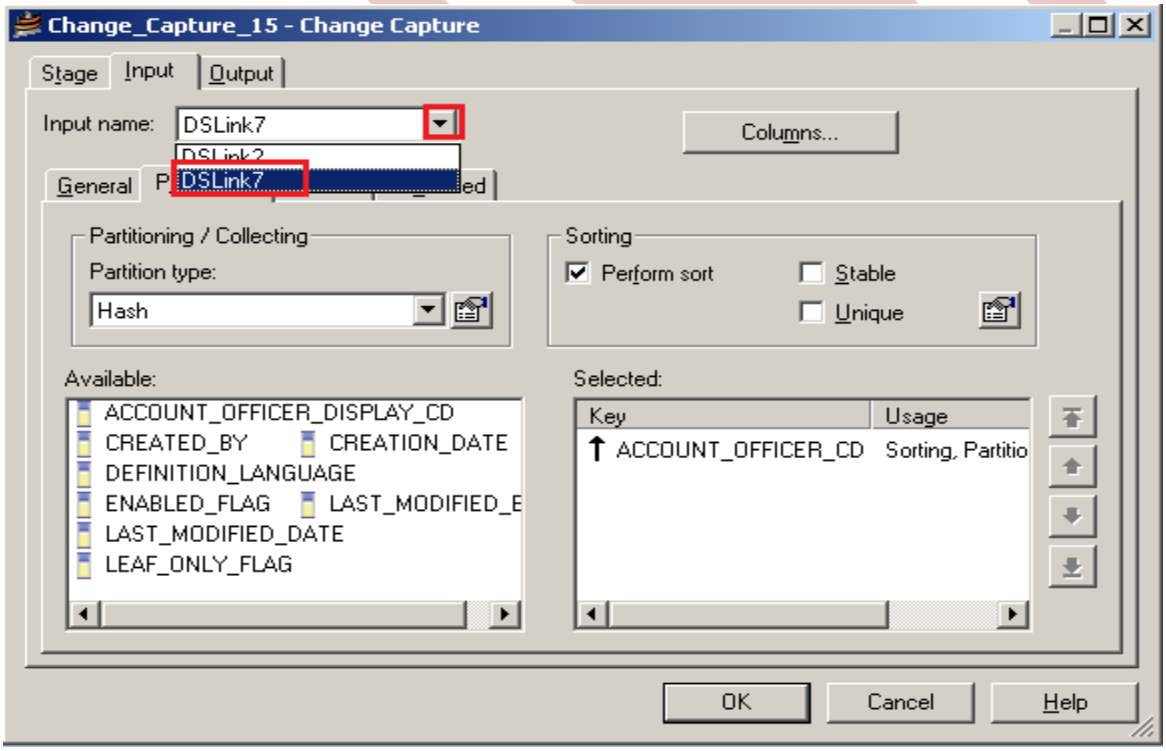
Step-23:- These are default values. We will change it, as it describes the type of operation one can perform**.**



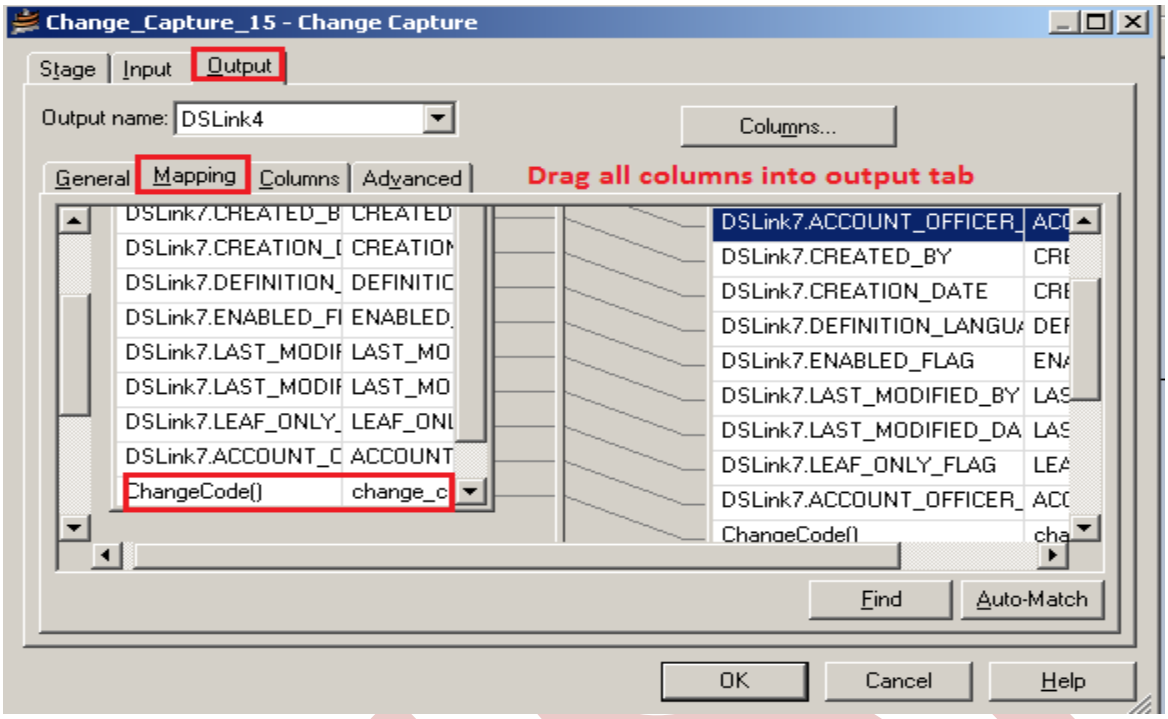
Step-24:- Go to Input tab and select Partitioning. Remember choose Hash if you use Key based job. When we select hash key then the sorting options will be visible.



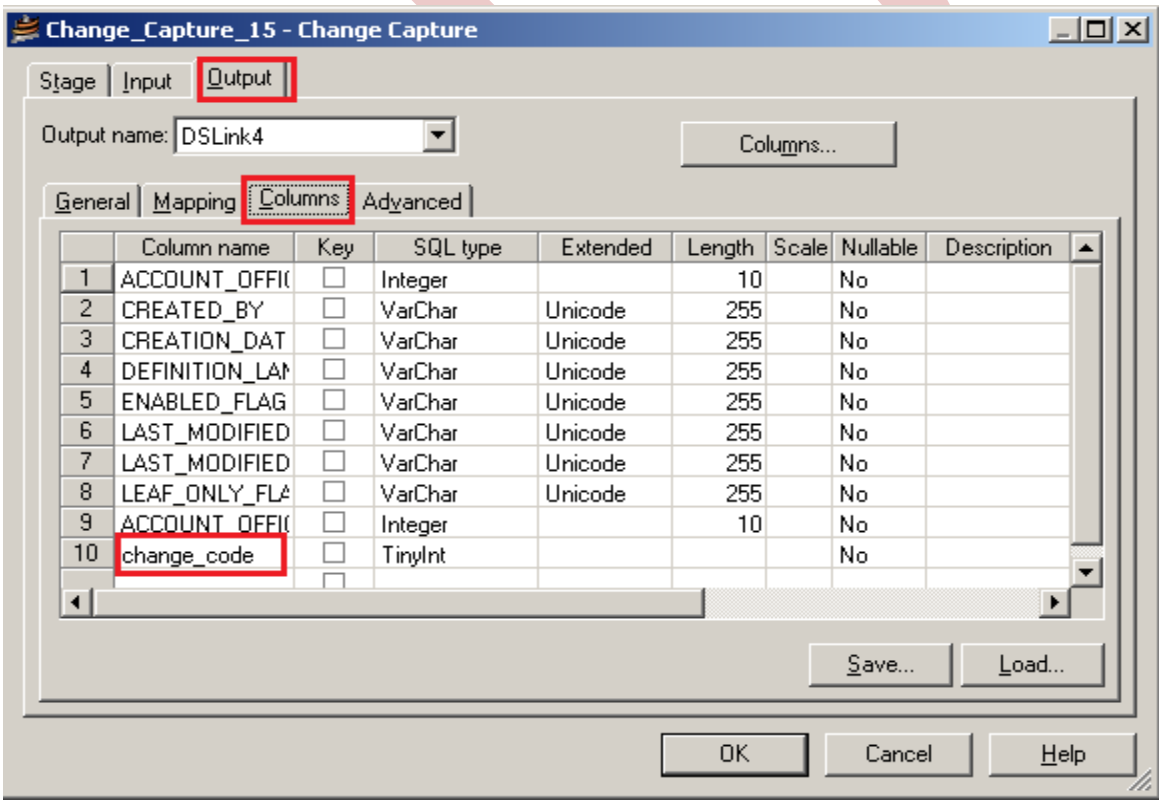
Step-25:- Apply same process as previous one, because we have to 2 input files.



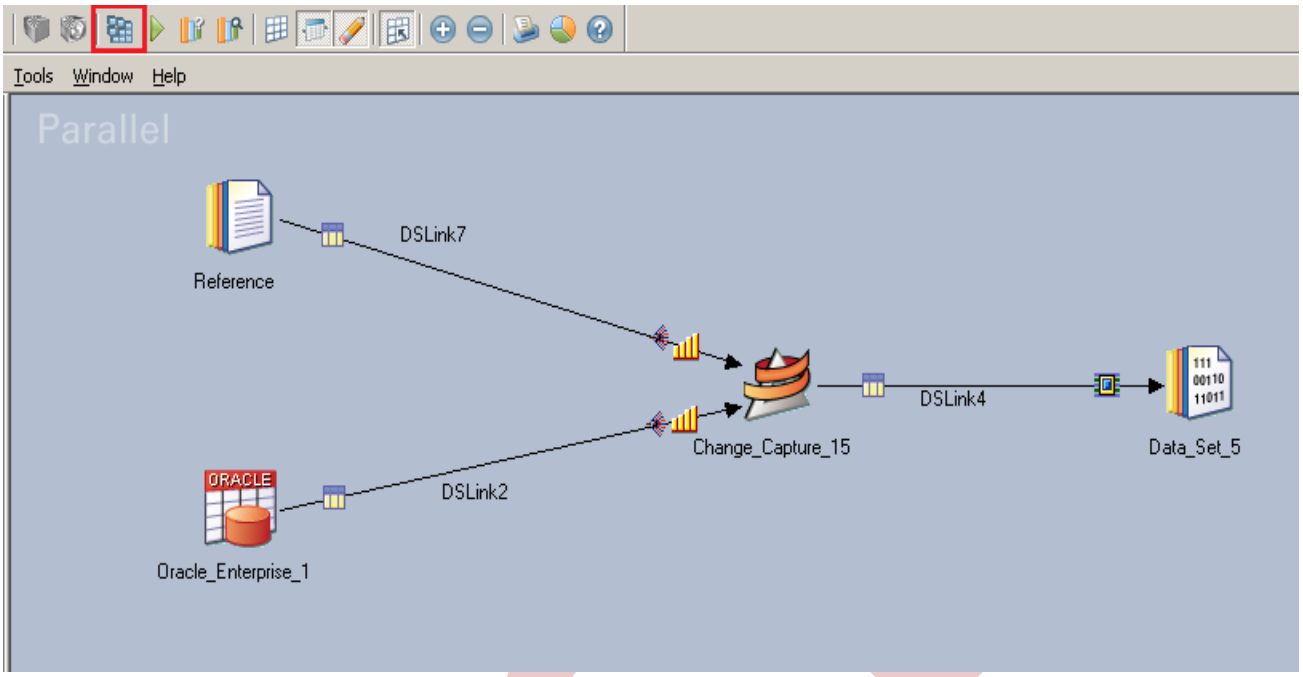
Step-26:- Now go to output tab then select all links and drag them into Out Link for sending data from change capture stage into output links.



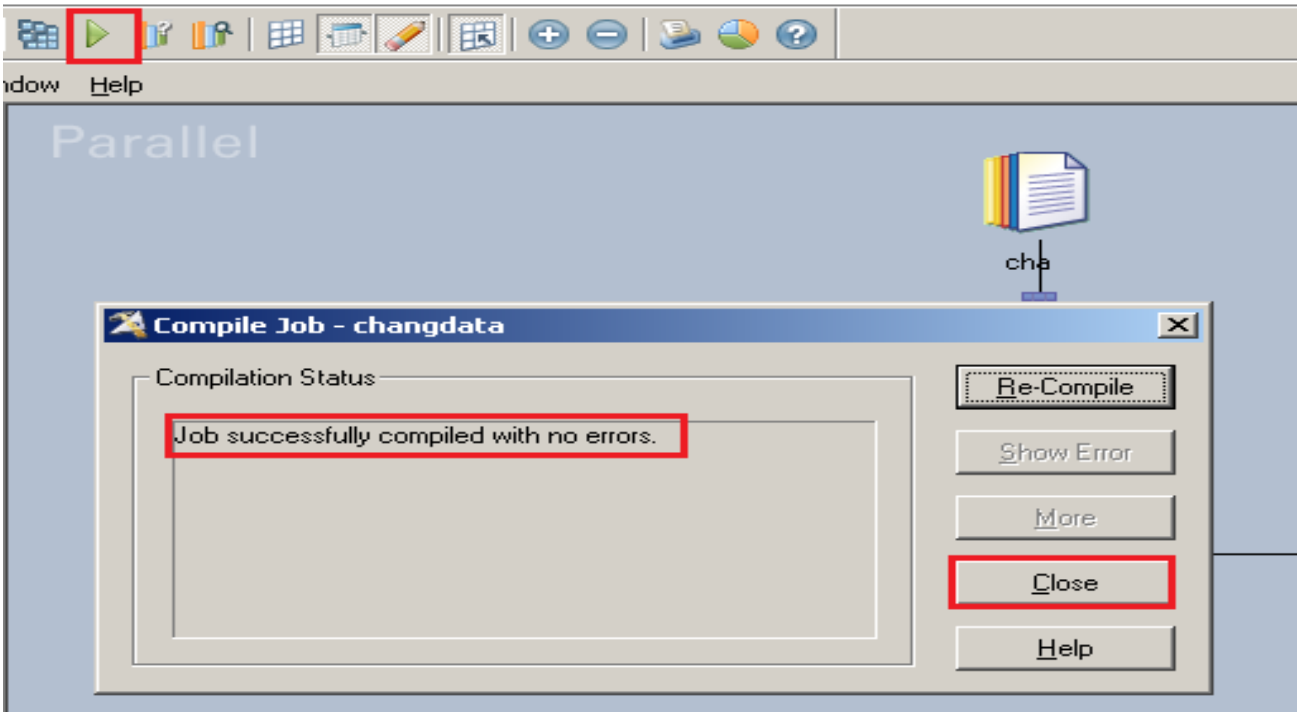
Step-27:- Select output column, there is one more column marked as red color. This column is required to capture the values of changing operation.



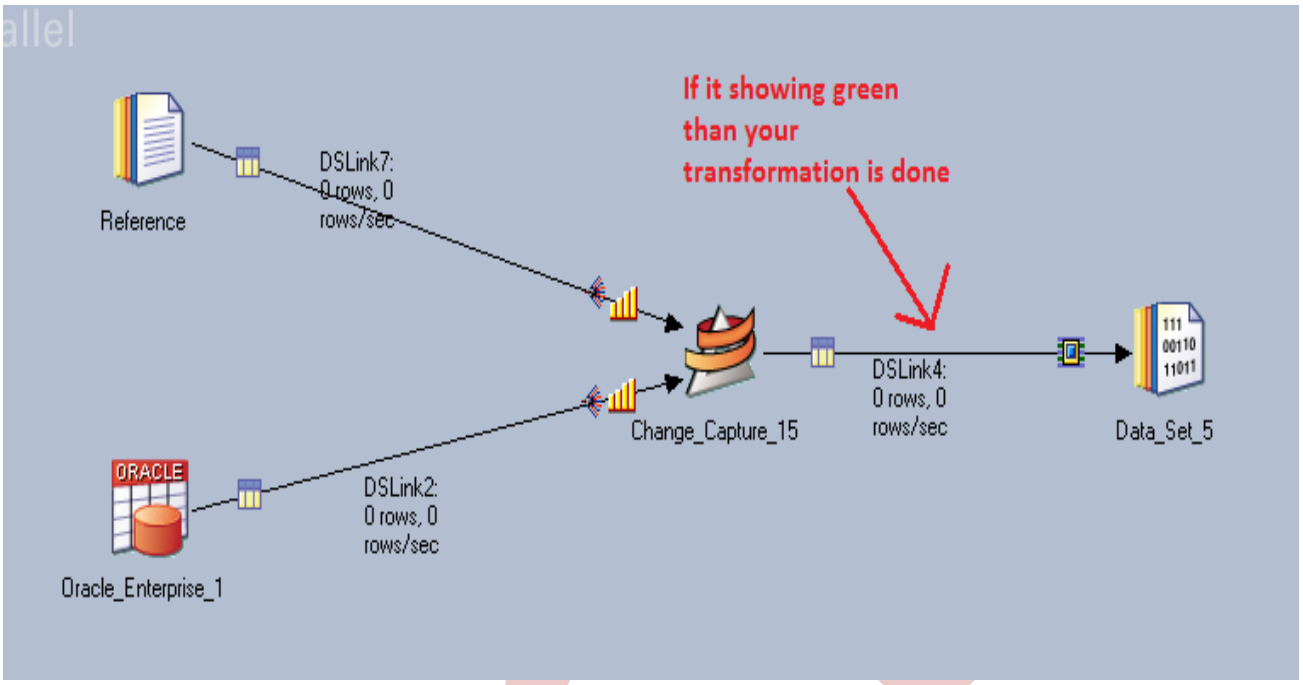
Step-28:- This shown like this and now click on button for compiling our job.



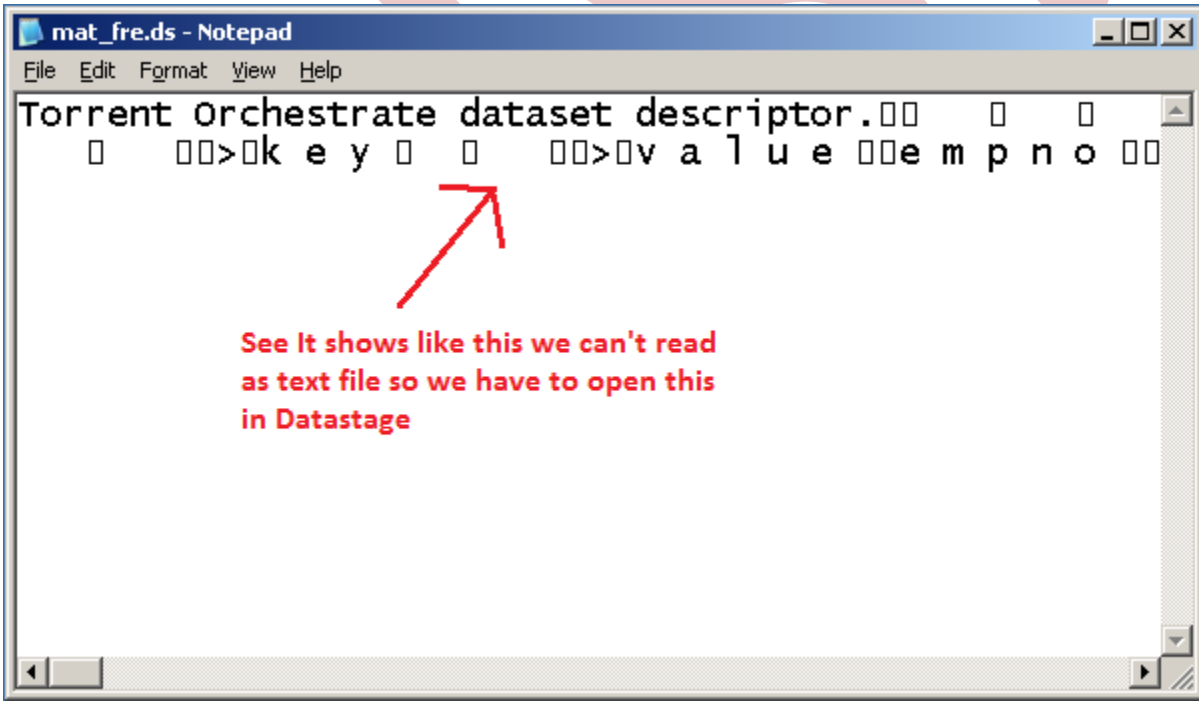
Step-29:- If this shown that means your compilation is done otherwise it shows error and now click on run button or (CTRL+F5).



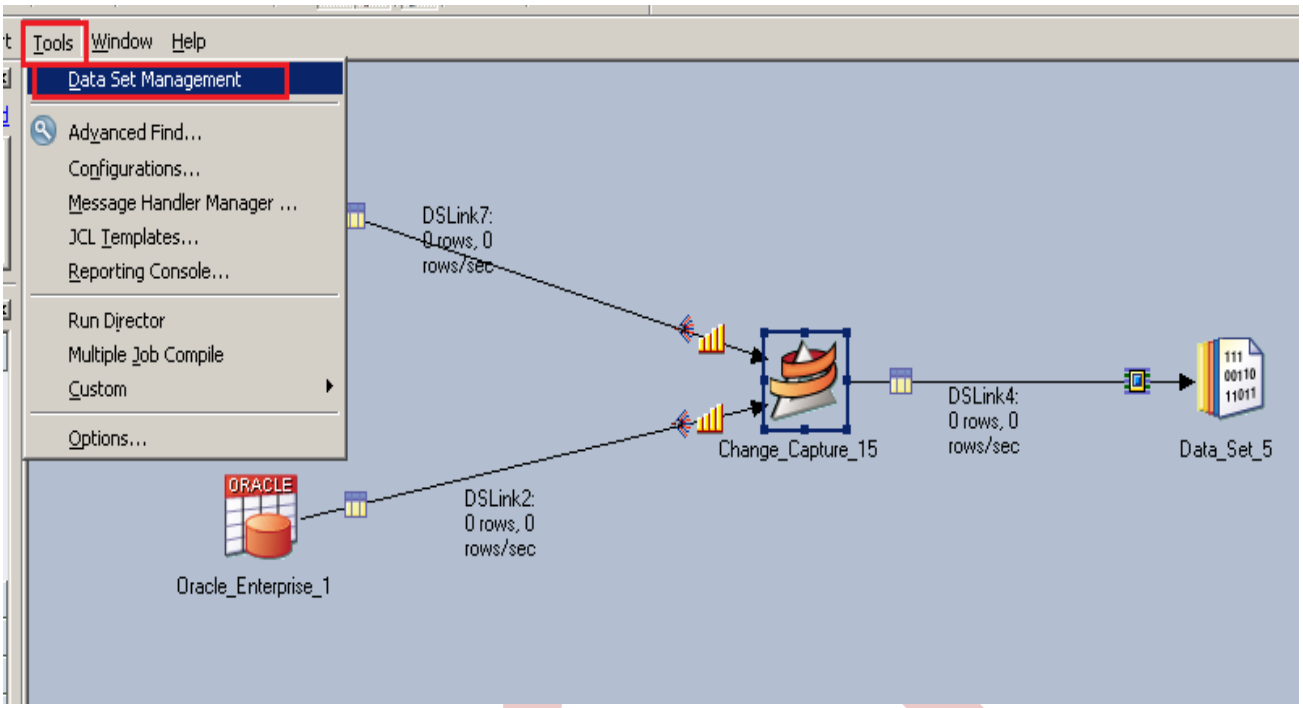
Step-30:- After click on run wait for a while than it shows GREEN line that means your job is successfully done otherwise if it shows RED Line that means not Done and BLUE Line means Under Process.



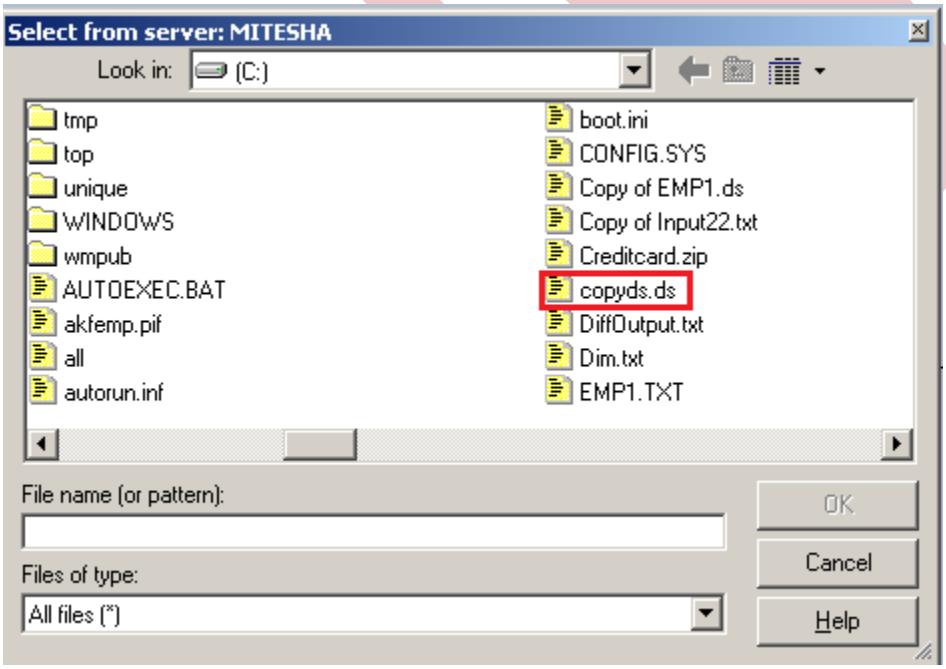
Step-31:- Now see we want to open our dataset output file but it shows like this because dataset file doesn't open on notepad because extension is .DS so we have to run this file in Datastage platform.



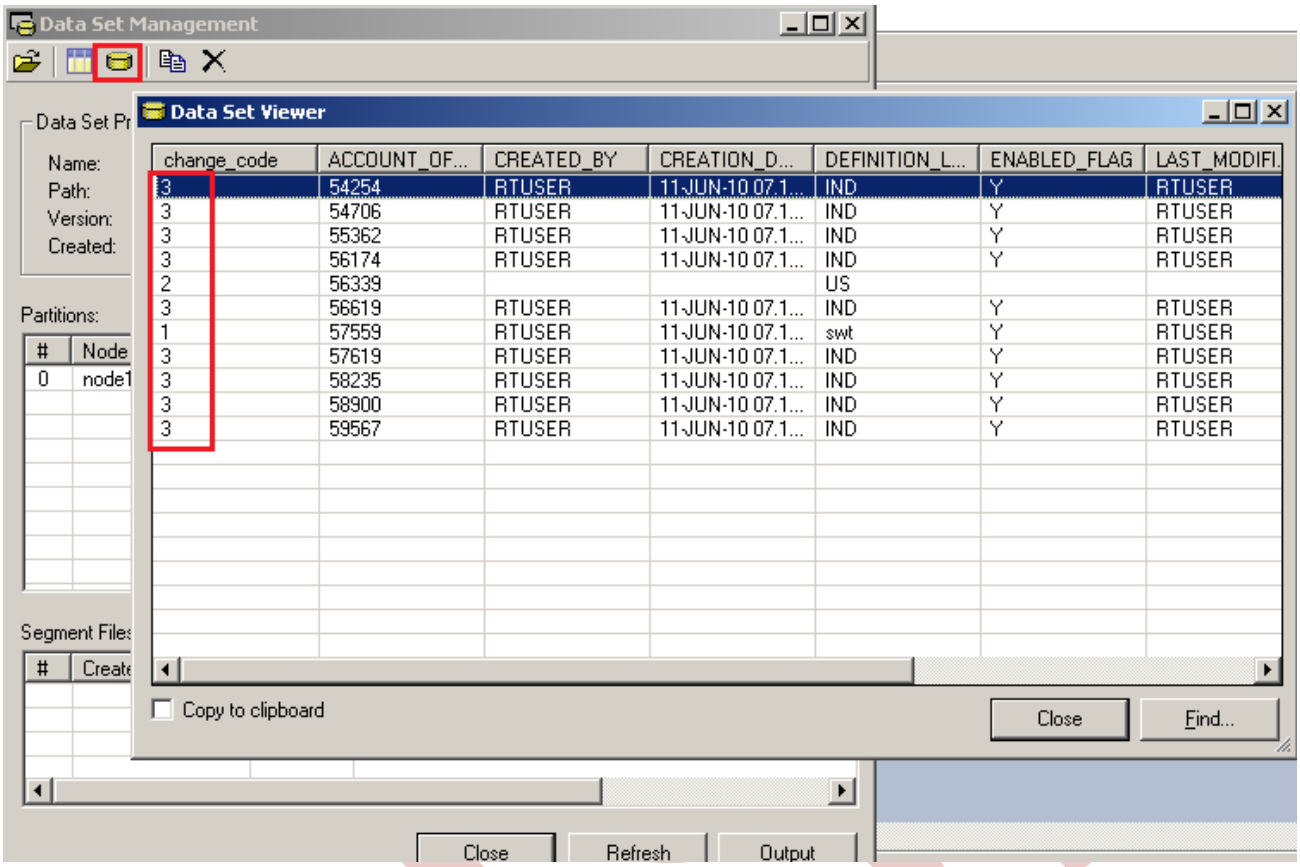
Step-32:- In Datastage designer go to tools option than click in this and choose Data Set Management file and wait for a while after some time there's open an dialog box. than specify your DS file and open.



Step-33:- This is dialog box automatically open 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.


Data Set Management
Data Set Pr
Name:
Path:
Version:
Created:
Partitions:
Node
0 node1
Segment Files
Create
Close
Refresh
Output
Data Set Viewer
change_code ACCOUNT_OF... CREATED_BY CREATION_D... DEFINITION_L... ENABLED_FLAG LAST_MODIFI.
3 54254 RTUSER 11-JUN-10 07.1... IND Y RTUSER
3 54706 RTUSER 11-JUN-10 07.1... IND Y RTUSER
3 55362 RTUSER 11-JUN-10 07.1... IND Y RTUSER
3 56174 RTUSER 11-JUN-10 07.1... IND Y RTUSER
2 56339 US
3 56619 RTUSER 11-JUN-10 07.1... IND Y RTUSER
1 57559 RTUSER 11-JUN-10 07.1... swt Y RTUSER
3 57619 RTUSER 11-JUN-10 07.1... IND Y RTUSER
3 58235 RTUSER 11-JUN-10 07.1... IND Y RTUSER
3 58900 RTUSER 11-JUN-10 07.1... IND Y RTUSER
3 59567 RTUSER 11-JUN-10 07.1... IND Y RTUSER
Copy to clipboard
Close
Find...