

## Practical No.4

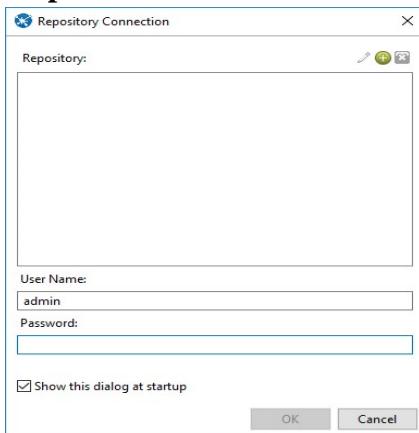
**Aim :** Perform Transformation on source Table and store the data to Output table in SQL.

Starting with Pentaho

**Step 1. Open Data Integration folder (C:\data-integration)**

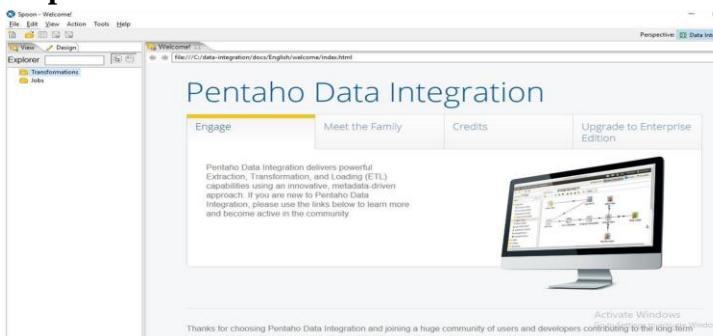
**Step 2. Double click on spoon(Windows Batch file)**

**Step 3. Click on cancel.**



1. Transforming Source Table and storing to Output Table in SQL

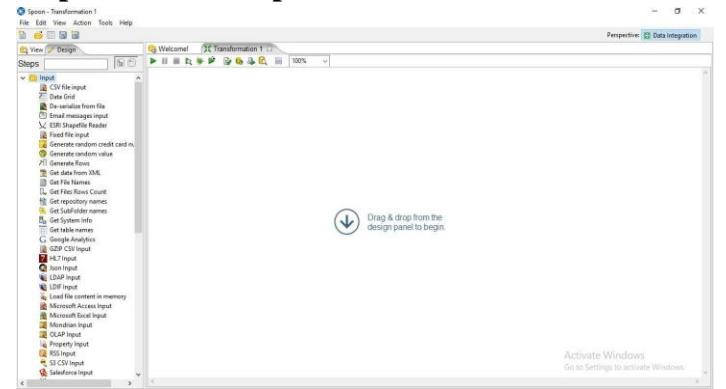
**Step1: Go to Pentaho**



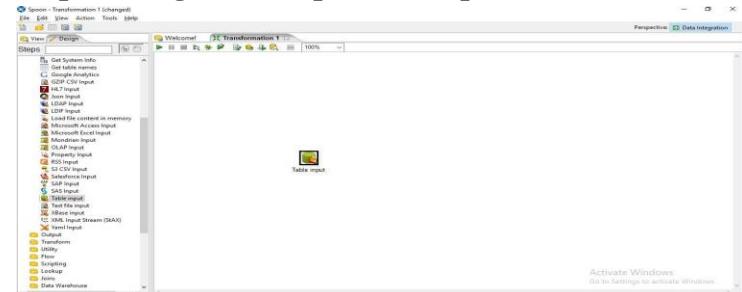
**Step2: Go to file->New->Transformation.**



**Step 3: Click on Input**



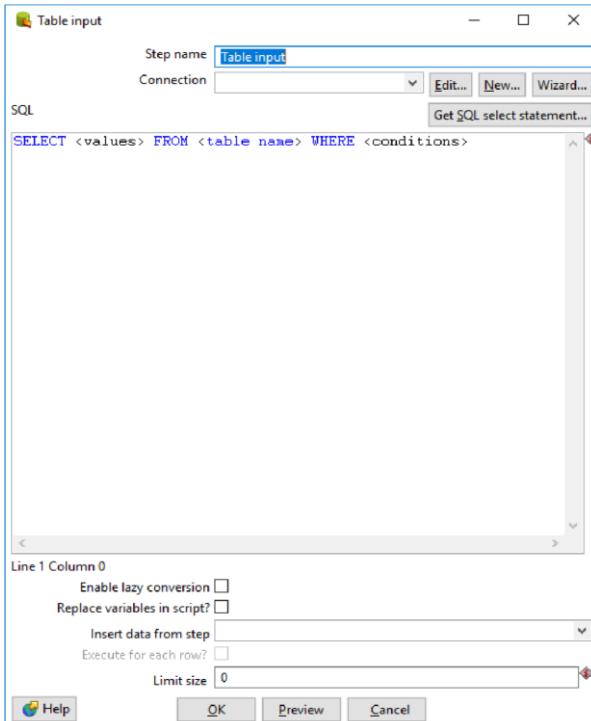
**Step 4: Drag Table Input on the panel**



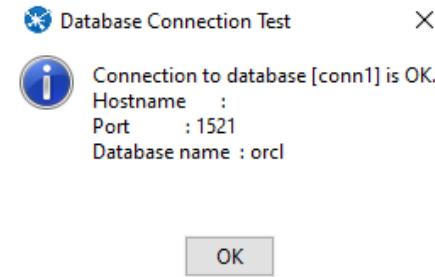
**Step 5: Double Click on Table Input and the following tab will appear.**

a. Click On New

# MCAL13 Advanced Database Management System Lab



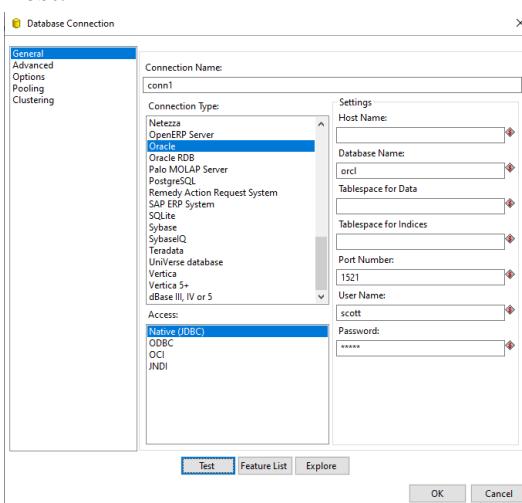
**Connection Successful**



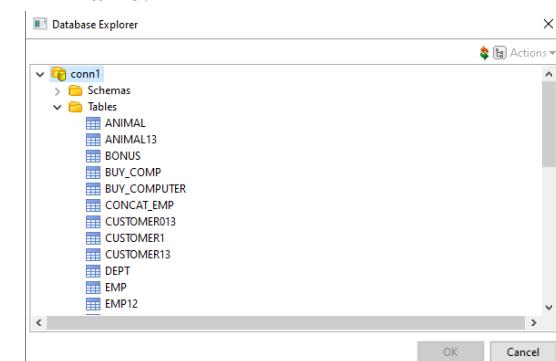
c. Click on get SQL select statement



- b. Select Oracle in Connection Type and enter Connection Name, Database Name, User Name and Password and click on Test.



d. Click on database -> table and select the table name.



e. Click on get Preview.

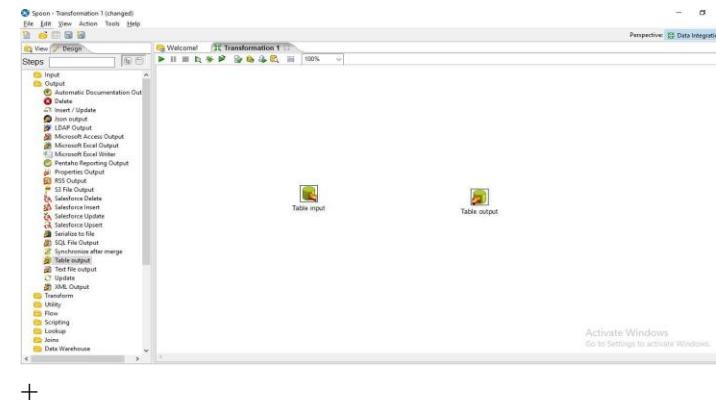


### Examine preview data

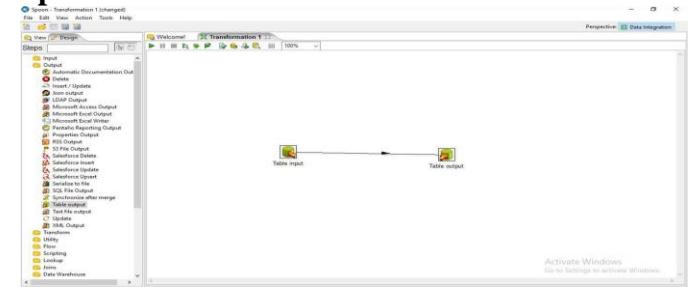
Rows of step: Table input (5 rows)

#	EMP_NO	FNAME	LNAME
1	1	Sonam	Singh
2	2	Pradnya	Suryavanshi
3	3	Komal	Malviya
4	4	Nidhi	Rai
5	5	Riya	Sharma

**Step 6: Click on Output and drag table Output.**

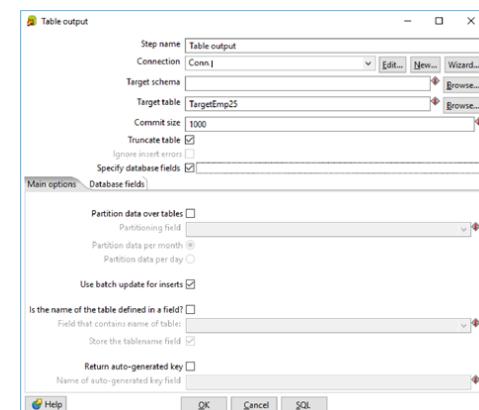


**Step 7: Hold the mouse Pointer on table input and select and drag the Output connector to the Table output.**

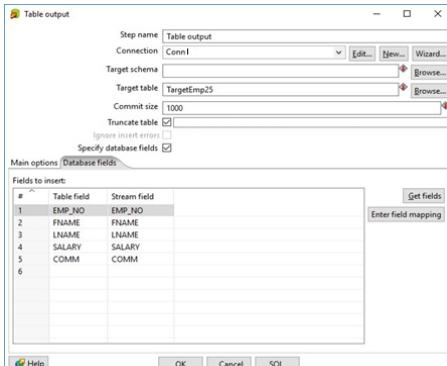


**Step 8: Double Click on target table**

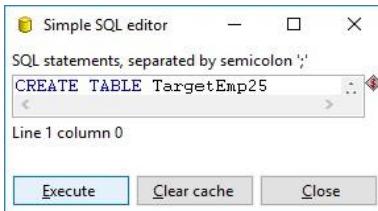
- Enter the Target table name and check the truncate table and Specify database fields Check Box.



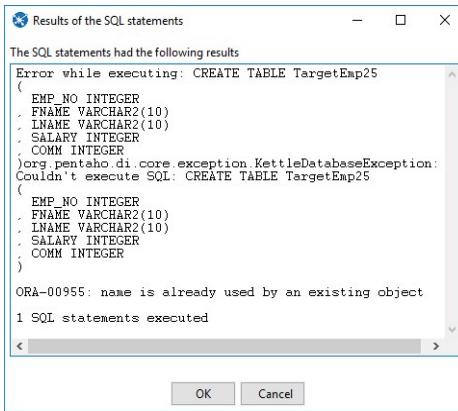
- Click On Database fields and click on Get fields.



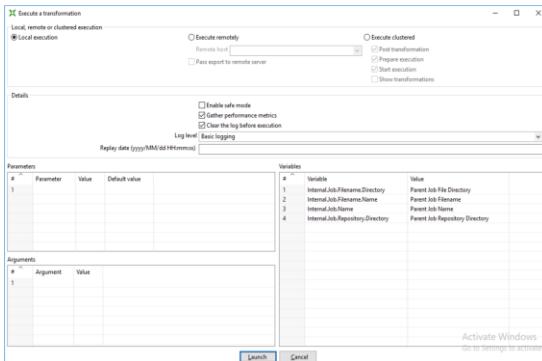
c. Click on SQL and click on Execute.



d. Click Ok



**Step 9: Click on Run Transformation and Click on Launch**

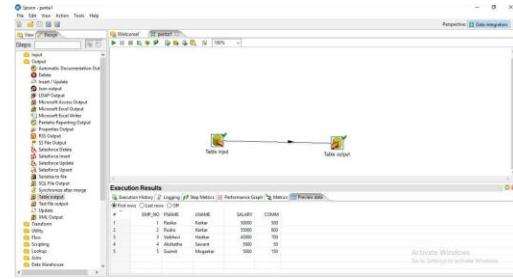


**Step 10: Click on Debug Transformation(Spider) and Click on Quick Launch**



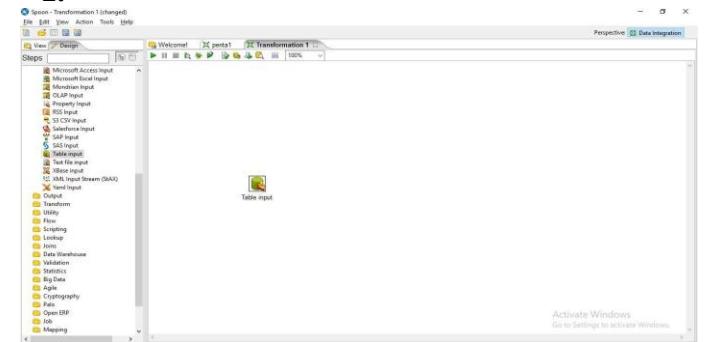
## Output:

The Green ticks on the table input and table output shows successful transformation.

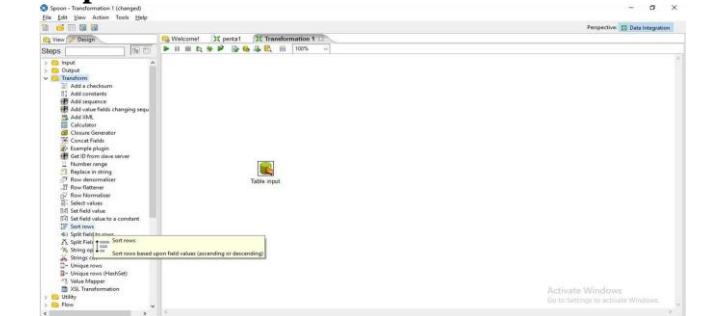


## 2. Sorting Operation and adding Sequence to Output Table.

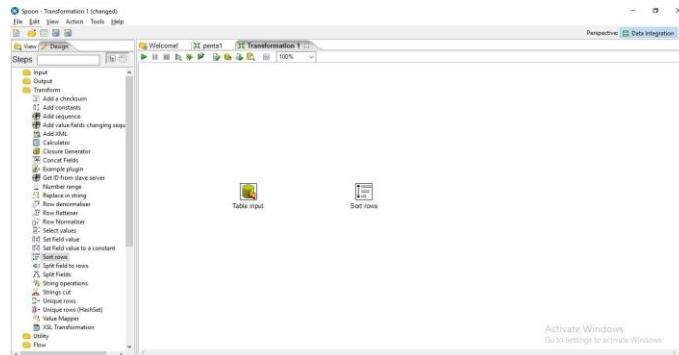
**Step 1: Perform the first 5 steps same as practical 1.**



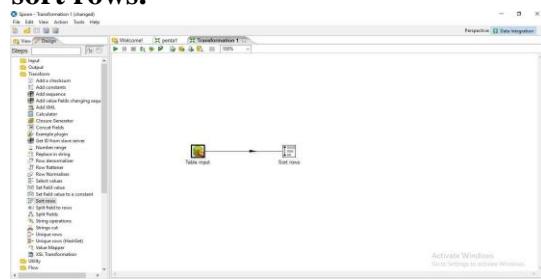
**Step 2: Click on Transform in the left**



**Step 3: Drag sort rows on the Panel**

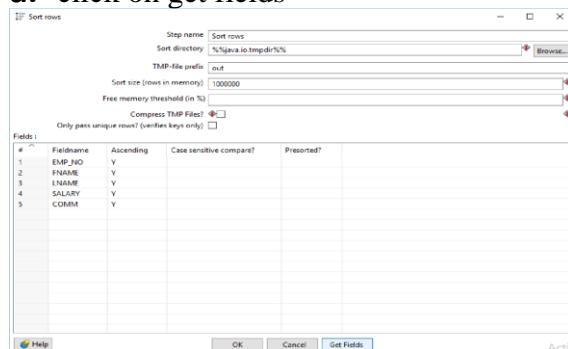


**Step 4:** Hold the mouse pointer on the Table Input and then drag the output connector to the sort rows.

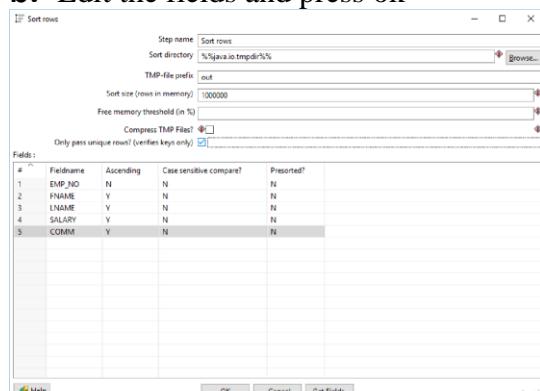


**Step 5: Double click on the sort rows**

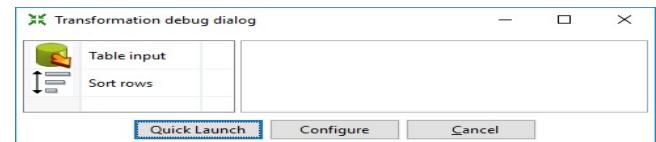
a. click on get fields



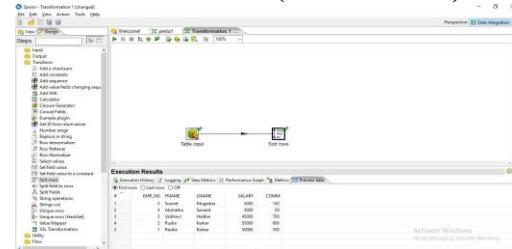
b. Edit the fields and press ok



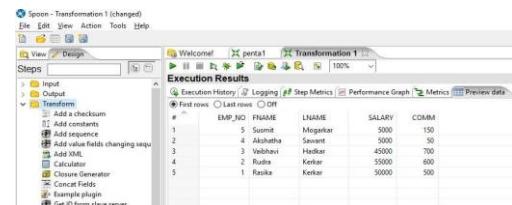
**Step 6: Click on Debug the Transformation and Click on Quick Launch.**



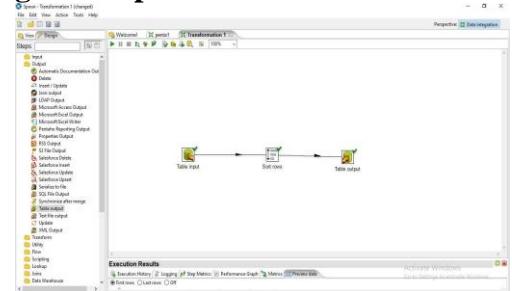
**Successful connections(Green Ticks)**



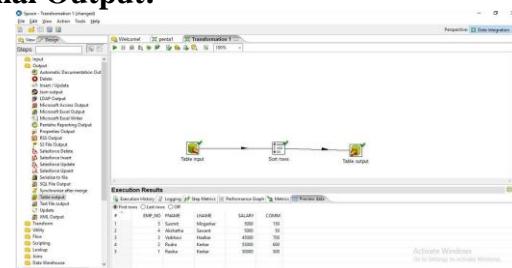
**Sorted Data**



**Step 7: hold the cursor on the sort row and then drag the output connector to the table output.**



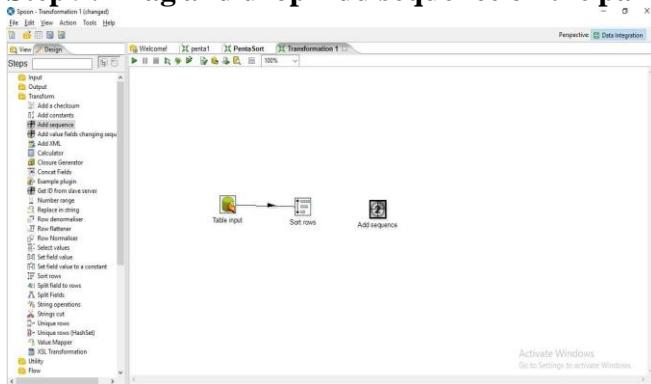
**Step 8: now perform step no 8 to step 10 from practical 1.  
Final Output:**



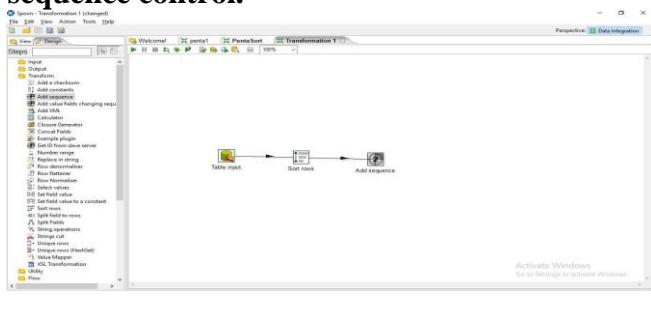
**SQL> select \* from emppp13;**

SQL> select * from emp_target_concat;				
EMP_NO	FNAME	LNAME	SALARY	COMM
1	Sonam	Singh	25000	500
2	Pradnya	Suryavanshi	20000	700
3	Komal	Malviya	17000	1000
4	Nidhi	Rai	30000	900
5	Riya	Sharma	15000	1100

## Step 9: Drag and drop Add Sequence on the panel



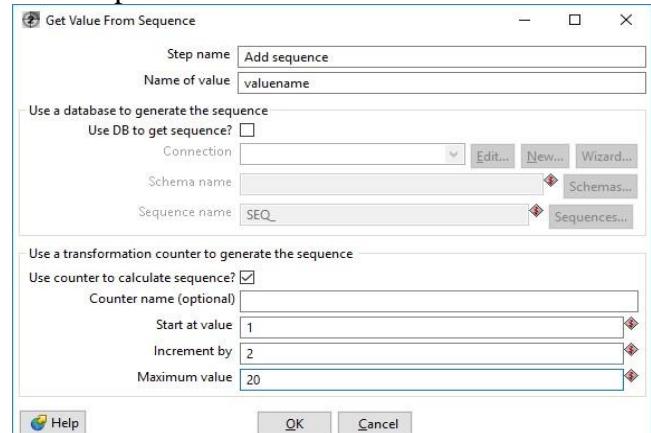
**Step 10: Hold the mouse pointer on the sort rows and drag the output connector to the Add sequence control.**



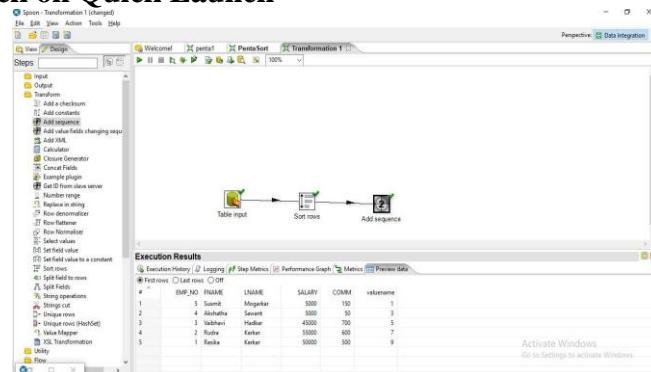
## Step 11: Double Click on Add Sequence Control

The screenshot shows the 'Get Value From Sequence' dialog box. It includes fields for 'Step name' (set to 'Add sequence'), 'Name of value' (set to 'valuename'), and options for generating a sequence or using a counter. The 'OK' button is visible at the bottom.

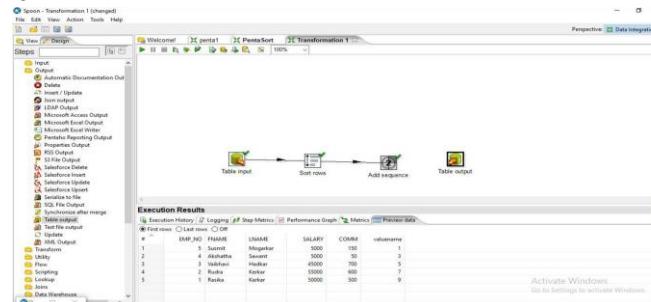
- Enter the values for Start at Values, Increment by and Maximum Values and press ok.



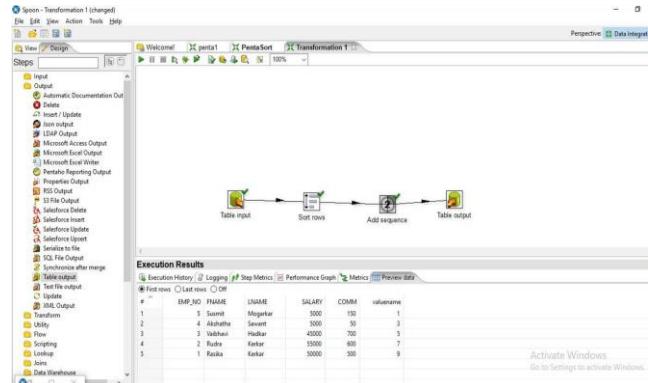
## Step 12: Click on Debug the Transformation and click on Quick Launch



## Step 13: Drag and drop the table output on the panel

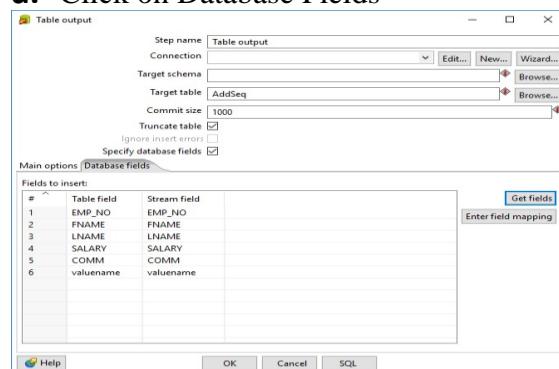


## Step 14: Hold the mouse pointer on the add sequence control and drag the output connector to the table output.

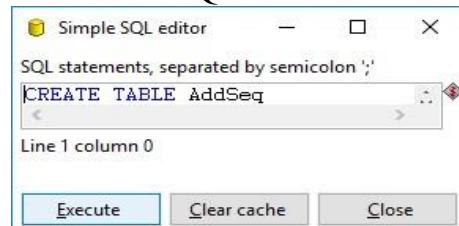


## Step 15: Double Click on the table output.

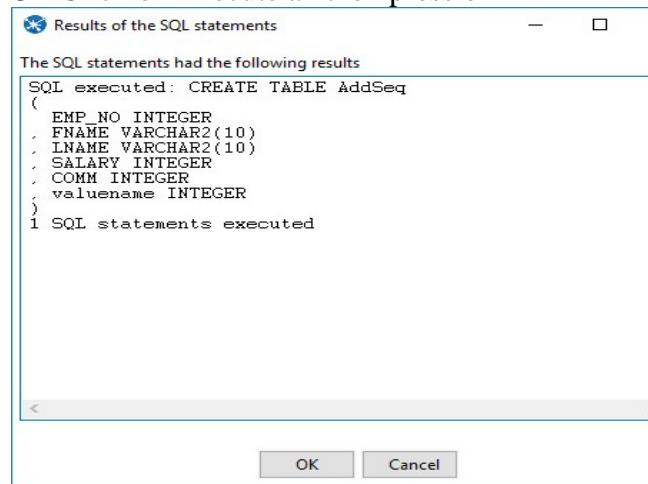
### a. Click on Database Fields



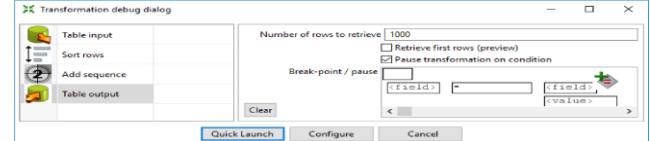
### b. Click on SQL



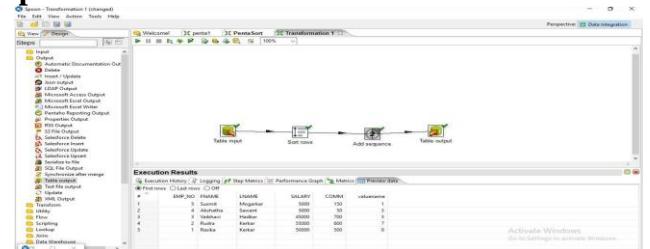
### c. Click on Execute an then press ok



## Step 16: Click on debug Transformation and then quick Launch



## Output



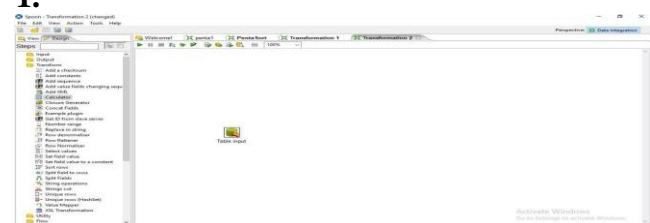
**SQL> select \* from empp13;**

**SQL> select \* from empp13;**

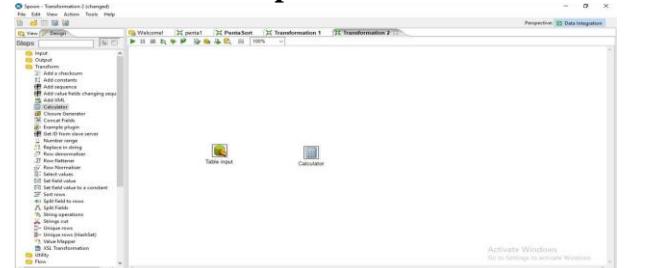
ENO	FNAME	LNAME	SALARY	COMM	VALUENAME
5	smit	bole	160000	12000	1
4	prateek	ahirrao	150000	15000	3
2	aadarsh	choudhari	200000	74000	5
1	simran	gupta	100000	44000	7

## 3. Calculator Operation

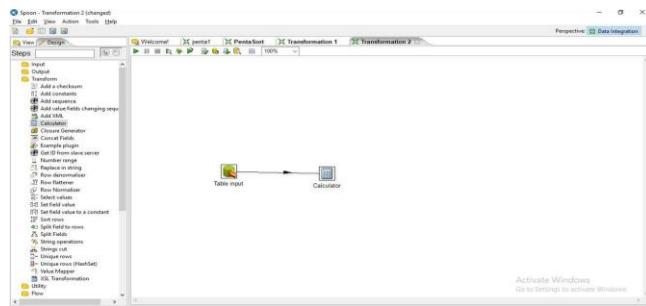
### Step 1: Perform first 5 steps same as practical 1.



### Step 2: Drag and Drop Calculator from transformation on the panel.

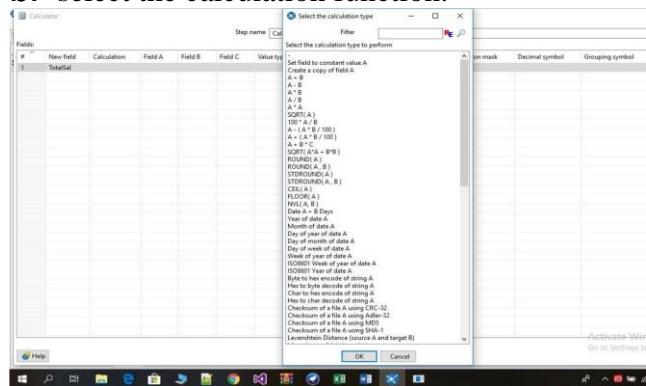


### Step 3: Hold the mouse pointer on the Table input and drag the output connector to the calculator control.

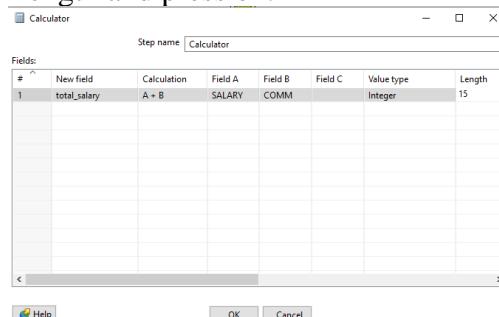


## Step 4: Double Click on the Calculator

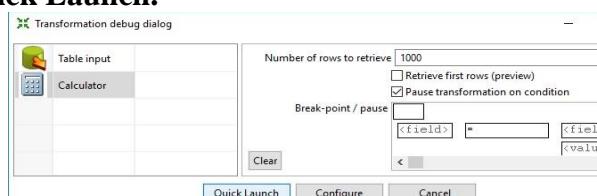
- specify the new field name
- select the calculation function.



- Enter Field A, Field B, Value type and Length and press ok.



## Step 5: Click on Debug transformation and click on Quick Launch.



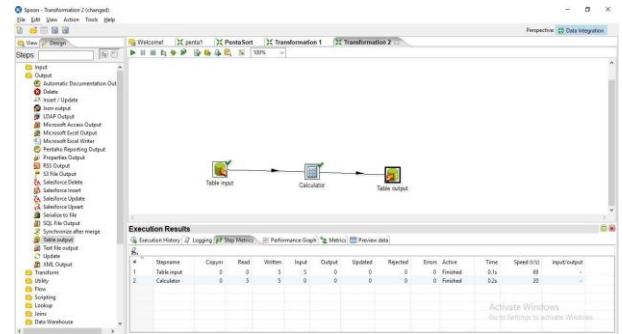
**Preview**

## Examine preview data

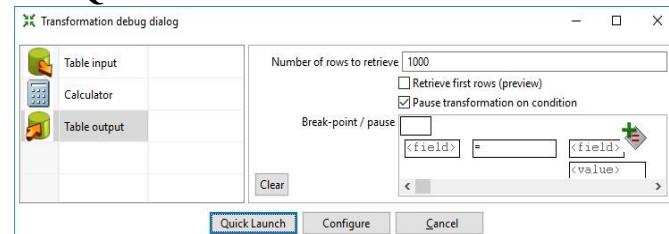
Rows of step: Calculator (4 rows)

#	ENO	FNAME	LNAME	SALARY	COMM	total_salary
1	5	smit	bole	160000	12000	00172000
2	4	prateek	ahirrao	150000	15000	00165000
3	2	aadarsh	choudhari	200000	74000	00274000
4	1	simran	gupta	100000	44000	00144000

**Step 6: Now Perform the steps 6-10 same as practical 1.**



## Step 7: Click on Debug Transformation and then click on Quick Launch



## Output:

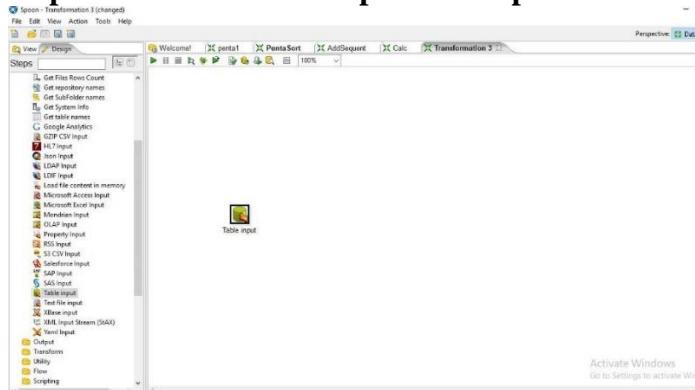


**SQL> select \* from emppp13;**

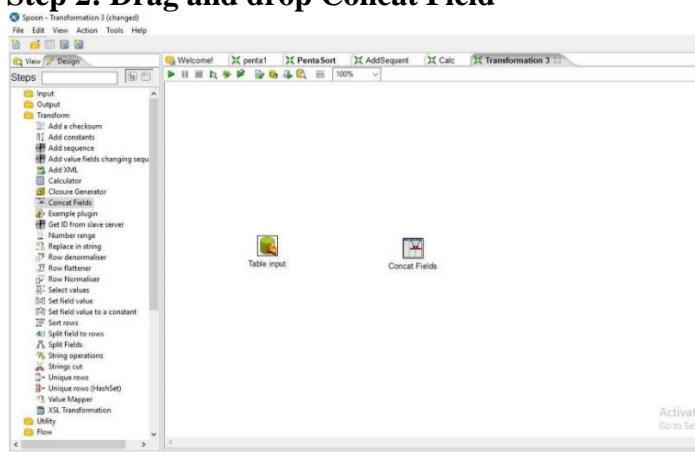
ENO	FNAME	LNAME	SALARY	COMM	TOTAL_SALARY
1	simran	gupta	100000	44000	144000
2	aadarsh	choudhari	200000	74000	274000
4	prateek	ahirrao	150000	15000	165000
5	smit	bole	160000	12000	172000

## d. Concatenation Operation

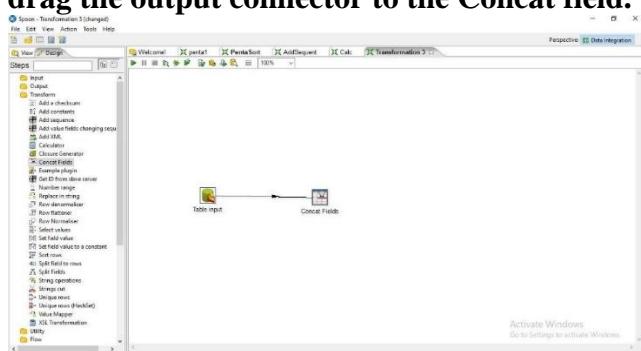
**Step 1:** Perform first 6 steps same as practical 1.



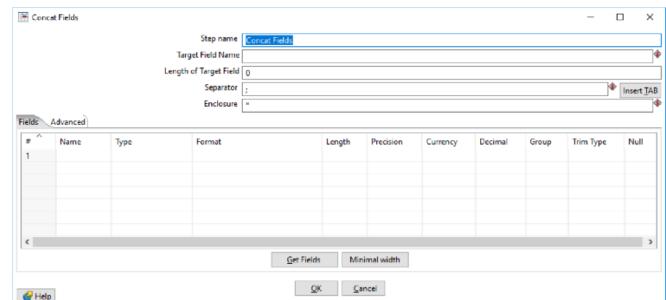
**Step 2: Drag and drop Concat Field**



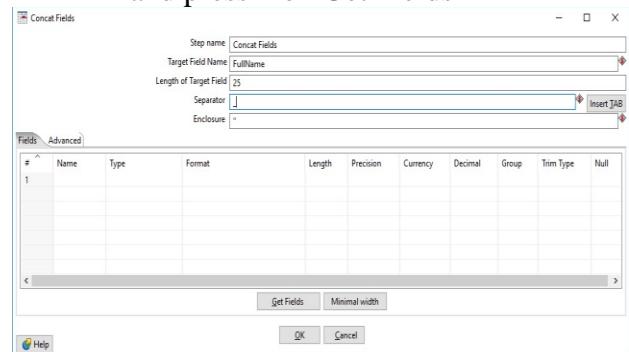
**Step 4: Hold mouse pointer on input table and drag the output connector to the Concat field.**



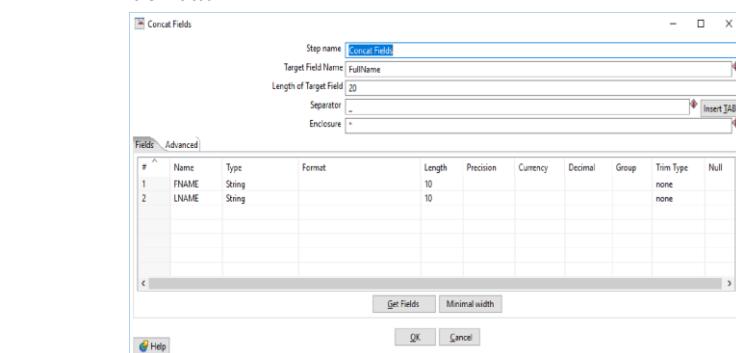
**Step 5: Double Click on Cncat Field**



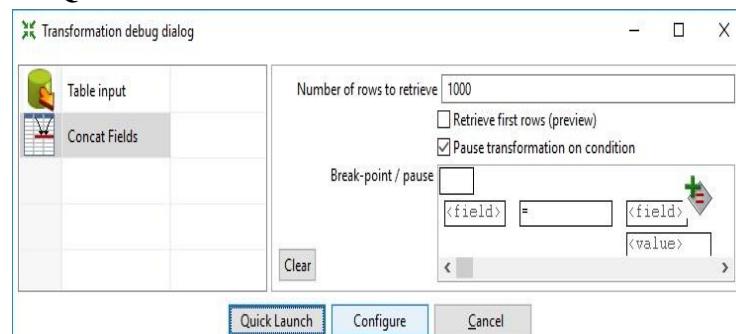
- a.** Enter the values for Target Field Name, Length of the Target Field and the Separator and press on Get Fields



- b.** Select the columns that you want to concat



**Step 6: Click on Debug transformation and click on Quick Launch.**



# MCAL13 Advanced Database Management System Lab

**Examine preview data**

Rows of step: Concat Fields (5 rows)

#	EMP_NO	FNAME	LNAME	SALARY	COMM	FullName
1	5	Susmit	Mogarkar	5000	150	Susmit _Mogarkar
2	4	Akshatha	Sawant	5000	50	Akshatha _Sawant
3	3	Vaibhavi	Hadkar	45000	700	Vaibhavi _Hadkar
4	2	Rudra	Kerkar	55000	600	Rudra _Kerkar
5	1	Rasika	Kerkar	50000	500	Rasika _Kerkar

**Close**

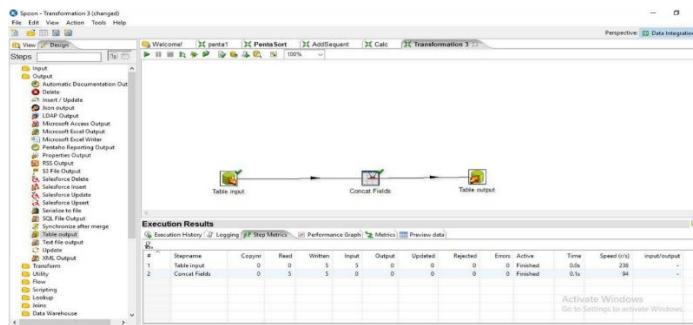
**Examine preview data**

Rows of step: Concat Fields (5 rows)

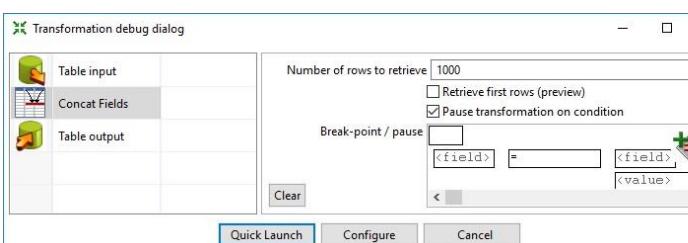
#	EMP_NO	FNAME	LNAME	SALARY	COMM	FullName
1	5	Susmit	Mogarkar	5000	150	Susmit _Mogarkar
2	4	Akshatha	Sawant	5000	50	Akshatha _Sawant
3	3	Vaibhavi	Hadkar	45000	700	Vaibhavi _Hadkar
4	2	Rudra	Kerkar	55000	600	Rudra _Kerkar
5	1	Rasika	Kerkar	50000	500	Rasika _Kerkar

**Close**

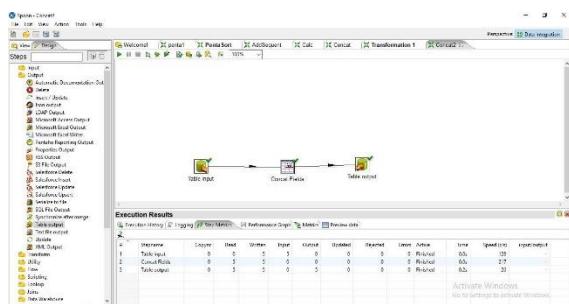
**Step 7:** Now Perform the steps 6-10 same as practical 1.



**Step 8:** Click on debug Transformation and click on Quick Launch



**Output:**

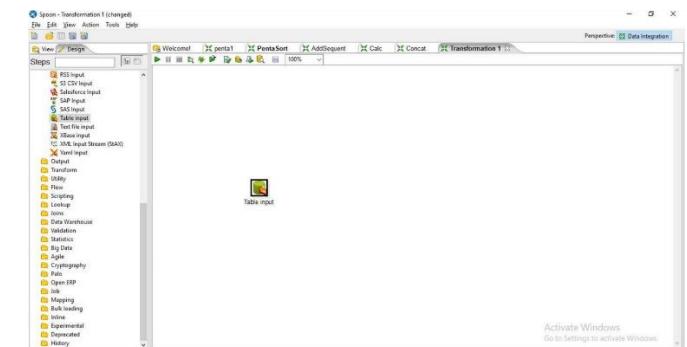


**SQL> select \* from Concat;**

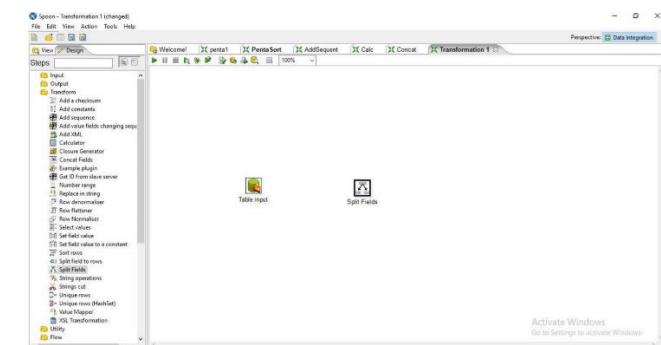
```
SQL> select * from emp_target_concat;
```

EMP_NO	FNAME	LNAME
SALARY	COMM	FULLNAME
25000	500	Sonam _Singh
20000	700	Pradnya _Suryavanshi
17000	1000	Komal _Malviya
30000	900	Nidhi _Rai
15000	1100	Riya _Sharma

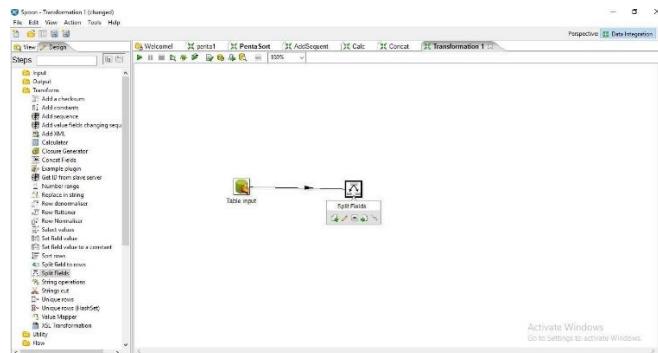
**Step 1:** Perform first 6 steps same as practical 1.



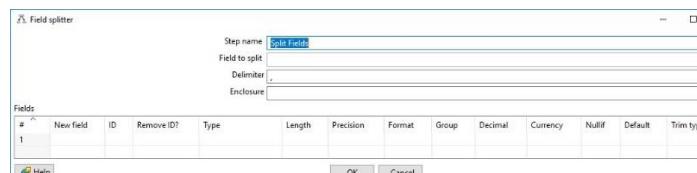
**Step 2:** Drag and drop Split Field on the Panel



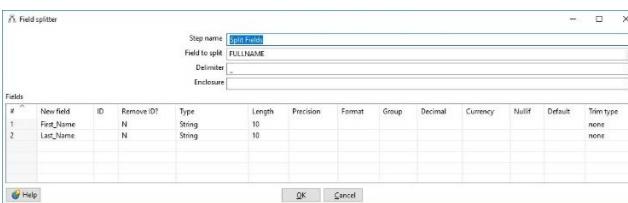
**Step 3:** Hold the mouse pointer over table input and drag it to the Split Field.



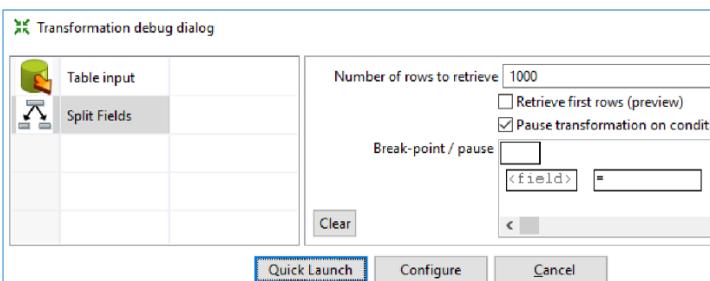
## Step 4: Double Click On Split Field.



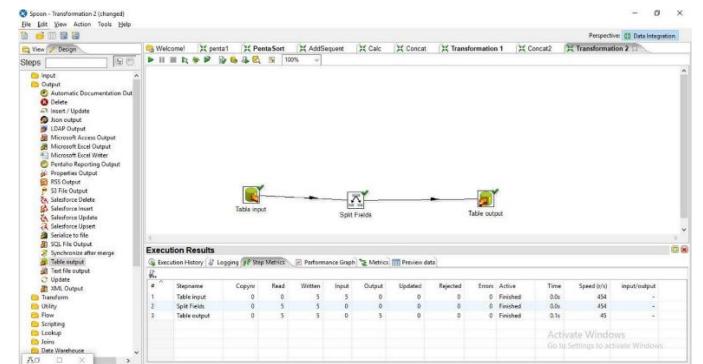
- a. Enter the values For Field to Split, Delimiter and the field names and press ok



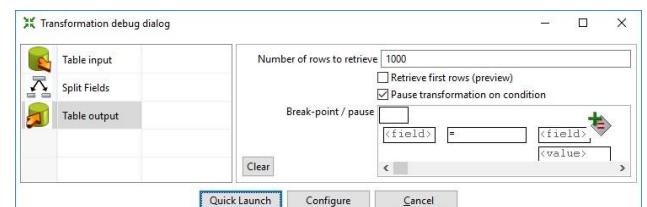
## Step 6: Click on Debug transformation and Click on Quick Launch.



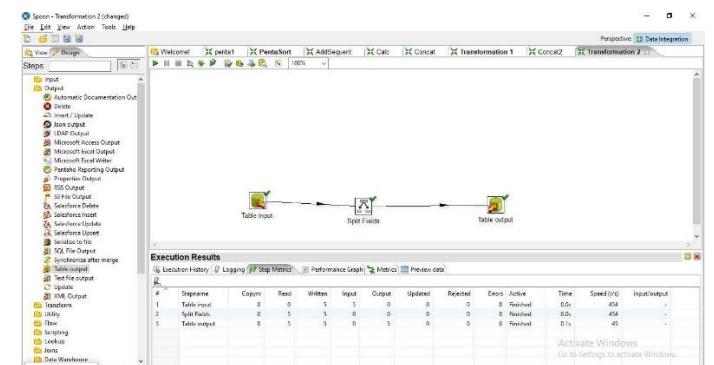
## Step 7: Perform steps 6-10 same as practical 1.



## Step 8: Click on Debug transformation and Click on Quick Launch.



## Output:



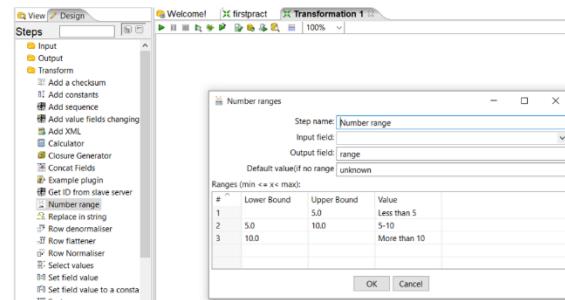
Execution Results													
#	Stepname	Copyto	Read	Written	Input	Output	Updated	Rejected	Error	Action	Time	Speed (MB)	Input/Output
1	Table input		0	0	5	0	0	0	0	Finished	0.0s	454	-
2	Split Fields		0	5	5	0	0	0	0	Finished	0.0s	454	-
3	Table output		0	5	5	0	0	0	0	Finished	0.1s	45	-

```
SQL> select * from emp_target_split;
EMP_NO FNAME LNAME
----- -----
1 Sonam Singh
25000 500 Sonam Singh
2 Pradnya Suryavanshi
20000 700 Pradnya Suryavansh
3 Komal Malviya
17000 1000 Komal Malviya

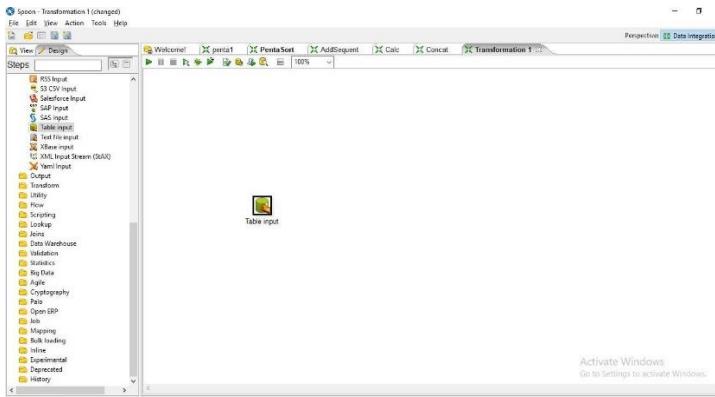
EMP_NO FNAME LNAME
----- -----
4 Nidhi Rai
30000 900 Nidhi Rai
5 Riya Sharma
15000 1100 Riya Sharma
```

```
SQL> select * from student13;

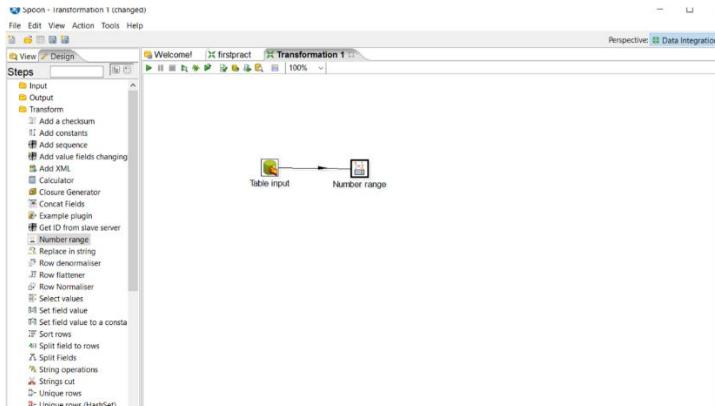
ROLL_NO PERCENTAGE
-----
101      55
102      59
103      68
104      75
105      83
106      35
107      88
```



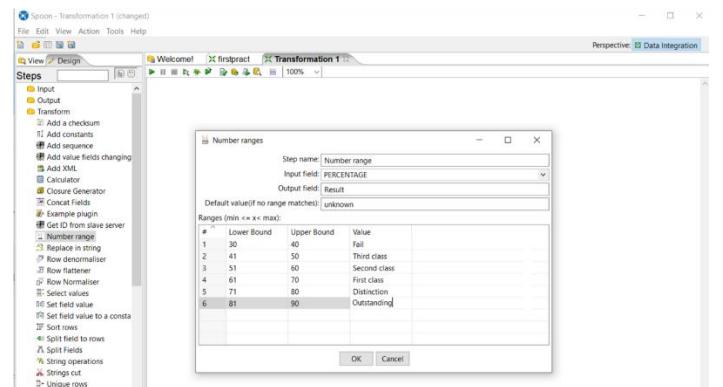
## Step 1: Perform first 6 steps same as practical 1.



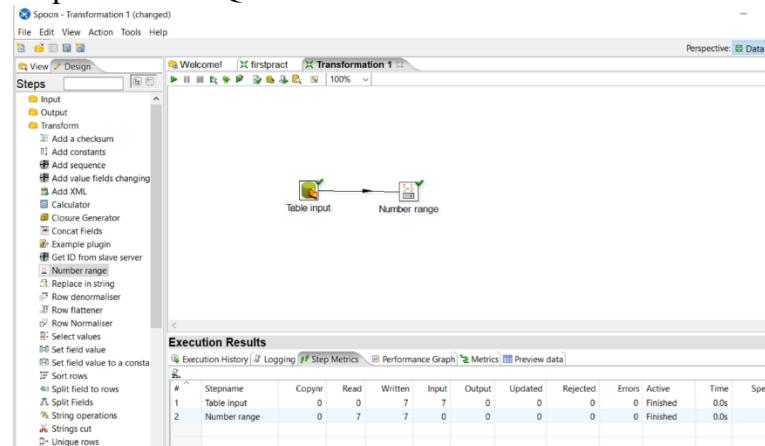
Step2: Drag and drop Number Range transformation and double click on it.



Step3: Set values for Input field, Output field, Lower bound, Upper bound and value according to the values specified in percentage column.



## Step4: Click on Quick launch



## Step 5: Click on Preview data

**Execution Results**

#	RDXL_NO	PERCENTAGE	Result
1	101	55	Second class
2	102	59	Second class
3	103	68	First class
4	104	75	Distinction
5	105	83	Outstanding
6	106	87	Outstanding
7	107	88	Outstanding

## 7. String Operation: Required SQL table:

```
SQL> select * from newemp13;
```

EMPN	ENAME	DEPT
E001	JOHN	D002
MANAGER		50000
E002	SMIT	D001
HR		40000
E003	STEVEKING	D002
		30000
EMPN	ENAME	DEPT
JOB		SALARY
E004	LEX HEAN	D003
MANAGER		50000
E005	BRUCE	D003
MANAGER		50000
E006	KEVINNASH	D002
SALES		50000

**Step 1: Perform first 6 steps same as practical 1.**

**Step 2: Drag and drop String operation transformation, make connection between table input and string operation and double click on it.**

**Step3: Set values for the parameters**

**Step 4: Launch the transformation**

#	EMPNO	ENAME	DEPT_NO	JOB	SALARY
1	E001	JOHN	D002	***Manager	50000
2	E002	SMIT	D001	*****Hr	40000
3	E003	STEVENKING	D002	<null>	30000
4	E004	LEX HEAN	D003	***Manager	14000
5	E005	BRUCE	D003	***Manager	17000
6	E006	KEVINNASH	D002	****Sales	18000

**Step 5: Drag and drop output table**

# MCAL13 Advanced Database Management System Lab

**Execution Results**

#	EMPNO	ENAME	DEPT_NO	JOB	SALARY
1	E001	JOHN	D002	***Manager	50000
2	E002	SMIT	D001	*****Hr	40000
3	E003	STEVENKING	D002	<null>	30000
4	E004	LEX HEAN	D003	***Manager	14000
5	E005	BRUCE	D003	***Manager	17000
6	E006	KEVINNASH	D002	****Sales	18000

Step 6: Launch the transformation

**Execution Results**

#	EMPNO	ENAME	DEPT_NO	JOB	SALARY
1	E001	JOHN	D002	***Manager	50000
2	E002	SMIT	D001	*****Hr	40000
3	E003	STEVENKING	D002	<null>	30000
4	E004	LEX HEAN	D003	***Manager	14000
5	E005	BRUCE	D003	***Manager	17000
6	E006	KEVINNASH	D002	****Sales	18000

Step 7: In SQL output table

```
SQL> select * from stringgg;
EMPN      ENAME          DEPT
-----      -----
JOB          SALARY
-----      -----
E001      JOHN           D002
***Manager          50000
E002      SMIT           D001
*****Hr            40000
E003      STEVENKING     D002
                      30000

EMPN      ENAME          DEPT
-----      -----
JOB          SALARY
-----      -----
E004      LEX HEAN        D003
***Manager          50000
E005      BRUCE           D003
                      50000
E006      KEVINNASH       D002
                      50000
```

6 rows selected.

8. Copy contents of.CSV file to the target table  
Step1:Drag and drop CSV file input on a panel

**Step 1: Input**

- CSV file input
- Data Grid
- De-serialize from file
- Email messages input
- ESRI Shapefile Reader
- Fixed file input
- Get credit card numbers
- Generate random value
- Generate Rows
- Get data from XML
- Get File Name
- Get Files Rows Count
- Get Group names
- Get Sub-Groups names
- Get System Info
- Get table info
- Google Analytics
- GZIP CSV Input
- HLT Input
- Image Input
- LDAP Input
- LDF Input
- Load content in memory
- Microsoft Access Input
- Microsoft Excel Input
- Montrion Input
- OLAP Input

Step2: Double click on it and open respective CSV file, Lazy conversation and Header row present? Options should be checked, click on Getfield and then ok.

**CSV Input**

Step name: CSV file input  
 Filename: D:\data mining\weka practicals\classification\buy\_com.csv  
 Delimiter: ,  
 Enclosure: "  
 NIO buffer size: 50000  
 Lazy conversion:   
 Header row present:   
 Add filename to result:   
 The row number field name (optional):  
 Running in parallel?:   
 New line possible in fields?:   
 File encoding:  
 Sample size: Enter the number of lines to sample: 100  
 Format: Trim by

In this step we can change data type and format of CSV files after doing setting click on ok  
Step3: Drag and drop Table output and connect it with CSV file.

**Steps**

- Input
- Output
- Automatic Documentation Output
- Delete
- Insert / Update
- JSON output
- LDAP Output
- Log file output
- Microsoft Excel Output
- Pentaho Reporting Writer
- Properties Output
- RSS Output
- S3 File Output
- Salesforce Delete
- Salesforce Insert
- Salesforce Update
- Salesforce Export
- Serialize to file
- SQL File Output
- Synchronize after merge
- Table output
- Text file output
- Update
- XML Output
- Unpivot

Step4: Double click on Table output and create new connection and follow same steps like before and launch the transformation.

# MCAL13 Advanced Database Management System Lab

The screenshot shows the Apache Nifi interface. At the top, there is a configuration window for a 'Table output' step. The 'Target table' is set to 'buycomp' and 'Commit size' is 1000. Below this, there are sections for 'Partition data over tables' and 'Main options'. A preview of the transformation flow is shown at the bottom, starting with a 'CSV file input' node connected to a 'Table output' node.

**Execution Results**

#	age	income	student	credit_rating	buys_computer
1	<=30	high	no	fair	no
2	<=30	high	no	excellent	no
3	31..40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	excellent	no
7	31..40	low	yes	excellent	yes
8	<=30	medium	no	fair	no
9	<=30	low	yes	fair	yes
10	>40	medium	yes	fair	yes
11	<=30	medium	yes	excellent	yes
12	31..40	medium	no	excellent	yes
13	31..40	high	yes	fair	yes

## Step 5: Output table created in SQL

```
SQL> select * from csv13;
AGE  INCOME STU CREDIT_RA BUY
-----
<=30  high  no   fair   no
<=31  high  no   excellent no
31&gt;40 high  no   fair   yes
>40   medium no   fair   yes
>40   low   yes  fair   yes
>40   low   yes  excellent no
31&gt;40 low   yes  excellent yes
<=30  medium no   fair   no
<=30  low   yes  fair   yes
>40   medium yes  fair   yes
>40   medium yes  excellent yes
<=30  medium yes  excellent yes

AGE  INCOME STU CREDIT_RA BUY
-----
31&gt;40 medium no   excellent yes
31&gt;40 high   yes  fair   yes
>40   medium no   excellent no
14 rows selected.
```

## 9. Implement the merge join transformation on

**tables Step1:** Drag two Data grid on panel rename them with Employee and Department respectively.

The screenshot shows the Apache Nifi interface. On the left, the 'Steps' panel is open, showing various types of nodes: Input, Data Grid, Output, Transform, Utility, Flow, Scripting, Lookup, Joins, Data Warehouse, Validation, Statistics, Big Data, Agile, Cryptography, Palo, Open ERP, Job, Mapping, Bulk loading, Inline, Experimental, and Parameterized. Two 'Data Grid' nodes are visible on the canvas, one labeled 'Employee' and one labeled 'Department'.

## Step2: Insert records into respective grids.

The screenshot shows three separate 'Add constant rows' dialog boxes for the 'Employee' table. Each dialog has a 'Meta' tab and a 'Data' tab. The 'Data' tab contains the following data:

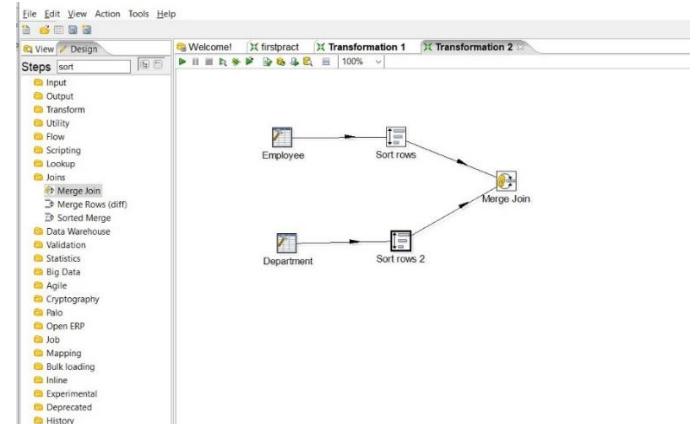
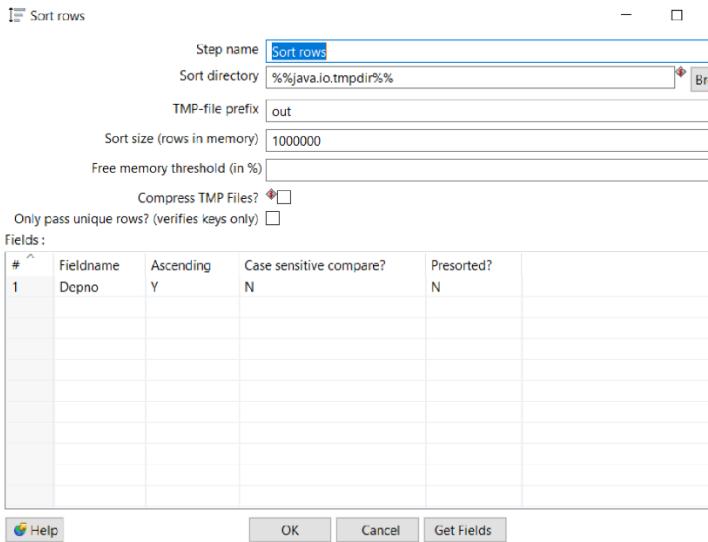
- Employee:**

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Set empty string?
1	Empid	Integer							N
2	Emprname	String							N
3	Age	Integer							N
4	Depno	Integer							N
- Dependent:**

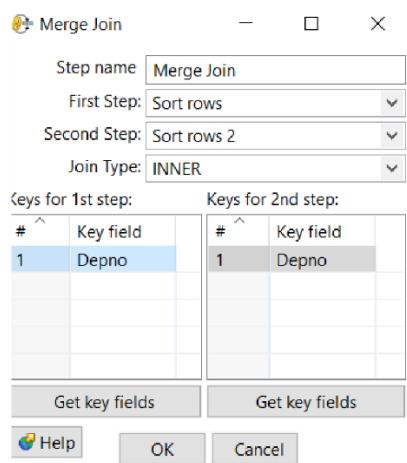
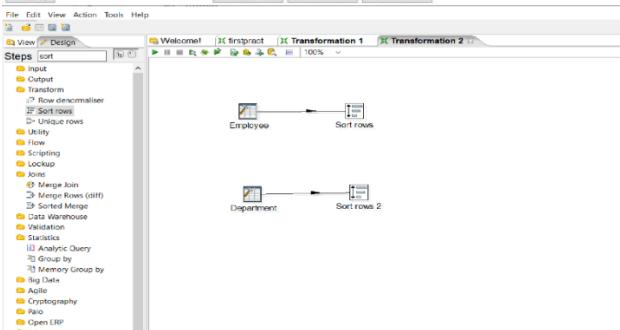
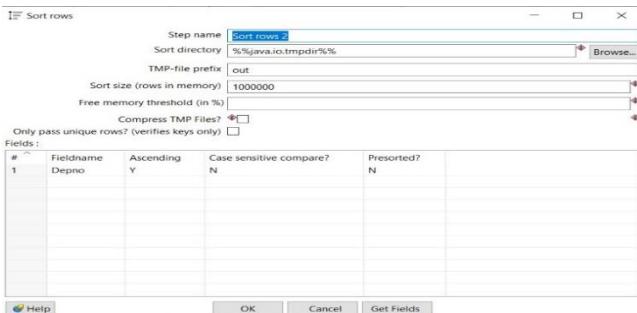
#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Set empty string?
1	Depno	Integer							N
2	Deprname	String							N
- Department:**

#	Depno	Deprname
1	101	sandip
2	102	vishwas
3	103	vaishu
4	104	sheetal

# MCAL13 Advanced Database Management System Lab

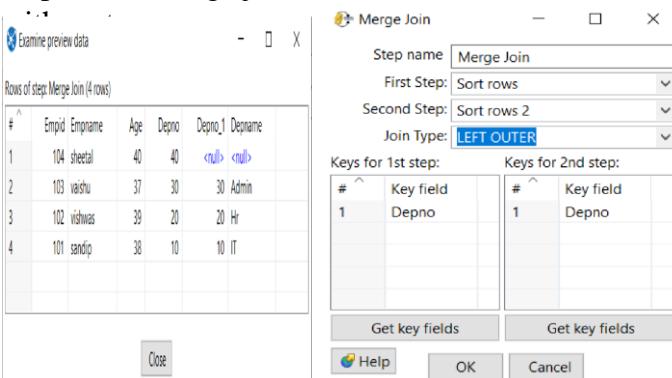


Step5: Double click on Merge join and made necessary settings First select First\_step, then Second\_step, select Join type INNER, and from get fields select Depno only, after that quick launch the transformation.

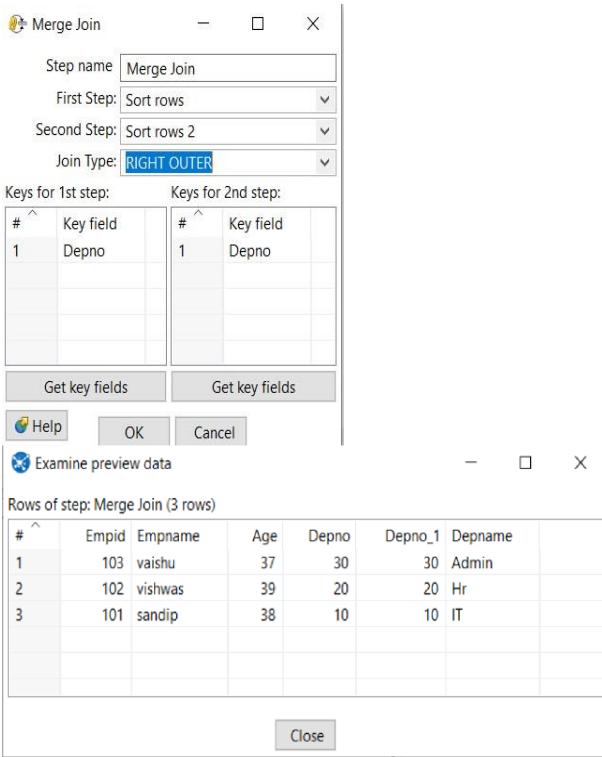


Step 6: Select Join Type LeftOuter,then Right outer and finally Full outer join

Step4: Add Merge join From Joins and connect it

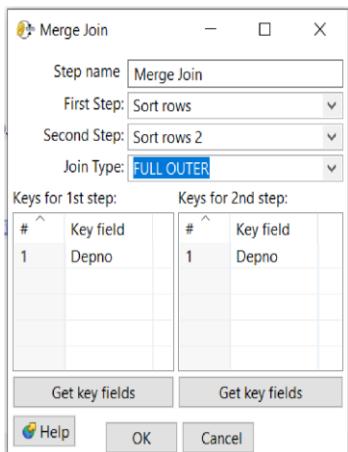


# MCAL13 Advanced Database Management System Lab



Rows of step: Merge Join (4 rows)

#	Empid	Emplname	Age	Depno	Depno_1	Depname
1	104	sheetal	40	40	<null>	<null>
2	103	vaishu	37	30	30	Admin
3	102	vishwas	39	20	20	Hr
4	101	sandip	38	10	10	IT



Rows of step: Merge Join (3 rows)

#	Empid	Emplname	Age	Depno	Depno_1	Depname
1	103	vaishu	37	30	30	Admin
2	102	vishwas	39	20	20	Hr
3	101	sandip	38	10	10	IT

Close

The background shows the Data Integration interface with various transformation steps and a 'Data Grid' component.

Step2: Double click on Data grid and create product table in Meta tab and enter records into it using Data tab.

The screenshot shows the Data Integration interface. A 'Data Grid' component is selected. A context menu is open over it, showing options like 'Add constant rows'. Below the grid, a 'Meta Data' table is displayed with the following columns and data:

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Set empty
1	Prod_ID	String							
2	Prod_name	String							
3	Price	Integer							
4	Check_Status	String							

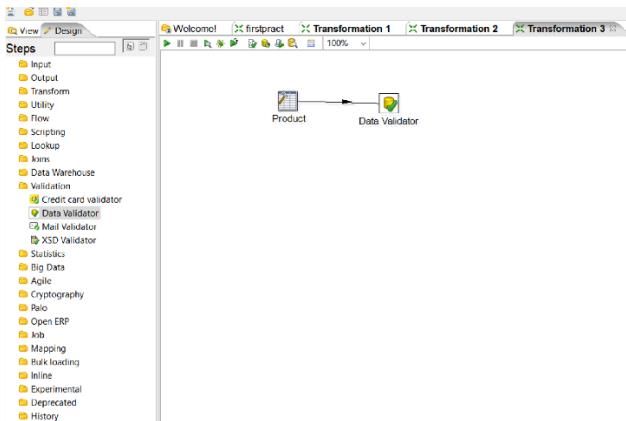
The screenshot shows the Data Integration interface. A 'Data Grid' component is selected. A context menu is open over it, showing options like 'Add constant rows'. Below the grid, a 'Meta Data' table is displayed with the following columns and data:

#	Prod_ID	Prod_name	Price	Check_Status
1	10001	Samsung	35000	Shifted
2	10002	Liono	33000	Cancelled
3	10003	HTC	30000	Shifted

**10. Implement different data validations on table.** Step1: Drag and drop Data grid on panel.

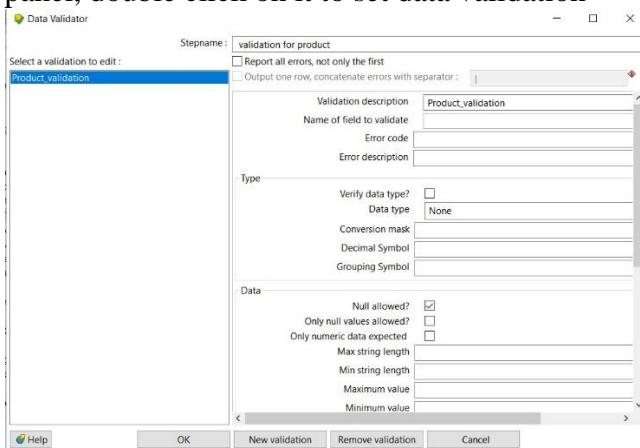
Step3: Go to Validation steps and select, Drag and Drop Data validator on panel and double click on it.

# MCAL13 Advanced Database Management System Lab

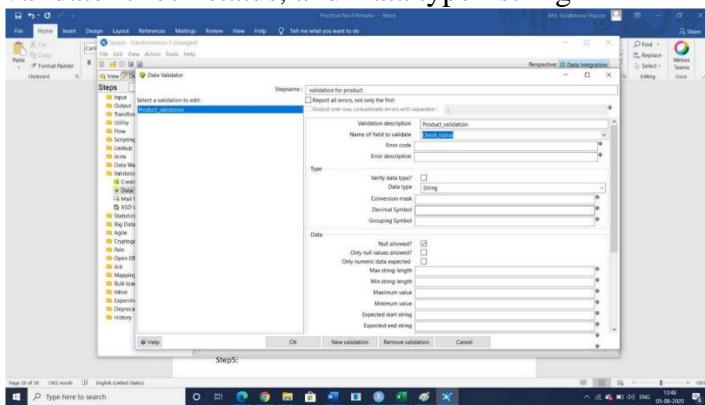


Step4: Click on new validation button and give new name to validation.

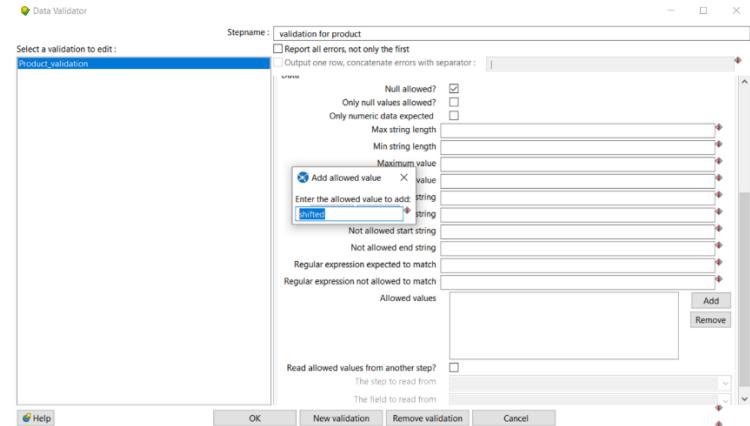
Once new validation is created it will appear on left panel, double click on it to set data validation



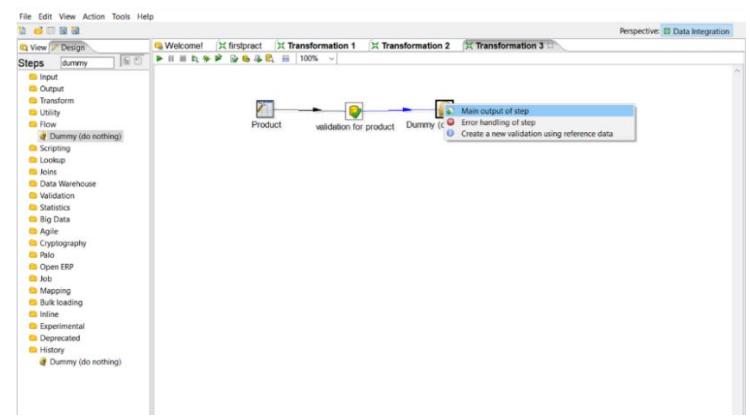
Step5: Set values for Name of field to validate=check status, and Data type =string



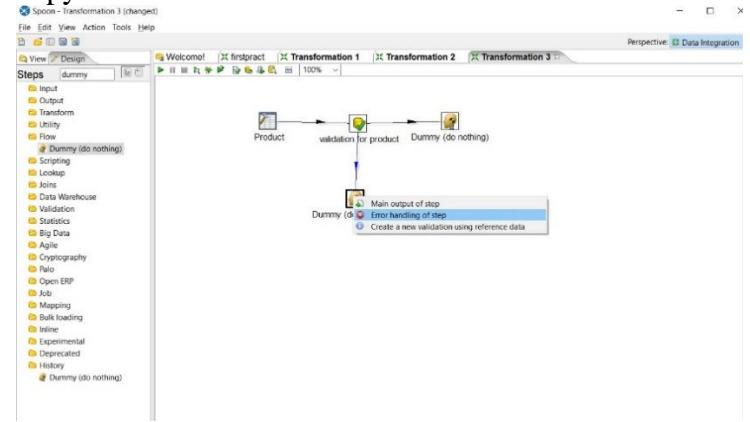
Step6: Click on add to set validation, set it 65 to Shifted and press Enter and click on ok.



Step7: Select Dummy file from Flow to hold output and connect it with Data validation, while connecting select option “ Main Output of step”



Step8: Select another dummy and connect it to ‘Validation for product’ and while selecting select ‘Error handling of step and in next window click on copy



Step 9: launch transformation by selecting one dummy file each.

