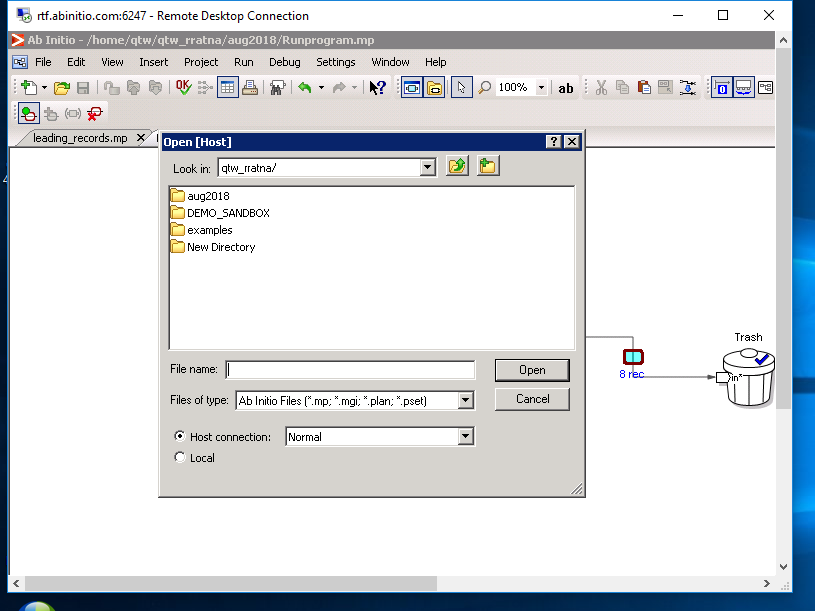
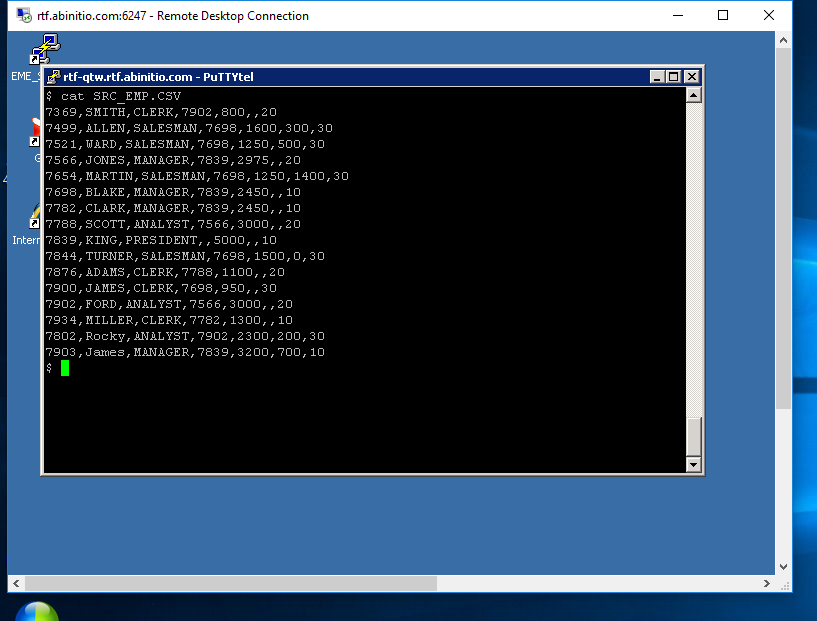
**DATASET AND TRANSFORM COMPONENT**

1)create a sandbox by the name DEMO\_SANDBOX and create the below graph in this new sandbox



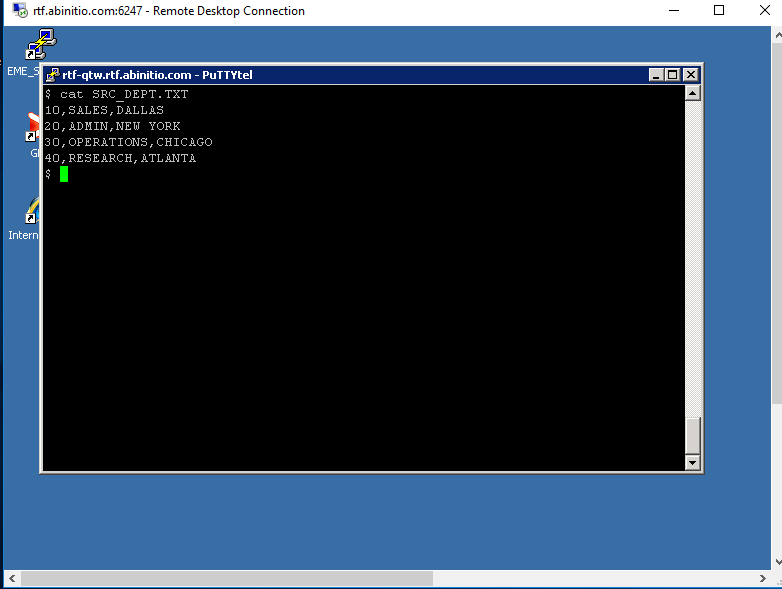
2)Login to EME shell and Create a source file with the following data. Name the file as **SRC\_EMP.CSV.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 7369 | SMITH | CLERK | 7902 | 800 |  | 20 |
| 7499 | ALLEN | SALESMAN | 7698 | 1600 | 300 | 30 |
| 7521 | WARD | SALESMAN | 7698 | 1250 | 500 | 30 |
| 7566 | JONES | MANAGER | 7839 | 2975 |  | 20 |
| 7654 | MARTIN | SALESMAN | 7698 | 1250 | 1400 | 30 |
| 7698 | BLAKE | MANAGER | 7839 | 2850 |  | 30 |
| 7782 | CLARK | MANAGER | 7839 | 2450 |  | 10 |
| 7788 | SCOTT | ANALYST | 7566 | 3000 |  | 20 |
| 7839 | KING | PRESIDENT |  | 5000 |  | 10 |
| 7844 | TURNER | SALESMAN | 7698 | 1500 | 0 | 30 |
| 7876 | ADAMS | CLERK | 7788 | 1100 |  | 20 |
| 7900 | JAMES | CLERK | 7698 | 950 |  | 30 |
| 7902 | FORD | ANALYST | 7566 | 3000 |  | 20 |
| 7934 | MILLER | CLERK | 7782 | 1300 |  | 10 |
| 7802 | Rocky | ANALYST | 7902 | 2300 | 200 | 30 |
| 7903 | James | MANAGER | 7839 | 3200 | 700 | 10 |



3)Create another file and name the file as **SRC\_DEPT.TXT.** the contents of the file should be as mentioned below

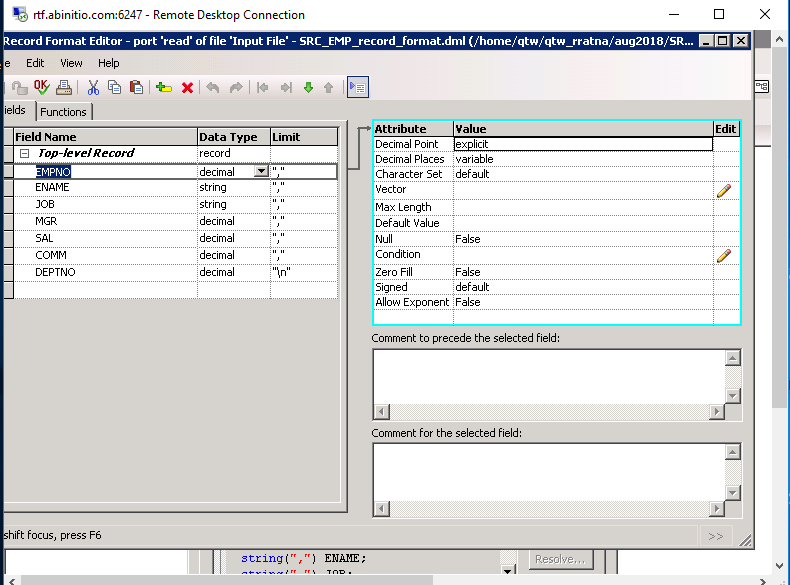
|  |  |  |
| --- | --- | --- |
| Deptno | Dname | Location |
| 10 | SALES | DALLAS |
| 20 | ADMIN | NEW YORK |
| 30 | OPERATIONS | CHICAGO |
| 40 | RESEARCH | ATLANTA |



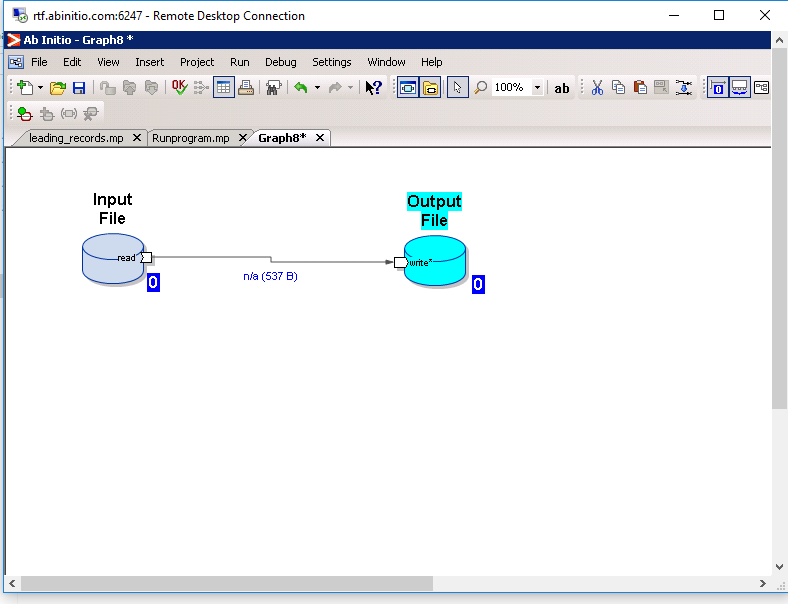
4)Open GDE if not open , and create an input file. the data source of the file should be **SRC\_EMP.CSV (created above)**. the record format of the file should be as below

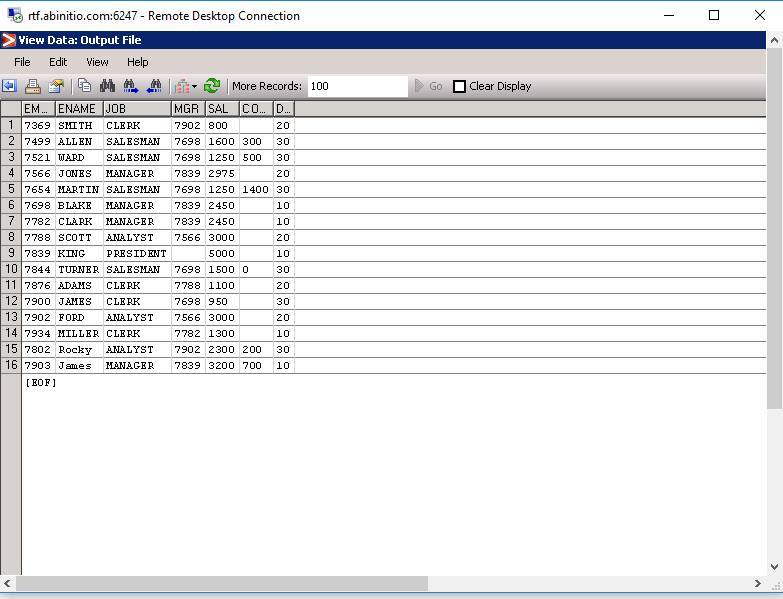
|  |  |
| --- | --- |
| Column Name | Data Type |
| EMPNO | INTEGER |
| ENAME | VARCHAR(10) |
| JOB | VARCHAR(10) |
| MGR | INTEGER |
| SAL | INTEGER |
| COMM | INTEGER |
| DEPTNO | INTEGER |

Save the above record format to be used later.

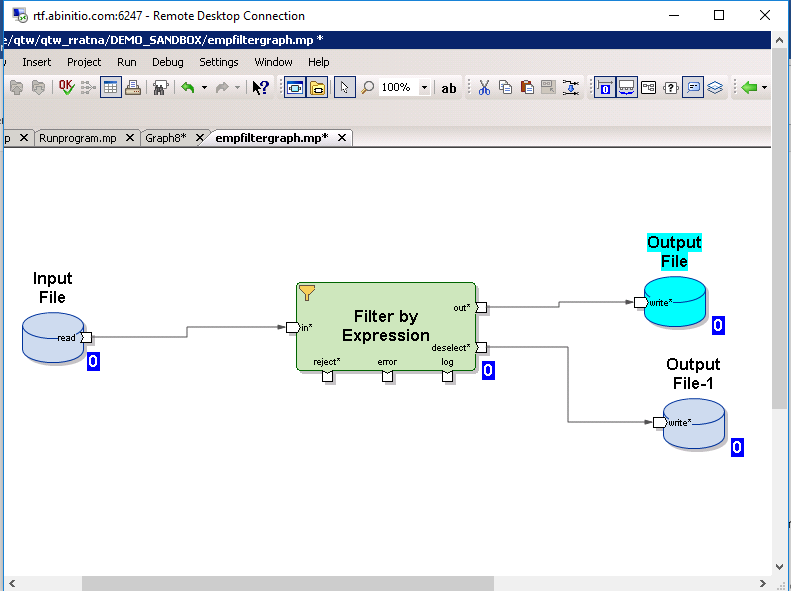


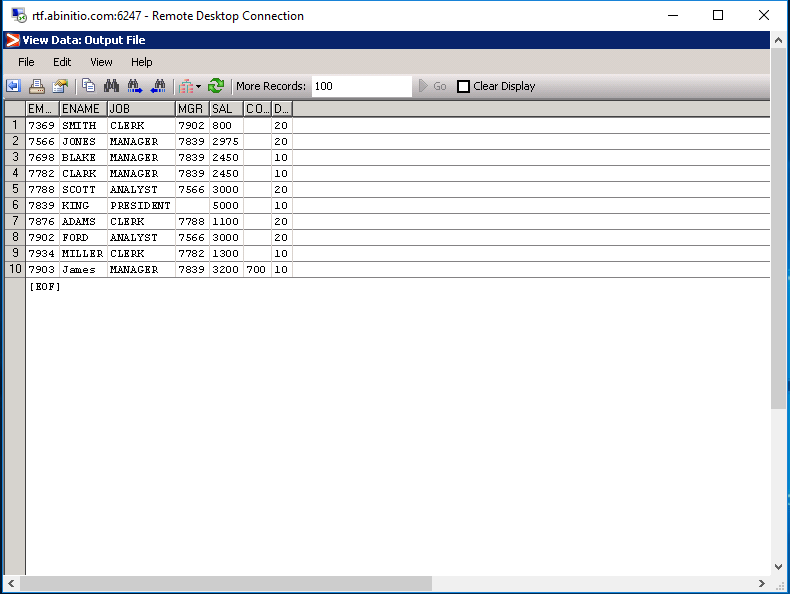
5)Load the data from SRC\_EMP.TXT to EMPOUT.txt. all records of the should get loaded

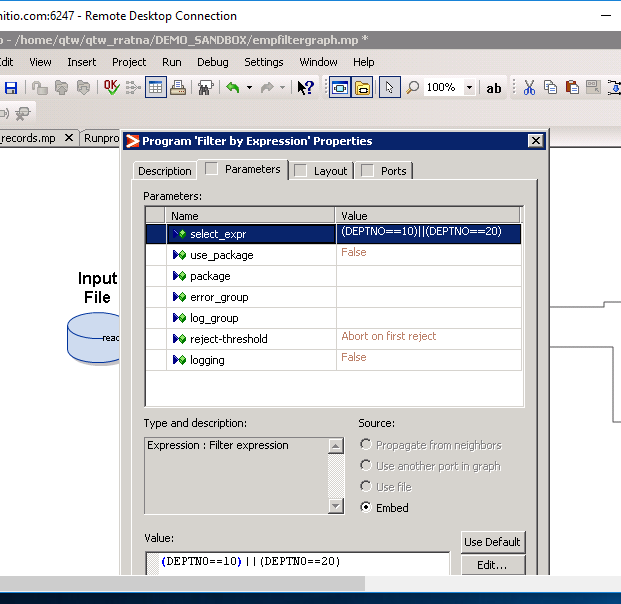




6)Create a graph that will load only records of deptno 10 and 20 from SRC\_EMP.txt to Emp1020.txt which is a target file . save the graph as empfiltergraph.mp.



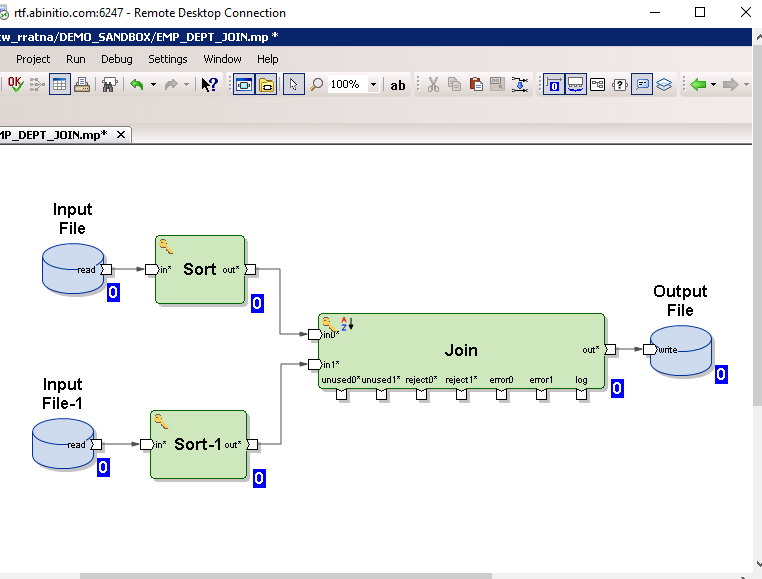


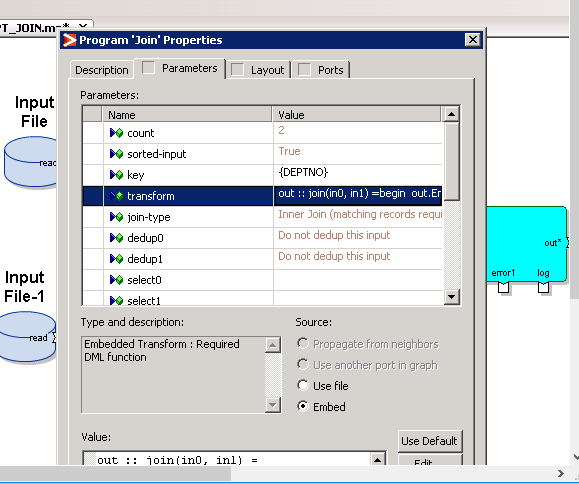


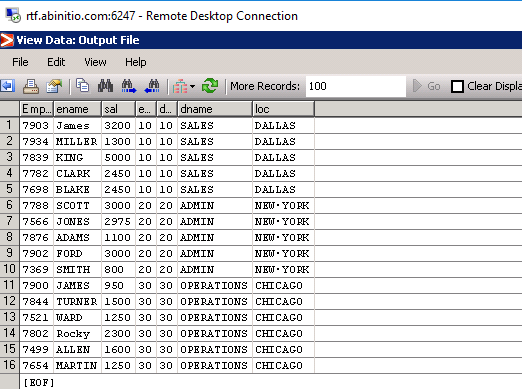
7)Create a Graph which will have the below target. Name the graph as empdept\_join.mp

Empno,ename,sal,emp\_deptno,dept\_deptno,dname,loc,

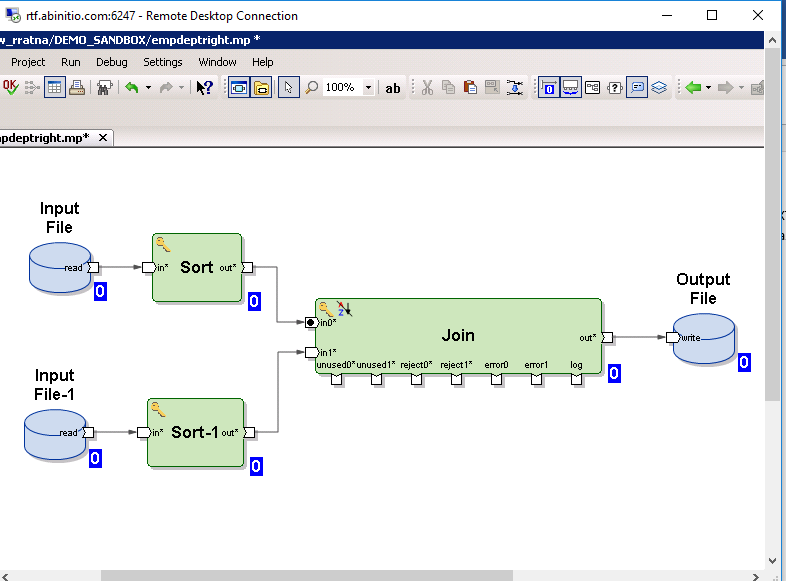
Consolidate data of SRC\_EMP.TXT and SRC\_DEPT.TXT into the above created file. load matching rows only

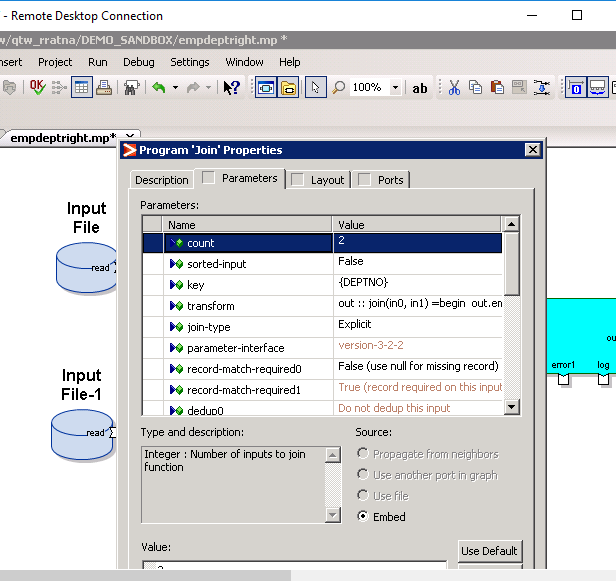


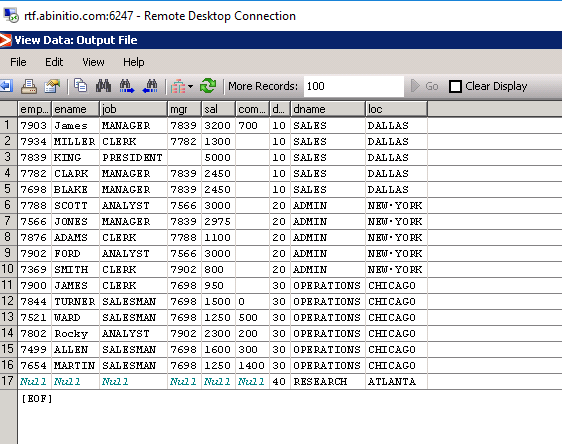




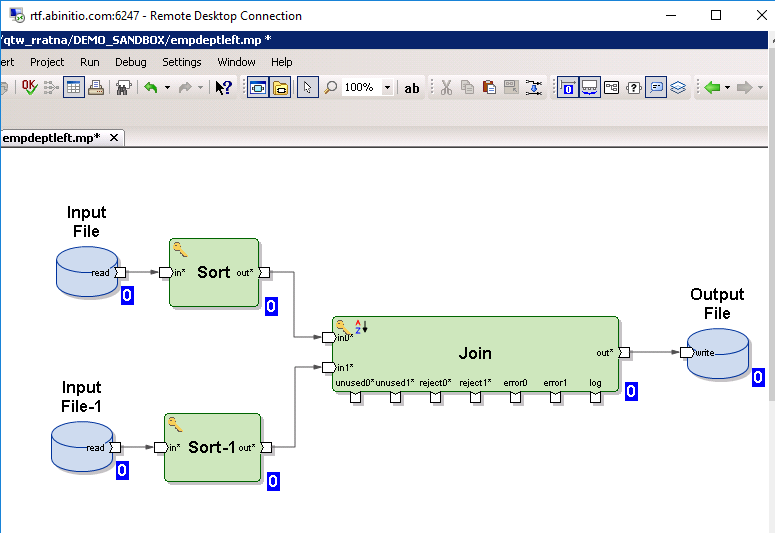
8) Create a graph which will load the matching records of SRC\_EMP.TXT and SRC\_DEPT.TXT and extra rows of SRC\_DEPT.txt for which there is no match in SRC\_EMP.TXT. save the graph as empdeptright.mp

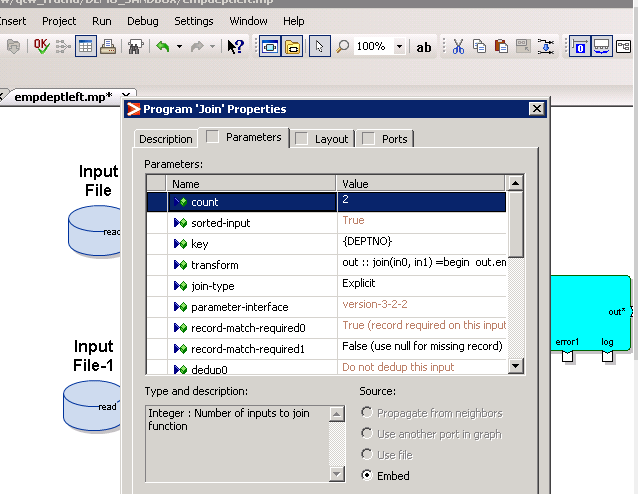




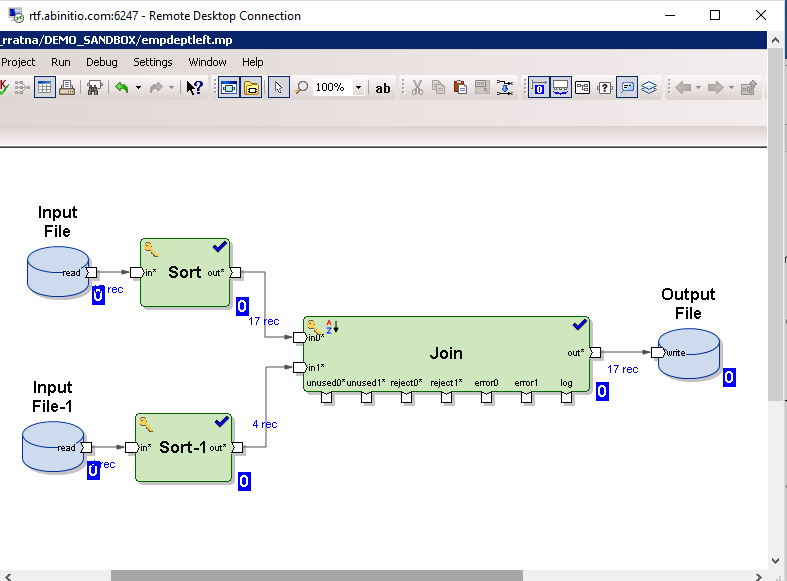


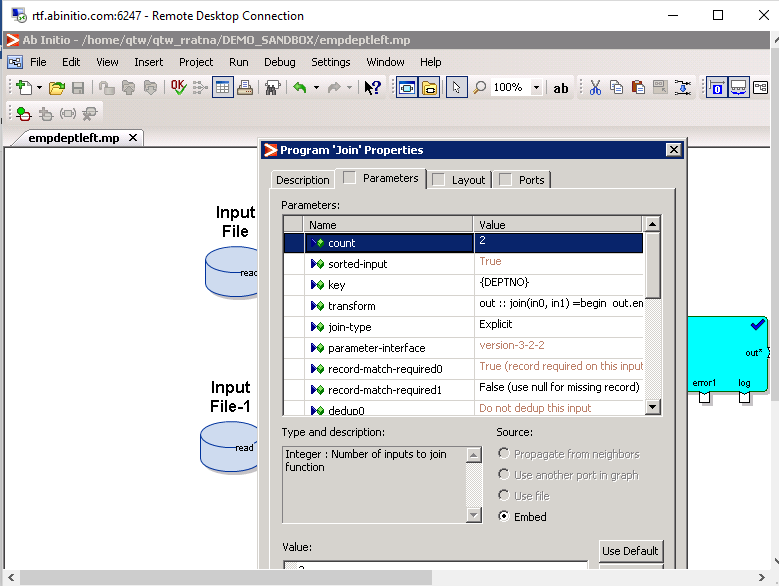
9)Create a graph which will load the matching records of SRC\_EMP.TXT and SRC\_DEPT.TXT and extra rows of SRC\_EMP.txt for which there is no match in SRC\_DEPT.TXT. save the graph as empdeptleft.mp

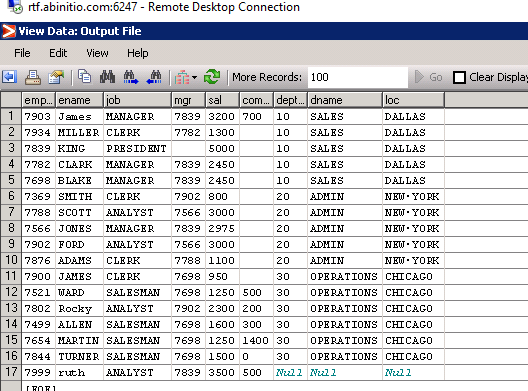




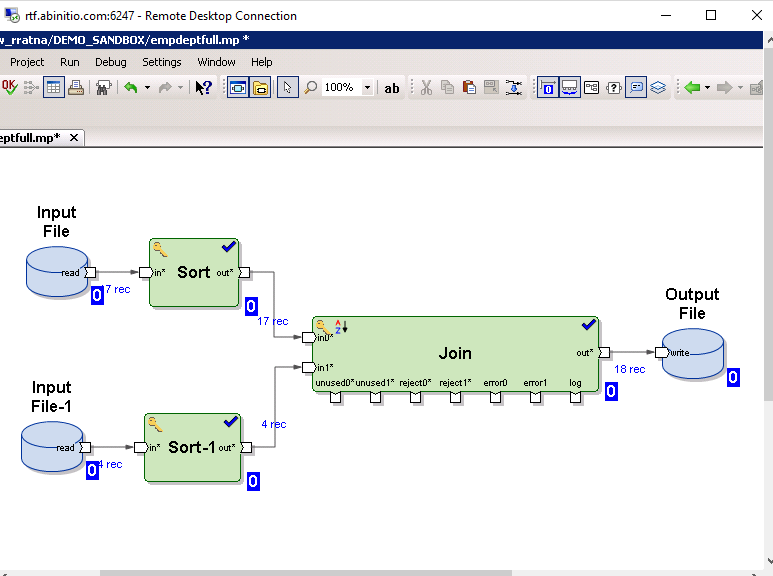
9)Create a graph which will load the matching records of SRC\_EMP.TXT and SRC\_DEPT.TXT and extra rows of SRC\_EMP.txt for which there is no match in SRC\_DEPT.TXT. save the graph as empdeptleft.mp



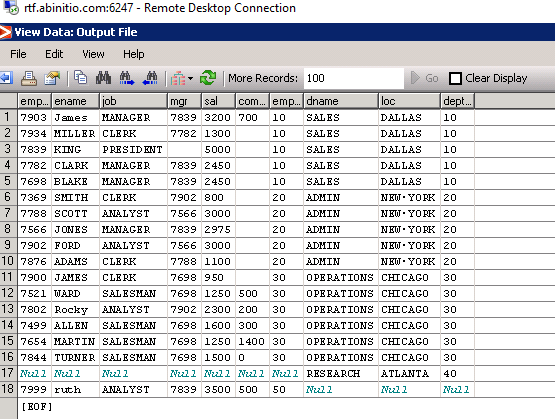




10)Create a graph that will load the matching records of SRC\_EMP.TXT and SRC\_DEPT.TXT and extra rows of SRC\_DEPT.txt and SRC\_EMP.TXT. save the graph as empdeptfull.mp



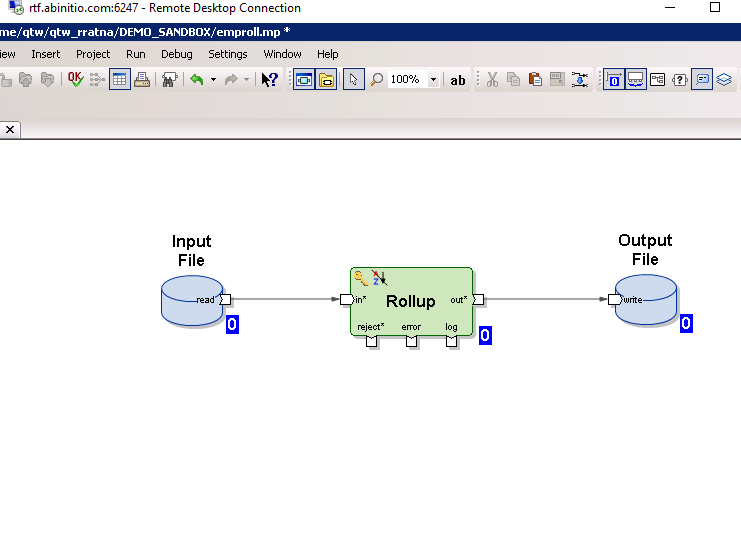


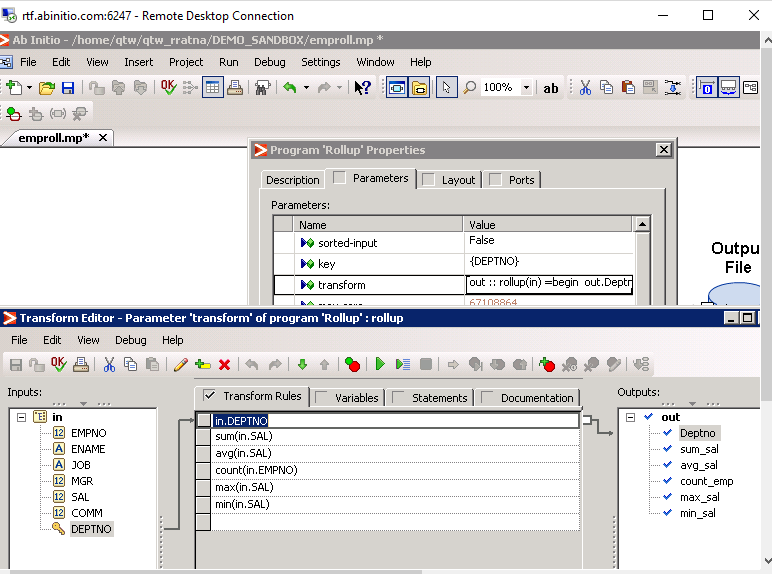


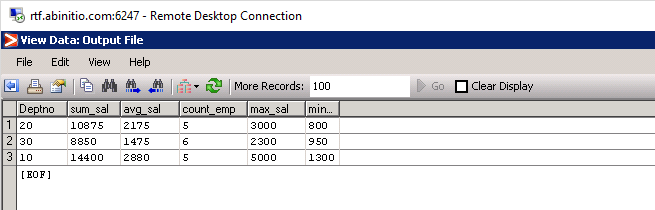
11)Create an output file component . name the file as emproll.txt. the record format of the file should be as below

Deptno, sum\_sal, avg\_sal, count\_emp, max\_sal, min\_sal

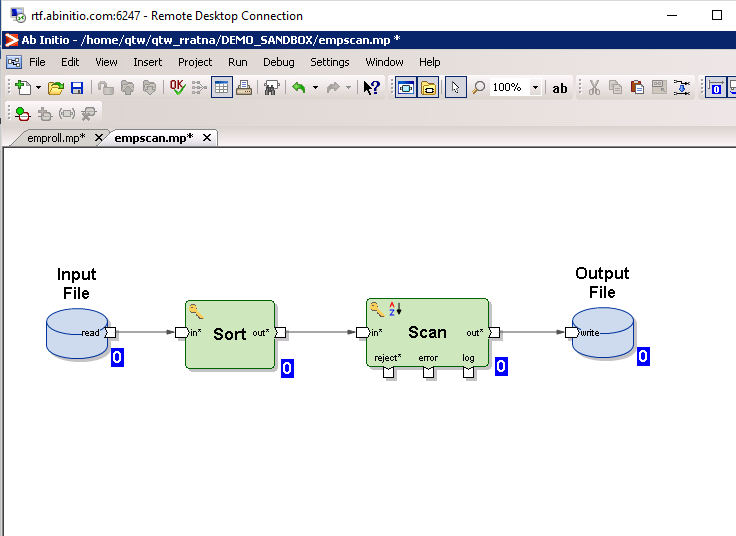
Calculate department wise sum of salary, average of salary, max of salary, min of salary and count empno. Map the source columns to right target columns

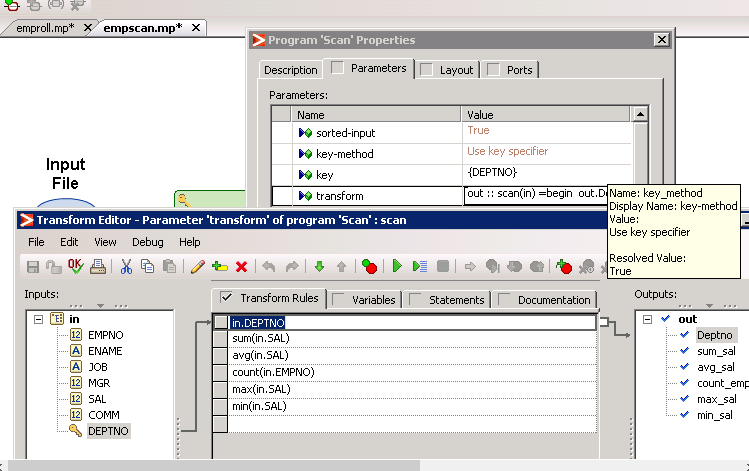




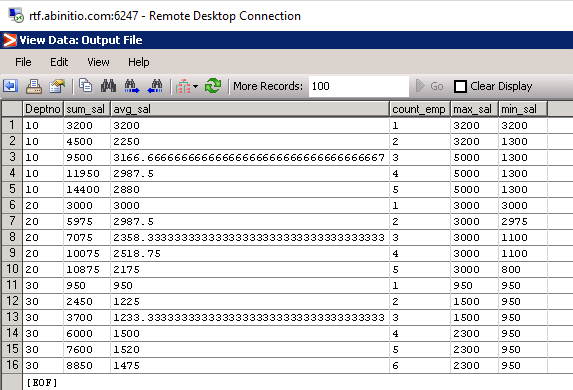


12)Perform the above activity using the SCAN component and observe how SCAN is different from ROLLUP. Write down your observation.

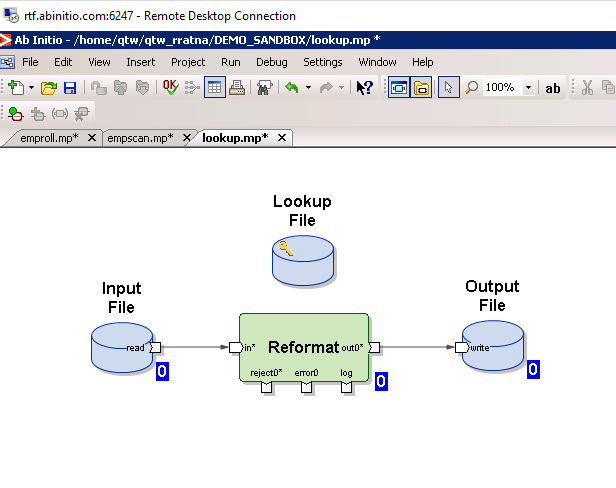


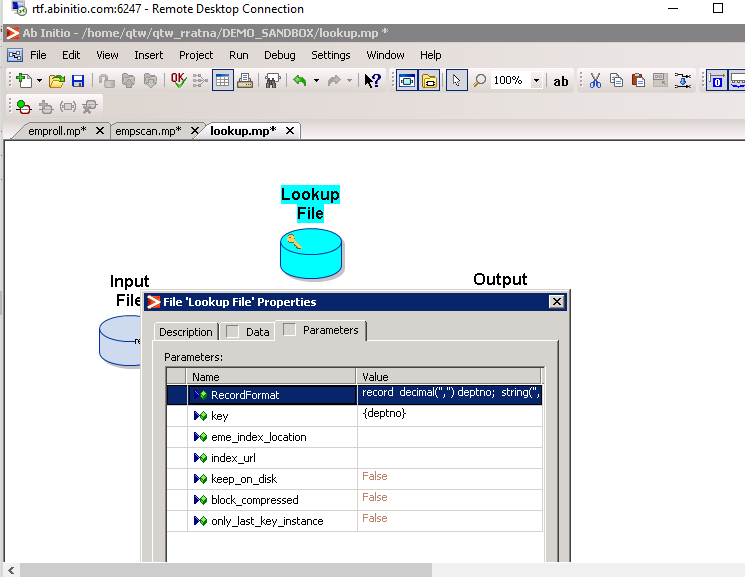


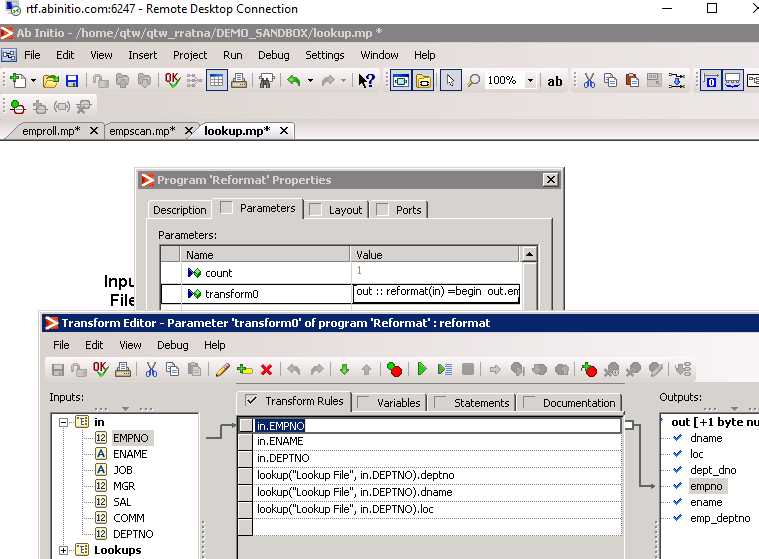
The output of the rollup is based on the group (the key we are using ) not for individual elements ,but in the case of scan the output is based on group of records,the output is generated for each and every element.

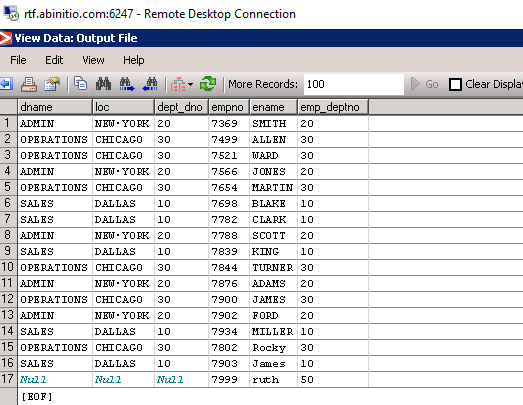


13)Create a new graph and in that create a Lookup on SRC\_DEPT.TXT file and use SRC\_EMP as source file. match the source records with lookup records. If the dept exists in lookup, lookup should return dname and location. Capture the details in the target file.









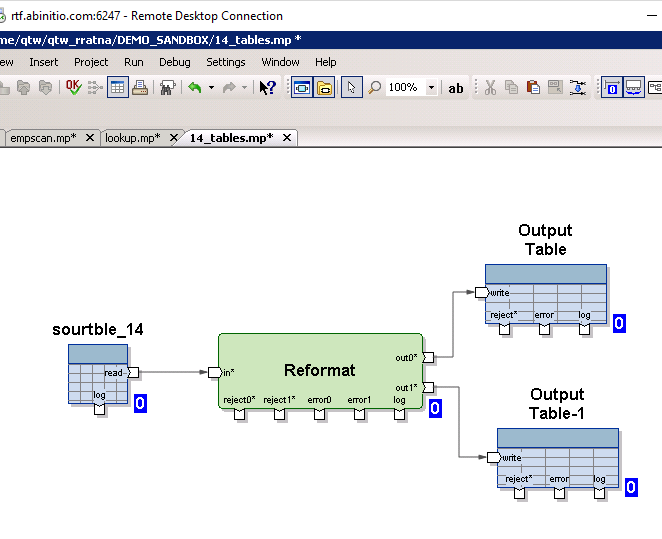
14)Create the source table with the data mentioned in SOURCE TABLE section. Create two targets.

Target 1 columns are ID and Name. Target 2 columns are ID and phone\_no.

Split the non-key columns to separate tables with key column in both / How to split the data of source table column-wise with respect to primary key. See the source and target tables below.

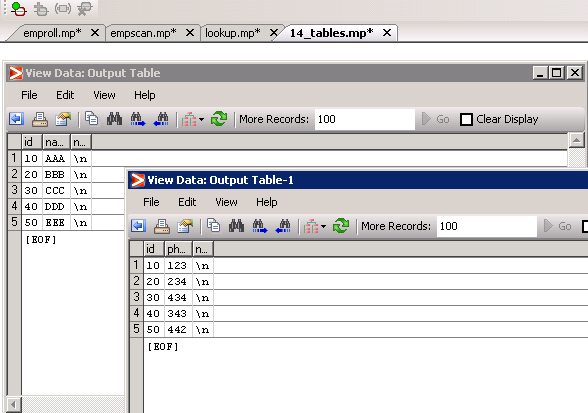
**source table:** ID is the key column, Name and Phone No are non-key columns

|  |  |  |
| --- | --- | --- |
| **Source Table** | **Target Table 1** | **Target Table 2** |
| |  |  |  | | --- | --- | --- | | **ID** | Name | Phone No | | 10 | AAA | 123 | | 20 | BBB | 234 | | 30 | CCC | 434 | | 40 | DDD | 343 | | 50 | EEE | 442 | | |  |  | | --- | --- | | ID | Name | | 10 | AAA | | 20 | BBB | | 30 | CCC | | 40 | DDD | | 50 | EEE | | |  |  | | --- | --- | | ID | Phone No | | 10 | 123 | | 20 | 234 | | 30 | 434 | | 40 | 343 | | 50 | 442 | |





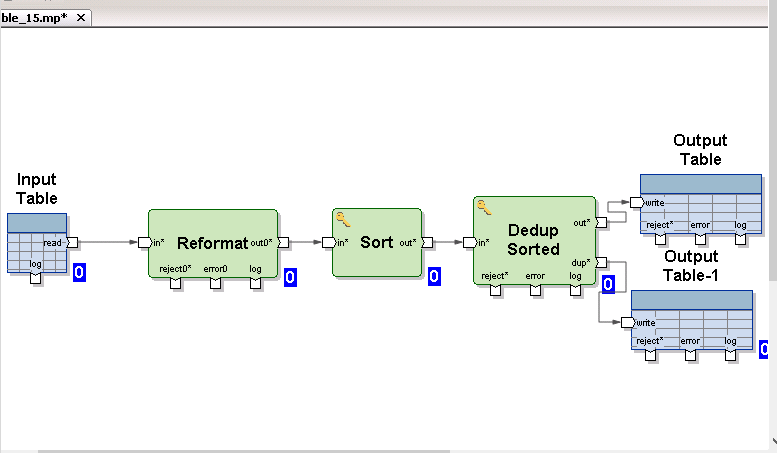


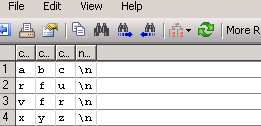


## 15)Identify and Separate duplicate rows into separate tables

How to segregate the duplicate and distinct rows from source table to separate target tables?

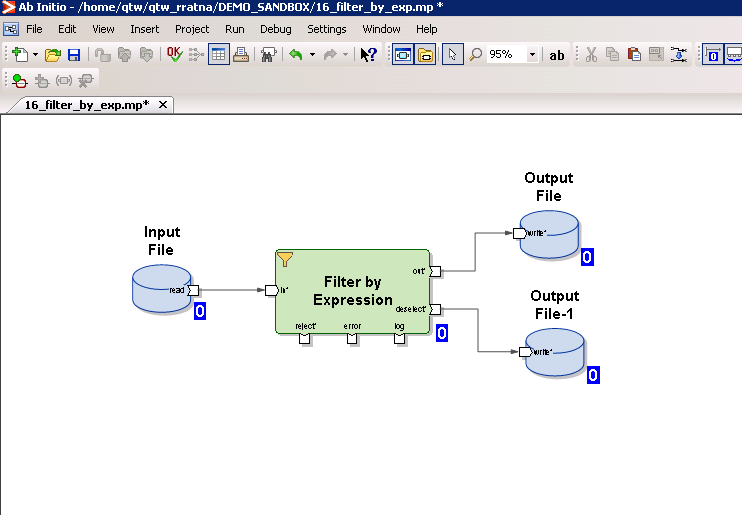
|  |  |  |
| --- | --- | --- |
| **Source Table** | **Target Table 1** | **Target Table 2** |
| |  |  |  | | --- | --- | --- | | **COL1** | COL2 | COL3 | | a | b | c | | x | y | z | | a | b | c | | r | f | u | | a | b | c | | v | f | r | | v | f | r | | Table containing all the unique rows   |  |  |  | | --- | --- | --- | | COL1 | COL2 | COL3 | | a | b | c | | x | y | z | | r | f | u | | v | f | r | | Table containing all the duplicate rows   |  |  |  | | --- | --- | --- | | COL1 | COL2 | COL3 | | a | b | c | | a | b | c | | v | f | r | |

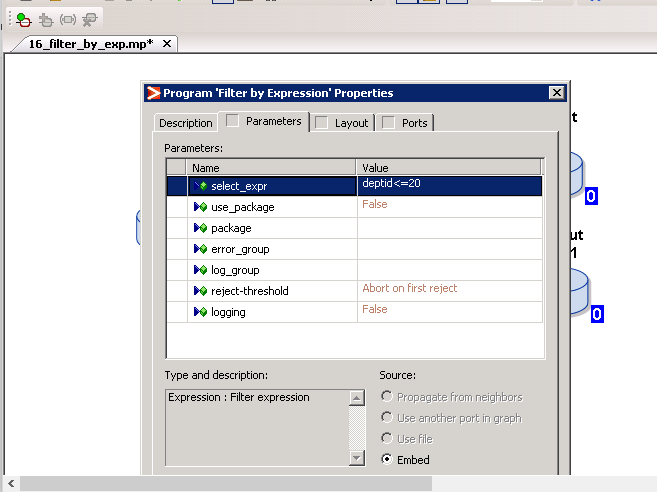


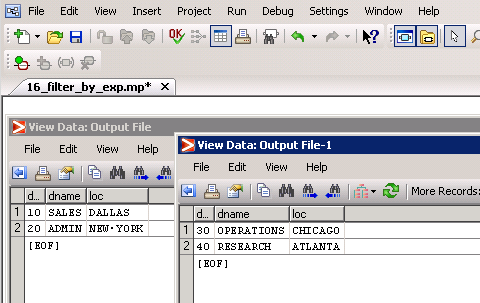
 

## 16)Separate rows on group basis

In SRC\_DEPT.TXT file, there are four departments (dept no 40,30,20,10). Separate the record to different target department wise. (Hint: use Filter by Expression)







17)Create a source file named  **Employees.txt.** insert 10 rows of your choice. The record format is as follows:

record format: empno,first\_name,last\_name,sal,job,deptno

target file: employee\_tgt

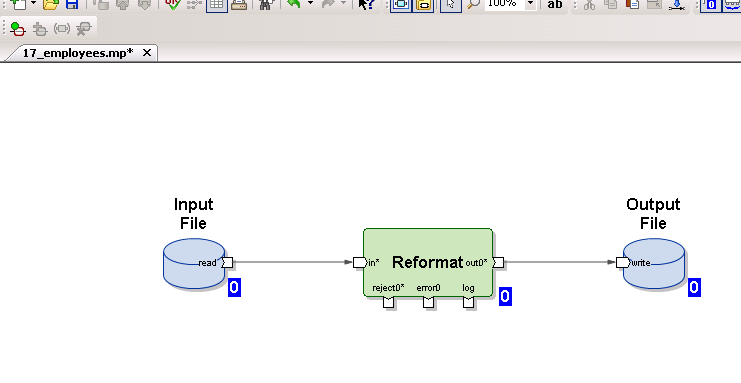
record format: tempno,tname,tsal,tjob,tdeptno

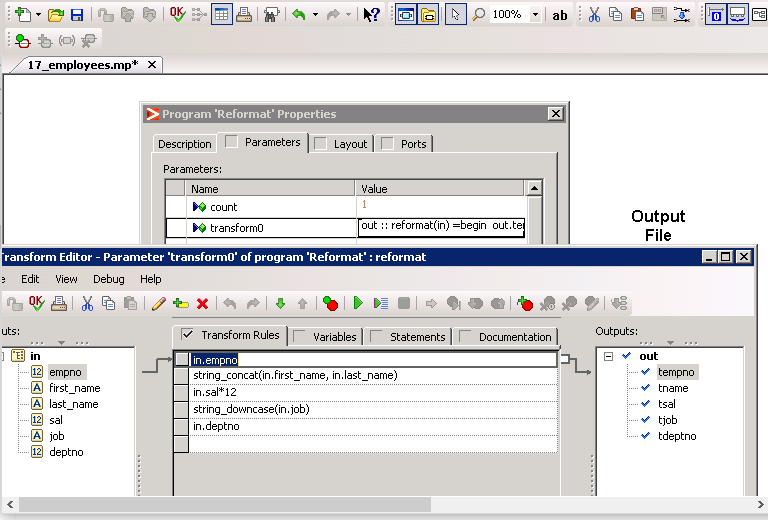
while loading the data into the target file, following changes need to be made:

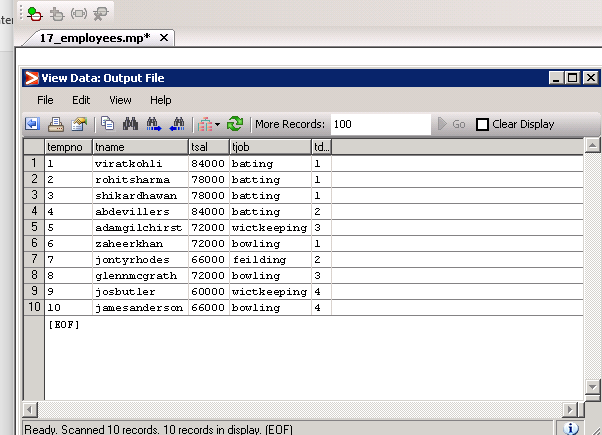
1. tname=first\_name+last\_name

2. tsal= sal\*12

3. tjob=lower(job)







## 18)Create 3 source files as defined below.

Source file: empnew.txt . define record format as per the data below

101,adam,7000,10

102,john,8000,20

103,mary,800,10

104,ana,7000,30

105,smith,9000,20

Source file: deptnew.txt

Record format: deptno, dname

10, sales

20, hr

40, admin

Source file: locnew.txt

Record format: empno, location

101, new delhi

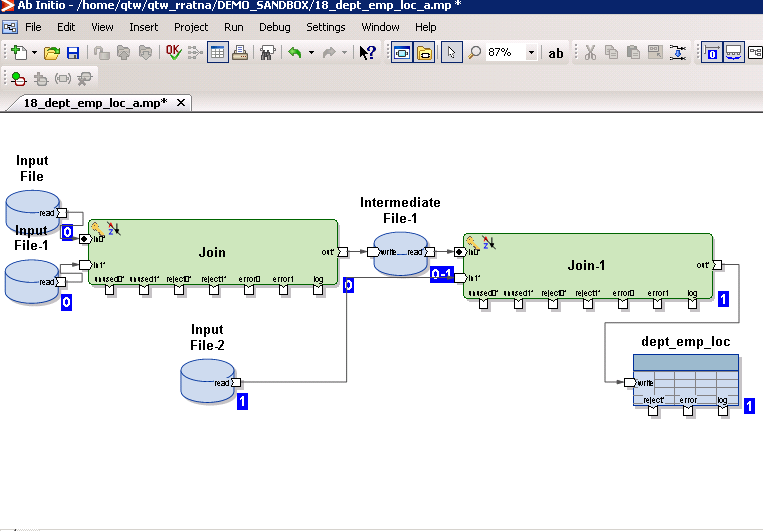
103, Hyderabad

104, ahmedabad

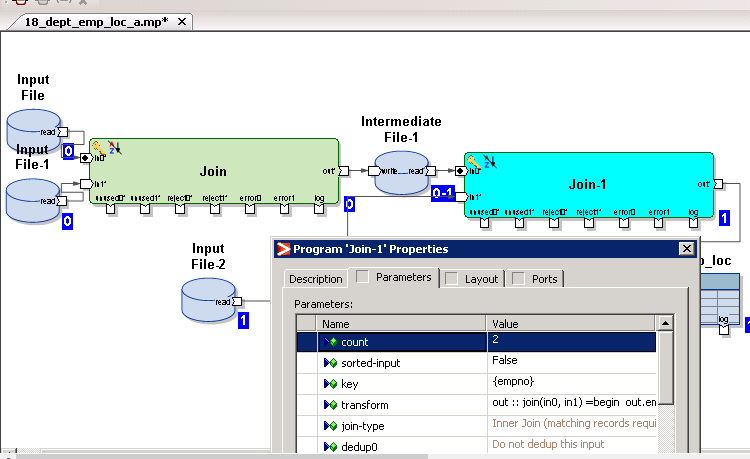
**Target table: dept\_emp\_loc**

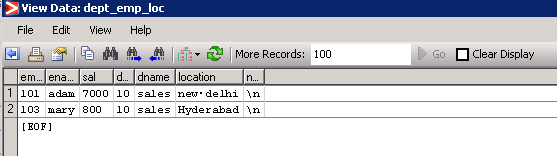
Record format: empno, ename, sal, deptno, dname, location

1. join the data of all the three sources and load into the target table

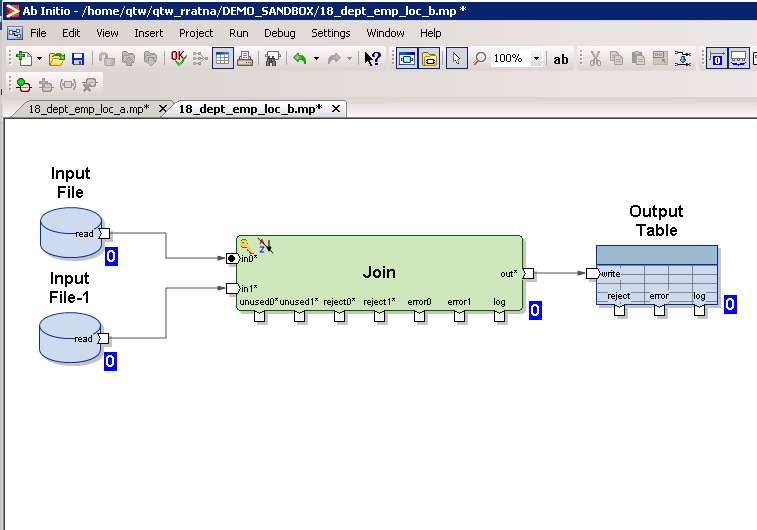


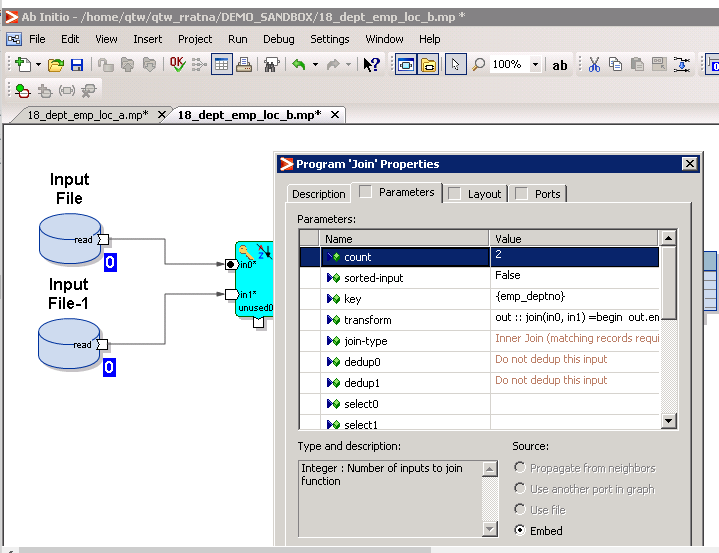


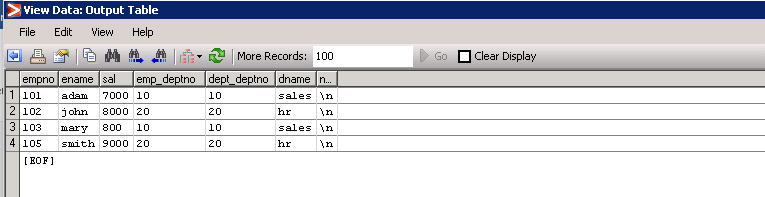




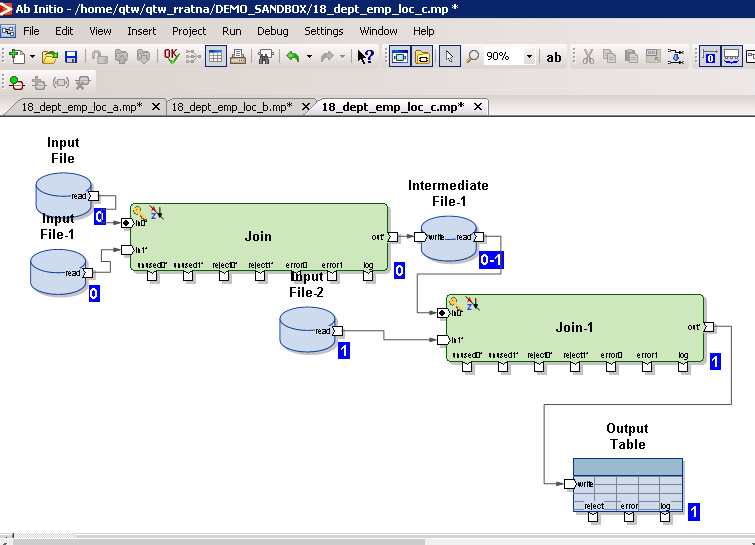
## b)Join the dept file with emp file and load only those records which are there in dept but not in emp.

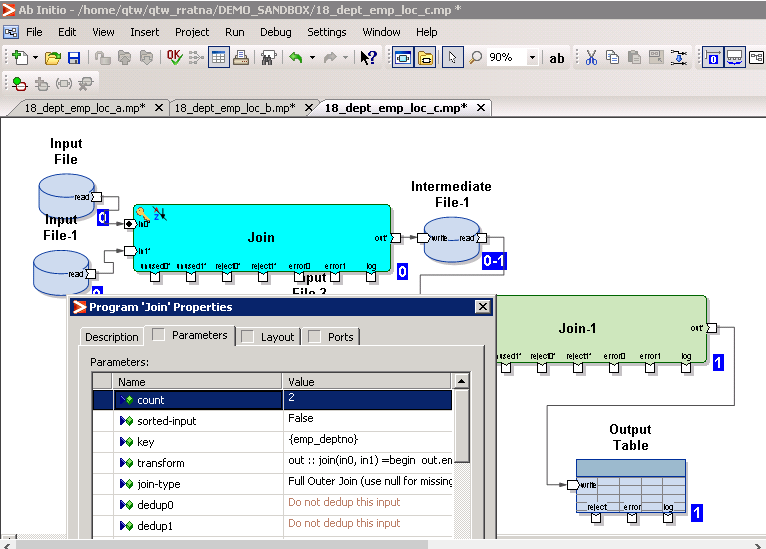


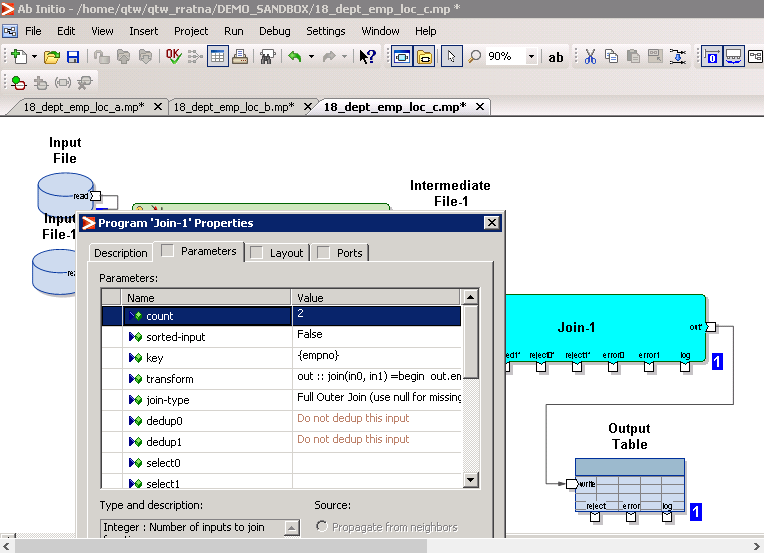


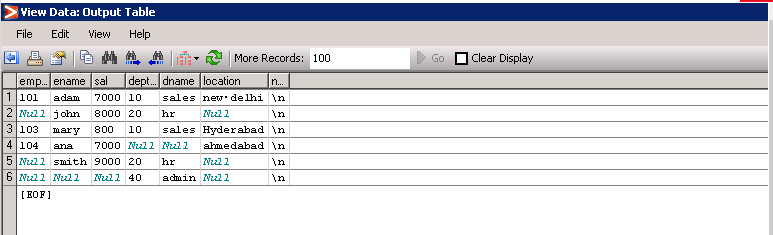


c)Perform a full outer join on all the 3 sources









19) Use the files empnew.txt and locnew.txt files created above. create another file name phonenew.txt

Records format: empno, phoneno

101,9898986599

102,6765356765

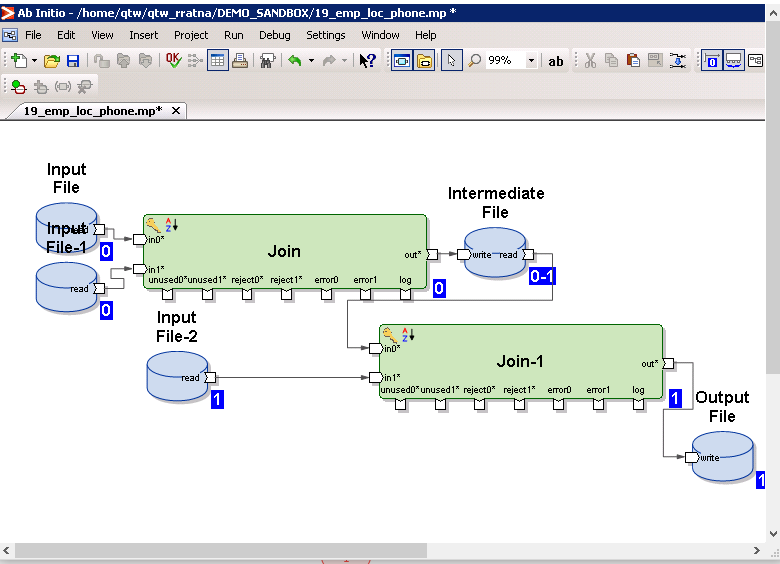
103,9098655345

105,9065354334

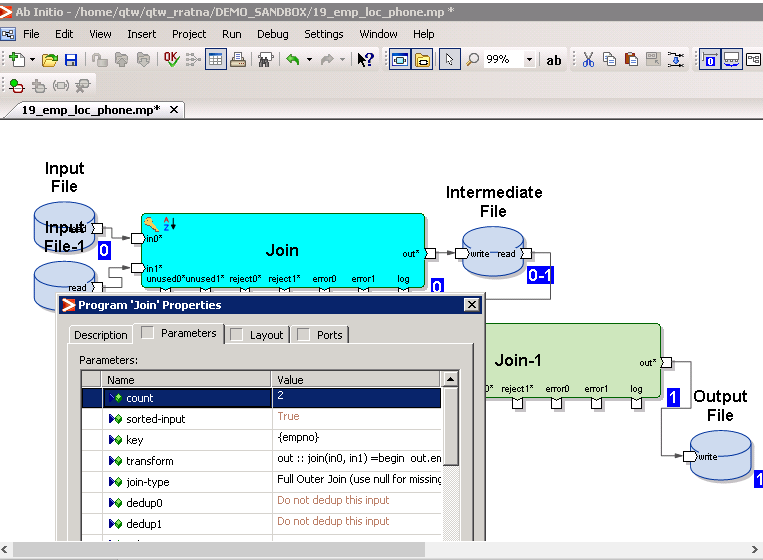
Join all the 3 sources to get the below results:

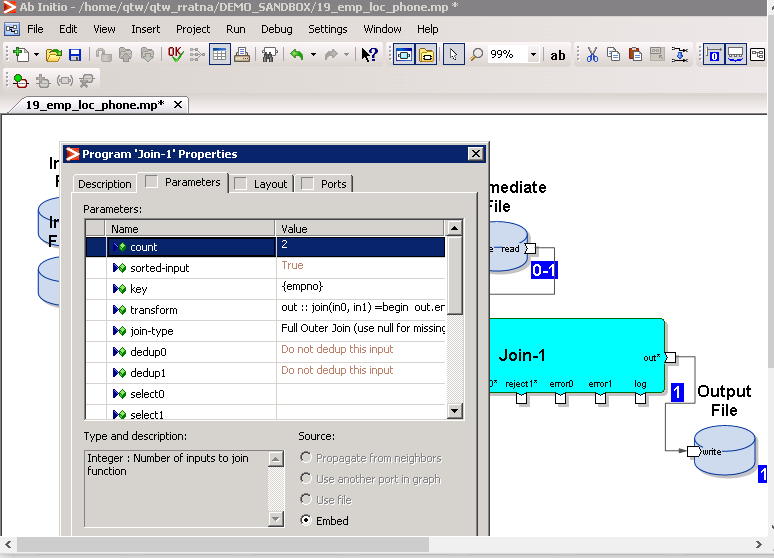
Empno, ename, sal, deptno ,location and phoneno.

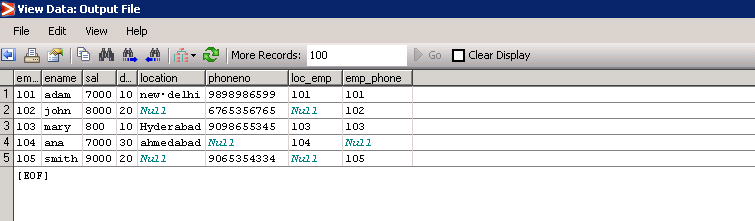
How many join components will accomplish the above task



It requires two joins to accomplish the above task







20)Let us assume the following Sales Data

ID PERSONAL\_SALES

10 40000

20 80000

30 40000

40 60000

50 NULL

60 50000

Create a graph to calculate the cumulative sum of PERSONAL\_SALES and load that into target file

The Expected Output in target should be

ID PERSONAL\_SALES O\_RETURN\_VALUE

10 40000 40000

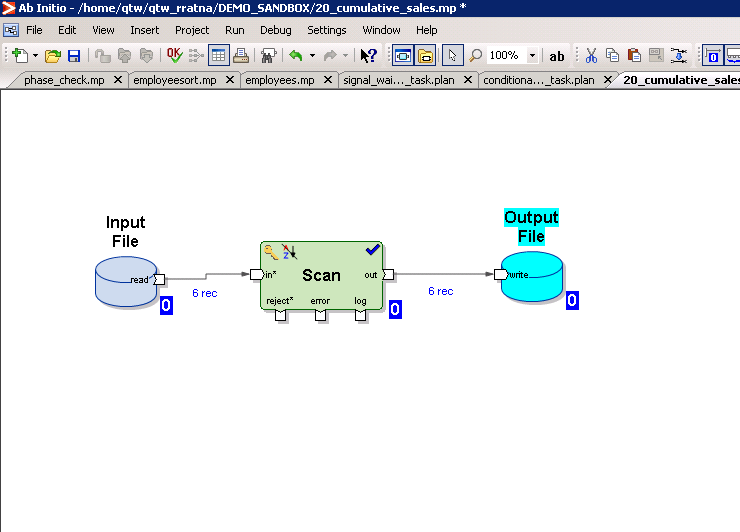
20 80000 120000

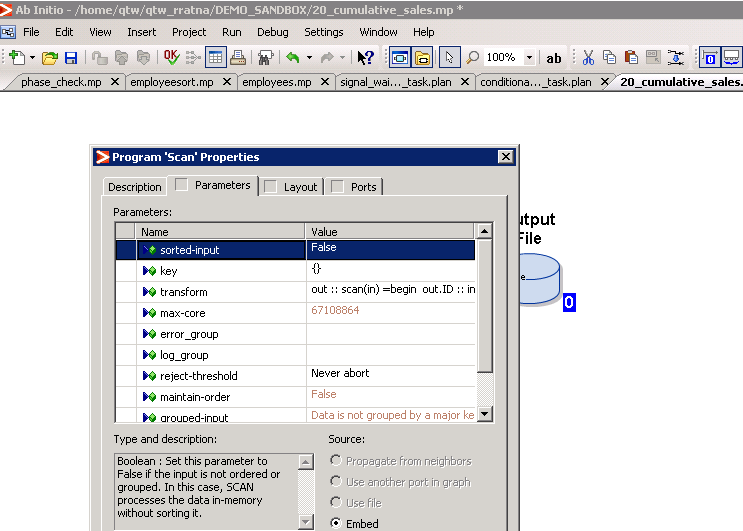
30 40000 160000

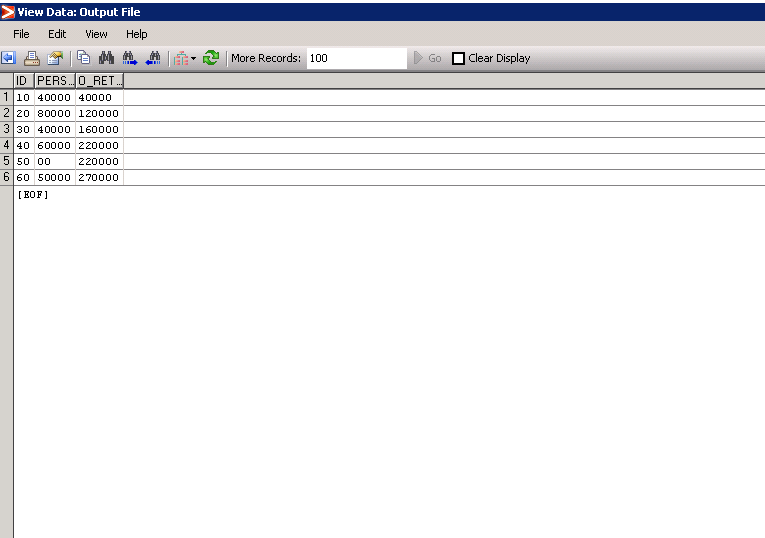
40 60000 220000

50 NULL 220000

60 50000 270000







21)create a source file named emp.txt with the below data:

dno,empno,sal,age

10,101,2000,25

20,102,3000,42

Create two lookup files as mentioned below

Dept.txt (lookup)

dno,dname,loc

10,maths,hyd

20,science,chen

Agegroup.txt (lookup)

lowerage, upperage, group

18 , 30, teenage

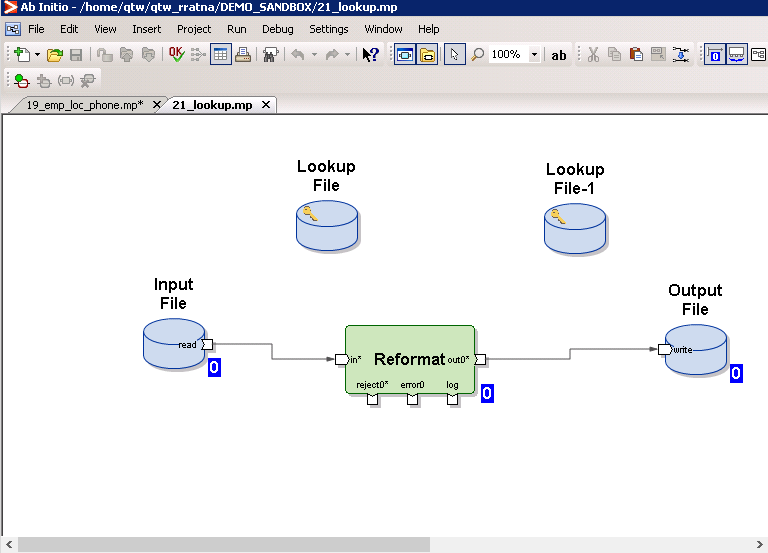
31, 100, elders

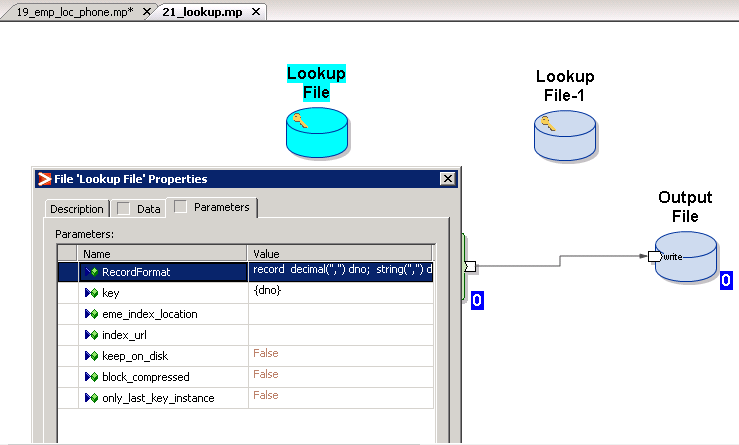
Draw a graph to populate the target file named dept\_emp\_agegroup.txt as below

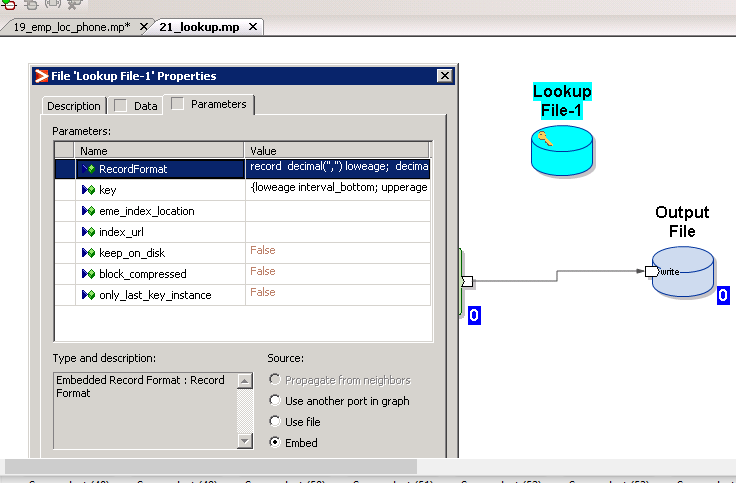
dno,dname,loc,eno,sal,age,group

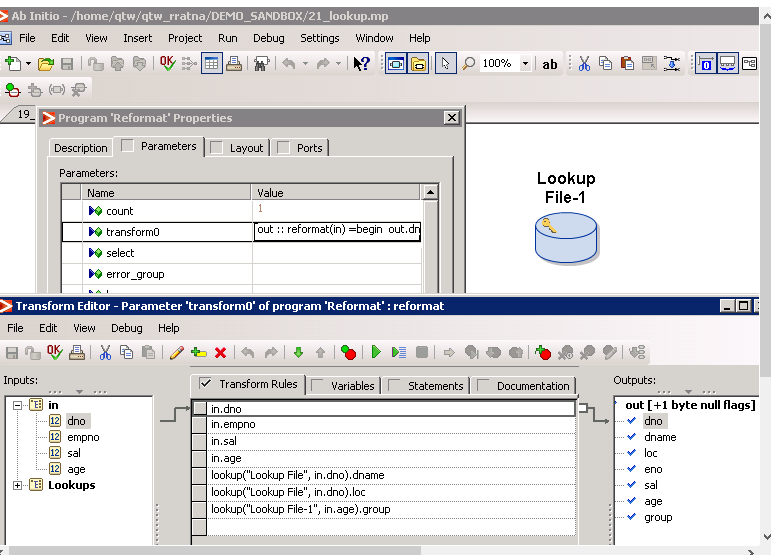
10, maths,hyd,101,2000,25,teenage

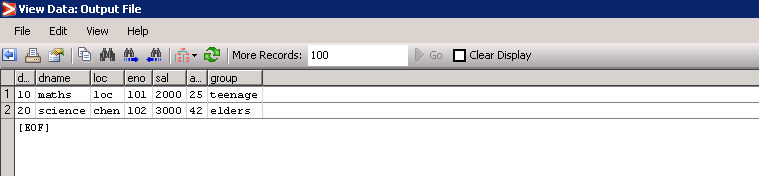
20, science, chn, 102,3000,42, elders











22) create a student.txt file with below data

stud\_id, sname, marks1, mark2, marks3

100, Ram, 90, 87, 76

Add some more records of your choice

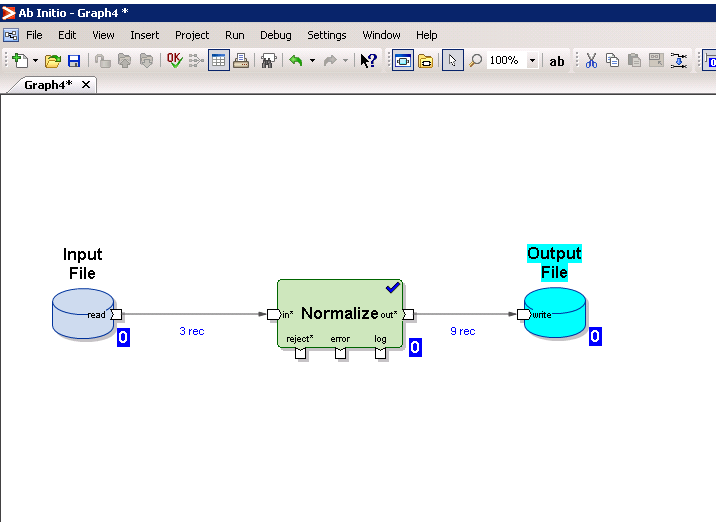
Create a suitable graph using appropriate components to get the below result

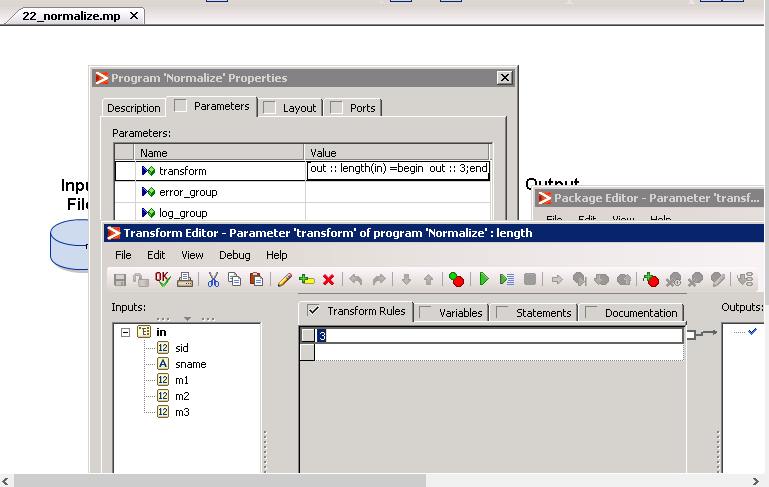
Stud\_id, name, marks

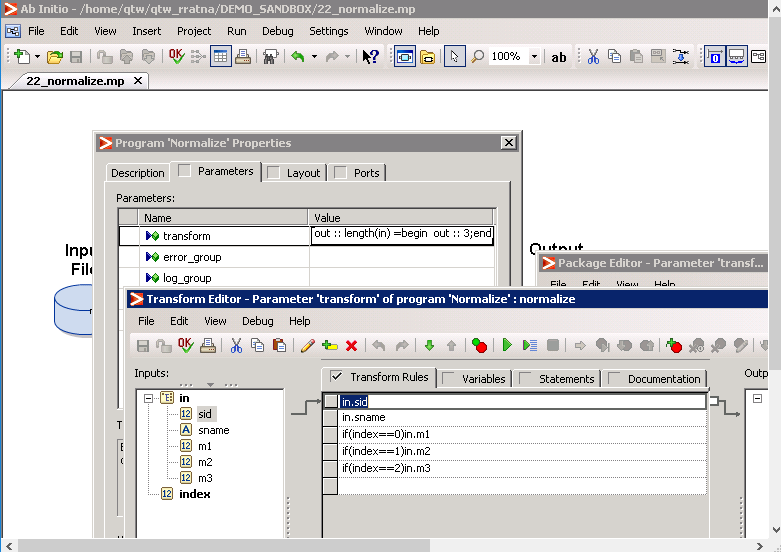
100, ram, 90

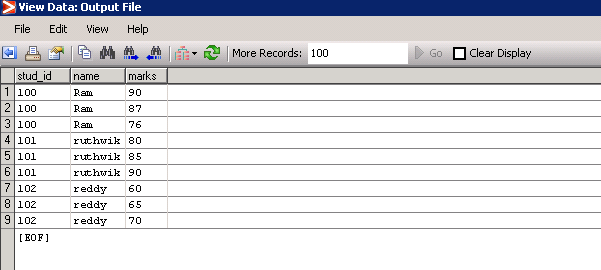
100, ram, 87

100, ram 76





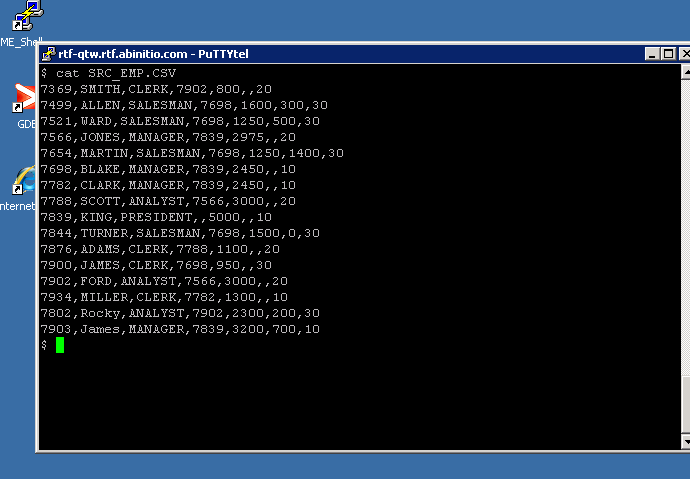




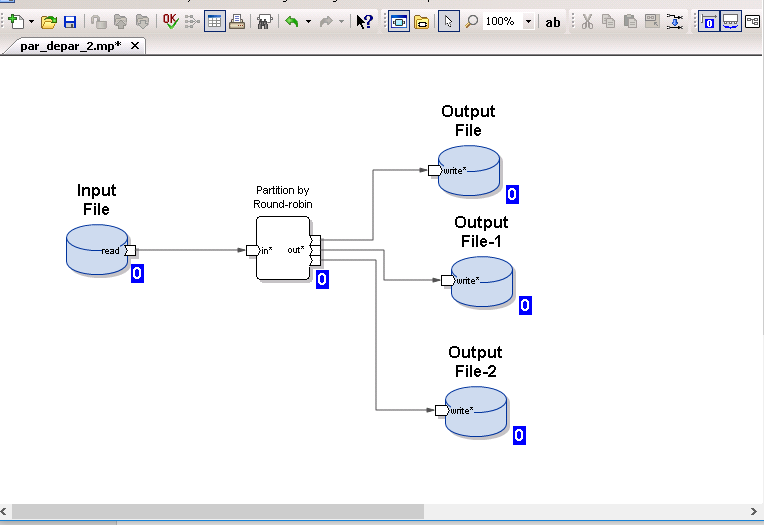
**PARTITION and DEPARTITION COMPONENTS:**

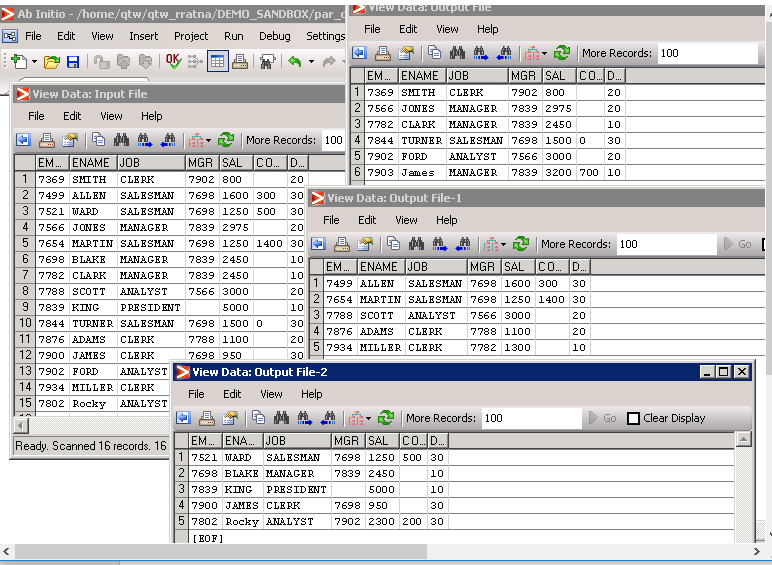
1. Login to EME shell and Create a source file with the following data. Name the file as **SRC\_EMP.CSV.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| EID | Ename | Title | Mgr\_ID | Sal | Comm | Dept |
| 7369 | SMITH | CLERK | 7902 | 800 |  | 20 |
| 7499 | ALLEN | SALESMAN | 7698 | 1600 | 300 | 30 |
| 7521 | WARD | SALESMAN | 7698 | 1250 | 500 | 30 |
| 7566 | JONES | MANAGER | 7839 | 2975 |  | 20 |
| 7654 | MARTIN | SALESMAN | 7698 | 1250 | 1400 | 30 |
| 7698 | BLAKE | MANAGER | 7839 | 2850 |  | 30 |
| 7782 | CLARK | MANAGER | 7839 | 2450 |  | 10 |
| 7788 | SCOTT | ANALYST | 7566 | 3000 |  | 20 |
| 7839 | KING | PRESIDENT |  | 5000 |  | 10 |
| 7844 | TURNER | SALESMAN | 7698 | 1500 | 0 | 30 |
| 7876 | ADAMS | CLERK | 7788 | 1100 |  | 20 |
| 7900 | JAMES | CLERK | 7698 | 950 |  | 30 |
| 7902 | FORD | ANALYST | 7566 | 3000 |  | 20 |
| 7934 | MILLER | CLERK | 7782 | 1300 |  | 10 |
| 7802 | Rocky | ANALYST | 7902 | 2300 | 200 | 30 |
| 7903 | James | MANAGER | 7839 | 3200 | 700 | 10 |

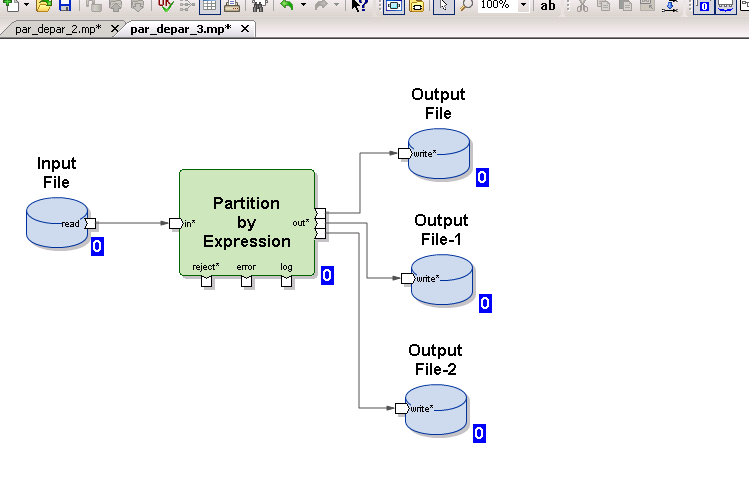


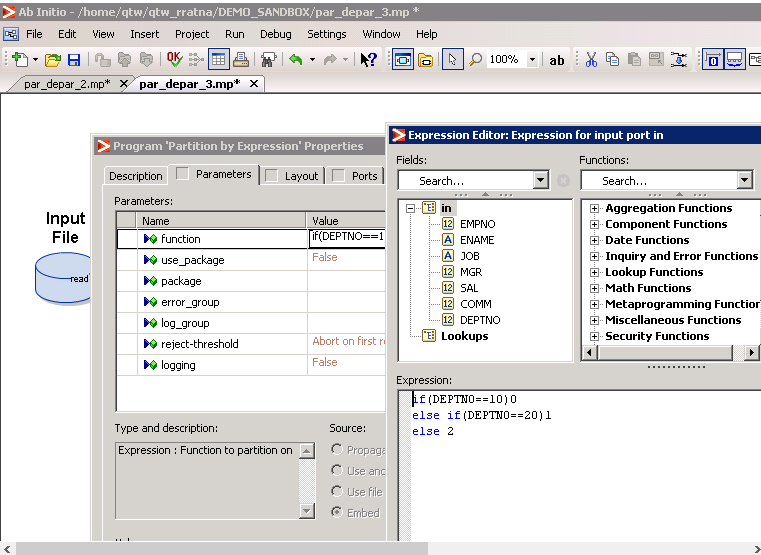
2)Create a Graph, which will use the above .csv file as input. The records of this file need to be sent to three targets in cyclic order. Eg: First record will go to first target, Second one will go to second target and third record will go to third target and then 4th to 1st, 5th to 2nd , 6th to 3rd and so on. Save the graph as CYCLIC\_LOAD\_GRAPH

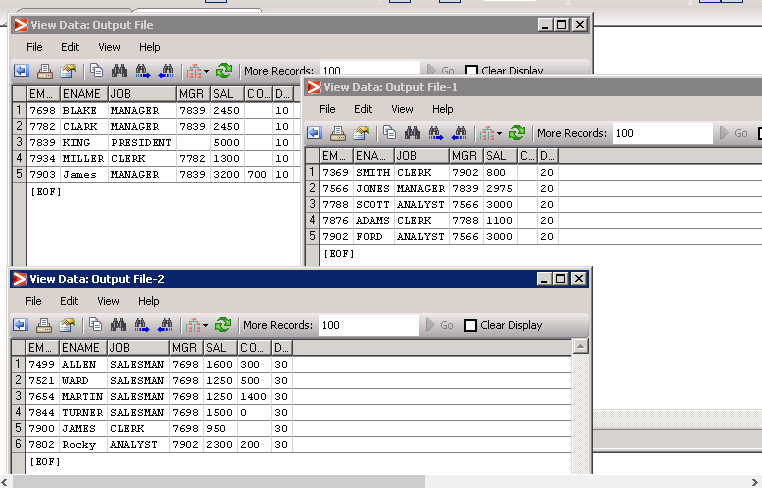




3)Create another Graph that will load the data according to the department number. Create 3 target files

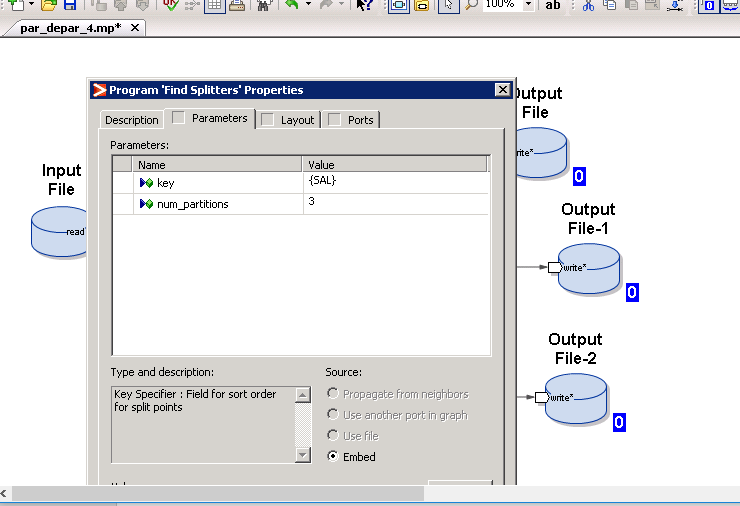


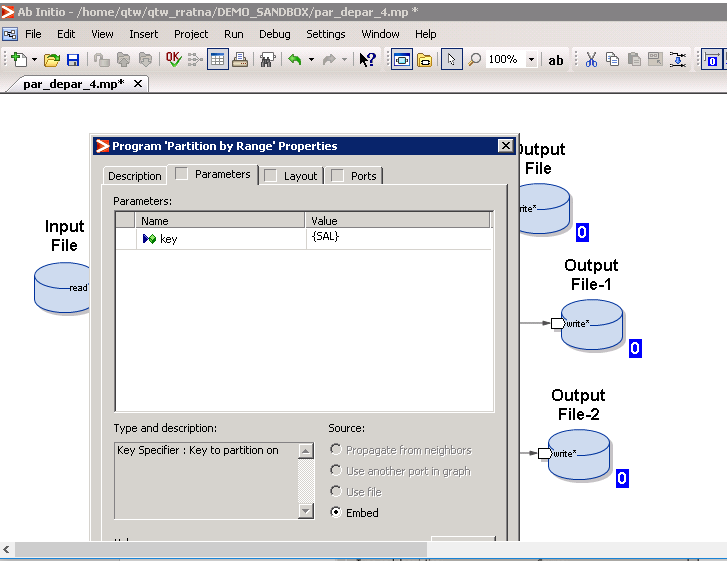


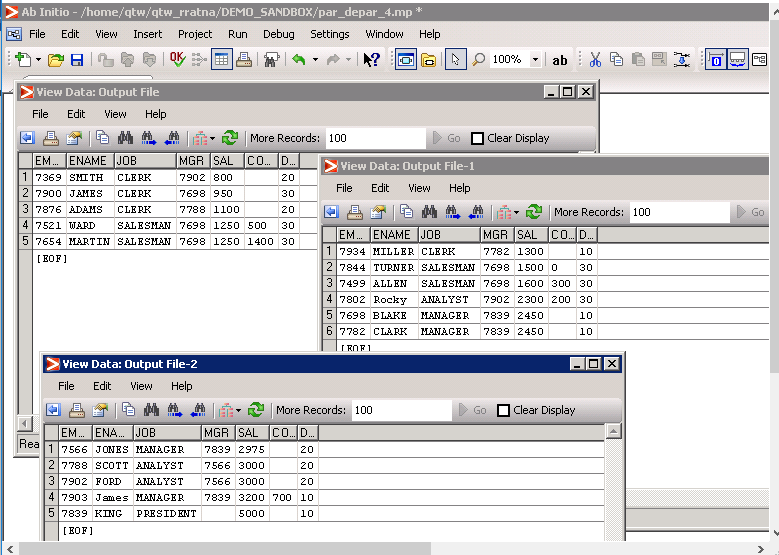


4)Use partition by range and load the data into 3 targets. Add watchers at the desired positions to know which records are created as split records . use salary as the Key column

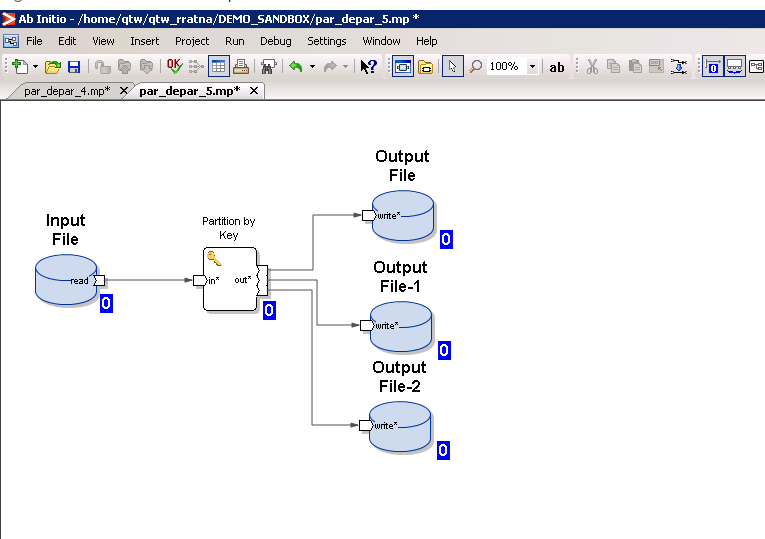


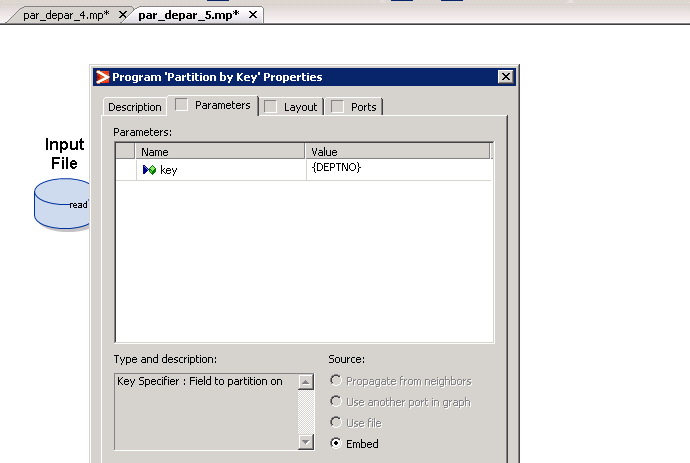


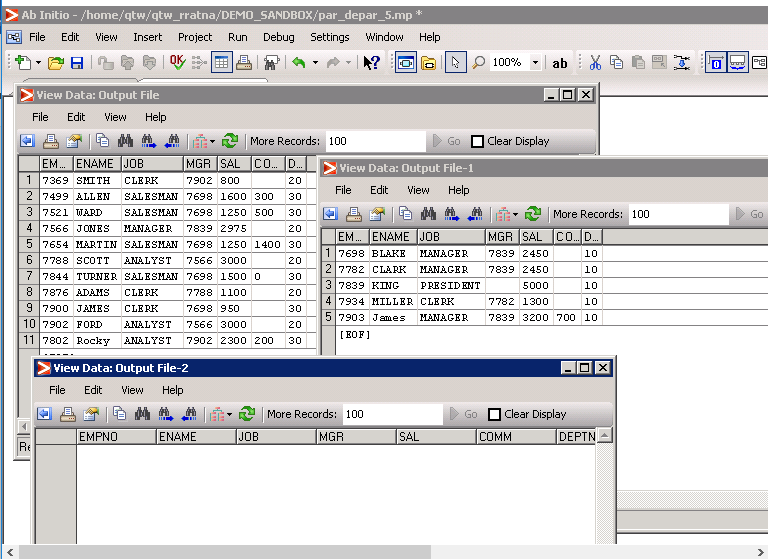




5)Using PARTITION by key component, try to load the data into 3 partitions. Observe if records have copied to all the 3 partitions? Which partition is free.

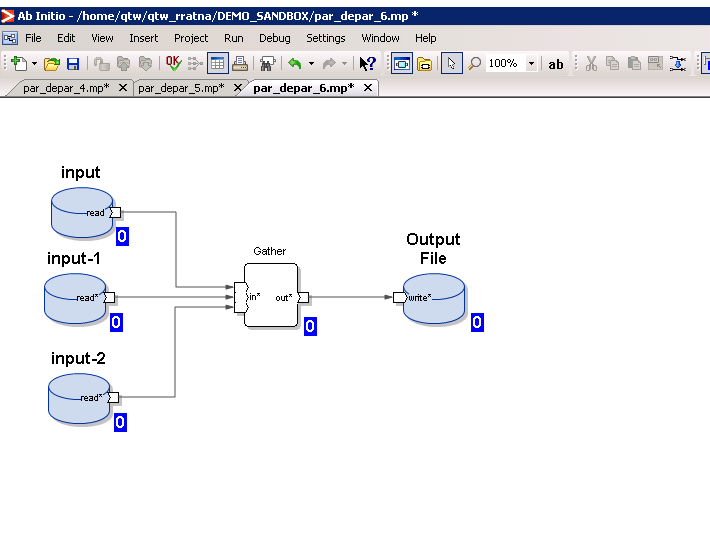


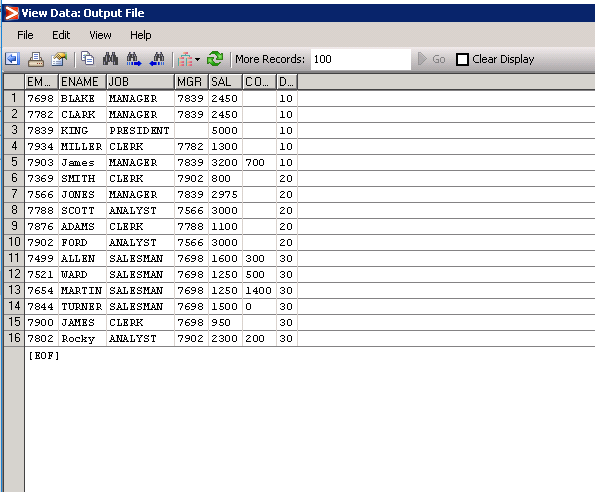




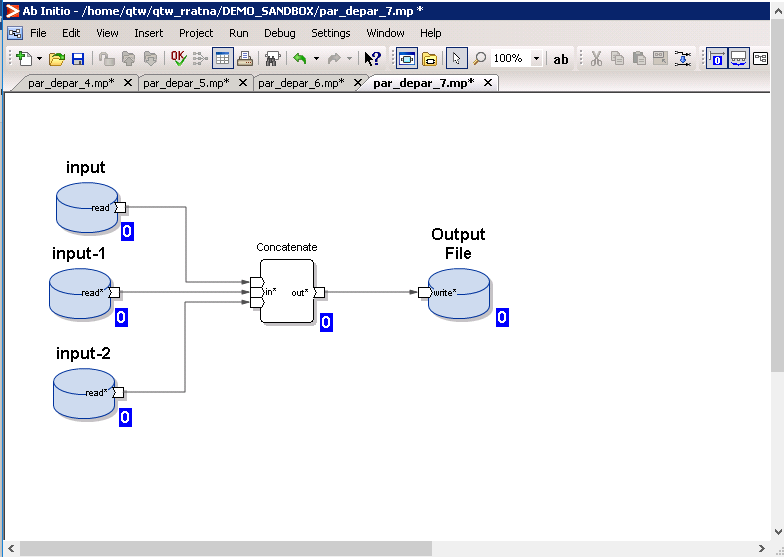
The records are not copied in to all three partitions ,the partition is free (no records) by using partition by key component.

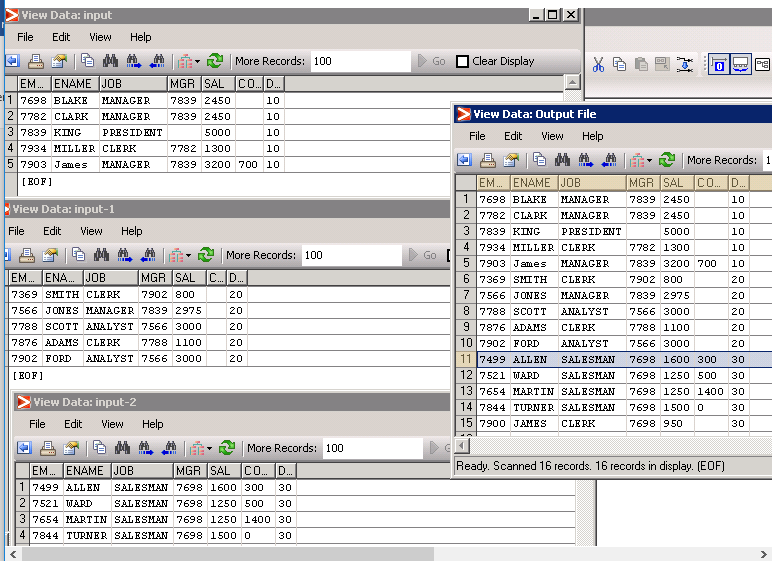
6)Use the 3 target files as source and create a new graph. This graph should read the data from the 3 source files, consolidate and load the data into empall.txt target.



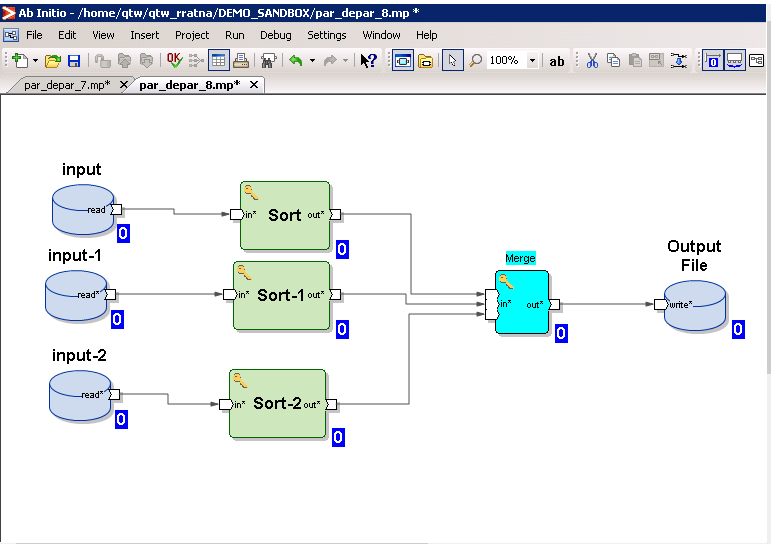


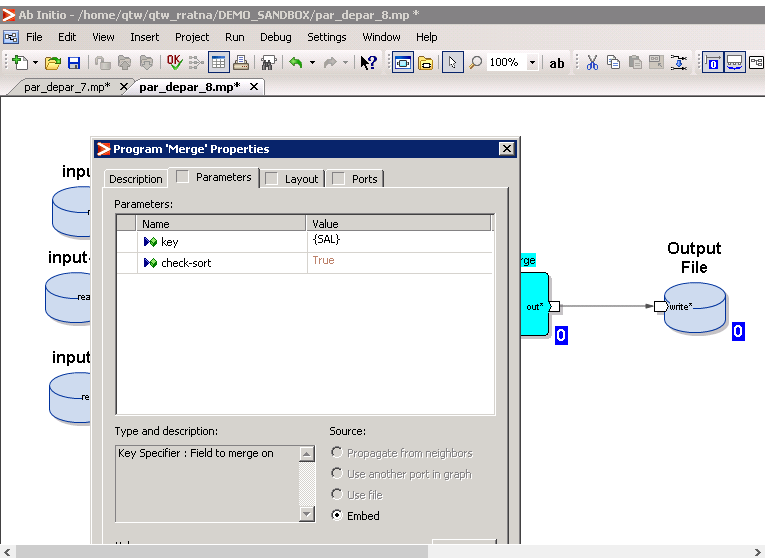
7)Create a graph which will read the data from 3 source files in a sequence.

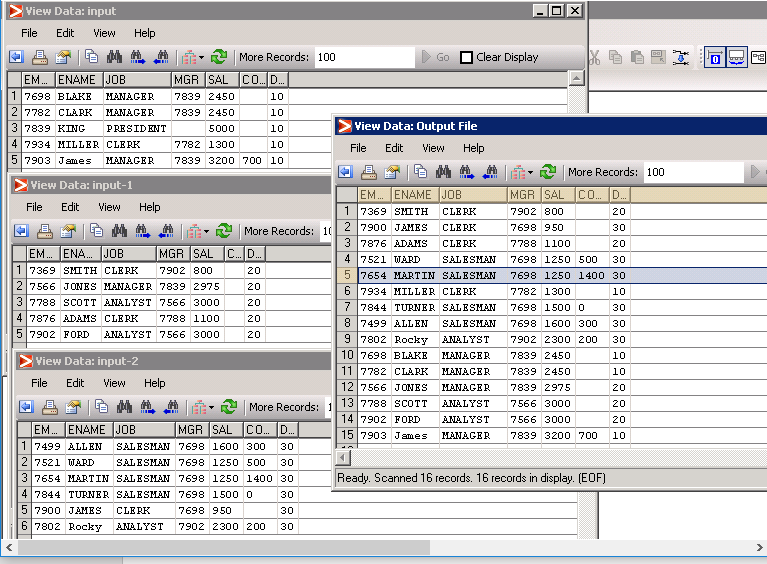




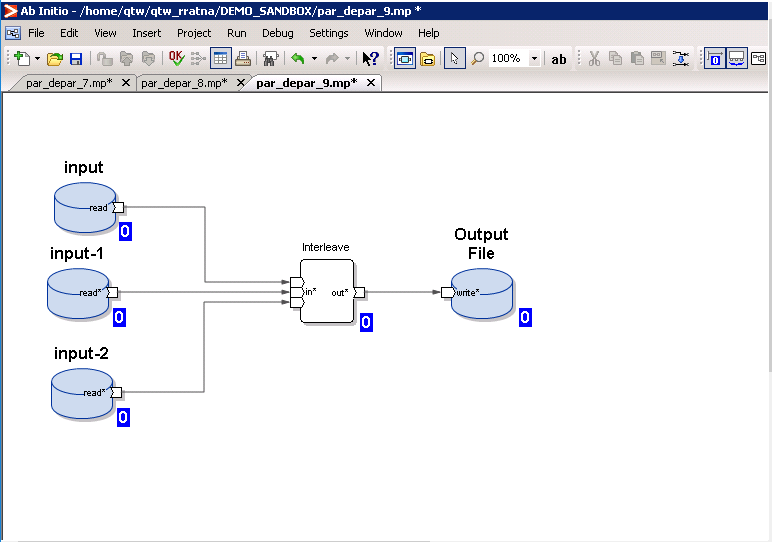
8)Create a graph which will load the data from 3 source files into the target. The target file must have data sorted

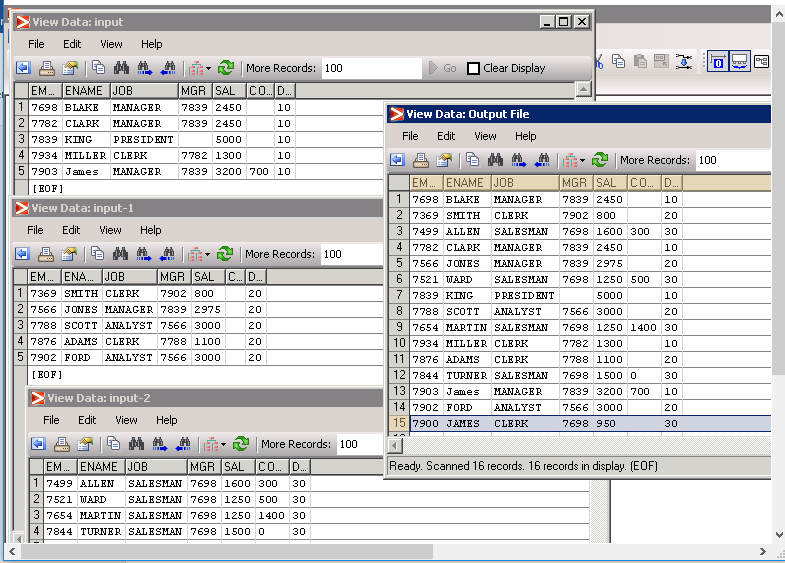




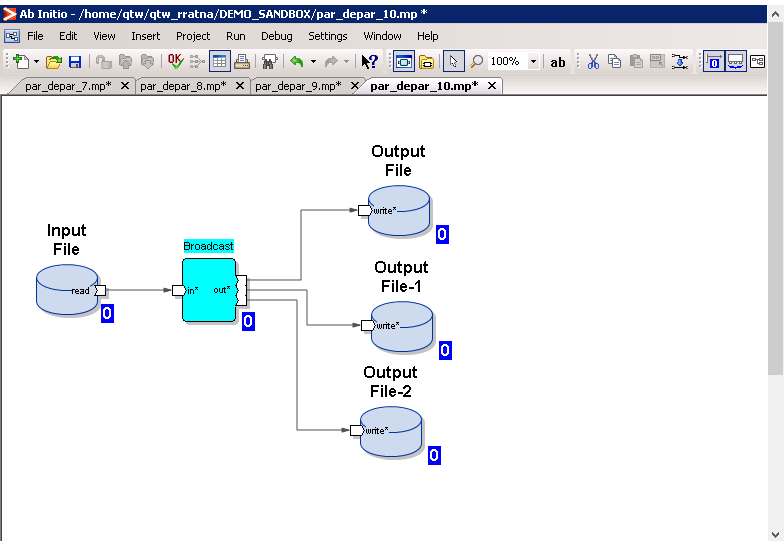


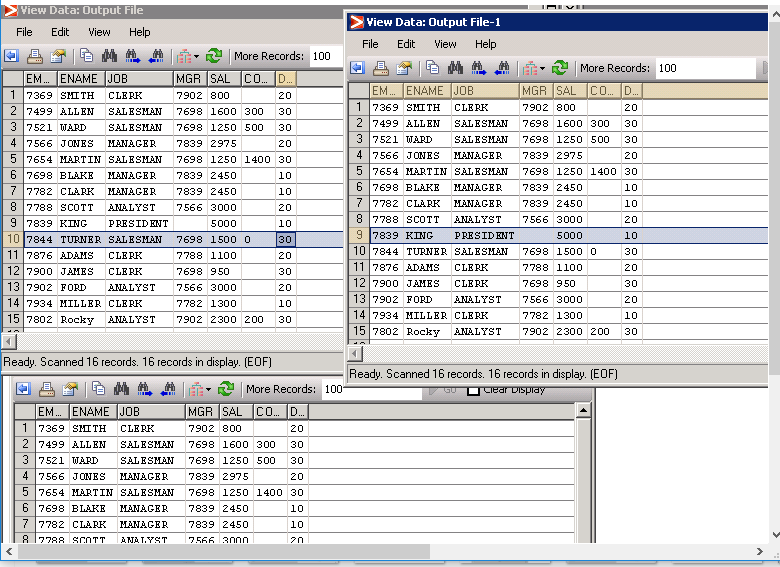
9)Create a graph that will load the data from 3 source files in a cyclic order





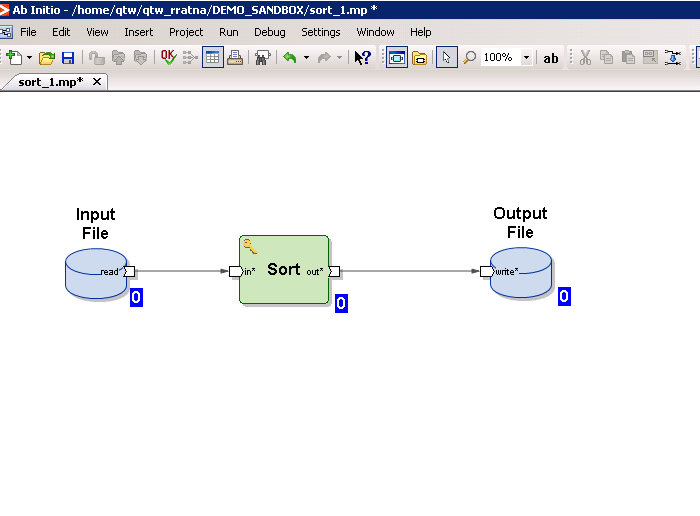
10)Create a graph which will read the data from SRC\_EMP.txt file and broadcast the same data to all the 3 targets

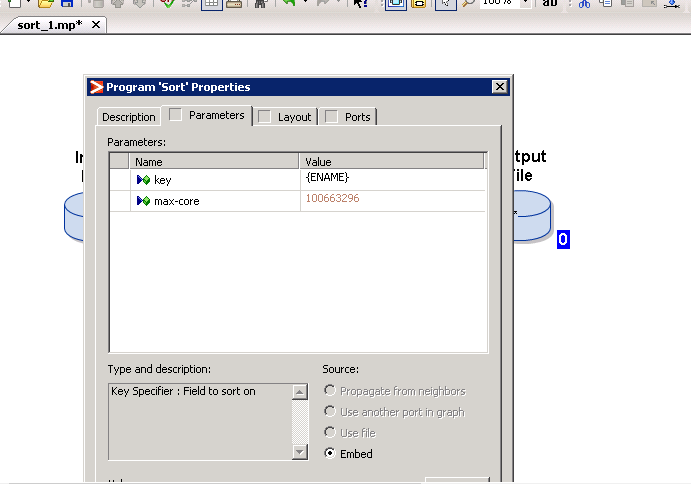


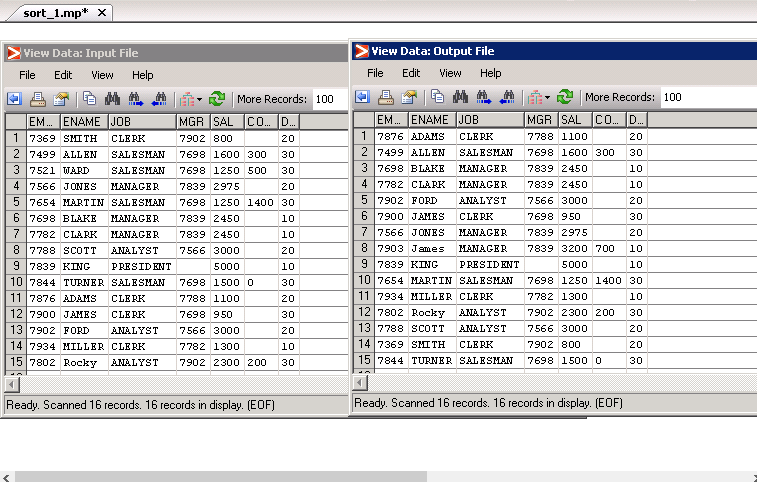


**SORT COMPONENTS:**

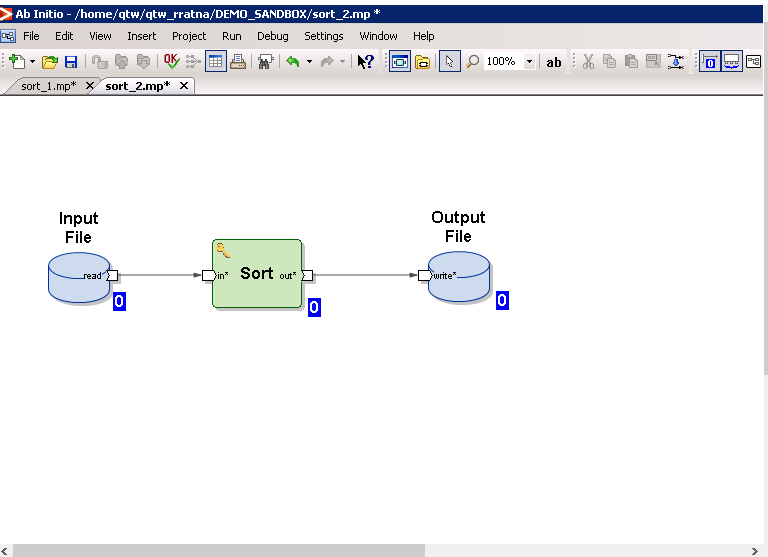
1. Use the EMP\_SRC as Source. Sort the data in ascending order of employee name and store the result in EMPSORT.txt output file (create the file if it does not exists).

****

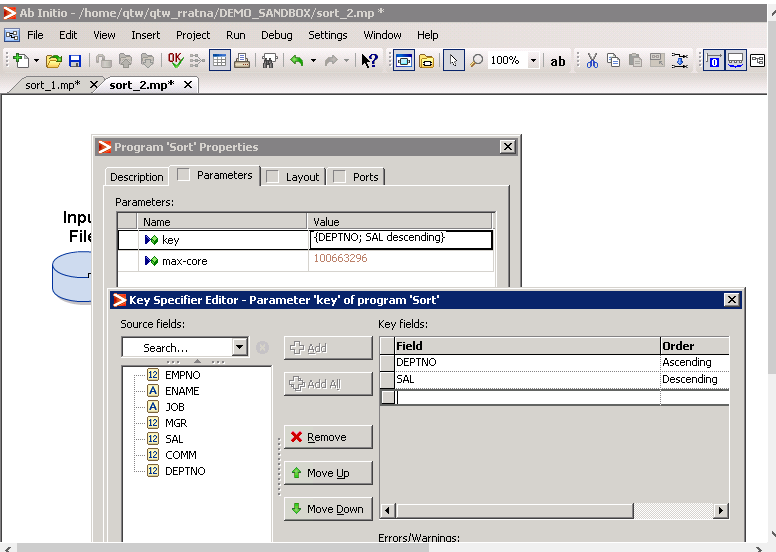
****

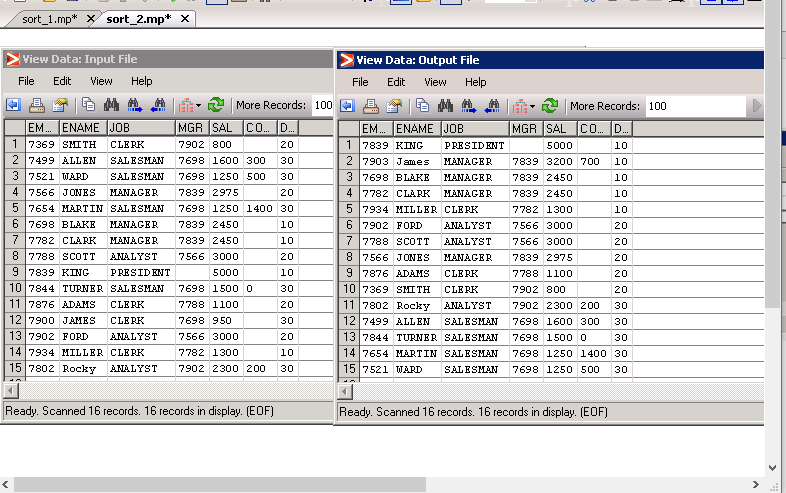
****

1. Create a graph and sort the data in such a way that it should display the department numbers in ascending order and for each department displays salary in descending order. (which is the best choice, SORT or SORT WITHIN GROUP)

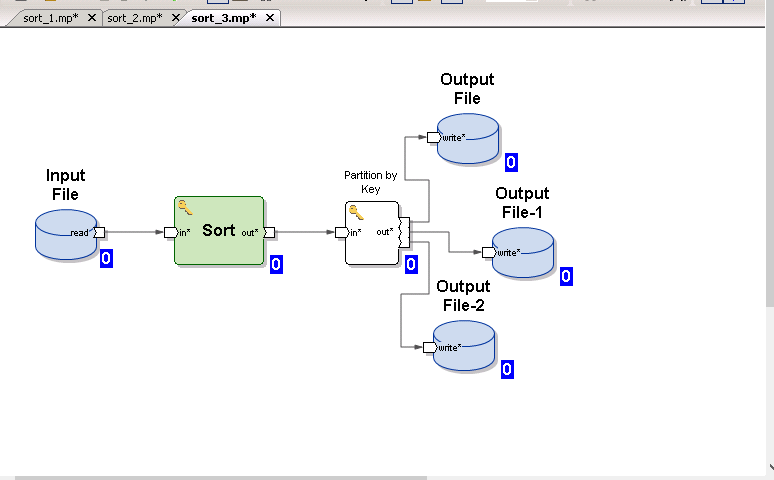
****

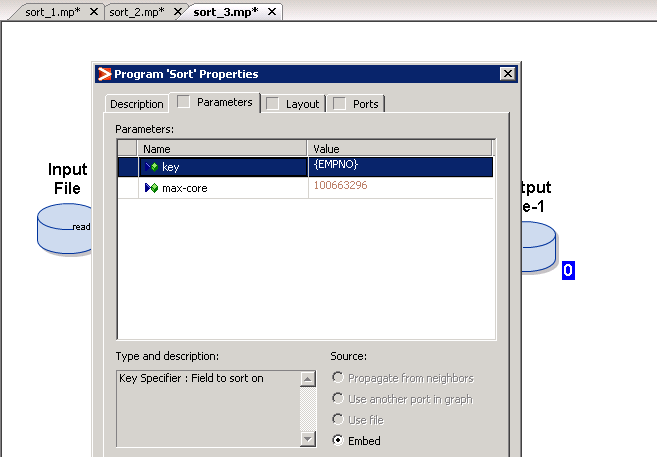
The best choice is to use sort component.

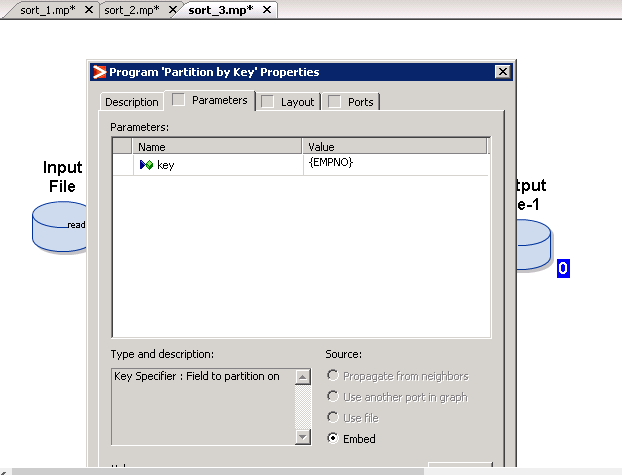
****

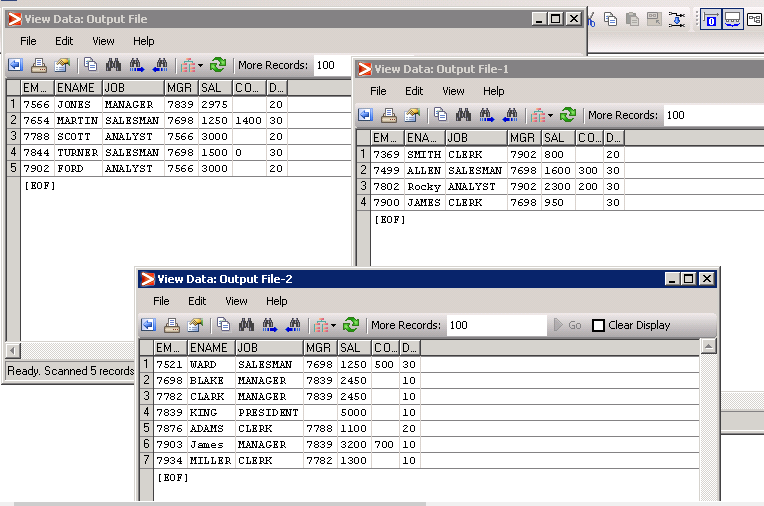
****

1. Create a graph to partition the data employee number wise (take 3 partitions ) and within the partition the data must be sorted on key component





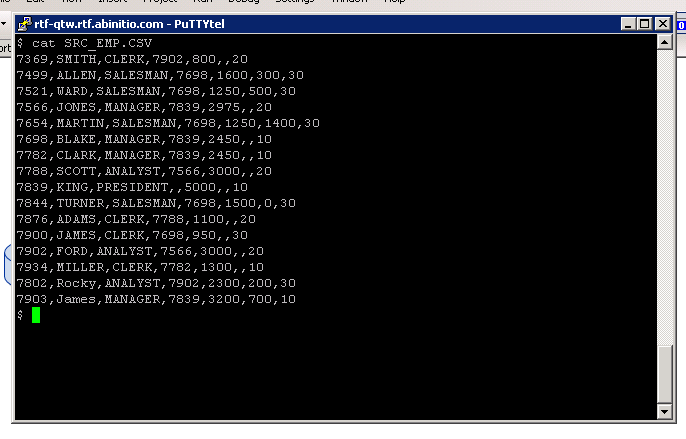




**DATABASE COMPONENTS:**

1. Create a source file by the name SRC\_emp.csv. insert the records as below

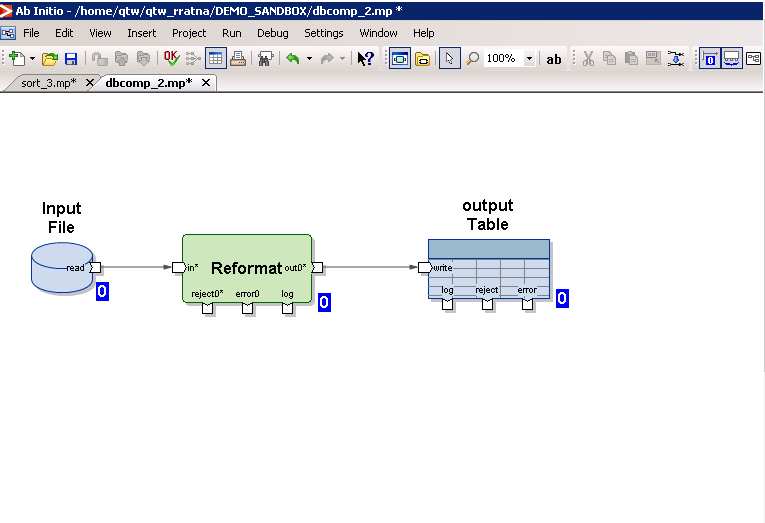
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| EID | Ename | Title | Mgr\_ID | DOJ | Sal | Comm | Dept |
| 7369 | SMITH | CLERK | 7902 | 17-Dec-80 | 800 |  | 20 |
| 7499 | ALLEN | SALESMAN | 7698 | 20-Feb-81 | 1600 | 300 | 30 |
| 7521 | WARD | SALESMAN | 7698 | 22-Feb-81 | 1250 | 500 | 30 |
| 7566 | JONES | MANAGER | 7839 | 2-Apr-81 | 2975 |  | 20 |
| 7654 | MARTIN | SALESMAN | 7698 | 28-Sep-81 | 1250 | 1400 | 30 |
| 7698 | BLAKE | MANAGER | 7839 | 1-May-81 | 2850 |  | 30 |
| 7782 | CLARK | MANAGER | 7839 | 9-Jun-81 | 2450 |  | 10 |
| 7788 | SCOTT | ANALYST | 7566 | 19-Apr-87 | 3000 |  | 20 |
| 7839 | KING | PRESIDENT |  | 17-Nov-81 | 5000 |  | 10 |
| 7844 | TURNER | SALESMAN | 7698 | 8-Sep-81 | 1500 | 0 | 30 |
| 7876 | ADAMS | CLERK | 7788 | 23-May-87 | 1100 |  | 20 |
| 7900 | JAMES | CLERK | 7698 | 3-Dec-81 | 950 |  | 30 |
| 7902 | FORD | ANALYST | 7566 | 3-Dec-81 | 3000 |  | 20 |
| 7934 | MILLER | CLERK | 7782 | 23-Jan-82 | 1300 |  | 10 |
| 7802 | Rocky | ANALYST | 7902 | 24-Feb-89 | 2300 | 200 | 30 |
| 7903 | James | MANAGER | 7839 | 21-Feb-90 | 3200 | 700 | 10 |

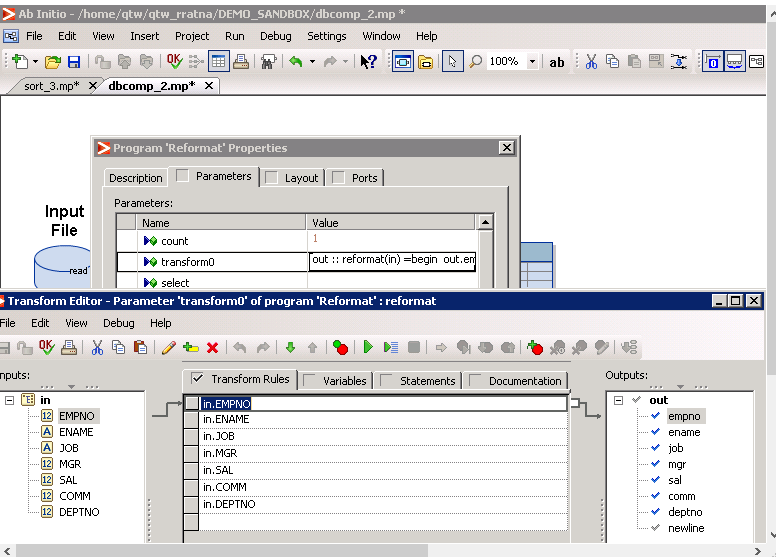


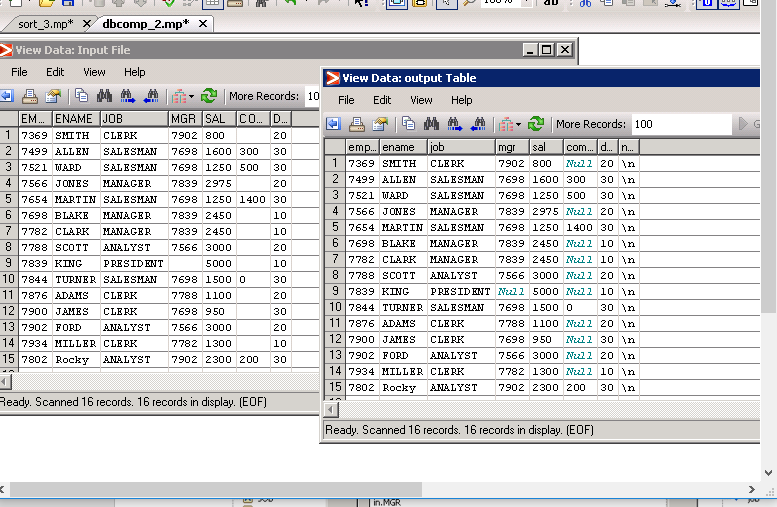
2)Create a graph and use RUNSQL component to create a table named EMPTAB with the same structure as that of the file above.

&

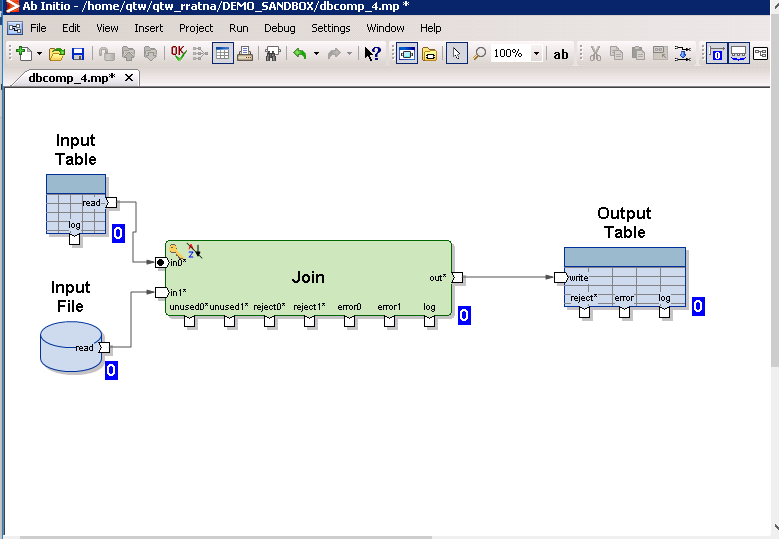
3)Create another graph and name it as file\_to\_table.mp. Use SRC\_EMP.TXT as input file and EMPTAB as target table. Load the data from file to the table

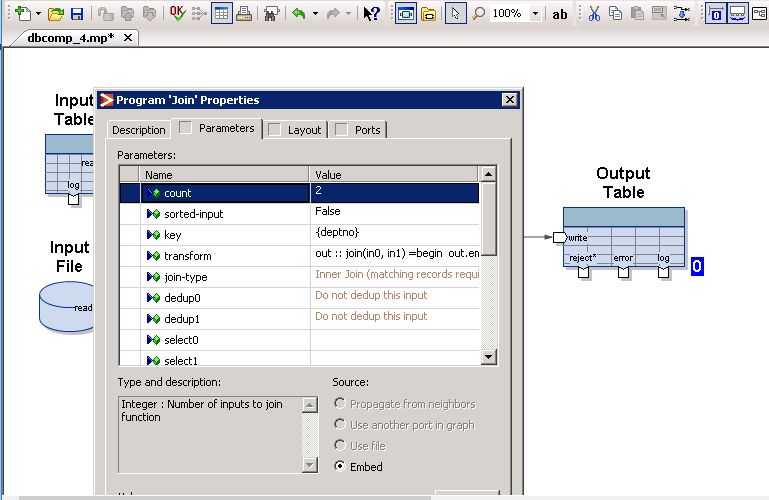


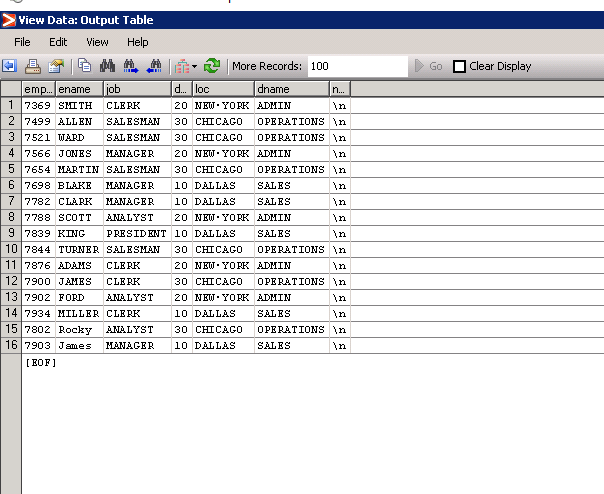




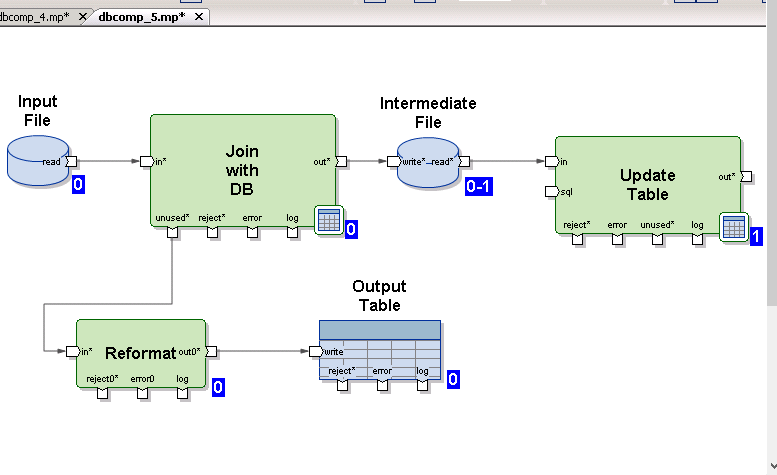
4)Use EMPTAB as input table and SRC\_DEPT. TXT as input file. Perform a join using table and file, and load the joined result in a new table called EMPDEPTJOIN. (create the table, if it does not exists)

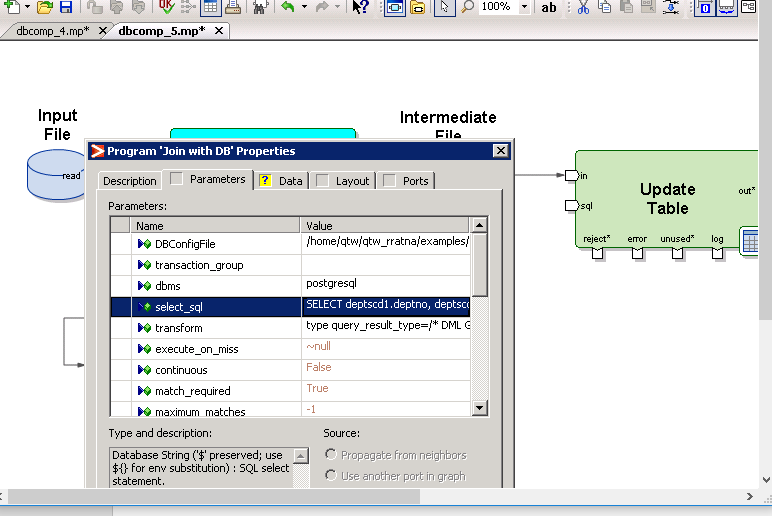


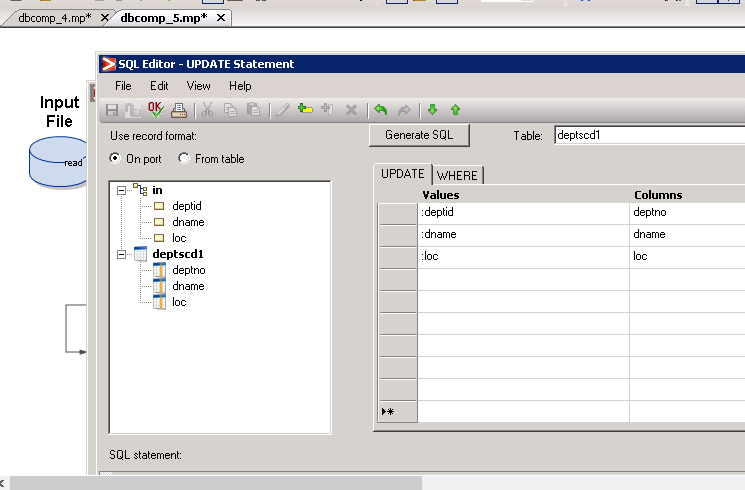


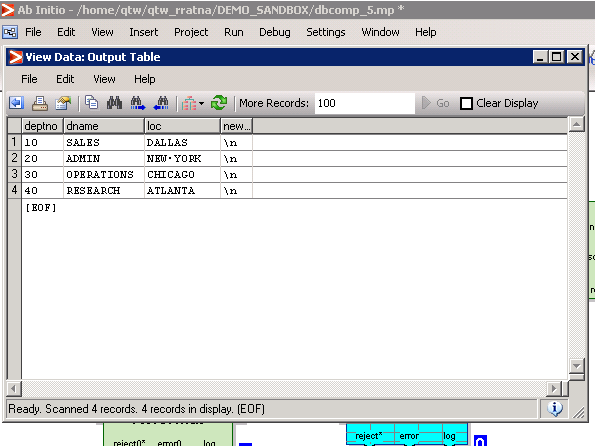


5)Create a table named DEPTSCD1. Create a graph which will read the data from the SRC\_DEPT.TXT FILE to DEPTSCD1 table. While doing so, it should check if the record exists in the table or not. If a particular deptno exists in the table, then that record should get updated and if the record does not exists, then that should get inserted as new record









6) Create a source file with the below data

Source file:

Dno,dname,loc

10,accts, hyd

Create a target table with the below record format

Target file:

Dno,dname,curr\_loc, prev\_loc

When you run the graph, the source data should get copied to target table as mentioned below

Dno🡺 dno

Dname🡺 dname

Loc🡺 currloc

Dno,dname,curr\_loc, prev\_loc

10,accts, hyd

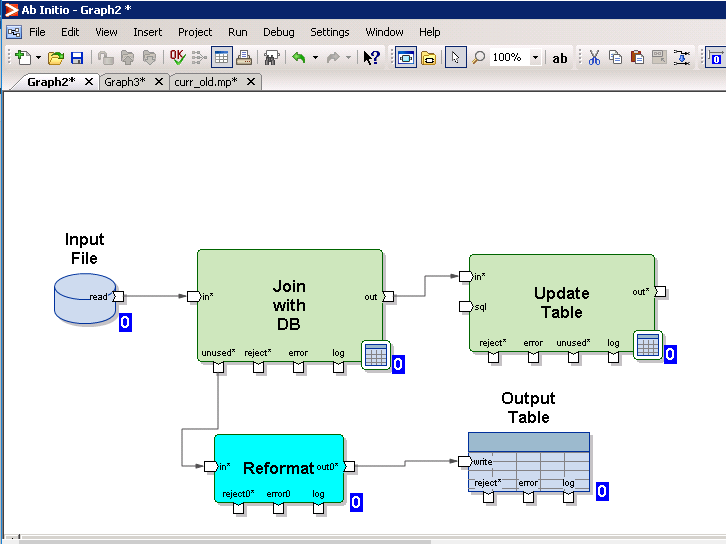
If the source record changes the target record should also change. For example if loc value changes in the source then in the target the curr\_loc should move to prev\_loc and curr\_loc should have new value coming from the source.

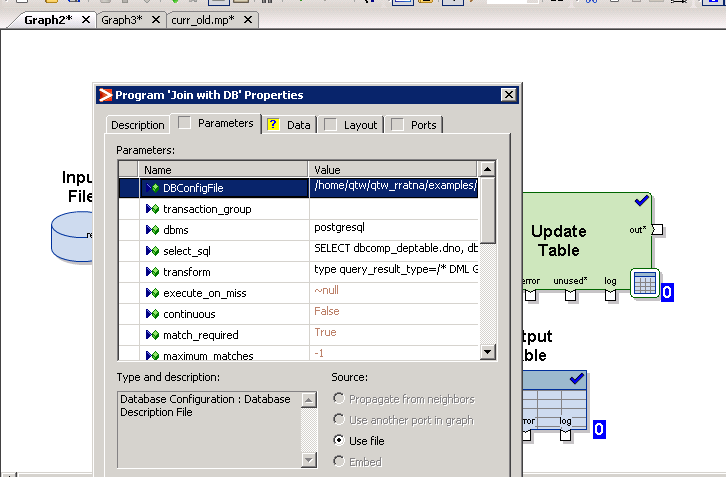
If the above record changes to pune , target file should look like

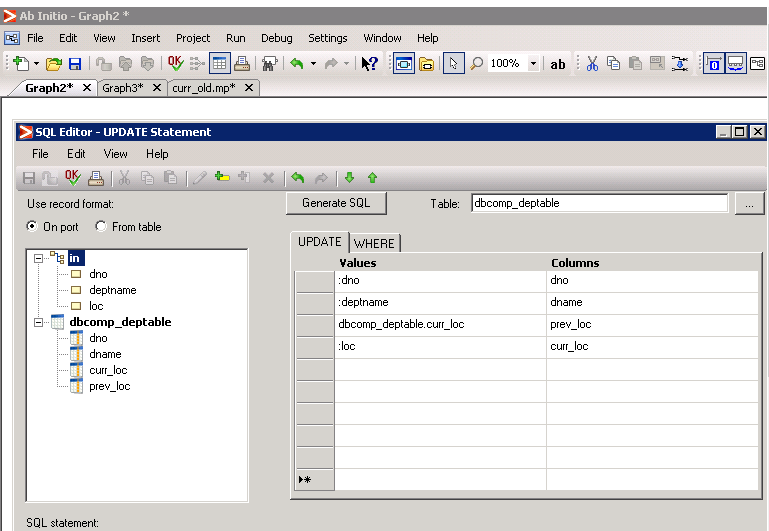
Target file

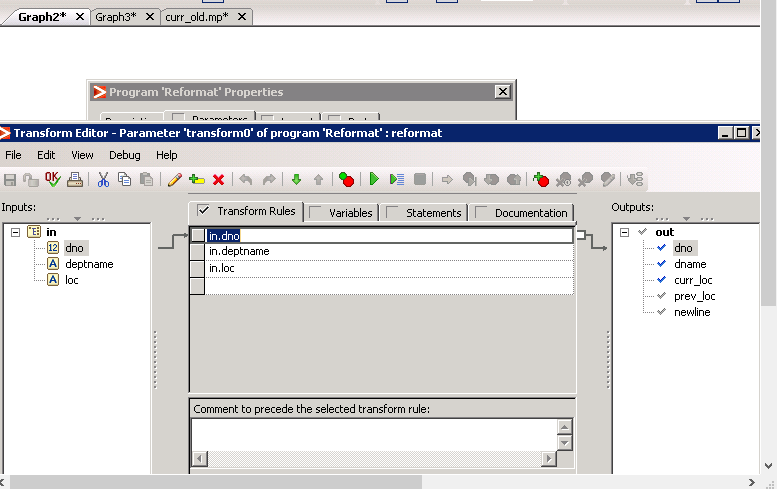
After the second run:

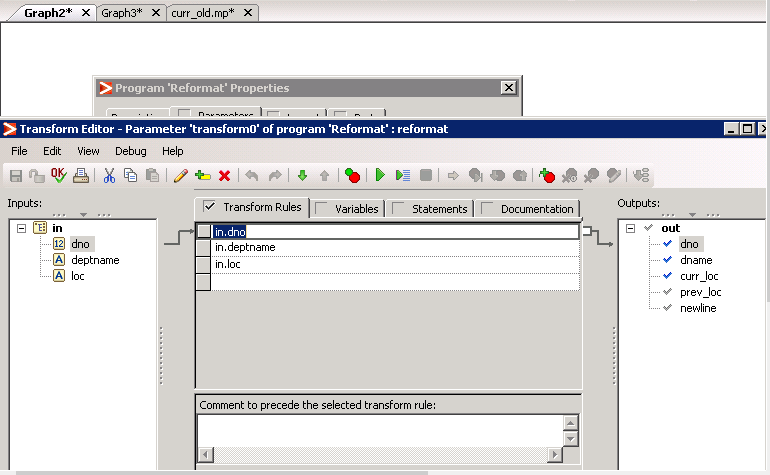
10,accts,pune , hyd

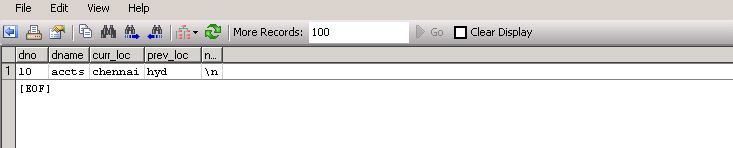






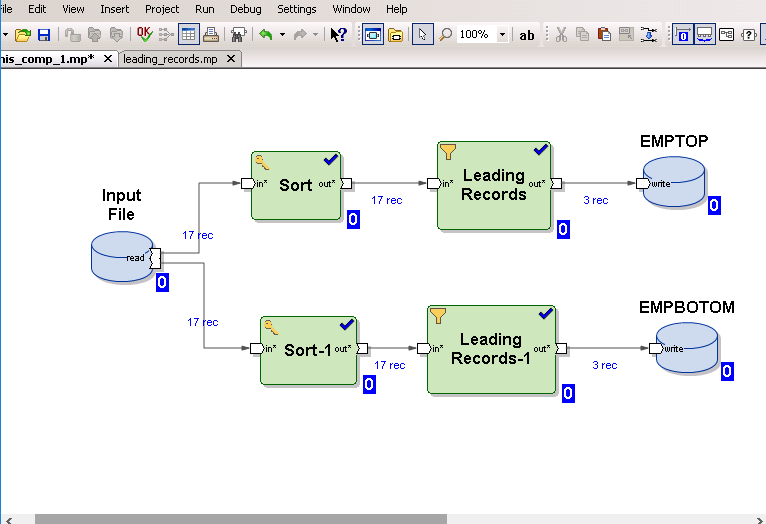


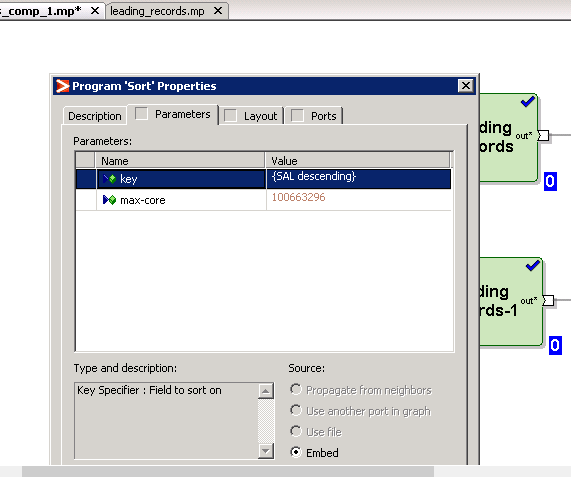


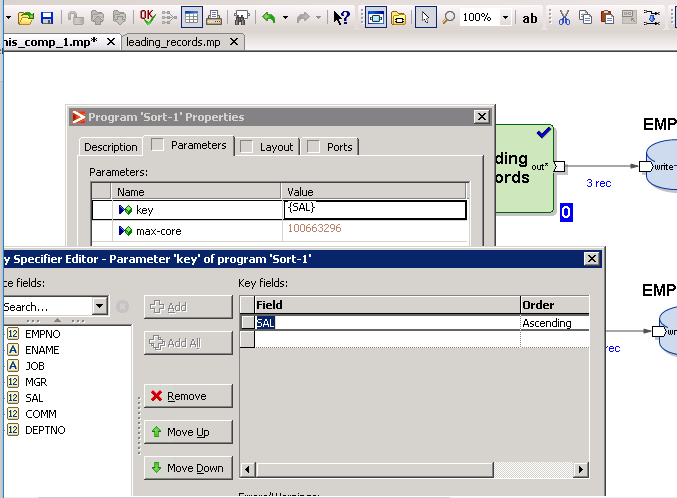


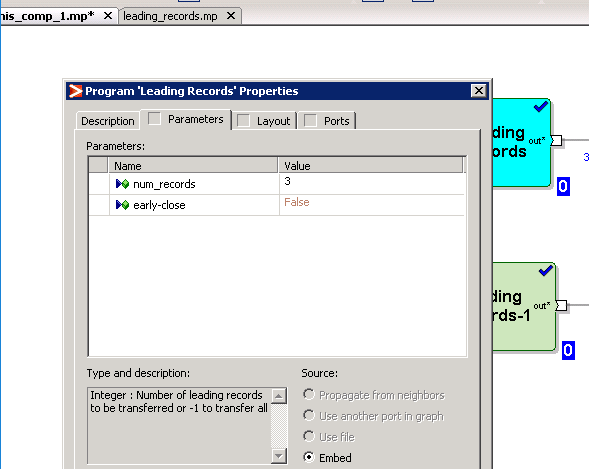
**MISCELLANEOUS COMPONENTS:**

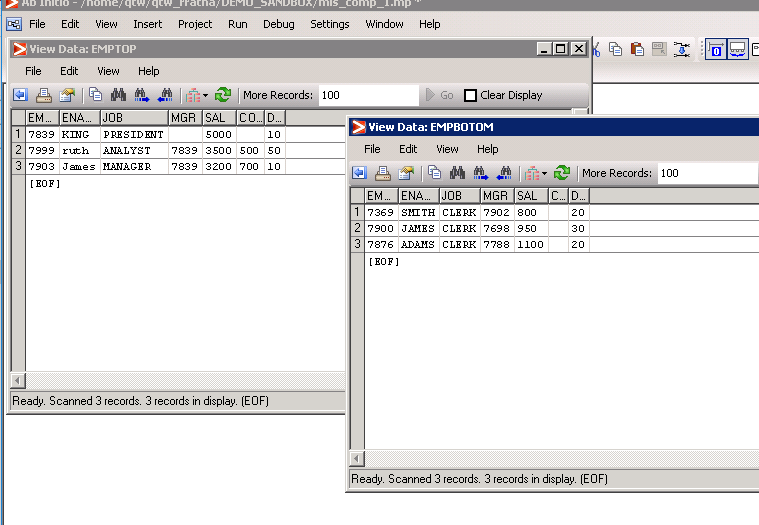
1. Use the source file EMP\_SRC.TXT. Develop a graph to load the target file named EMPTOP\_N.TXT with top-3 salary records and EMPBOTOM\_N.TXT with bottom-3 salary records from EMP\_SRC.txt file



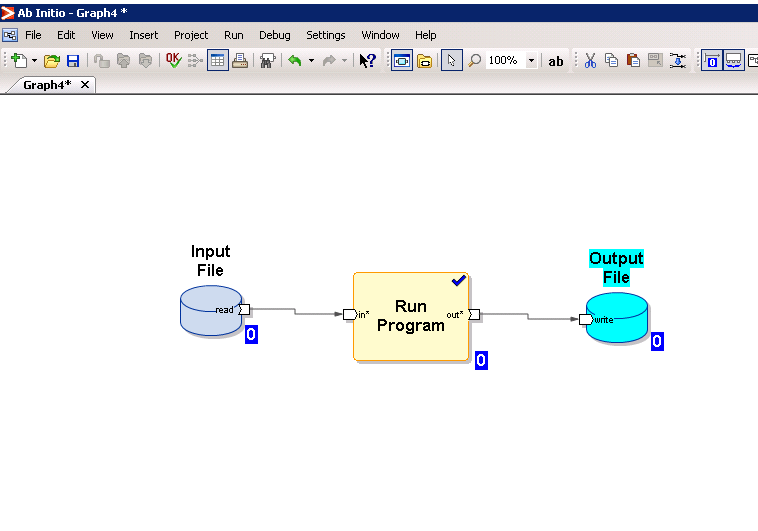


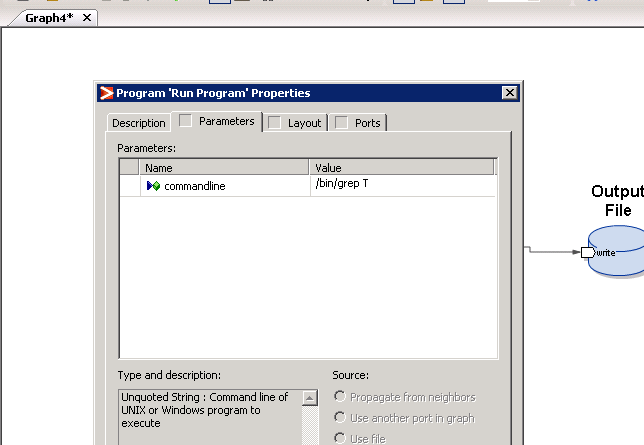






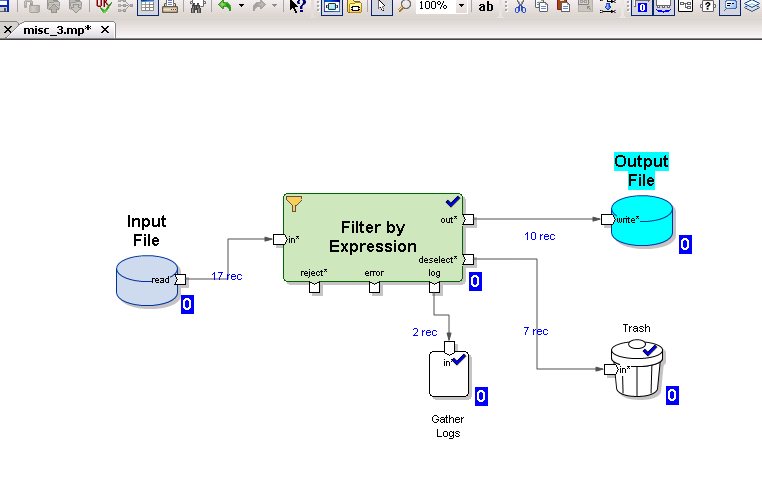
2)Use RUN PROGRAM component to read all the records that contains the letter ‘T’ from the EMP\_SRC.TXT. The search should be case insensitive

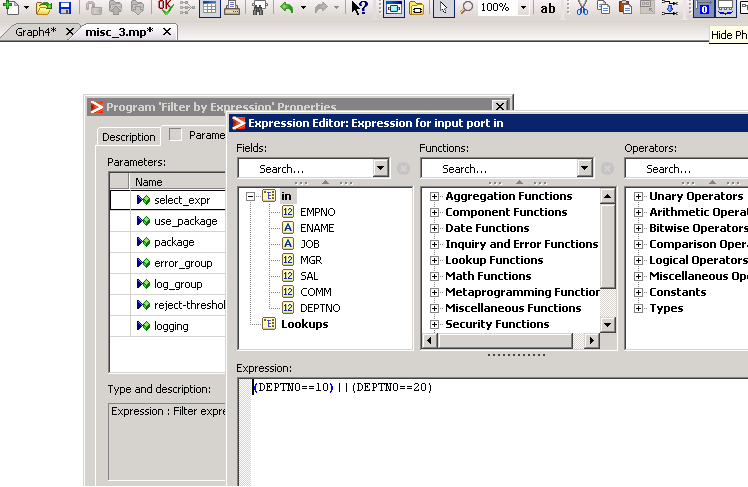


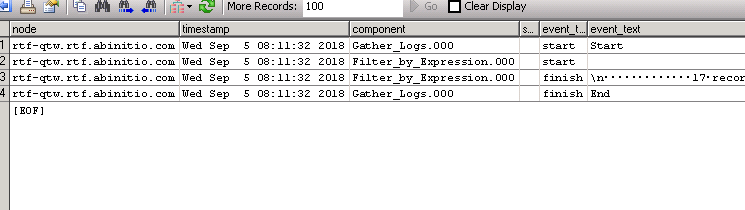


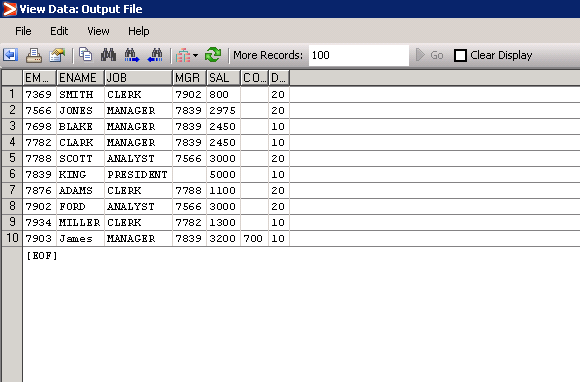


3)Use EMP\_SRC.TXT as source file. draw a graph to load only dept10 and 20 records to the target. Other records should go to Trash. Use a gather log component to know how much time the component has taken to execute





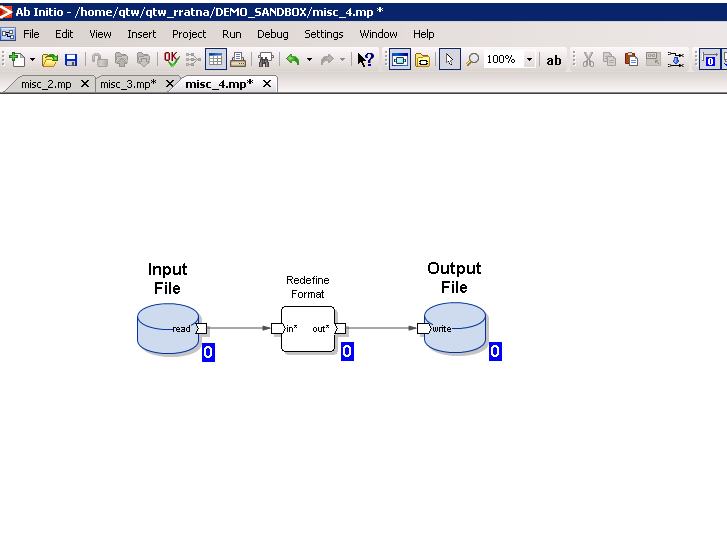


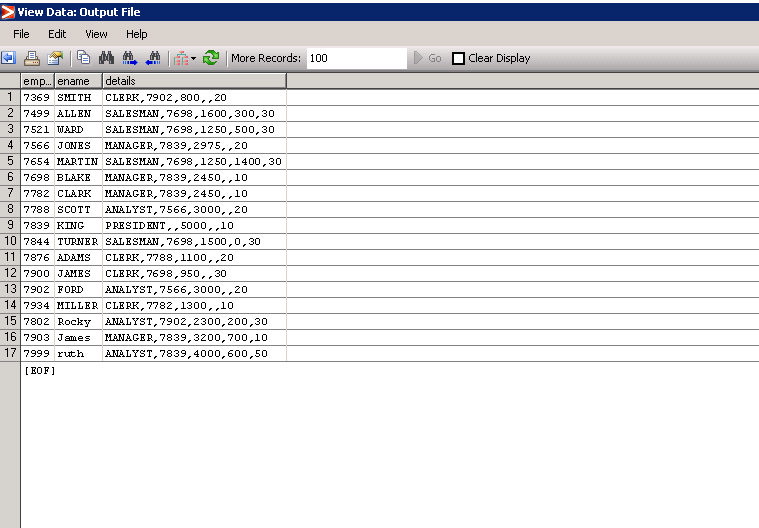


4)Source file is EMP\_SRC.txt. target file contains the following columns

Empno, ename, details

Use the redefine component to move the data from source to the target. Check how many columns data is moved and how is it organized





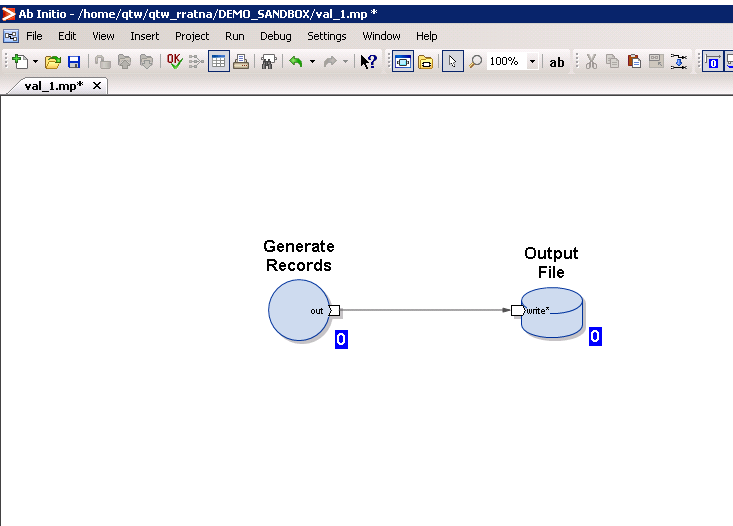
**VALIDATION COMPONENT**

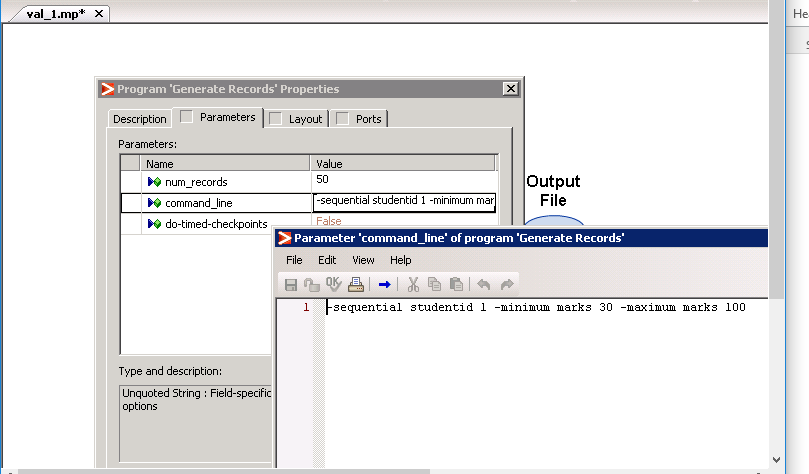
1. Use the generate record component to generate 50 records in the RECGEN.TXT file

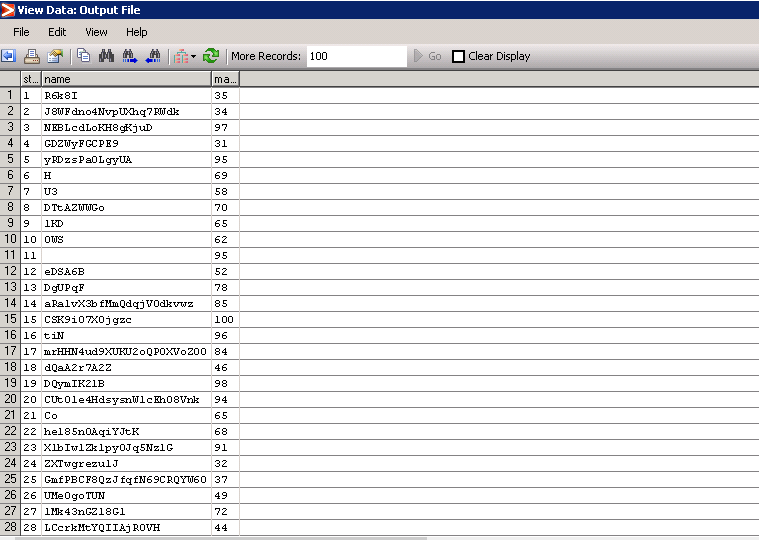
The file has the below structure:

Studentid, name, marks

* Studentid values must be sequential and should start with 1
* Marks values must be between 30 to 100

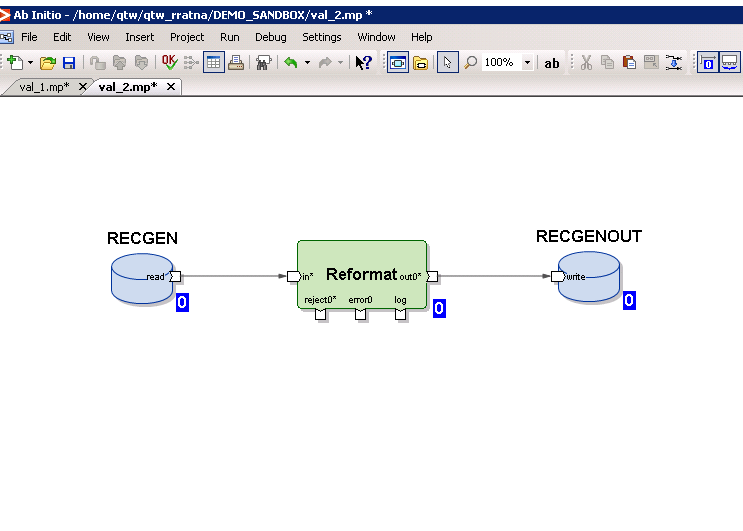


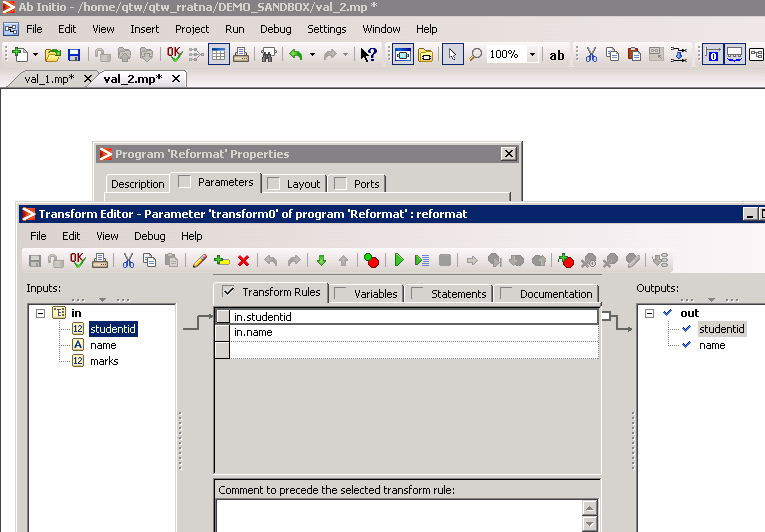


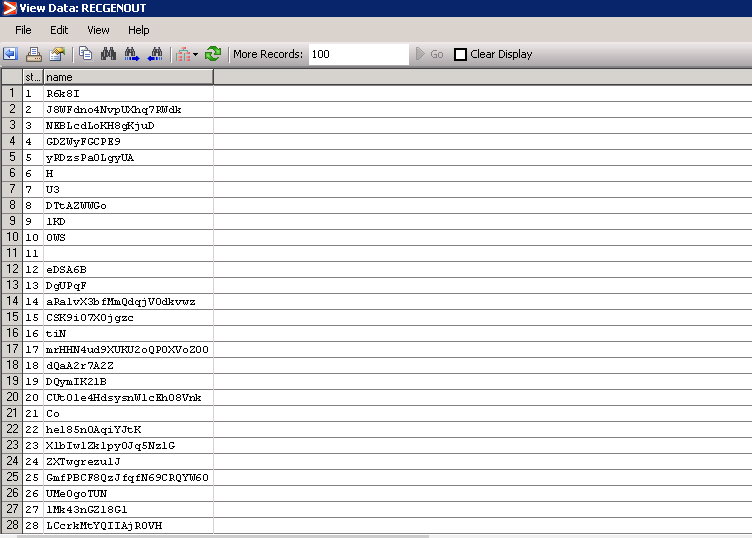


2)Use the RECGEN.TXT file created above and load the data into the RECGENOUT.TXT file which has the below record format

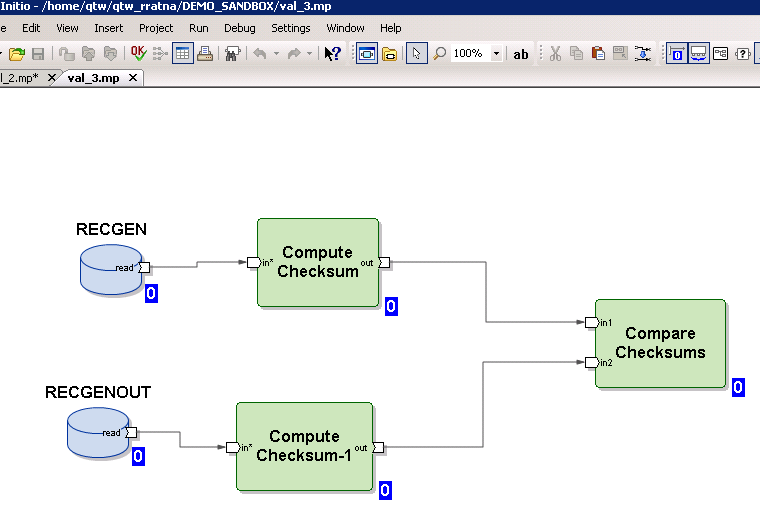
Studentid, name







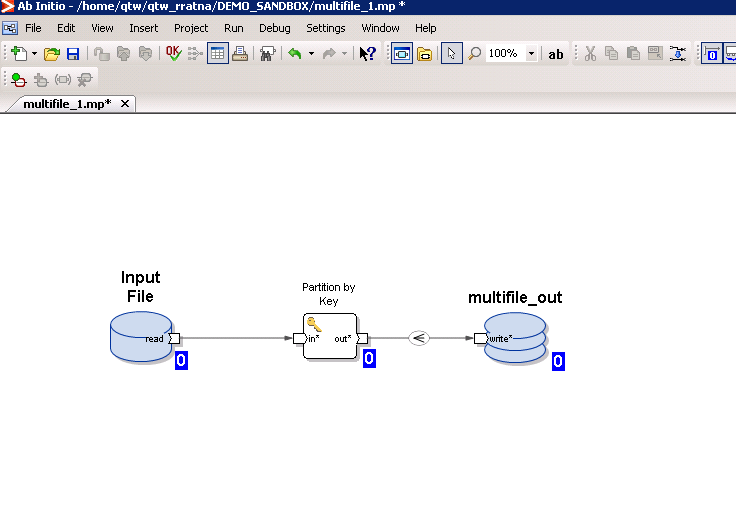
3)Create a graph which uses RECGEN.TXT and RECGENOUT.txt file as source. Compare both the files to check if they are identical.

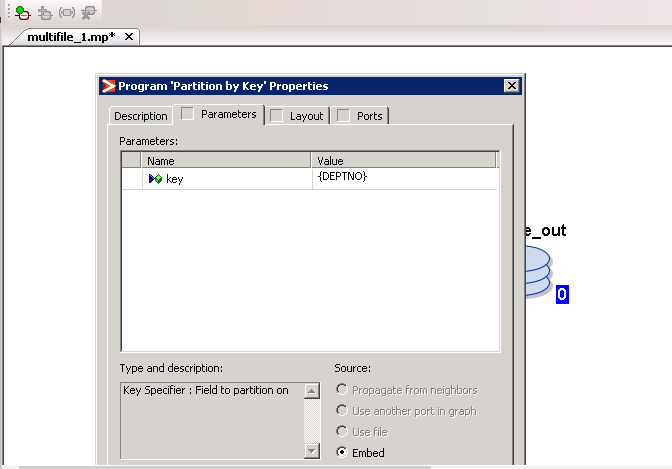


Graph did not execute ,so the files are not identical.

**MULTIFILE**

1. Create a 3-way multifile system and use the same as a target. To this target, load the data from EMP\_SRC.TXT file. use any partition component of your choice





2)Use the above multifile as source and load the data into a target file. give any name of your choice.

