CHITKARA UNIVERSITY

PRACTICAL FILE
OF
Data Warehousing
&
ETL Technologies - Lab
(CA130)

SUBMITTED TO :-

Mrs. Deepika Chaudhary

SUBMITTED BY :

Yatin Bajaj

1810992624

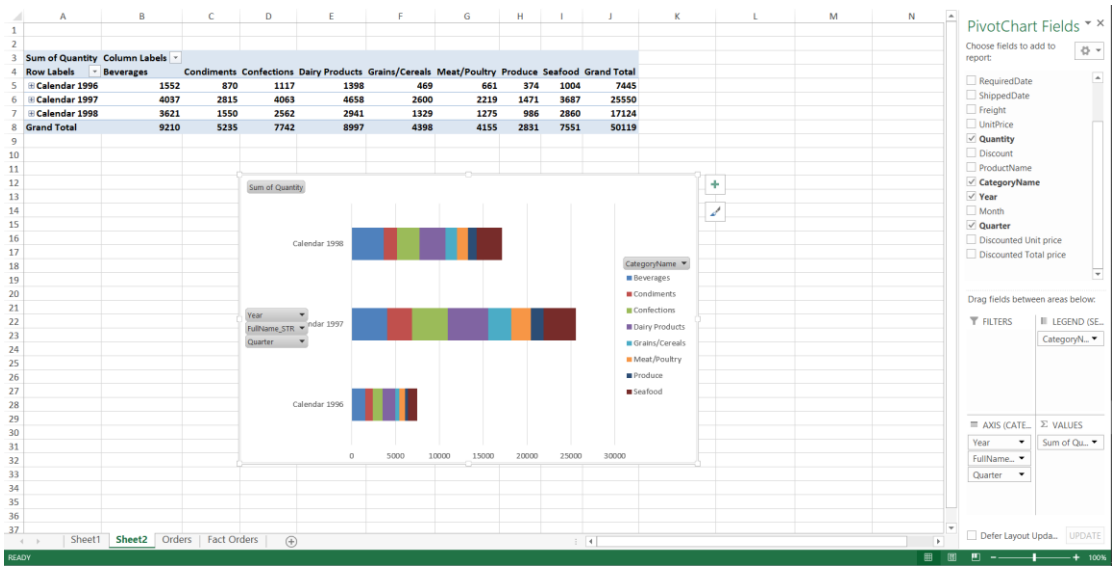
BCA – 5Z-UCA

Experiment 1: How to illustrate data analysis using Pivot Table and Charts and Use of functions (count, Average, Product etc.) and Exercise questions.

Solution)

Steps:

1) Collect data in excel sheet on which we are going to analyze.
2) Select the data and under “Insert” tab click on pivot table and pivot chart.
3) Now click on “OK” and a new existing window will appear consisting of Pivot chart fields.
4) Drag and drop the attributes from there into Axis, values and legends respectively.
5) The graph will automatically configure according to the data. Now the data will be more organized and easy to analyze.



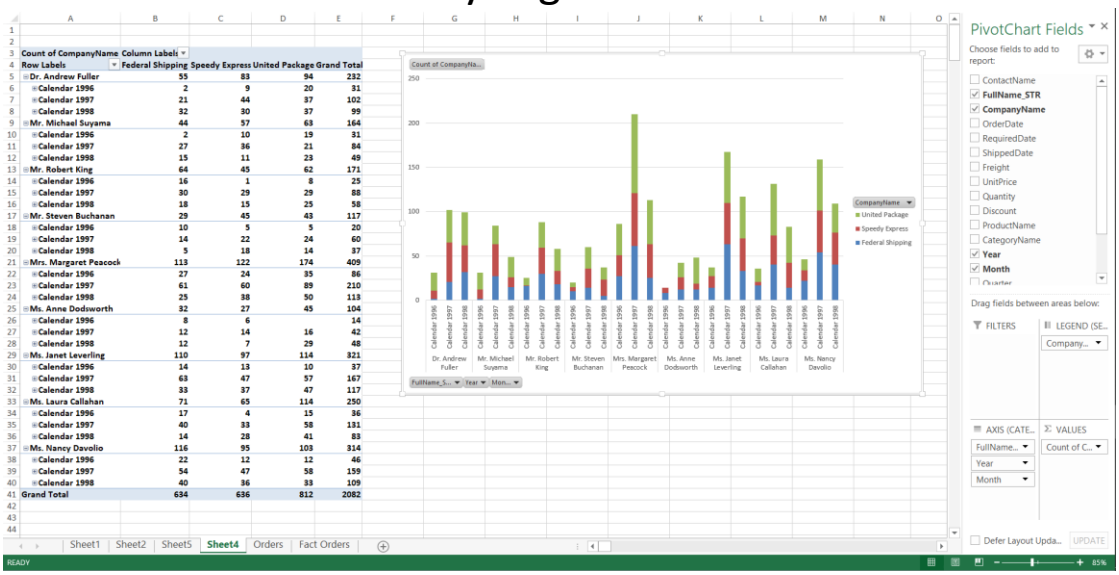
Experiment 2: Use of functions such as refresh, filters , drill down option of pivot tale and Exercise questions on Pivot table and chart implementation.

Solution)

Steps:

1) Collect the data in excel sheet on which we are going to analyze.
2) Select the data and under “Insert” tab click on Pivot table and pivot chart.
3) Now click on “OK” and a new or existing window will appear consisting of Pivot chart fields.
4) Drag and drop the attributes from there into Axis, Values and Legends respectively.
5) **Functions:** To use this function, click on “Value field settings” and select options under “summarize by”. Same goes for product, max, min, etc.
6) **Drill down:** Left click on the selected data value and select “show details”. It will open a drill down dashboard in new sheet.

Pivot chart exercise to find the average speed of a roller coaster from analyzing data.

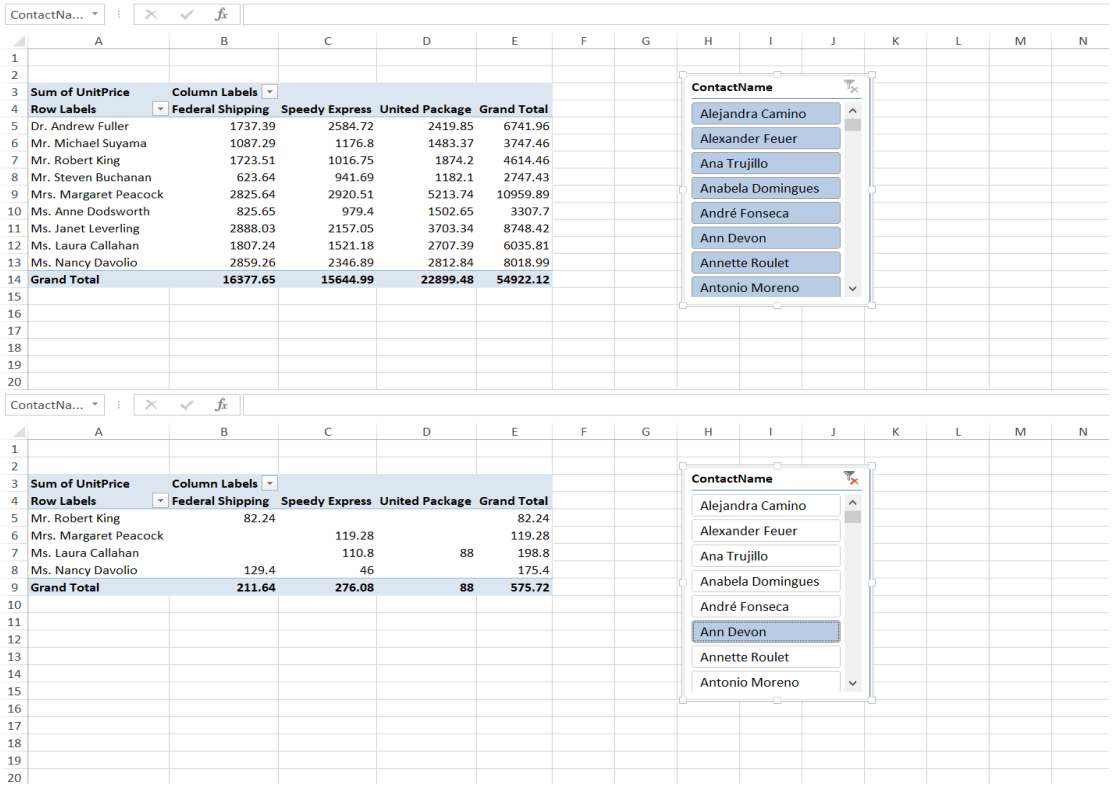


Experiment 3: How to perform slicing and dicing operation in MS Excel.

Solution)

Steps:

1) Collect the data in excel sheet on which we are going to analyze.
2) Select the data and under “Insert” tab click on Pivot table and pivot chart.
3) Now click on “OK” and a new or existing window will appear consisting of Pivot chart fields.
4) After arranging the data in pivot data fields, click on any cell in the pivot table.
5) Under “Insert” Tab click on Insert slicer and select the desired field.
6) Slicer option will give a button like dashboard to slice and dice the data according to user requirement.
7) It makes loads of data easy to analyze by categories.



Sum of UnitPrice	Column Labels			
Row Labels	Federal Shipping	Speedy Express	United Package	Grand Total
Dr. Andrew Fuller	1737.39	2584.72	2419.85	6741.96
Mr. Michael Suyama	1087.29	1176.8	1483.37	3747.46
Mr. Robert King	1723.51	1016.75	1874.2	4614.46
Mr. Steven Buchanan	623.64	941.69	1182.1	2747.43
Mrs. Margaret Peacock	2825.64	2920.51	5213.74	10959.89
Ms. Anne Dodsworth	825.65	979.4	1502.65	3307.7
Ms. Janet Leverling	2888.03	2157.05	3703.34	8748.42
Ms. Laura Callahan	1807.24	1521.18	2707.39	6035.81
Ms. Nancy Davolio	2859.26	2346.89	2812.84	8018.99
Grand Total	16377.65	15644.99	22899.48	54922.12

ContactName: Alejandra Camino, Alexander Feuer, Ana Trujillo, Anabela Domingues, André Fonseca, Ann Devon, Annette Roulet, Antonio Moreno

Sum of UnitPrice	Column Labels			
Row Labels	Federal Shipping	Speedy Express	United Package	Grand Total
Mr. Robert King	82.24			82.24
Mrs. Margaret Peacock		119.28		119.28
Ms. Laura Callahan		110.8	88	198.8
Ms. Nancy Davolio	129.4	46		175.4
Grand Total	211.64	276.08	88	575.72

ContactName: Alejandra Camino, Alexander Feuer, Ana Trujillo, Anabela Domingues, André Fonseca, Ann Devon, Annette Roulet, Antonio Moreno

Experiment 4: How to perform the VLook UpHLook Up operation in MS Excel

Solution) Steps for VLOOKUP:

1)To implement VLOOKUP, get the data on which we want to use VLOOKUP.

2) From the data, write the fields name in the side sheet and write their article number as it's shown in the output.

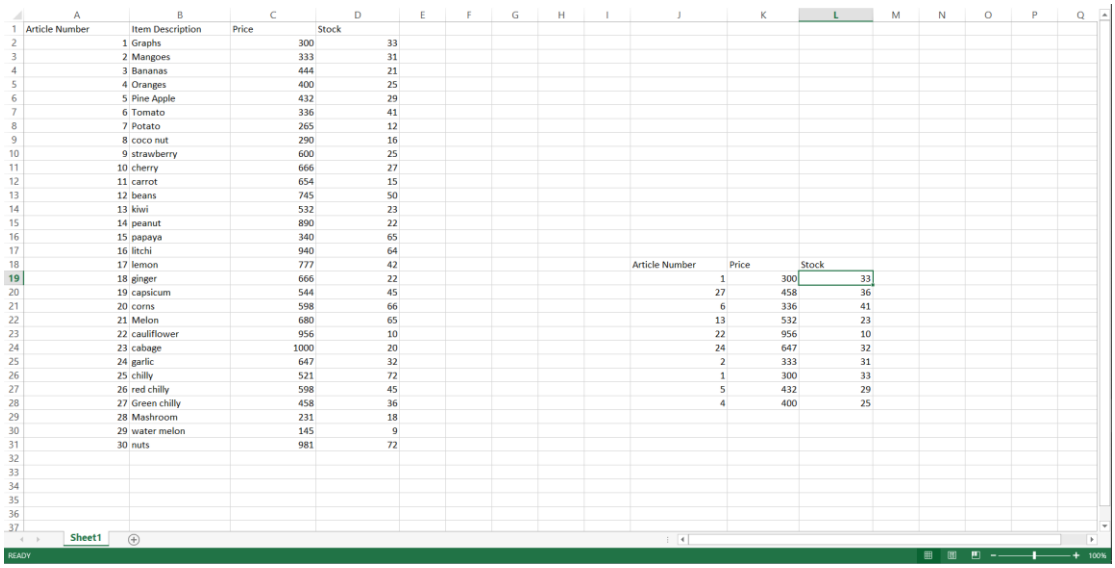
3) Use the VLOOKUP formula in the cell where you want to fetch the data.

VLOOKUP(lookupvalue,table_array,col_index_num,Range Lookup)

4) Specify the data which you want VLOOKUP to use for search in table_array.

5) Specify column number and range value. False for exact match and True for Approximate match.

6) Press **Enter** to find the result. To fetch more than one result just copy the above result and paste in other cells.

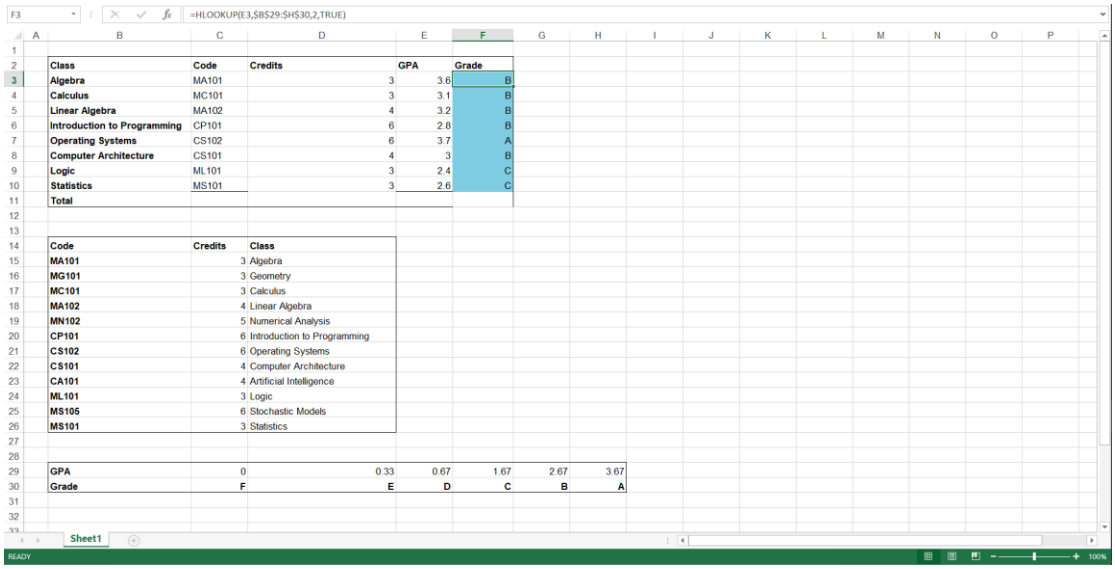


Article Number	Item Description	Price	Stock
1	Graphs	300	33
2	Mangoes	333	31
3	Bananas	444	21
4	Oranges	400	25
5	Pine Apple	432	29
6	Tomato	336	41
7	Potato	265	12
8	coco nut	290	16
9	strawberry	600	25
10	cherry	666	27
11	carrot	654	15
12	beans	745	50
13	kiwi	532	23
14	peanut	890	22
15	papaya	340	65
16	litchi	940	64
17	lemon	777	42
18	ginger	666	22
19	capsicum	544	45
20	corns	598	66
21	Melon	680	65
22	cauliflower	956	10
23	cabage	1000	20
24	garlic	647	32
25	chilly	521	72
26	red chilly	598	45
27	Green chilly	458	36
28	Mashroom	231	18
29	water melon	145	9
30	nuts	981	72

Article Number	Price	Stock
1	300	33
27	458	36
6	336	41
13	532	23
22	956	10
24	647	32
2	333	31
1	300	33
5	432	29
4	400	25

Steps for HLOOKUP:

The steps are same as of VLOOKUP. HLOOKUP works on horizontal cells.



Class	Code	Credits	GPA	Grade
Algebra	MA101	3	3.6	B
Calculus	MC101	3	3.1	B
Linear Algebra	MA102	4	3.2	B
Introduction to Programming	CP101	6	2.8	B
Operating Systems	CS102	6	3.7	A
Computer Architecture	CS101	4	3	B
Logic	ML101	3	2.4	C
Statistics	MS101	3	2.6	C
Total				

Code	Credits	Class
MA101	3	Algebra
MG101	3	Geometry
MC101	3	Calculus
MA102	4	Linear Algebra
MN102	5	Numerical Analysis
CP101	6	Introduction to Programming
CS102	6	Operating Systems
CS101	4	Computer Architecture
CA101	4	Artificial Intelligence
ML101	3	Logic
MS105	6	Stochastic Models
MS101	3	Statistics

GPA	0	0.33	0.67	1.67	2.67	3.67
Grade	F	E	D	C	B	A

Experiment 5: How to Install and Run Pentaho Data Integration IDE and Weka Installation

Solution) Pentaho is business intelligence(BI) software that provides data integration, OLAP services, reporting, information dashboard, data mining and ETL capabilities.

Steps to install Pentaho by Hitachi Vantara:

1) To install and run Pentaho Data Integration, we must have JDK installed in our system.

2) Now download the Pentaho setup, run the installer.

3) Make sure the PATH is set to JAVA in environment variables.

4) Now under data-integration folder, click on “set-pentaho-env.bat” and run the batch file.

5) After setting the environment, run the batch file “spoon.bat”, it will open Pentaho Data integration software.

Experiment 6: Create a Simple PDI Transformation for transforming data from:

a. From Text File to Excel file.
b. From Excel file to Text File
c. From Text to XML File

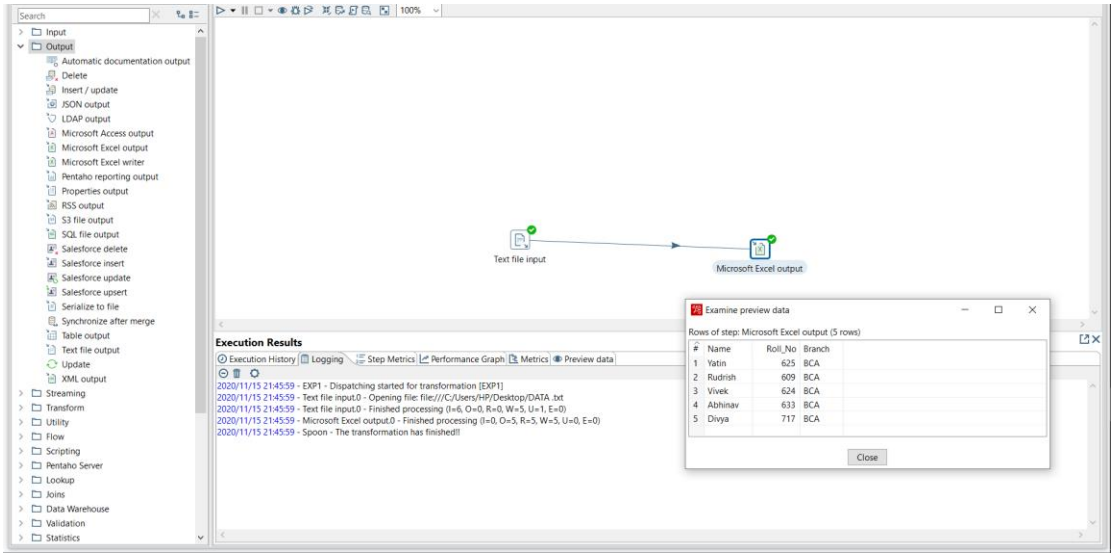
a) From Text file to Excel File

Solution-a)

Steps for text to excel:

1. Click on Input tab and select Text input, drag and drop it to the area.
2. Click on Output tab and select Excel output, drag and drop it to the area.
3. Select the input text, double click on it and add the source file, get the fields.
4. Draw the Hop from Input to the output file, in the excel output save the location and get the fields.
5. Now click on Run and see the preview.

Output:



b) From Excel to text file

solution)

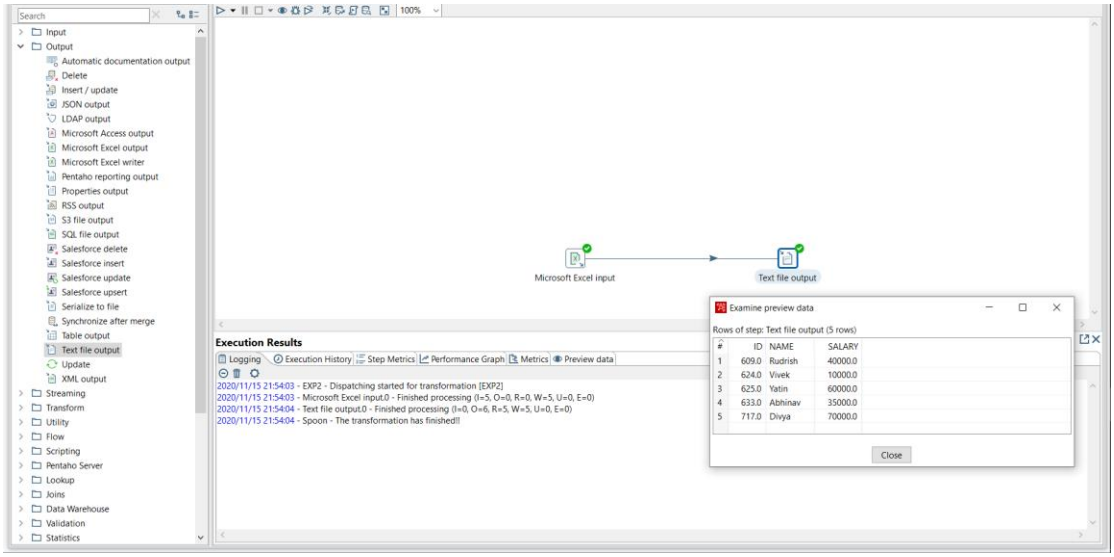
1.Click on Input tab and select Excel input, drag and drop it to the area.

2.Click on Output tab and select text output, drag and drop it to the area.

3. Select the input excel, double click on it and add the source file, get the fields

4. Draw the Hop from Input to the output file, in the text output save the location and get the fields.

5. Now click on Run and see the preview.



Experiment 7:

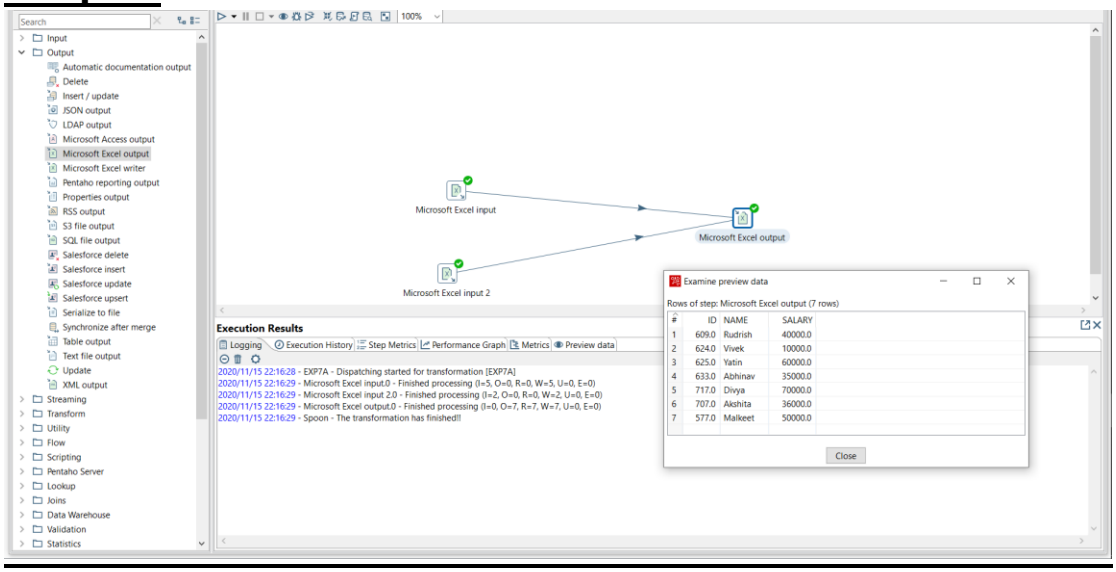
a. Create a Simple PDI Transformation for transforming data from Multiple Excel File to One Excel File.
b. Create a Simple PDI Transformation for transforming data from multiple text File to One Text File.
c. Create a Simple PDI Transformation for transforming data from Multiple text File to One Excel File

Solution

a) Steps:

1) Drag and Drop Microsoft Input files and place in the area.
2) Add the data source files in them and get sheets, fields.
3) Drag and drop the Excel output file and place in the area.
4) Add Hop from input to output files and get sheet field , preview rows and save location.
5) Click on run and preview the output.

Output:

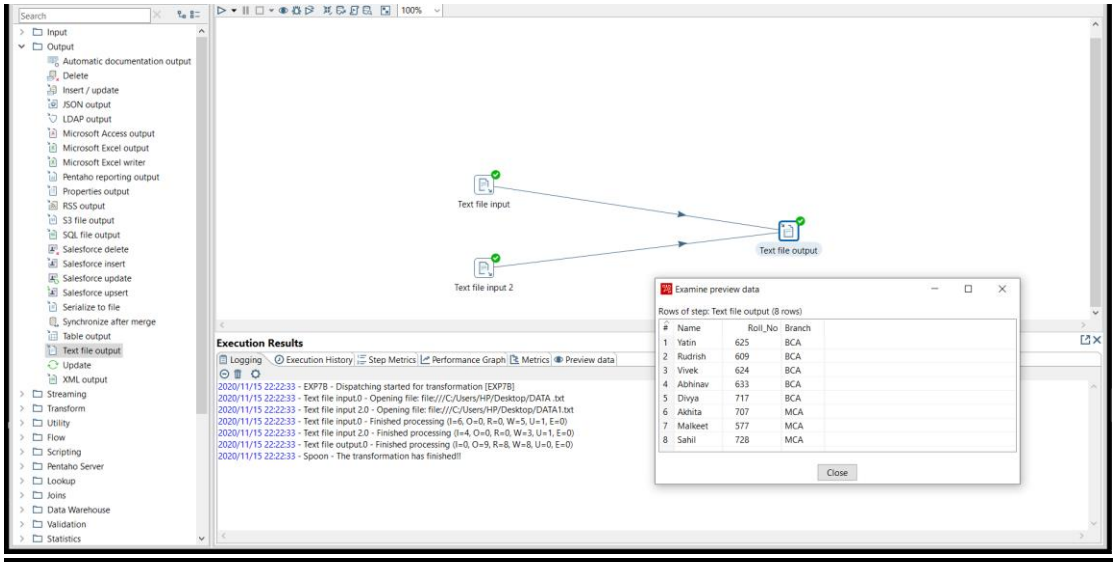


b) Create a Simple PDI Transformation for transforming data from Multiple text file to One text File.

Solution b) Steps:

1) Drag and Drop Text Input files and place in the area.
2) Add the data source files in them and get sheets, fields.
3) Drag and drop the Text output file and place in the area.
4) Add Hop from input to output files and get sheet field , preview rows and save location.
5) Click on run and preview the output.

Output:

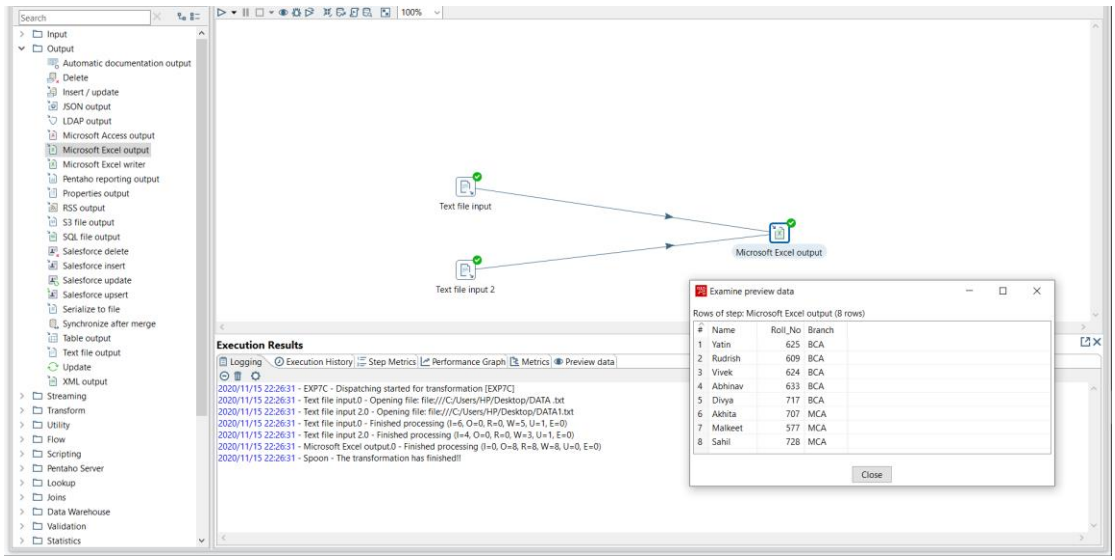


c) Create a Simple PDI Transformation for transforming data from Multiple text file to One Excel File.

Solution)

1) Drag and Drop Text Input files and place in the area.
2) Add the data source files in them and get sheets, fields.
3) Drag and drop the Text output file and place in the area.
4) Add Hop from input to output files and get sheet field , preview rows and save location.
5) Click on run and preview the output.

Output:



Experiment 8:

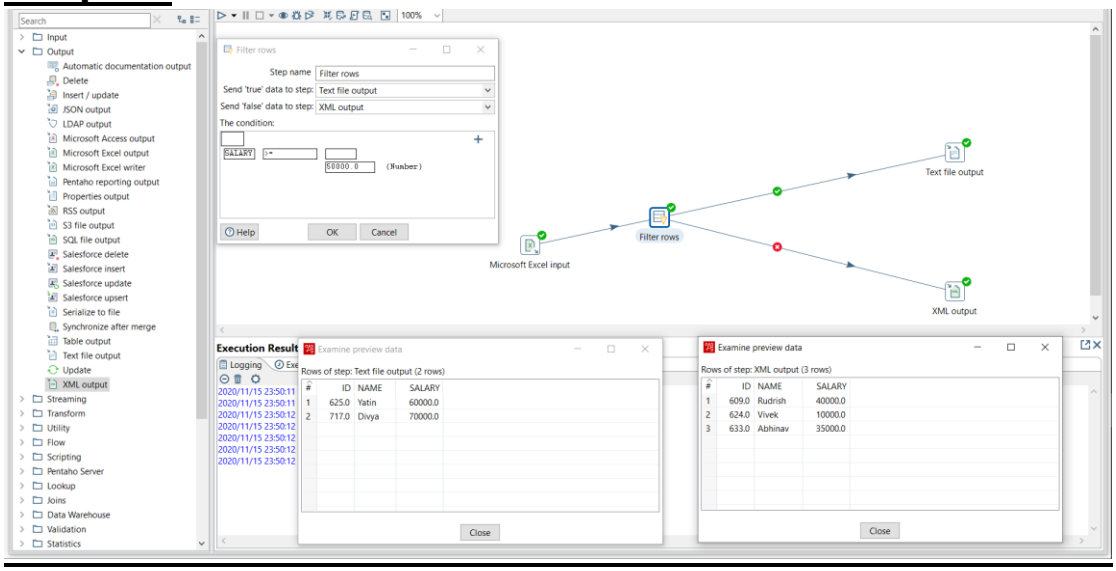
a. Use of Filter rows for dividing Source data to two destinations.
b. Distribution of data from one excel file to text file and excel file
c. Distribution of data from one Text file to Multiple excel files.

a) Use of Filter rows for dividing Source data to two destinations.

Sol a) Steps:

1) Drag and Drop Text Input files and place in the area.
2) Add the data source files in them and get sheets,fields.
3) Drag and drop the Text output file and place in the area.
4) Add filter rows and divert the true and false result to the output files
5) Add Hop from input to output files.
6) Click on run and preview the output.

Output:

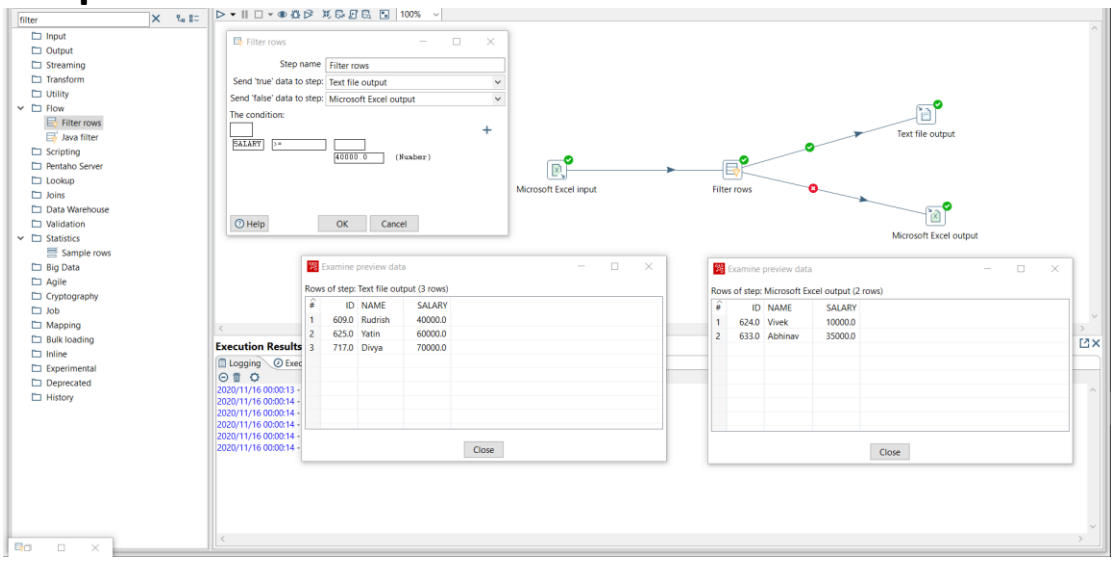


b) Distribution of data from one excel file to text file and excel file.

Solution) Steps:

1) Drag and drop the Excel input file and Excel output and Text Output.
2) Add the Excel file in input , get fields and draw a hop to the files.
3) After adding hope to the output files, a message popup will appear, click on “Distribute” to each of the output files.
4) Now, Click on “Get fields” in both the outputs. Click on save.
5) Run the transformation and see the output.

Output:



c) Distribute data from one Text file to multiple Excel files.

Solution) Steps:

1) Drag and drop the Text input file and Excel output files.

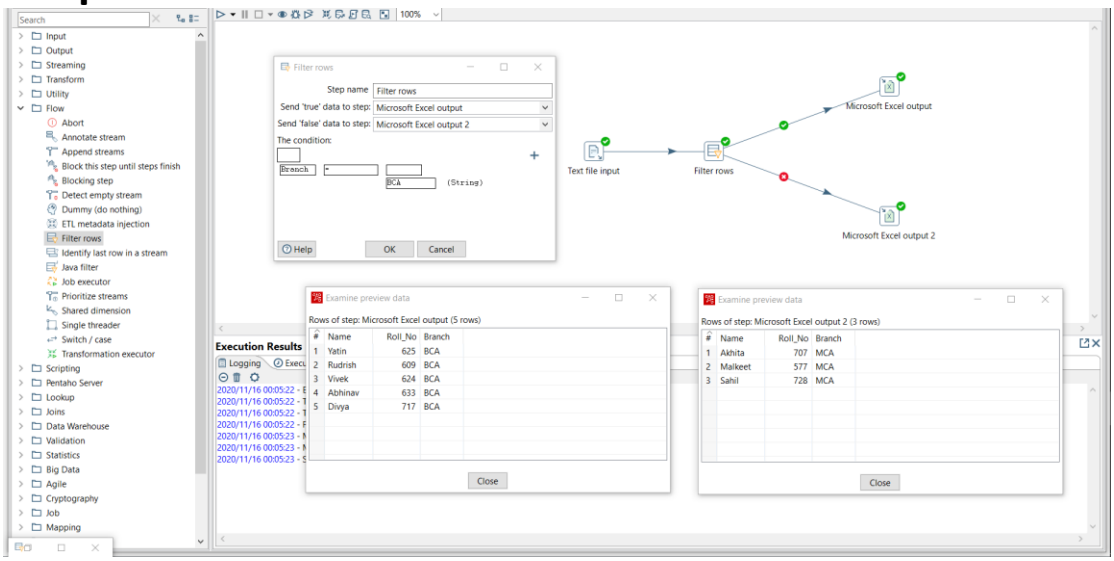
2) Add the Text file in input , get fields and draw a hop to the files.

3) After adding hope to the output files, a message popup will appear, click on “Distribute” to each of the output files.

4) Now, Click on “Get fields” in both the outputs. Click on save.

5) Run the transformation and see the output.

Output:



Experiment 9: Create Database connection in PantaHo with MySQL and Create a Simple PDI Transformation for transforming data from:

a. From Database table to Text File.

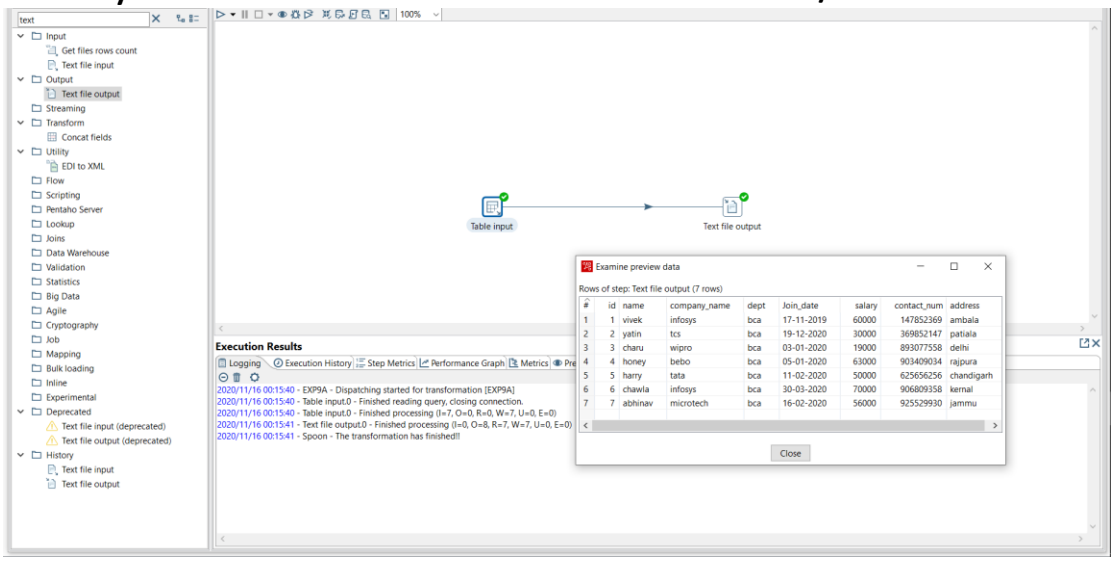
b. From Database table to Excel file

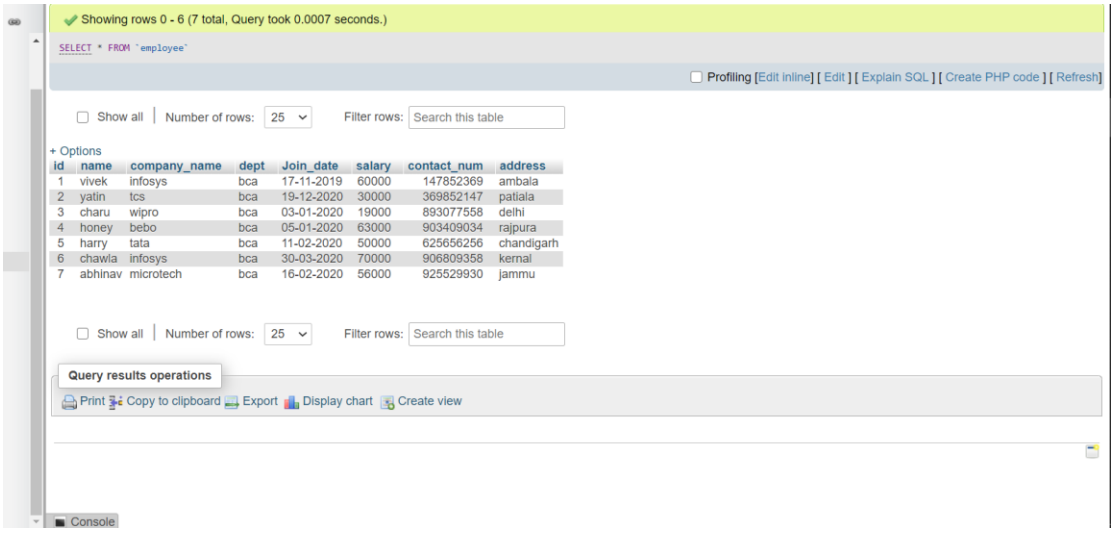
From Access to Excel file.

Solution)

a.

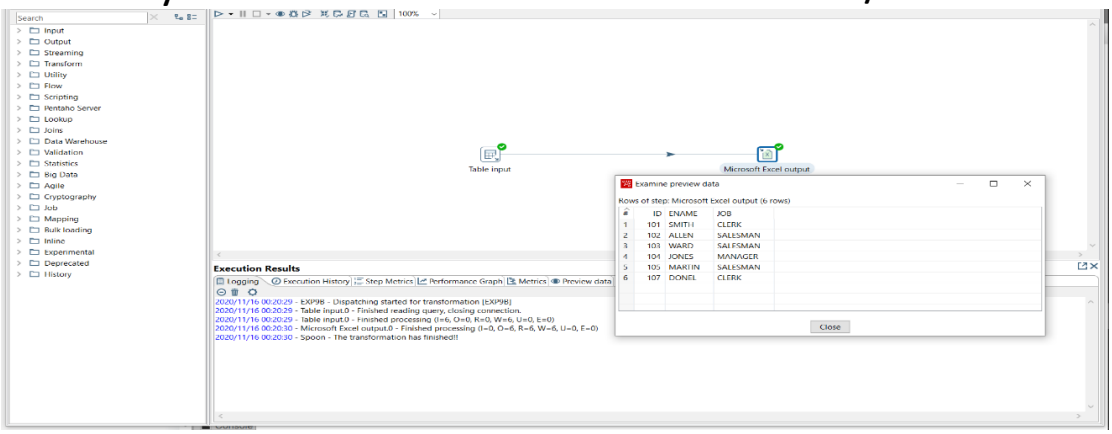
1) Drag and drop the Database table input and Output text file.
2) Double click on table input then, click on connection name (new).
3) Select database type , enter appropriate database name and provide additional info.
4) Test weather the connection is established or not.
5) Now , click on setup wizard and select the table that needs to be converted into Text file.
6) Draw a hop from Database table to Text File, get desired fields by double clicking on text file output and get fields , save location.
7) Run the transformation and preview the data, if any error occurs check the connection /drivers.





b.

1) Drag and Drop the Database table input and Output excel file.
2) Double click on table input then, click on connection name (new).
3) Select database type , enter appropriate database name and provide additional info.
4) Test weather the connection is established or not.
5) Now , click on setup wizard and select the table that needs to be converted into excel file.
6) Draw a hop from Database table to excel File, get desired fields by double clicking on excel file output and get fields, save location.
7) Run the transformation and preview the data, if any error occurs check the connection /drivers.



Experiment 10: Create a Simple PDI Transformation for transforming data from:

a. From Excel file to Database table.

b. From Text file to Database table.

Solution)

a.

1) Drag and Drop the excel file input and Output database table.

2) Double click on excel input, and add the data source file in them and get sheets fields and then preview rows .

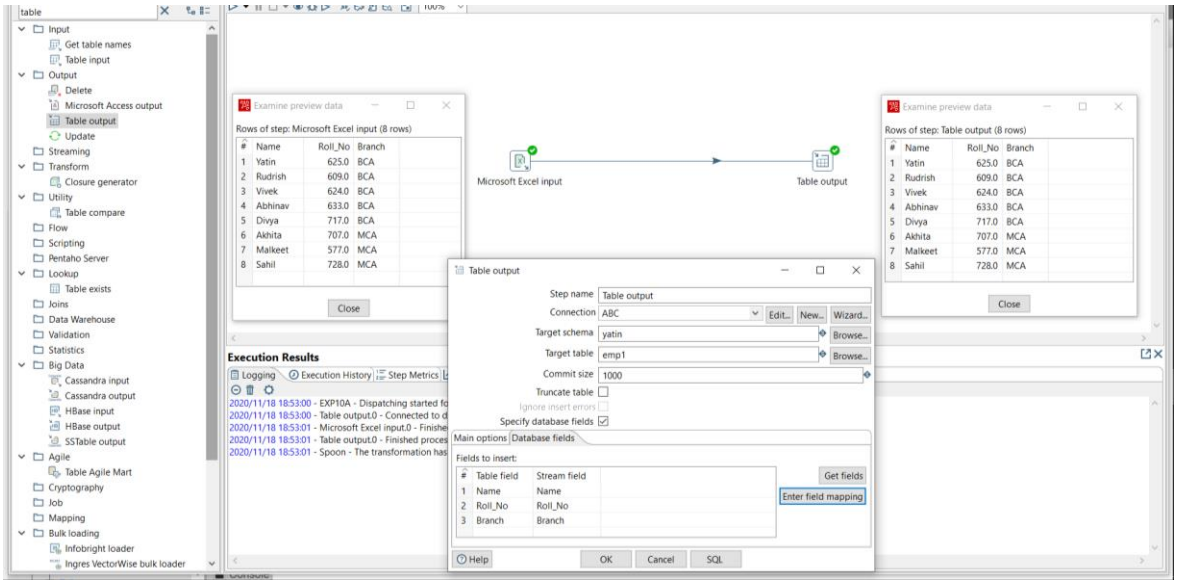
3) Draw a hop from excel input to table output .

4) Select database type , enter appropriate database name and provide additional info.

5) Test weather the connection is established or not.

6) Now , select the table and get fields and perform mapping between excel and table fields.

7) Run the transformation and preview the data, if any error occurs check the connection /drivers.



b.

1) Drag and Drop the text file input and Output database table.
2) Double click on text input, and add the data source file in them and get fields and then preview rows .
3) Draw a hop from text input to table output .
4) Select database type , enter appropriate database name and provide additional info.
5) Test weather the connection is established or not.
6) Now , select the table and get fields and perform mapping between text and table fields.
7) Run the transformation and preview the data, if any error occurs check the connection /drivers.