Business Intelligence Solution Providers

Specialized in creating talent resource pool

IBM DataStage 8.5 Beginner's Guide - Lab#11 Aggregate Transformation

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 describes in step by step manner how to aggregation at different levels. Data is coming from multiple flat files sources and loading into Target tables **_Join our professional training to learn from Experts._**

Transformation	Aggegration Transformat						
Source Name	Flat File, Flat File						
Target Name	Oracle						
JOIN Condition	SRC_BRANCH_CD.BRANCH_CD=SRC_CREDITCARD.BRANCH_CD						
Aggegration Condition	GroupBy(Branch_Name, Branch_cd), SUM(charge_off_bal)						

| Flat File Source Details			Flat File2 Source Details			Oracle Details Target		
Entity Name	Data Types	Is Null	Entity Name	Data Types	Schema	Destination Entity Name	DataTypes	Is Null
N_ACCT_SKEY	NUMBER(14,0)	N	BRANCH_CD	NUMBER(5)	src_creditcard	BRANCH_CD	NUMBER(14,0)	N
COMMON_COA_ID	NUMBER(14,0)	y	CREATED_BY	VARCHAR2(30	src_creditcard	BRANCH_NAME	VARCHAR2(100	y
CUSTOMER_ID	NUMBER(14,0)	y	CREATION_DATE	TIMESTAMP(6	src_creditcard	SUM	NUMBER(14,0)	y
GL_ACCOUNT_ID	NUMBER(14,0)	y	DEFINITION_LANGUAGE	VARCHAR2(4)	src_creditcard			
ORG_UNIT_ID	NUMBER(14,0)	y	ENABLED_FLAG	VARCHAR2(1)	src_creditcard			
PRODUCT_ID	NUMBER(14,0)	y	LAST_MODIFIED_BY	VARCHAR2(30	src_creditcard			
AMRT_TERM	NUMBER(14,0)	y	LAST_MODIFIED_DATE	TIMESTAMP(6	src_creditcard			
V_CCY_CODE	VARCHAR2(20 B	y	LEAF_ONLY_FLAG	VARCHAR2(1)	src_creditcard			
CHARGE_OFF_BAL	NUMBER(14,0)	y	BRANCH_DISPLAY_CD	VARCHAR2(10	src_creditcard			
COMPOUND_BASIS_CD	NUMBER(14,0)	y			src_creditcard			
CUR_DELQ_BAL	NUMBER(14,0)	y			src_creditcard			
IDENTITY_CODE	NUMBER(14,0)	y			src_creditcard			
ISSUE_DATE	VARCHAR2(100	y			src_creditcard			

Aggregator stage:-

The Aggregator stage is a processing stage. It classifies data rows from a single input link into groups and computes totals or other aggregate functions for each group. The summed totals for each group are output from the stage via an output link.

Source Data Model:-
We using this financial data model and we perform operation on Src_Branch_cd and SRC_Creditcard

Steps to apply Aggregate 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.
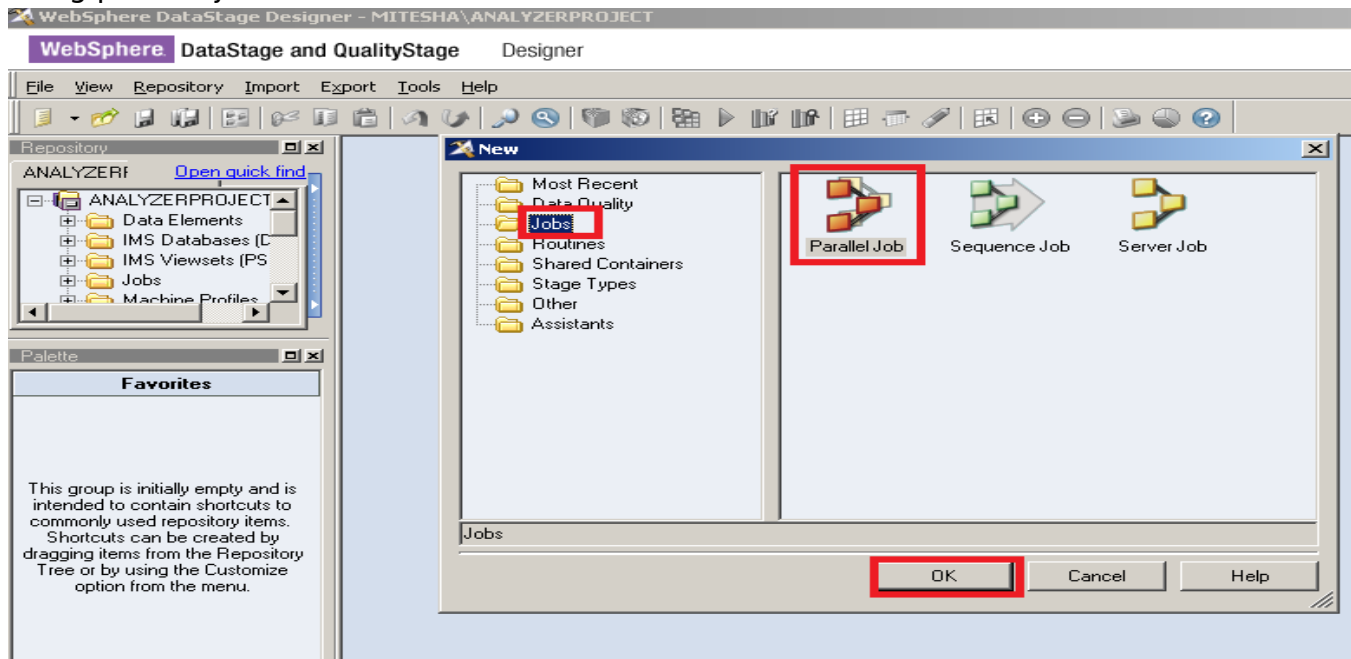

Step-3:- First go on File Palette then scroll the cursor and choose Sequential File and drag it to Parallel pane and rename it. This file is used for input data.
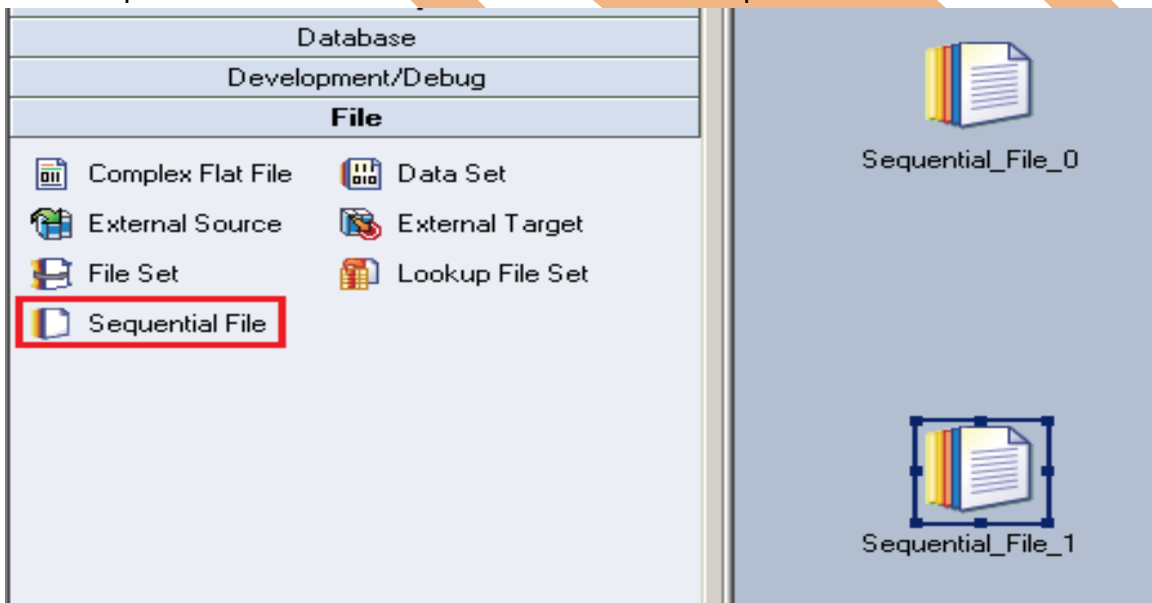

Step-4:- Choose Oracle Enterprise from database Palette and drag it to parallel pane. This is used for loading data into oracle from flat files.
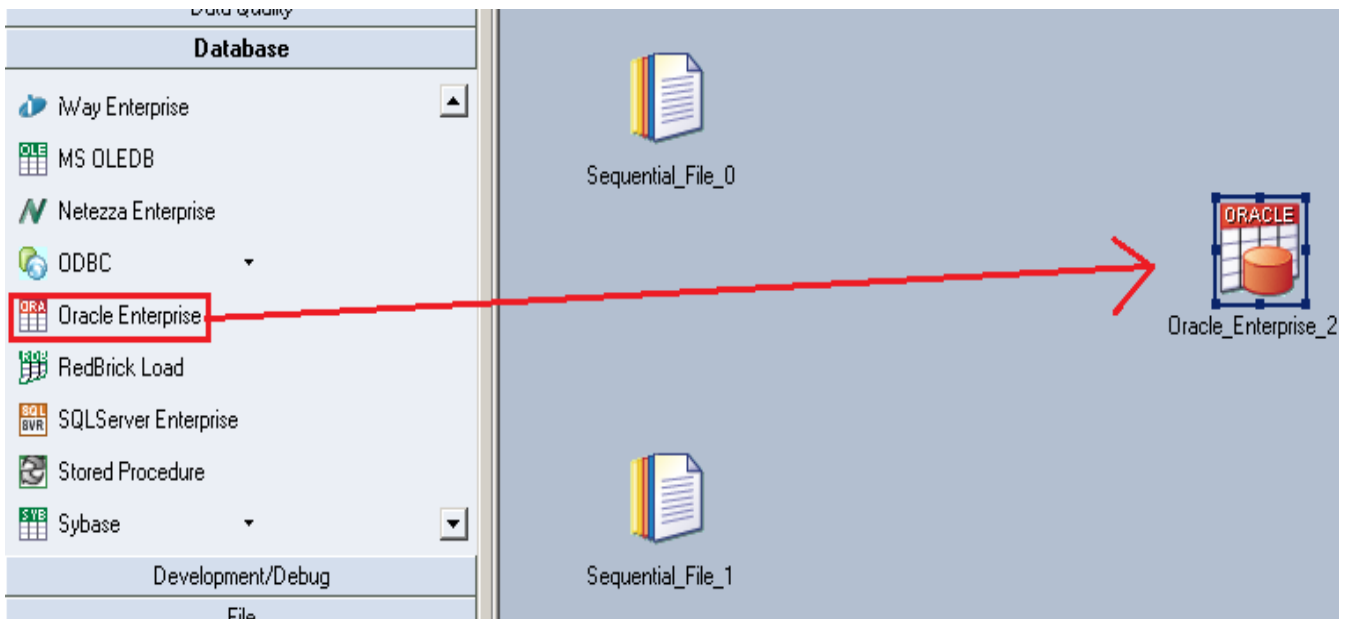
Step-5 Choose an Aggregator and join stage from processing palette. This aggregator process is used when we want to aggregate our data and join stage is used for joining two tables based on key column.
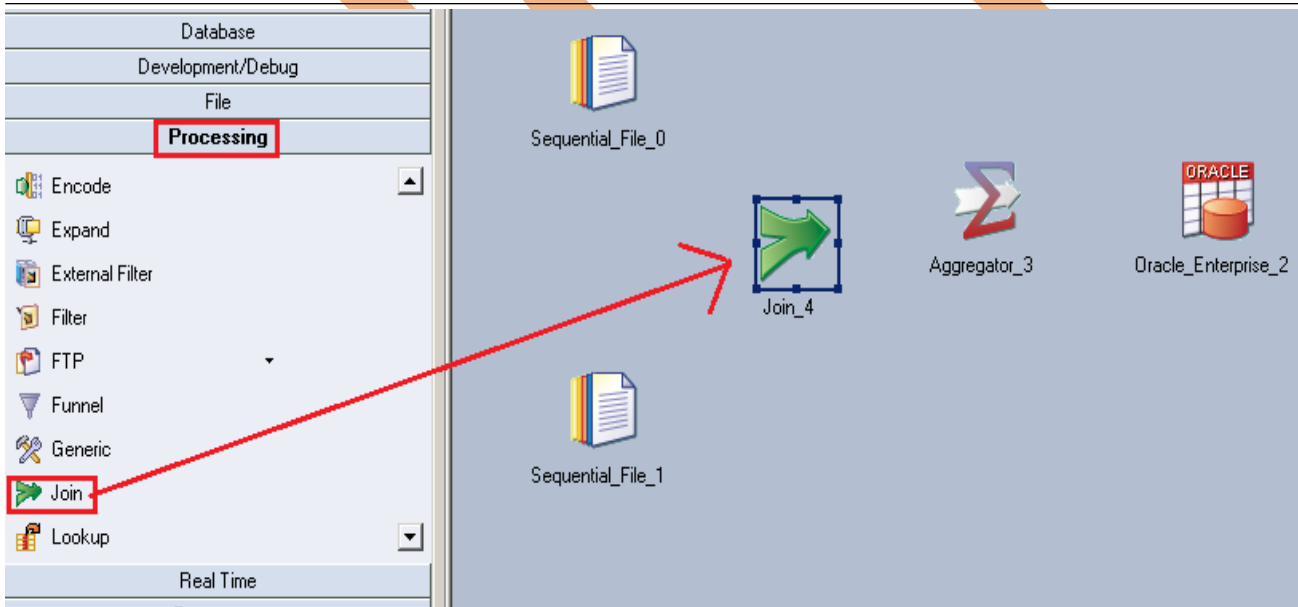
Step-6:- Now make an connection with these stages that's why we have to use this link stage for connecting these stages.
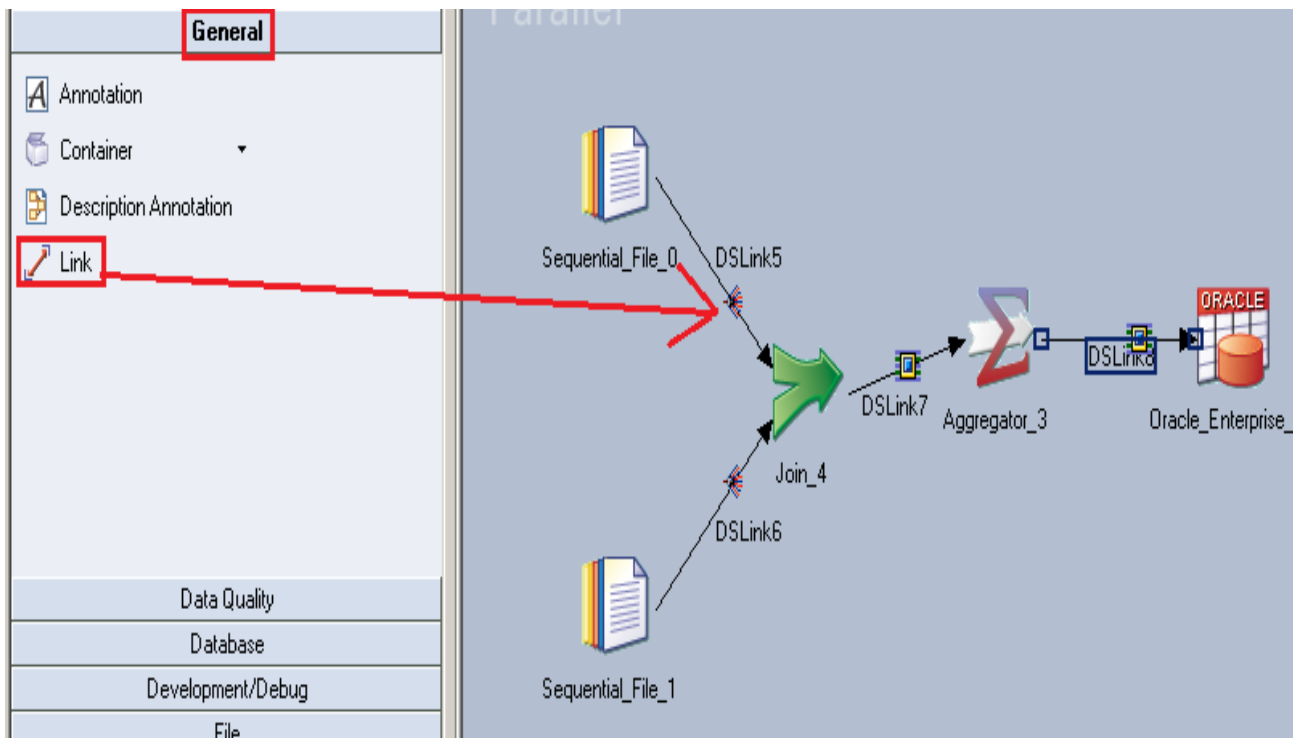
Step-7:- Double click on Input file and browse an input file that can be .txt, .csv any test file.
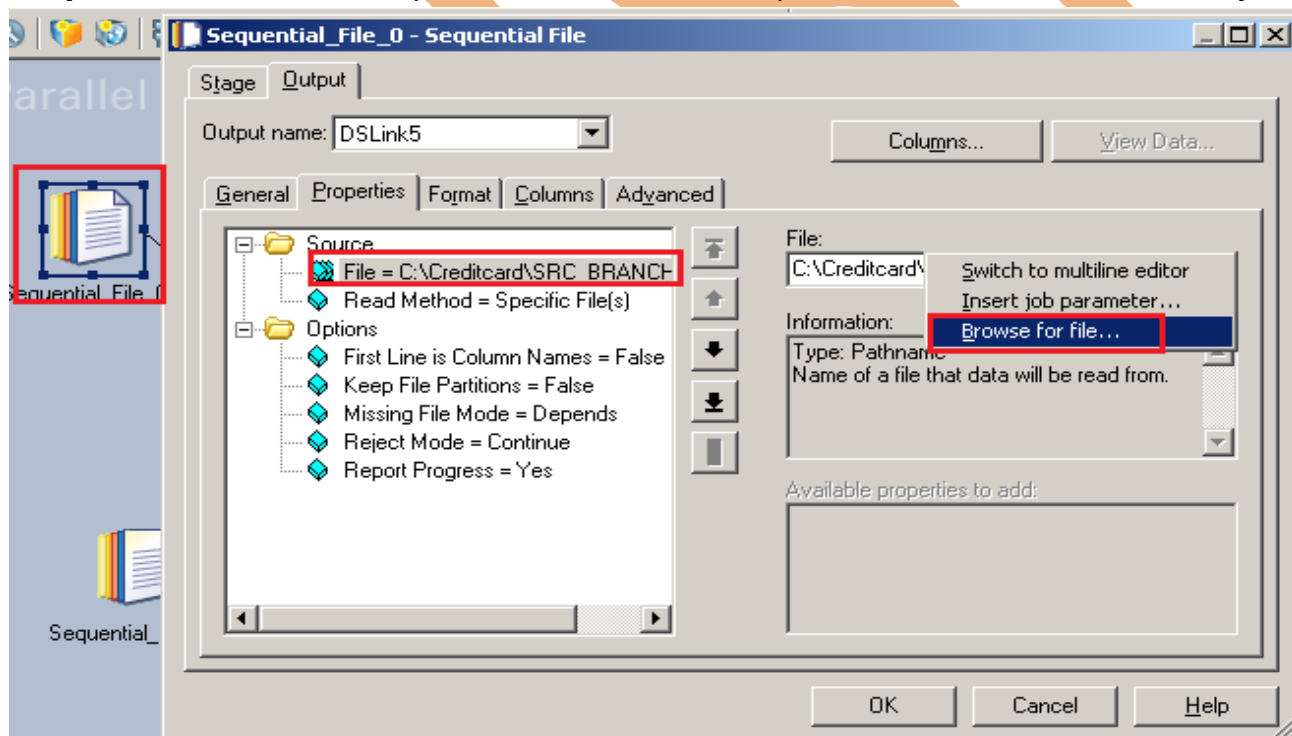
Step-8:- Go to Format tab and Remove double Quote because we don't need them as it comes from source flat files. Click OK.
Note:- If you use date as a input than specify here which type of format you used
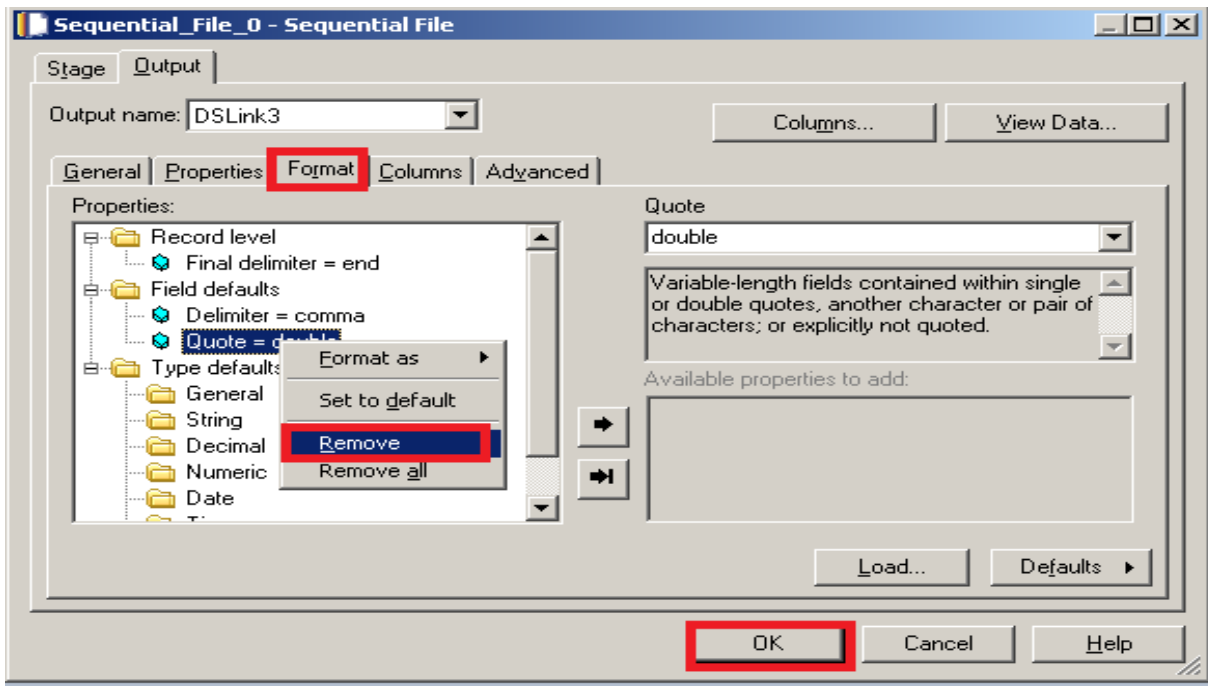
Step-9 :- Here I'm using this date format so we specify here than we able to load the date from flat file.
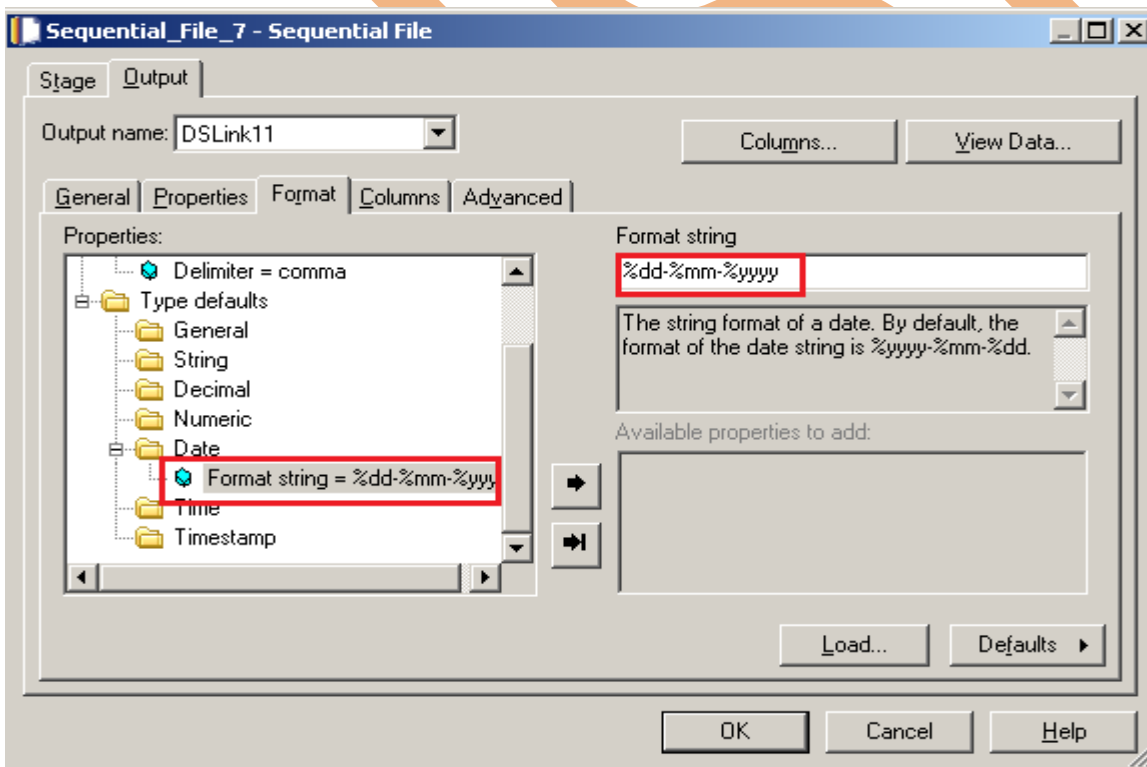

Step-10 Go to Columns tab and enter same column names as declared in Input file and put data type, length then click on View Data. This will show your Input Data.
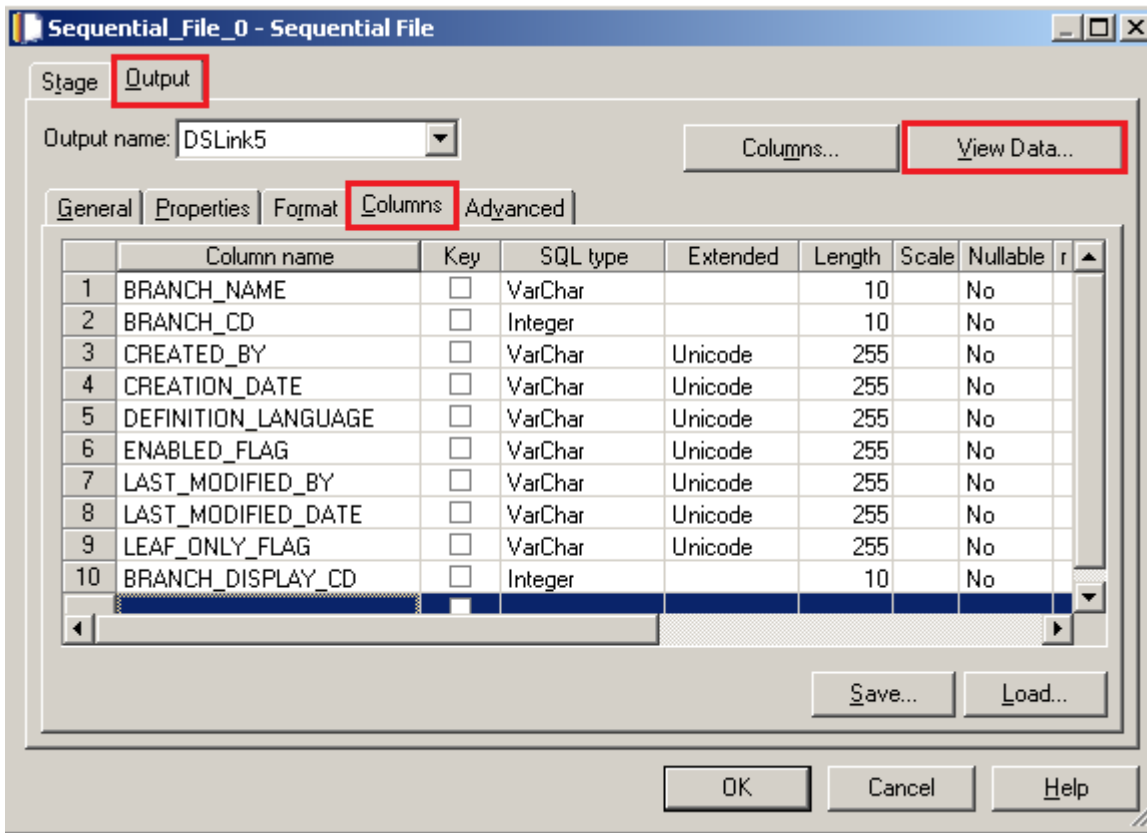
Step-11:- Now click on View Data. Your data view should be as given below, it shows the you made successfully connection between input file and DataStage
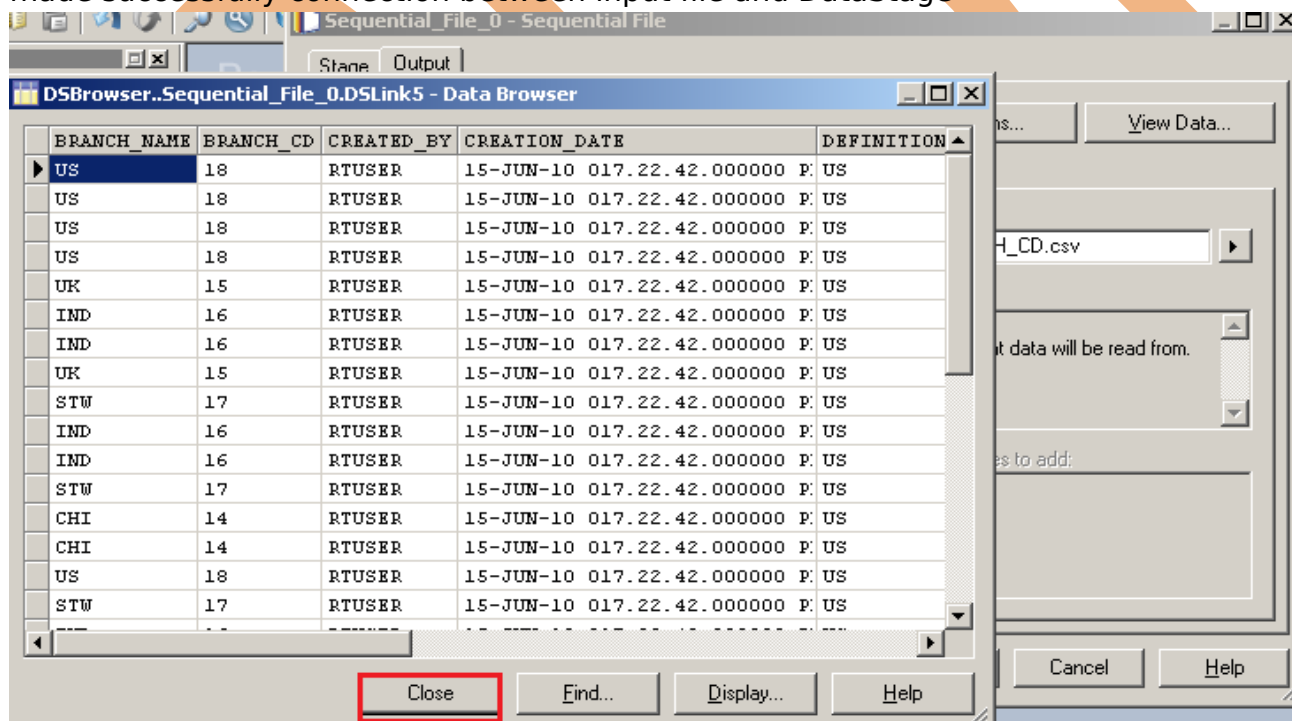
Step-12:- After choosing file, select 'True' from First line in column Names dropdown for removing first line from Input file.
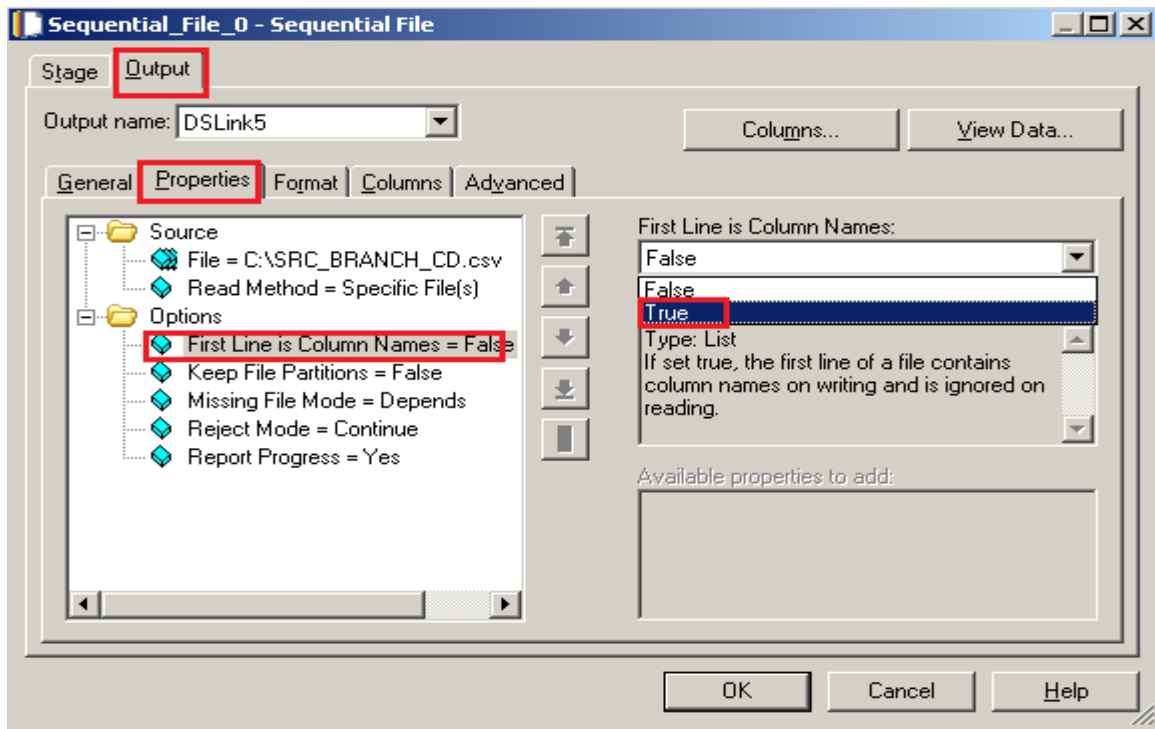
Step-13:- Double click on second Input file and browse an input file that can be .txt, .csv any test file and apply same process as apply in previous input file.
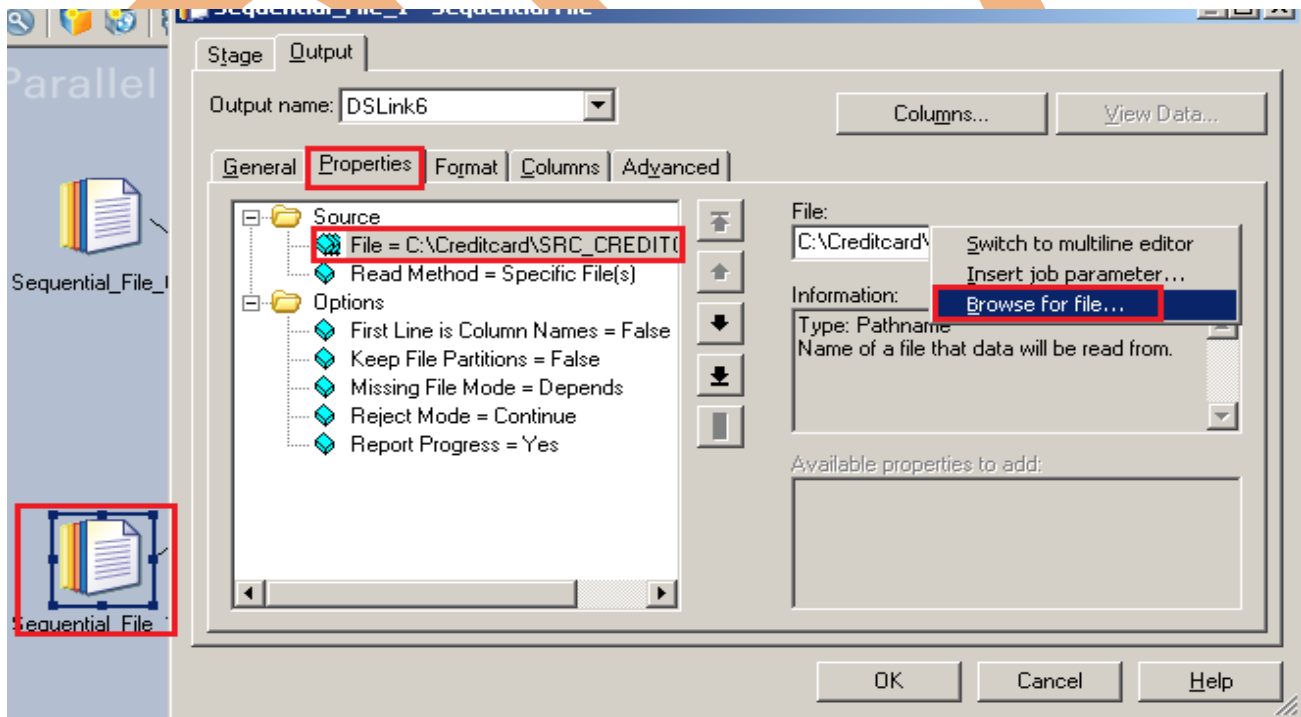
Step-14:- Go to Format tab and Remove double Quote because we don't need them then Click on OK.

Note:- If you use date as a input than specify here which type of format you used
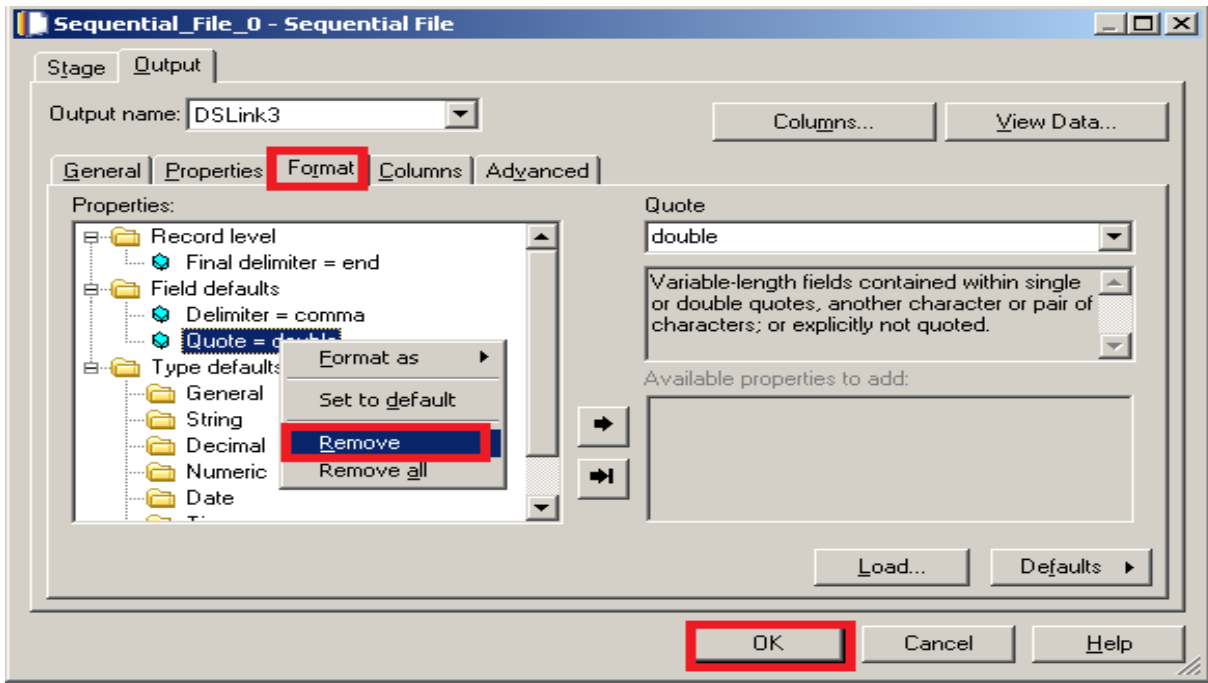
Step-15 :- Here I'm using this date format so we specify here than we able to load the date from flat file.
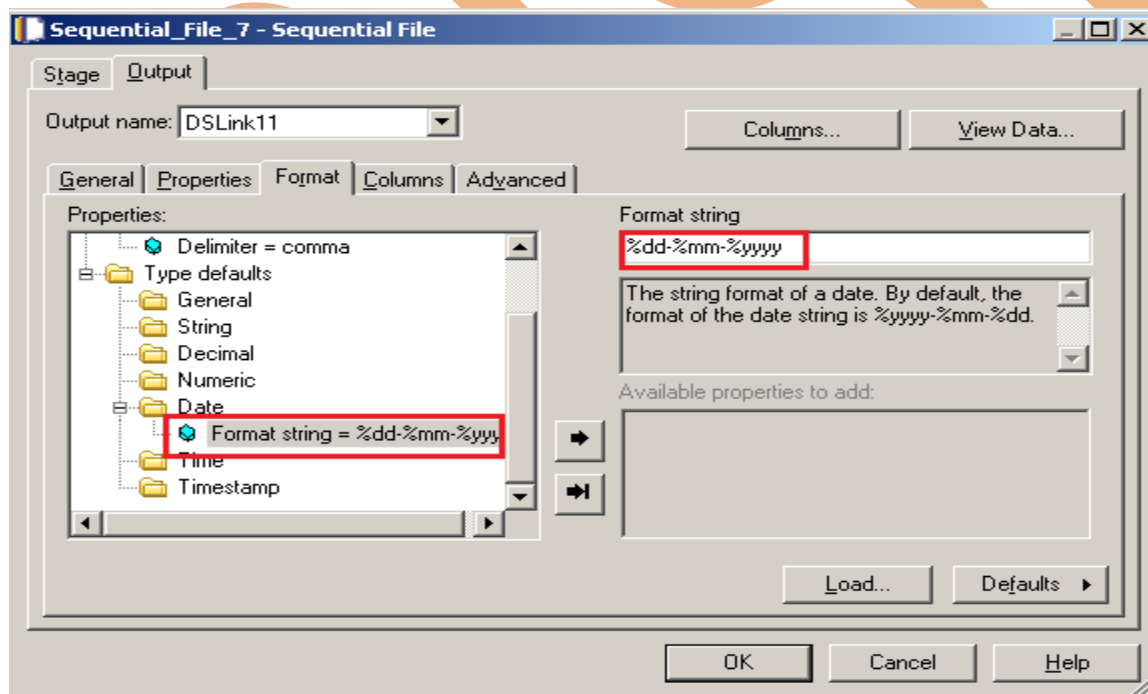
Step-16:- Go to Columns tab and enter same column names as declared in Input file and put data type, length then click on View Data. This will show your Input Data.
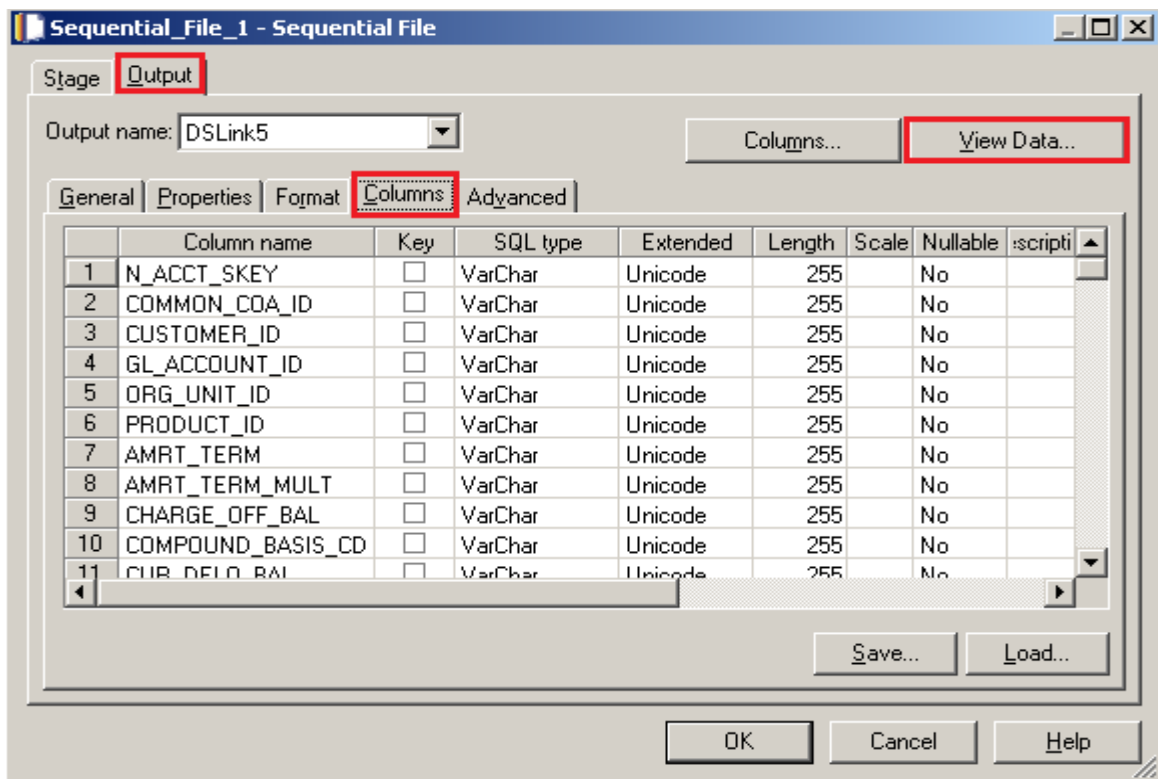
Step-17:- Now click on View Data. Your data view should be as given below, it shows the you made successfully connection between input file and DataStage
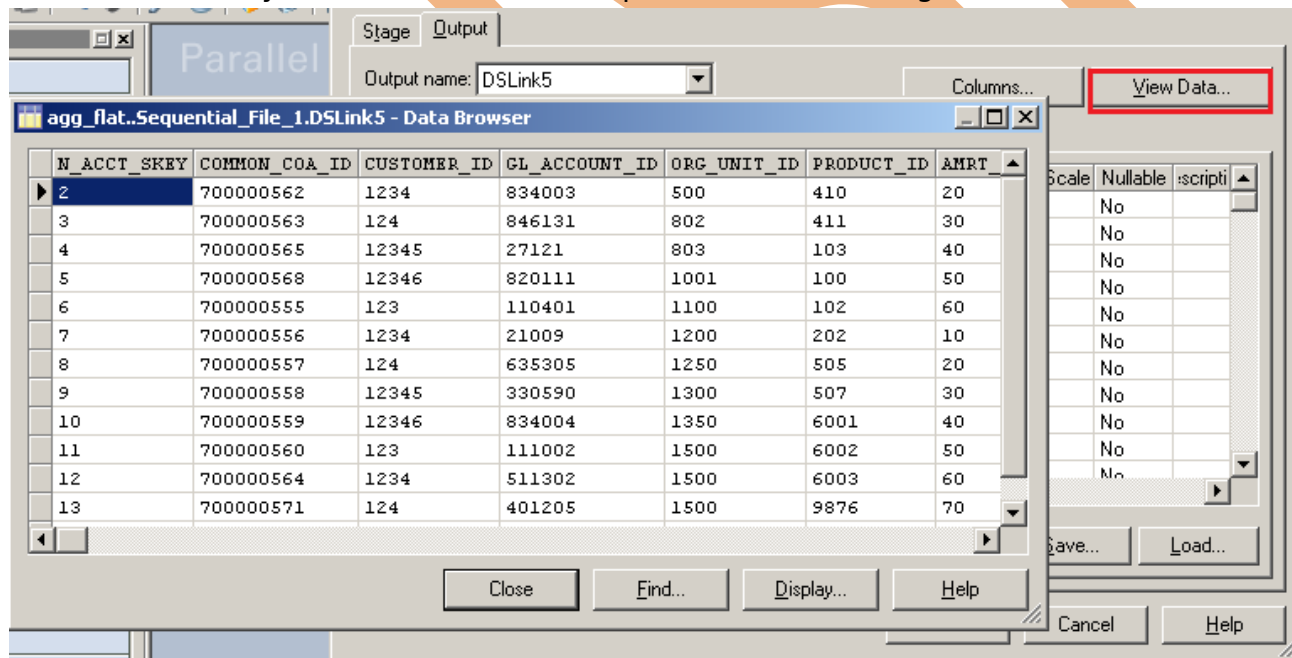
Step-18:- After choosing file, select 'True' from First line in column Names dropdown for removing first line from Input file.
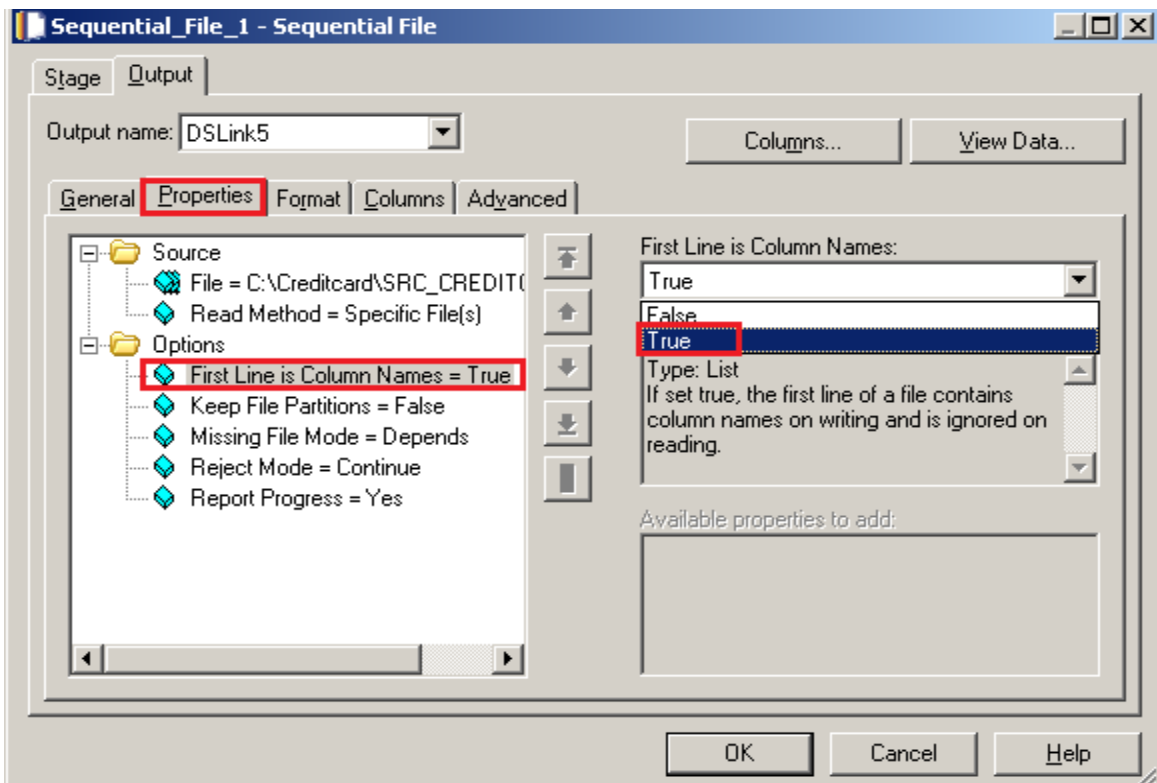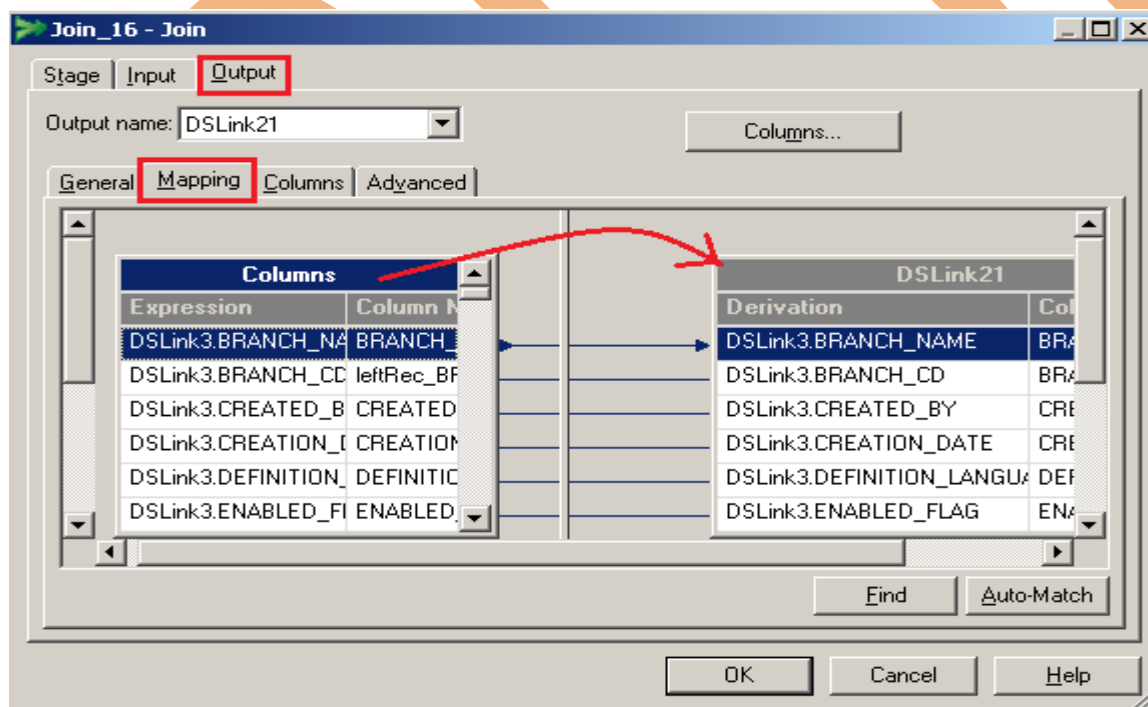
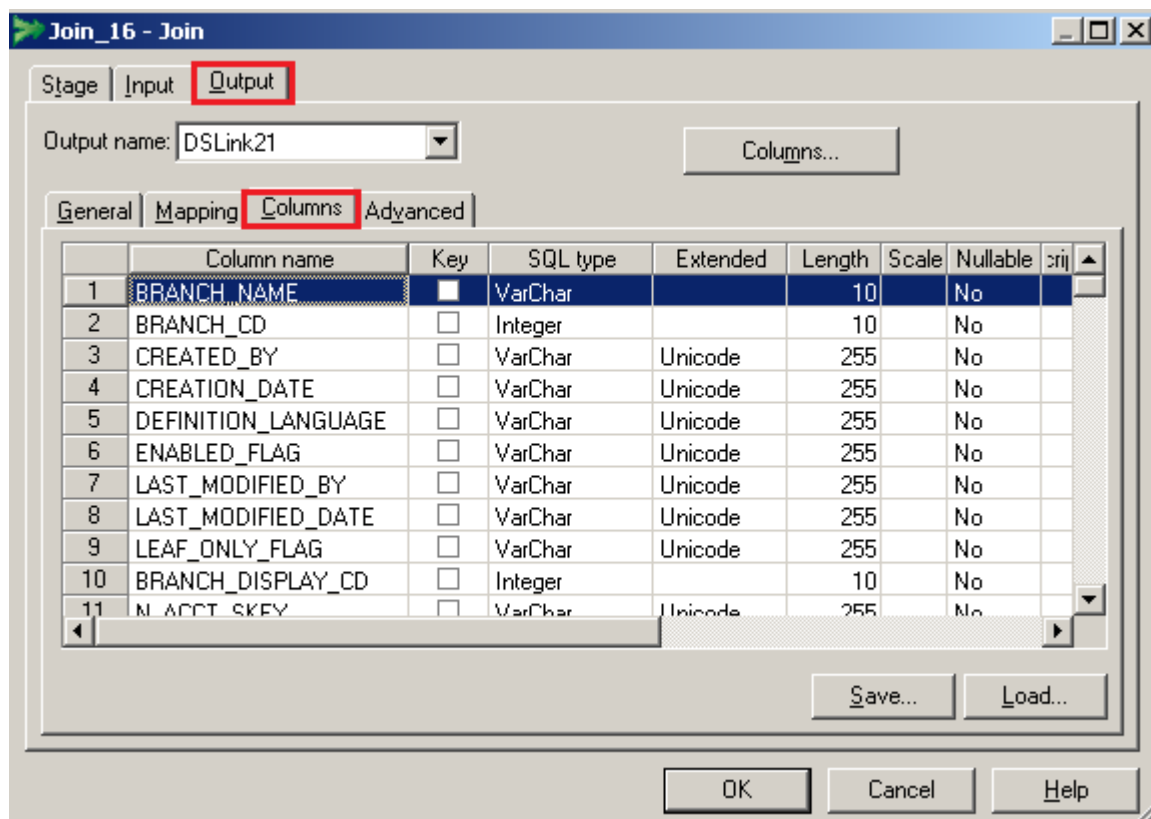**Step-19:-** Now double click on Join stage and make an key column its automatically retrieve by datastage.
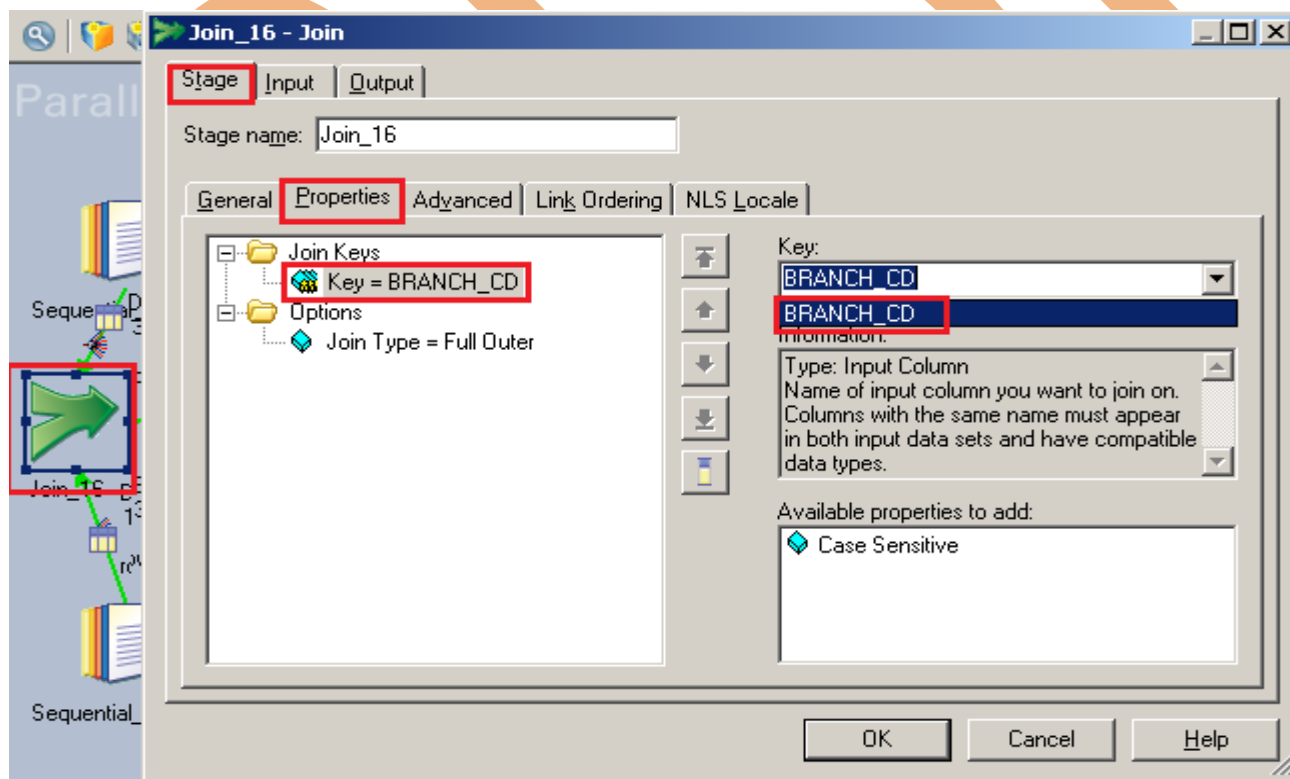
**Step-20:-** Now there are basically four types of joins, in this we used inner join.
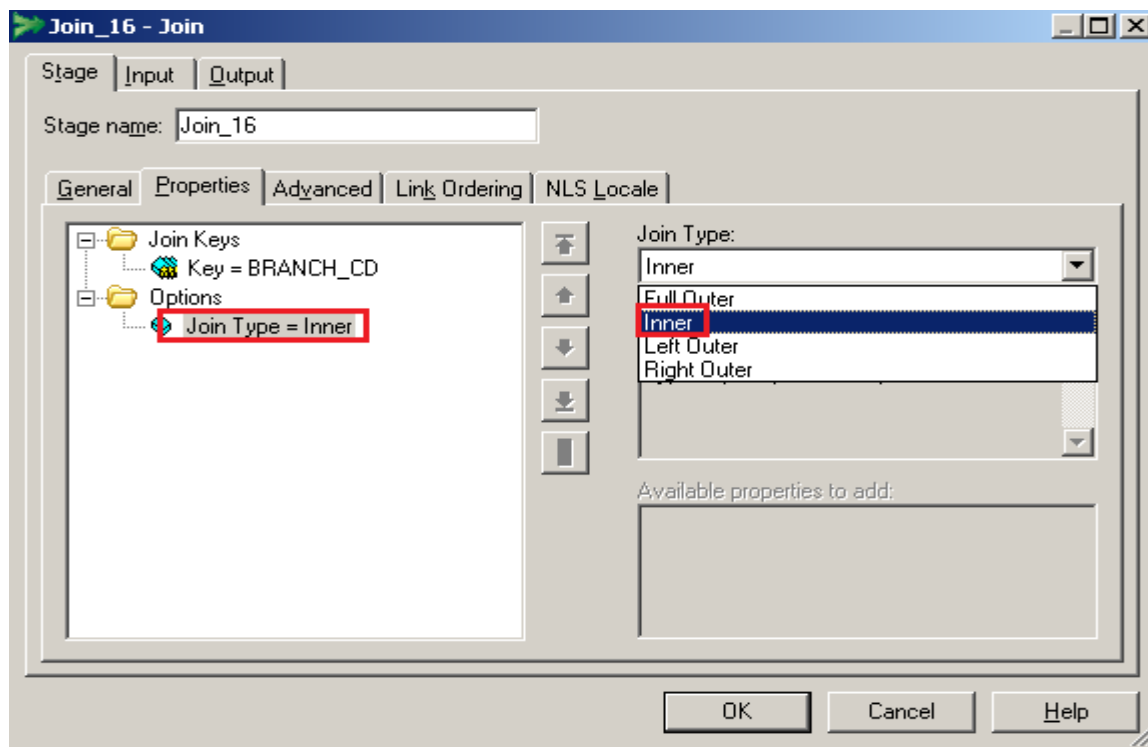
Step-21:- Now Go to output tab and create mapping, select all columns and drag and drop in output space.
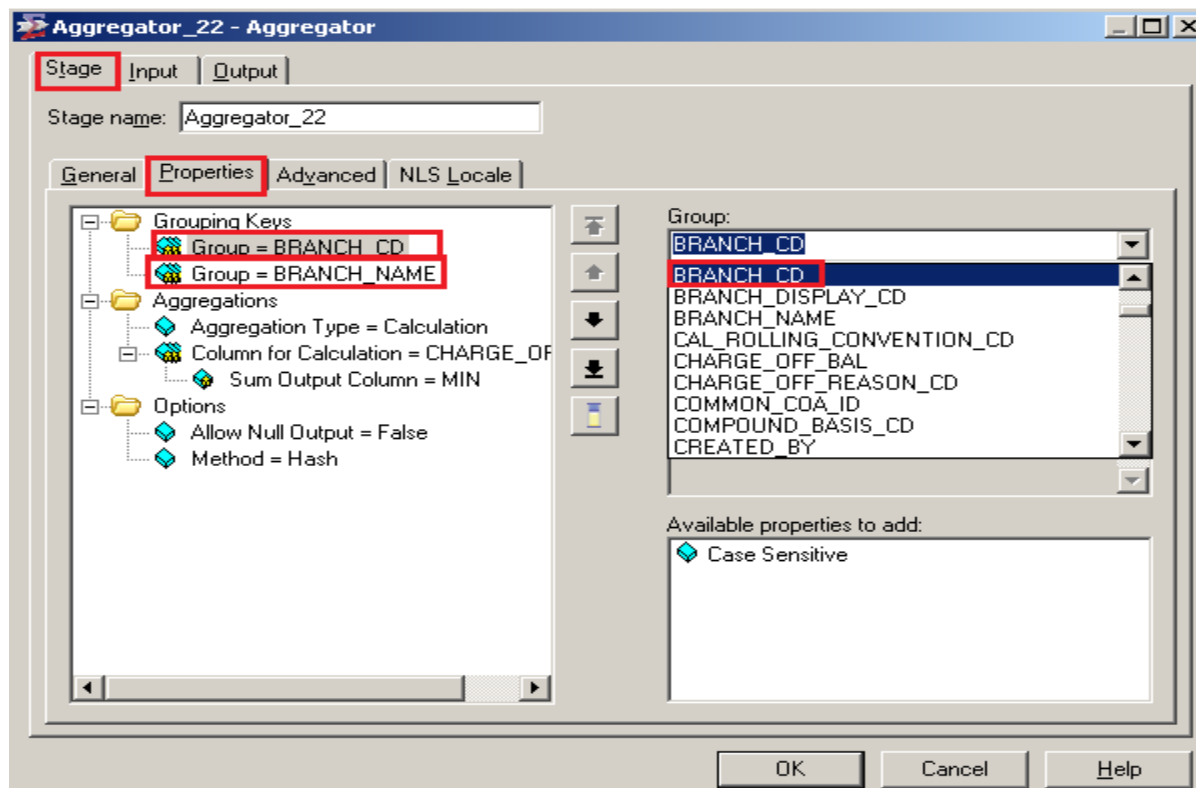

**Step-22:-** Click on Input and see all columns are available or not and click on OK.

Step-23:- Now go to aggregator stage and see here group and calculation space available, now here choose group column to be applied aggrrgataion.
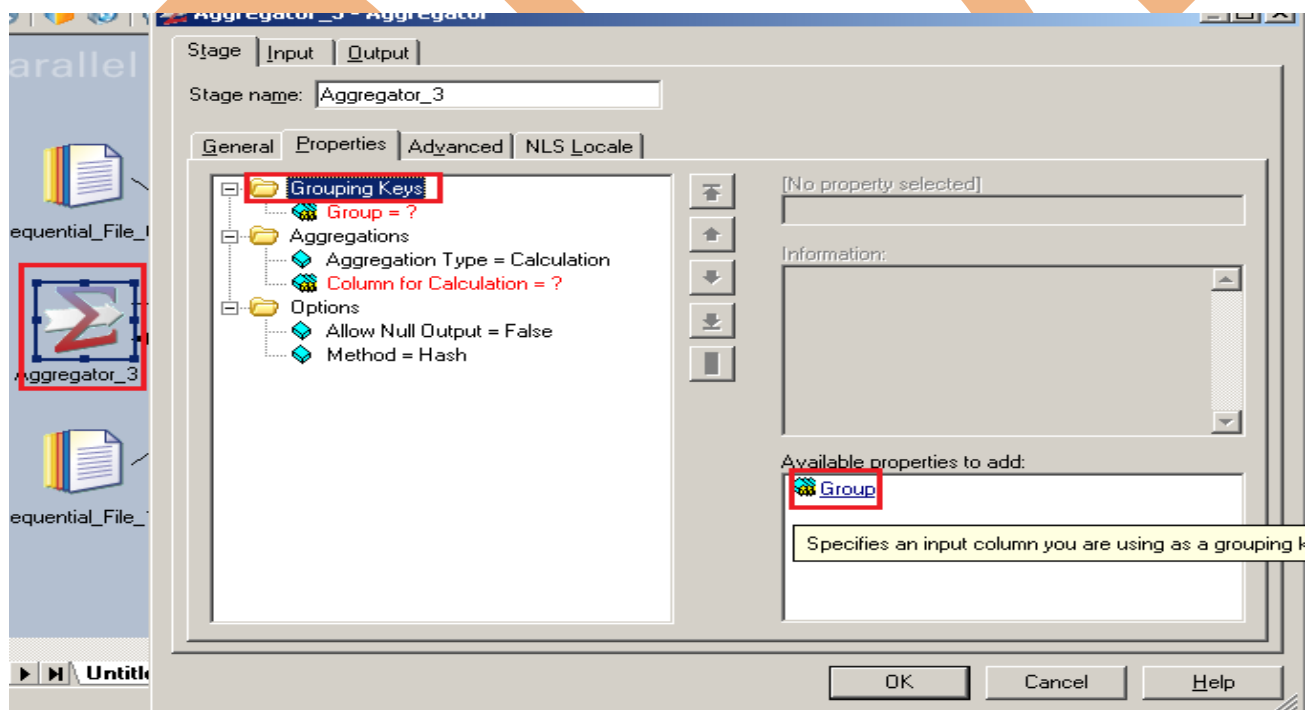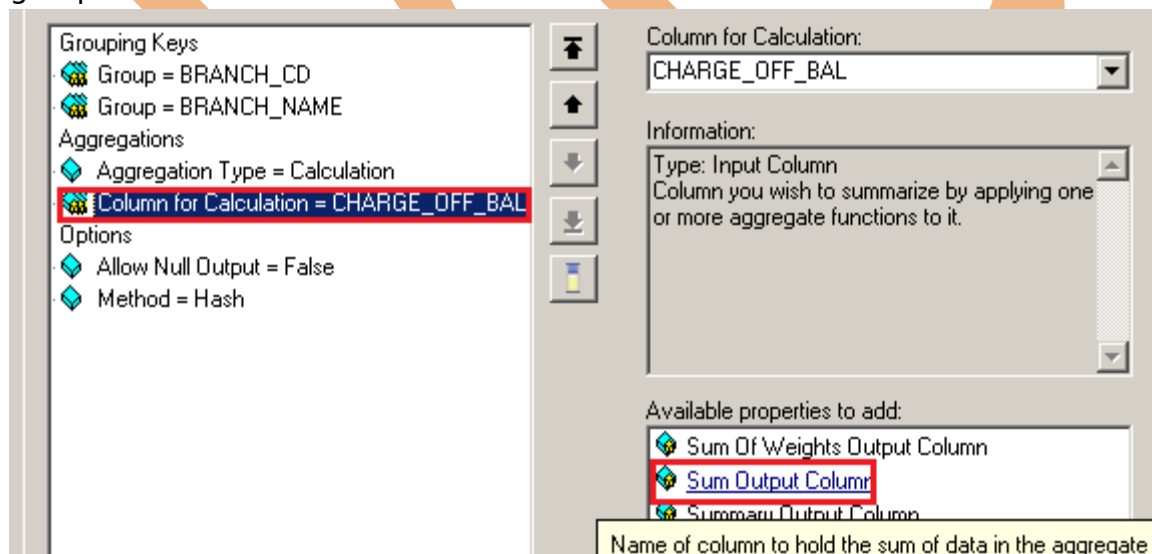

**Step-24:-** see here we choose two columns as a group column.

Step-25 :- We want to sum all charge_off_bal as are requirement based on Brach_Name, choose this column and also select which operation you want to perform and give column name for that particular calculation.
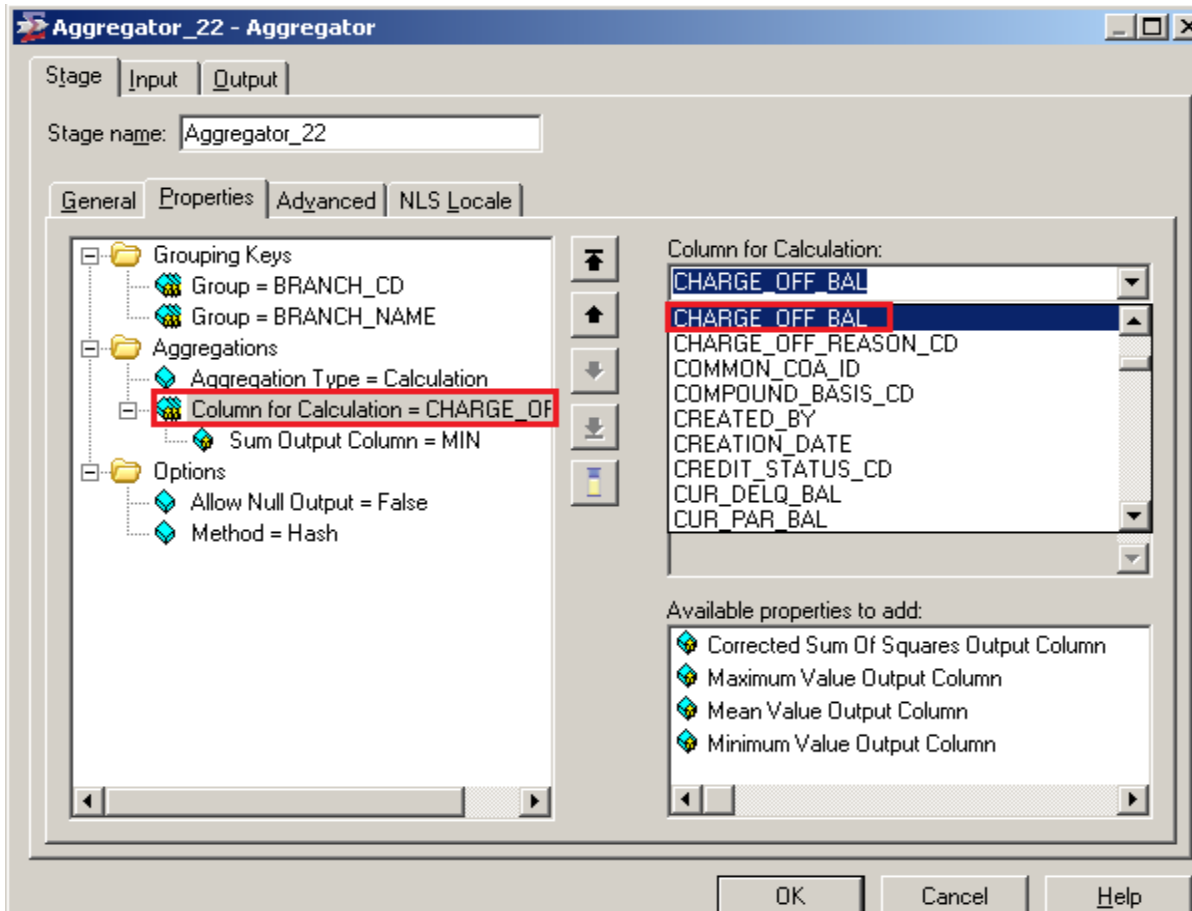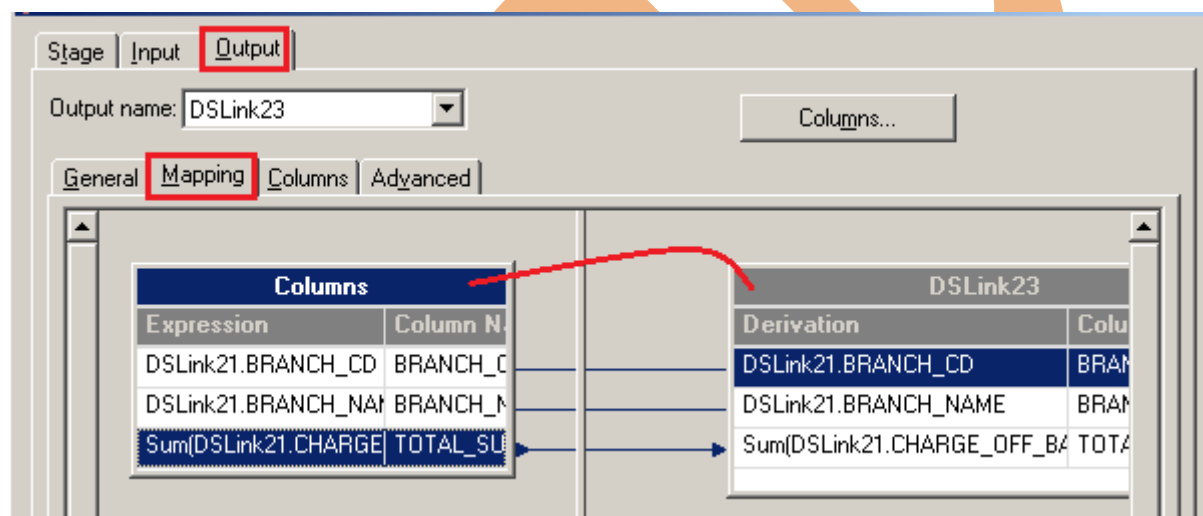
Step-26:- See here we select sum output column that means sum all charge_off_cal based on group column name.
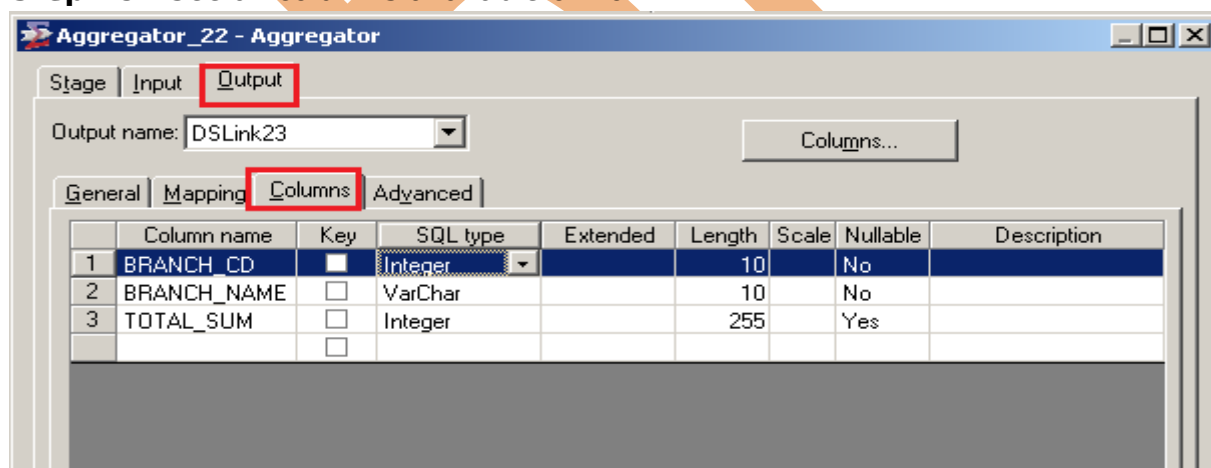

Step-27:- Put output column name for that result.
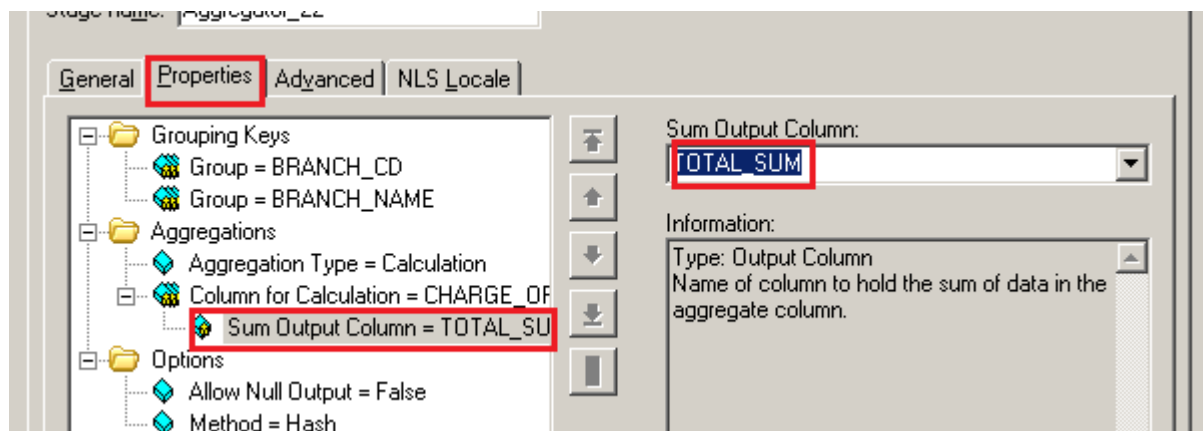
**Step-28:-** Now create mapping into this go to output tab drag and drop all columns.

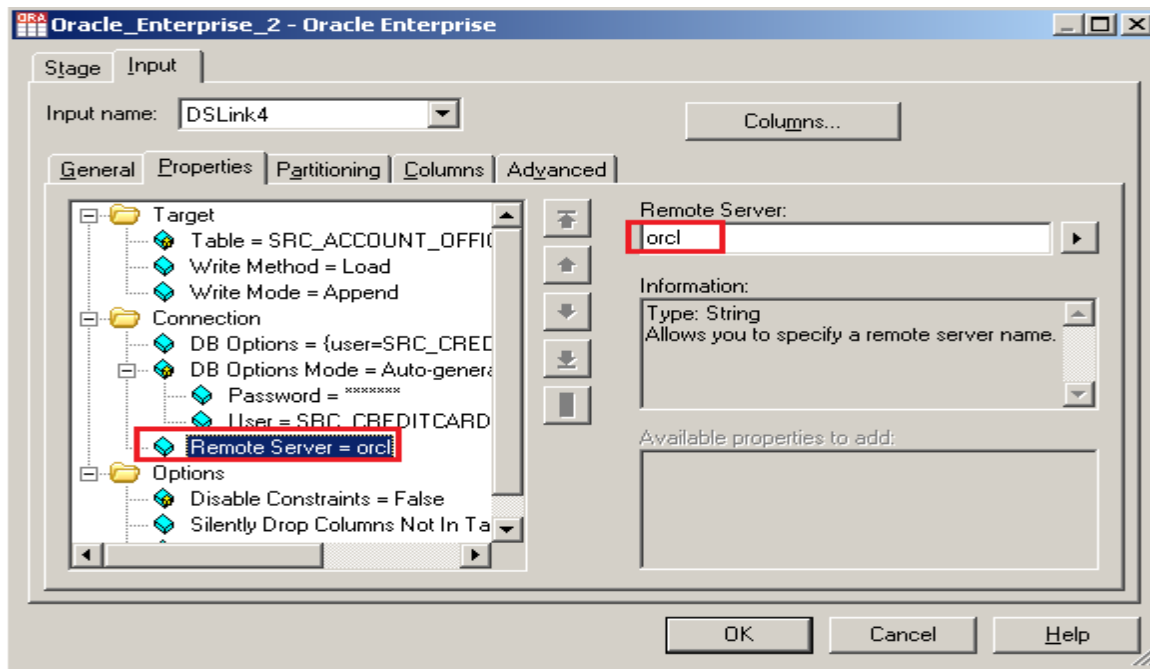**Step-29**:- See all columns available or not.

Step-30 :- Now double click on Oracle enterprise stage as given below. Specify the output table. There are multiple options available, if we directly want to load data then we simple use write method as load otherwise we manually put queries and also we have to specify oracle username/password.
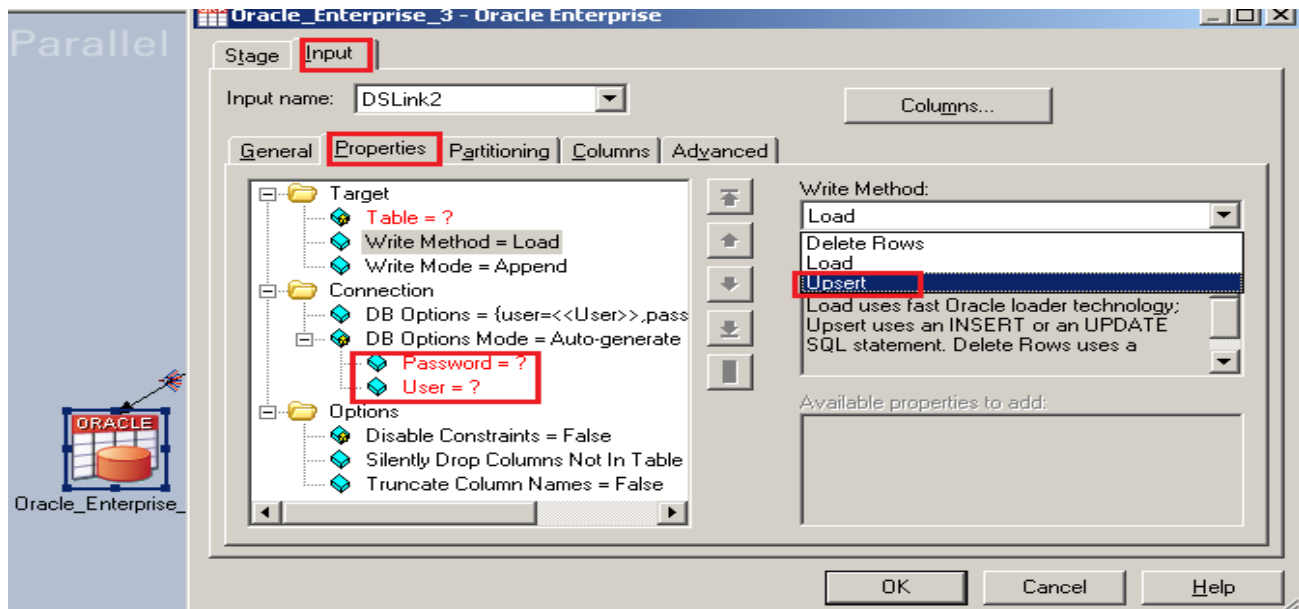

Step-31:- There is one more option here once you click on connection tab then remote server options are popup then we have to specify our server name of oracle.
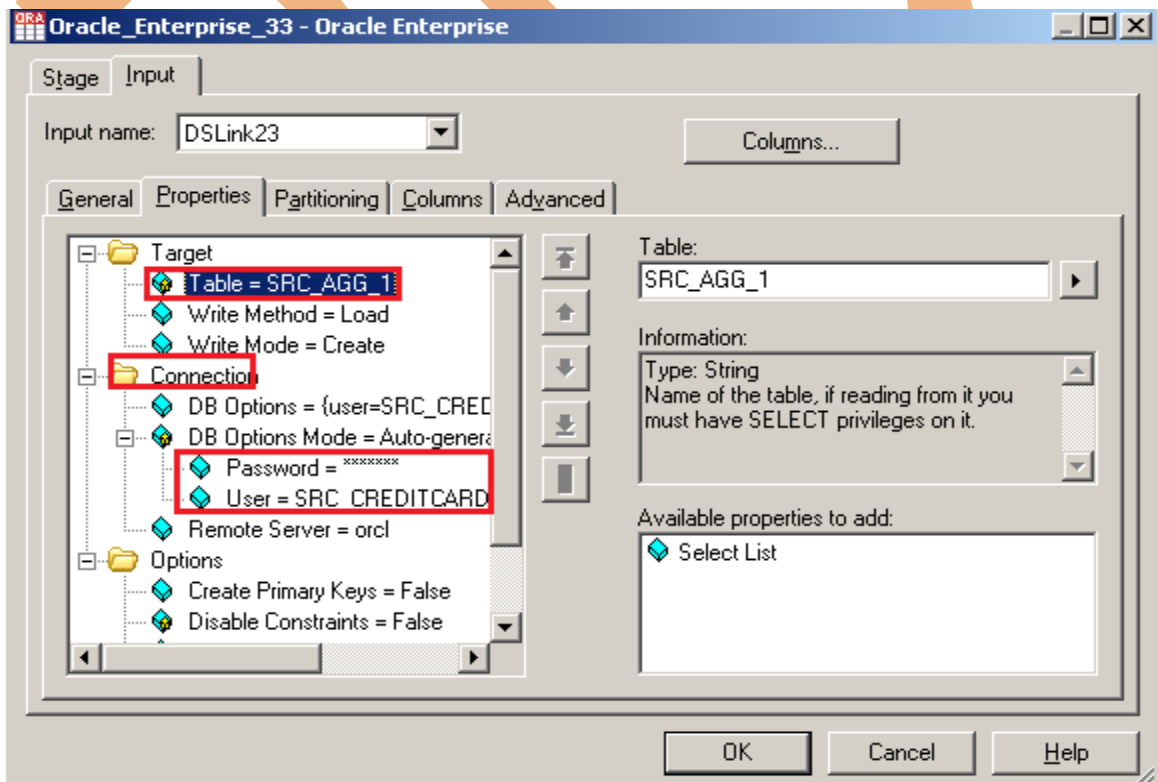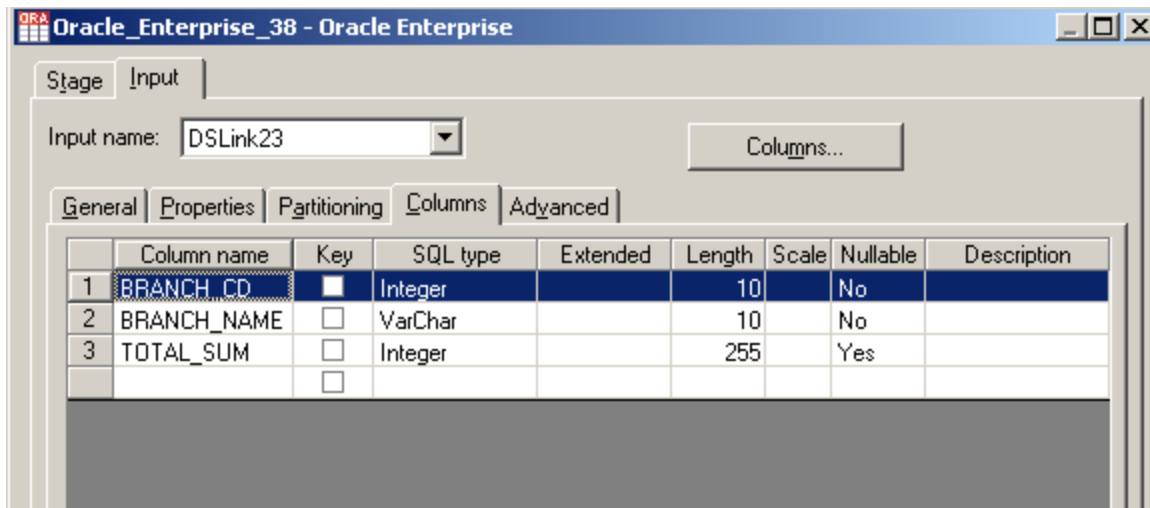
Step-32:- See below, my SID is ORCL and username/password and table name. Ensure that you specify correct information then only you will be able to load the data into table.
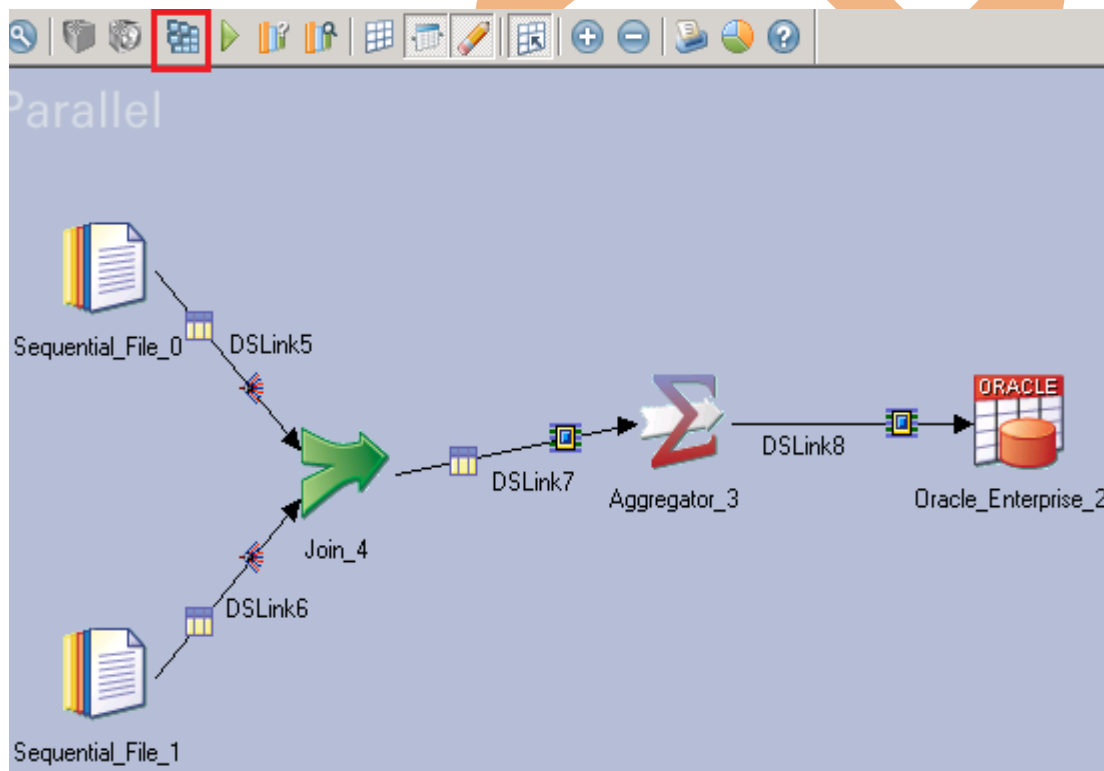

Step-33:- In the Write Mode we select Create, it creates a table in Oracle as Target table.
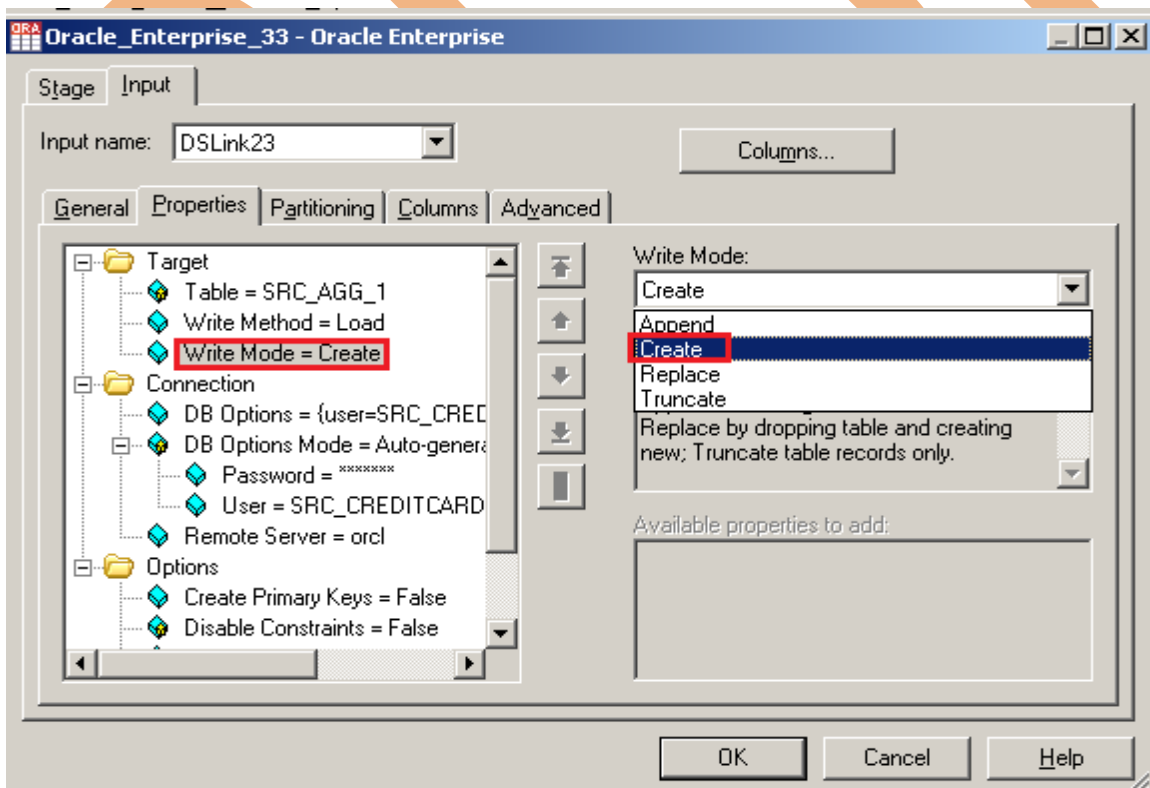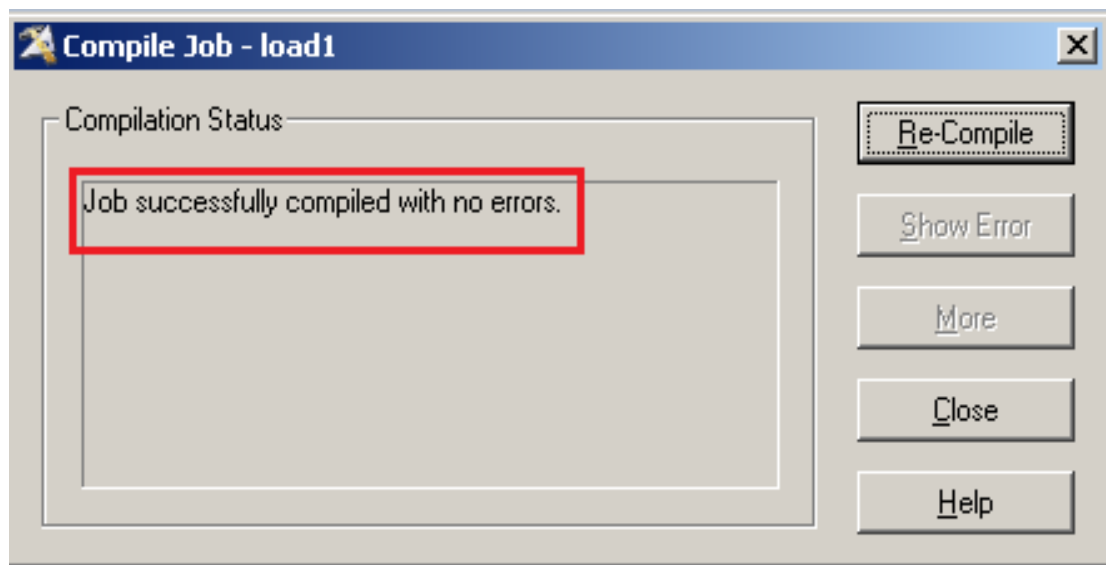

Step-34:- See below, all columns available or not.
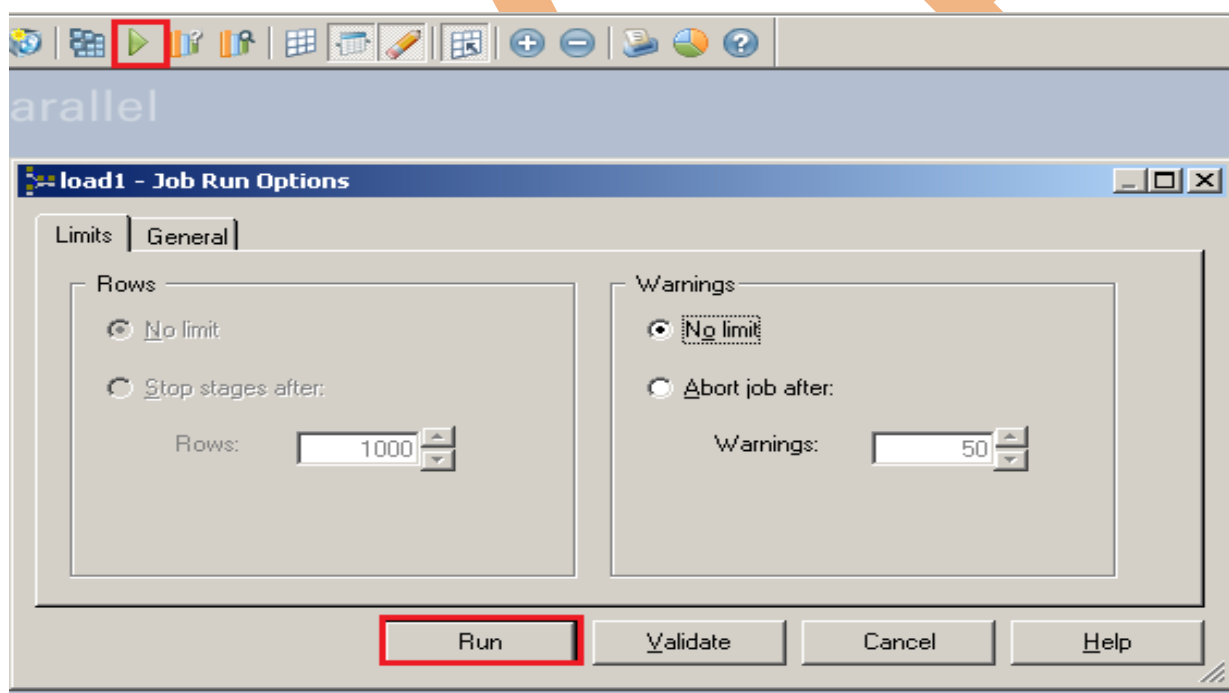

**Step-35**:- The mapping is ready click on execute button

Step-36:- Job executed successfully and no error found, click on run button or (CTRL+F5).

Step-37:- Now click on Run tab and notice some options available here. We have selected No limit of warnings and we can validate before run the job.

Step-38:- 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.
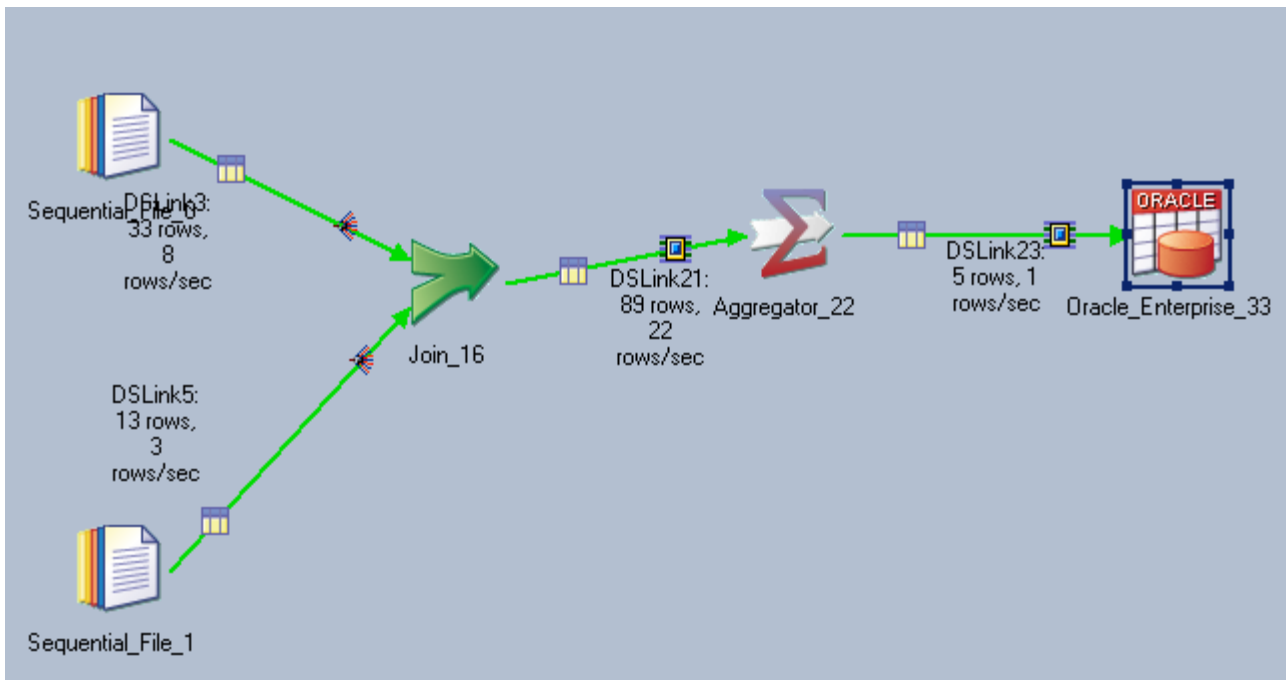
Step-39:- Now Go to SQL Console and connect with the same login credentials as I mentioned in oracle enterprise stage src_creditcard/password so you can choose your own username and password. Before loading, remember to check structure of this file is available and then you should be able to load the data. then simple query:
<SELECT * FROM SRC_AGG_1;>

	BRANCH_CD	BRANCH_NAME	SUM
1	14	CHI	12844.0
2	15	UK	12000.0
3	17	STW	103572.0
4	16	IND	139240.0
5	18	US	10192.0