ADBMS PRACTICAL NO: 4

AIM: Implementation of ETL transformation with Pentaho.

THEORY:

Pentaho Reporting is a suite (collection of tools) for creating relational and analytical reporting. Using Pentaho, we can transform complex data into meaningful reports and draw information out of them. Pentaho supports creating reports in various formats such as HTML, Excel, PDF, Text, CSV, and xml. Pentaho can accept data from different data sources including SQL databases, OLAP data sources, and even the Pentaho Data Integration ETL tool.

ETL Process Transformation:

As the name suggests ETL stands for Extract Transform and Load. Just like the name applies, ETL tool Extracts data from the source. Transform the data while in transit and then it loads the data in to Specified database.

Features of Pentaho

Pentaho Reporting primarily includes a Reporting Engine, a Report Designer, a Business Intelligence (BI) Server. It comes loaded with the following features −

Report Designer − Used for creating pixel perfect report.

Metadata Editor − Allows you to add user-friendly metadata domain to a data source. **Report Designer and Design Studio** − Used for fine-tuning of reports and ad-hoc reporting. **Pentaho user console web interface** − Used for easily managing reports and analyzing views. **Ad-Hoc reporting interface** − Offers a step-by-step wizard for designing simple reports. Output formats include PDF, RTF, HTML, and XLS.

A complex scheduling sub-system − Allows users to execute reports at given intervals. **Mailing** − Users can email a published report to other users.

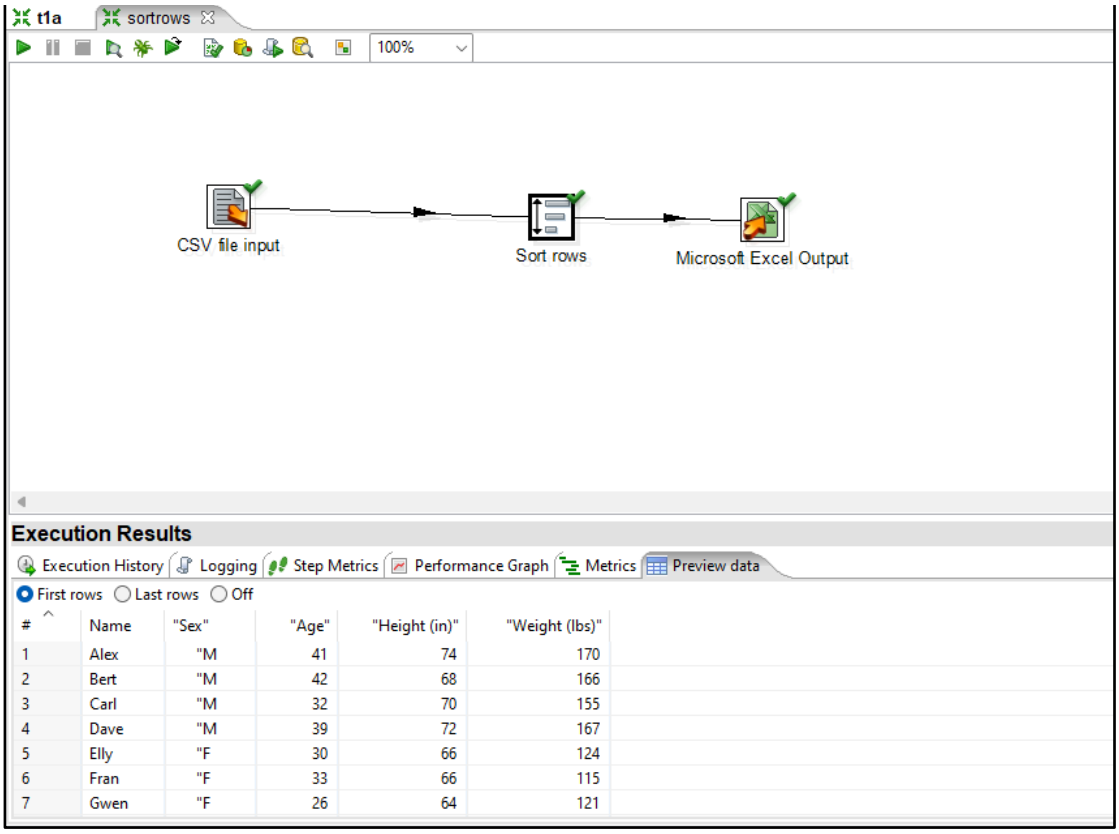
Connectivity − Connectivity between the reporting tools and the BI server, which allows to publish the content directly to the BI server.

1: Convert Csv to Excel and Apply transformations.

Step 1: Create new transformation, in input drag & drop the csv file.
Step 2: Specify the path of input file
Step 5: Check Preview Data
Step 4: Now, drag & drop csv output file from output section
Step 5: Connect Hops from table input to Microsoft excel output

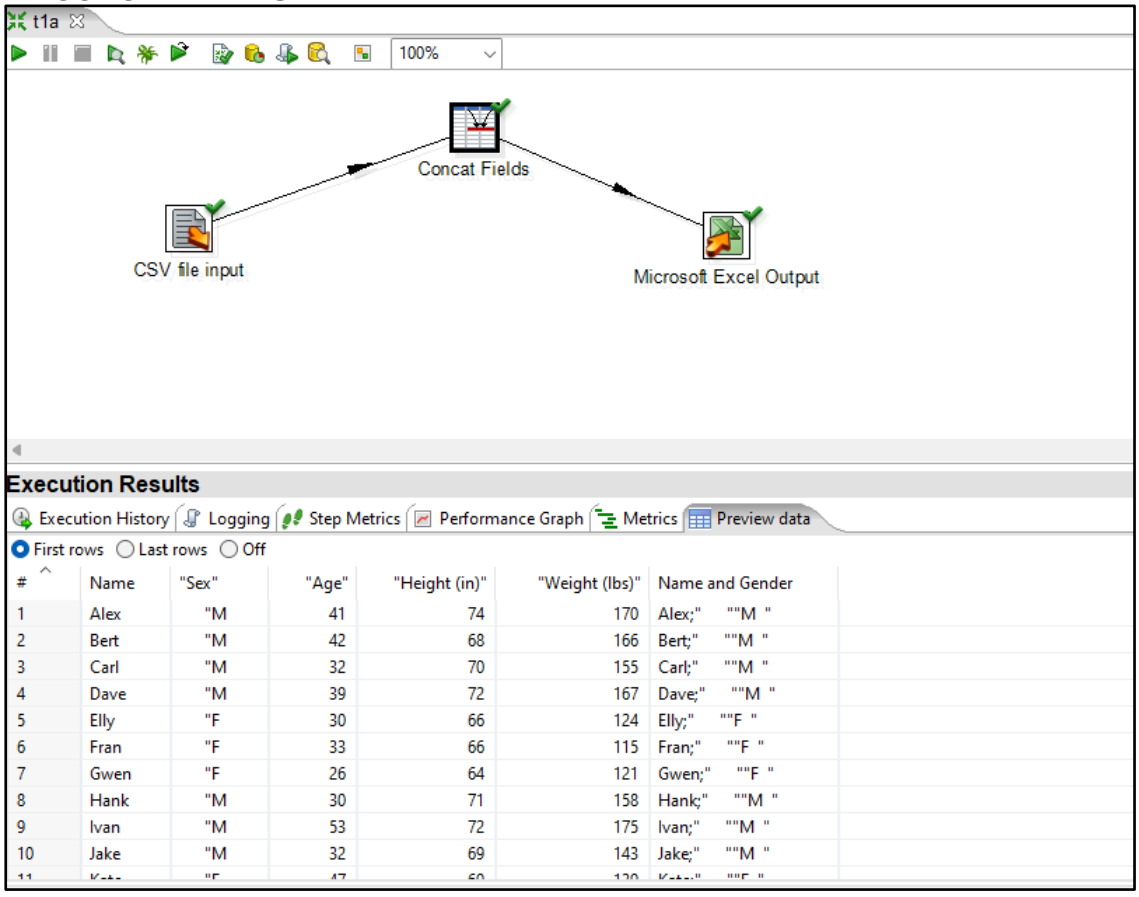
TRANSFORMATIONS:

1. SORT ROWS



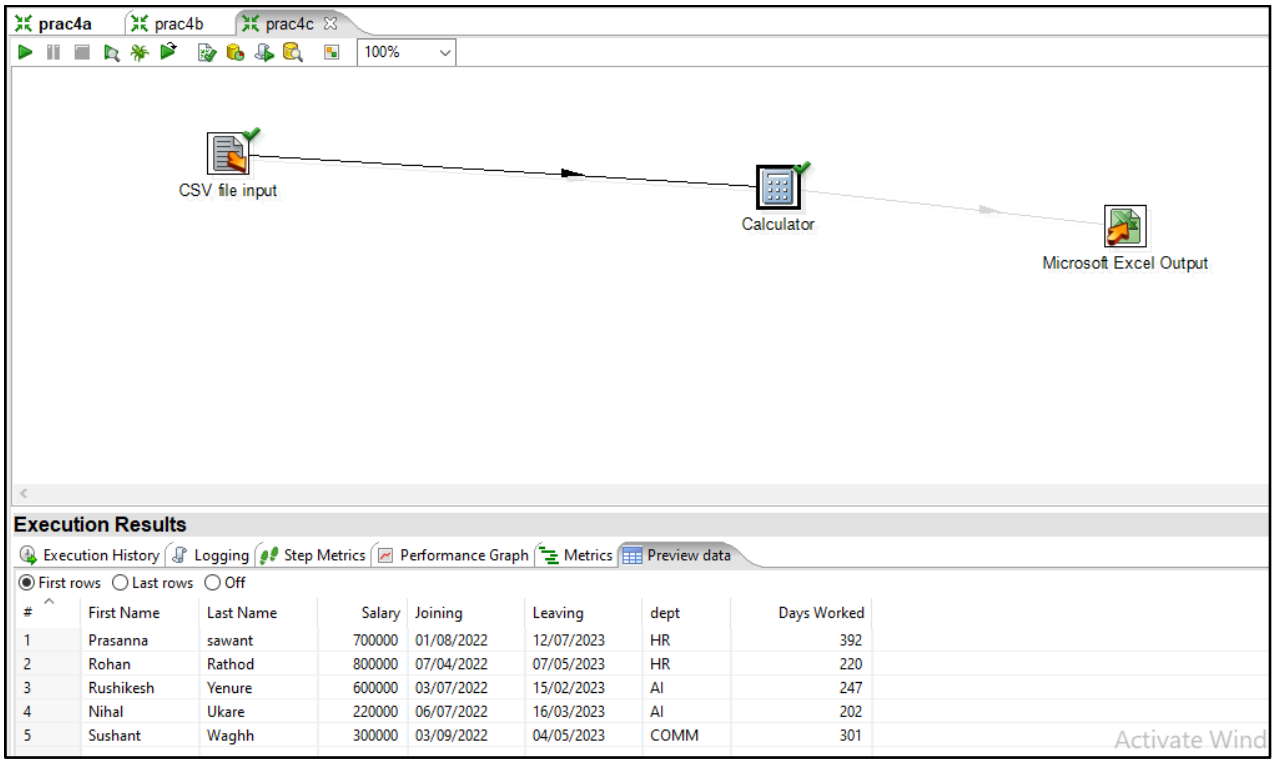
#	Name	"Sex"	"Age"	"Height (in)"	"Weight (lbs)"
1	Alex	"M	41	74	170
2	Bert	"M	42	68	166
3	Carl	"M	32	70	155
4	Dave	"M	39	72	167
5	Elly	"F	30	66	124
6	Fran	"F	33	66	115
7	Gwen	"F	26	64	121

2. CONCAT FIELDS



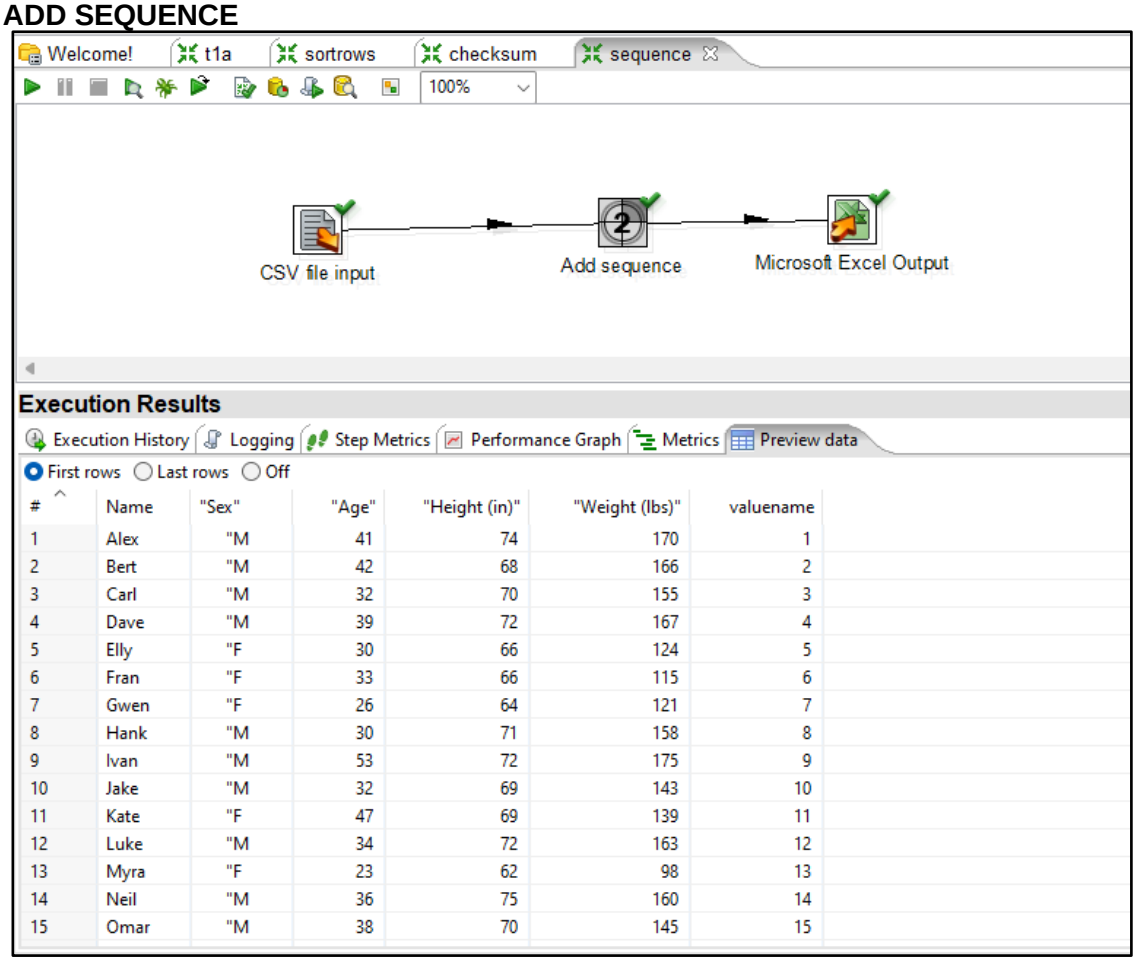
#	Name	"Sex"	"Age"	"Height (in)"	"Weight (lbs)"	Name and Gender
1	Alex	"M	41	74	170	Alex;" ""M "
2	Bert	"M	42	68	166	Bert;" ""M "
3	Carl	"M	32	70	155	Carl;" ""M "
4	Dave	"M	39	72	167	Dave;" ""M "
5	Elly	"F	30	66	124	Elly;" ""F "
6	Fran	"F	33	66	115	Fran;" ""F "
7	Gwen	"F	26	64	121	Gwen;" ""F "
8	Hank	"M	30	71	158	Hank;" ""M "
9	Ivan	"M	53	72	175	Ivan;" ""M "
10	Jake	"M	32	69	143	Jake;" ""M "

3. CALCULATOR



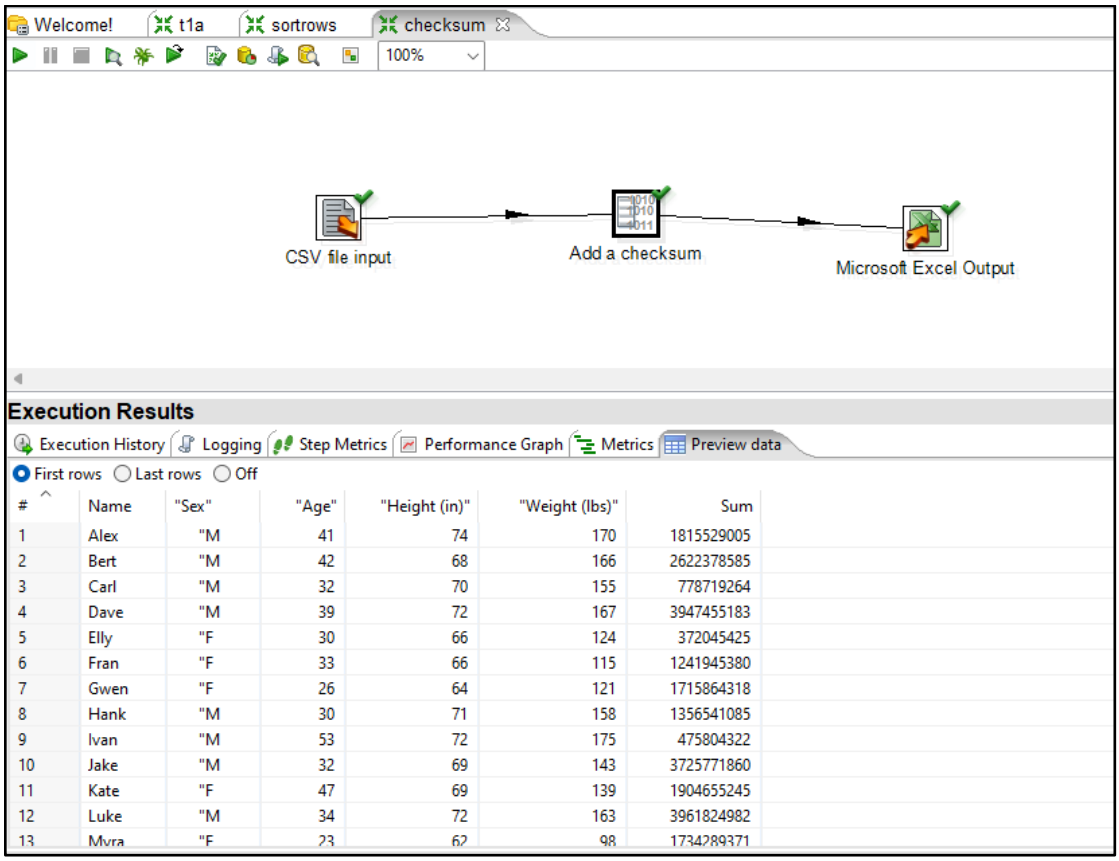
#	First Name	Last Name	Salary	Joining	Leaving	dept	Days Worked
1	Prasanna	sawant	700000	01/08/2022	12/07/2023	HR	392
2	Rohan	Rathod	800000	07/04/2022	07/05/2023	HR	220
3	Rushikesh	Yenure	600000	03/07/2022	15/02/2023	AI	247
4	Nihal	Ukare	220000	06/07/2022	16/03/2023	AI	202
5	Sushant	Waghh	300000	03/09/2022	04/05/2023	COMM	301

4. ADD SEQUENCE



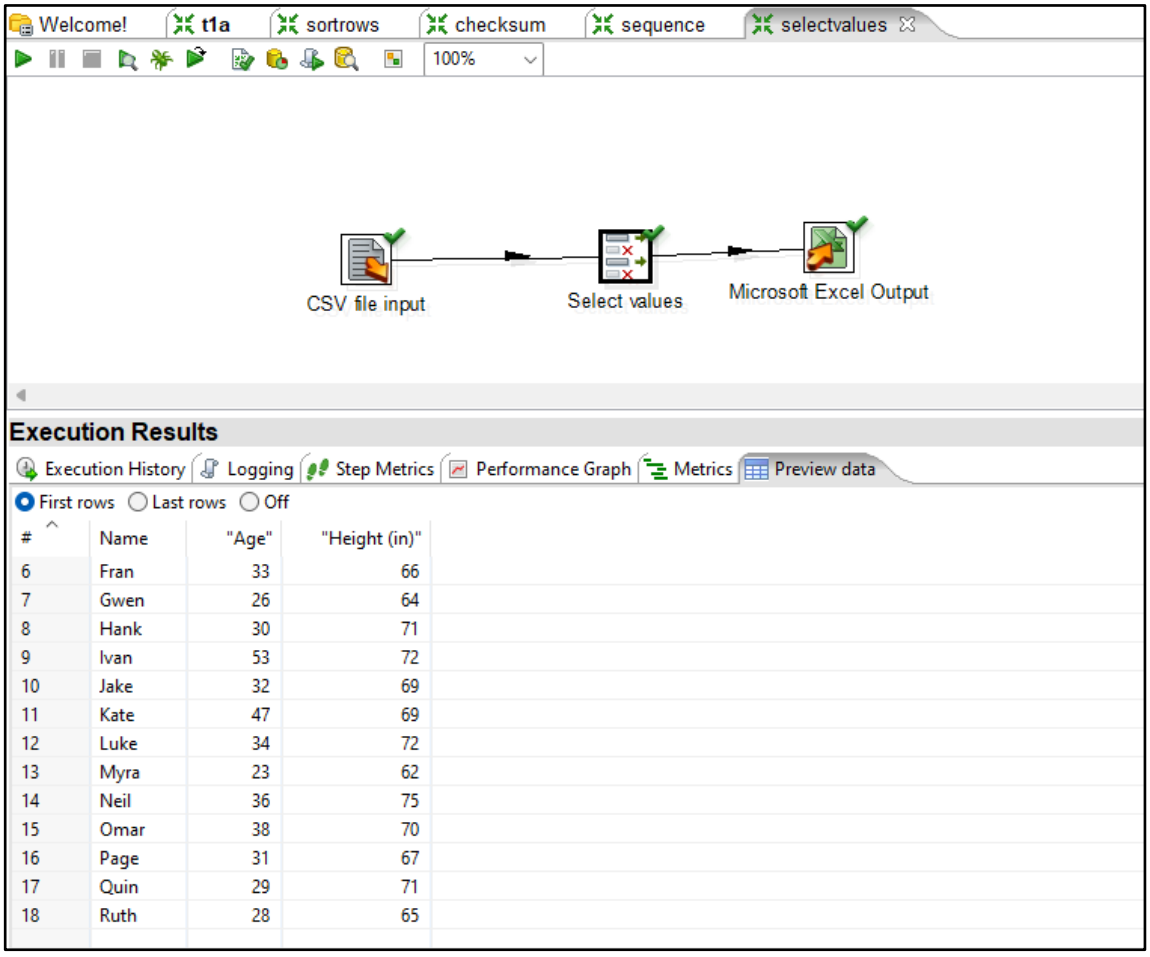
#	Name	"Sex"	"Age"	"Height (in)"	"Weight (lbs)"	valuename
1	Alex	"M	41	74	170	1
2	Bert	"M	42	68	166	2
3	Carl	"M	32	70	155	3
4	Dave	"M	39	72	167	4
5	Elly	"F	30	66	124	5
6	Fran	"F	33	66	115	6
7	Gwen	"F	26	64	121	7
8	Hank	"M	30	71	158	8
9	Ivan	"M	53	72	175	9
10	Jake	"M	32	69	143	10
11	Kate	"F	47	69	139	11
12	Luke	"M	34	72	163	12
13	Myra	"F	23	62	98	13
14	Neil	"M	36	75	160	14
15	Omar	"M	38	70	145	15

5. ADD A CHECKSUM



#	Name	"Sex"	"Age"	"Height (in)"	"Weight (lbs)"	Sum
1	Alex	"M	41	74	170	1815529005
2	Bert	"M	42	68	166	2622378585
3	Carl	"M	32	70	155	778719264
4	Dave	"M	39	72	167	3947455183
5	Elly	"F	30	66	124	372045425
6	Fran	"F	33	66	115	1241945380
7	Gwen	"F	26	64	121	1715864318
8	Hank	"M	30	71	158	1356541085
9	Ivan	"M	53	72	175	475804322
10	Jake	"M	32	69	143	3725771860
11	Kate	"F	47	69	139	1904655245
12	Luke	"M	34	72	163	3961824982
13	Myra	"F	23	62	98	1734289371

6. SELECT VALUES



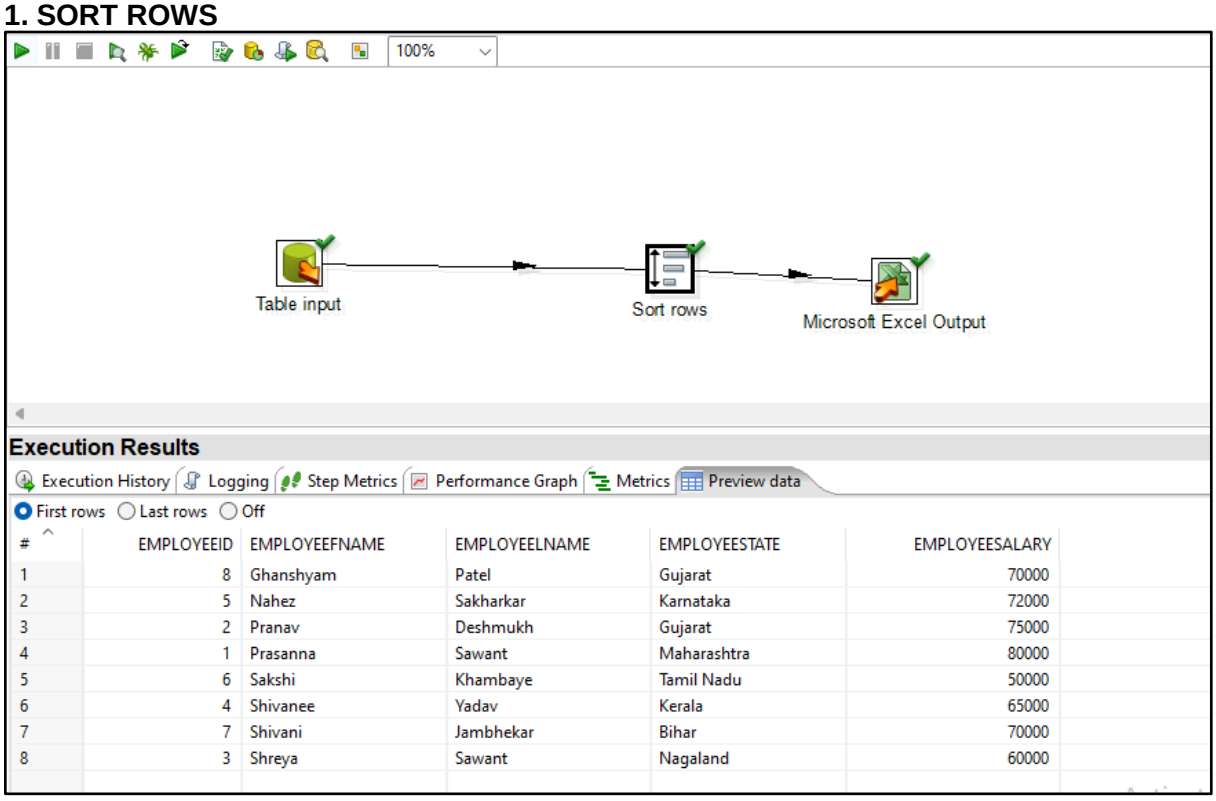
#	Name	"Age"	"Height (in)"
6	Fran	33	66
7	Gwen	26	64
8	Hank	30	71
9	Ivan	53	72
10	Jake	32	69
11	Kate	47	69
12	Luke	34	72
13	Myra	23	62
14	Neil	36	75
15	Omar	38	70
16	Page	31	67
17	Quin	29	71
18	Ruth	28	65

2: Transform Oracle Table Details into Microsoft Excel File

Step 1: Connect your oracle database
Step 2: Select a table from your database.
Step3: Connect Hops from table input to Microsoft excel output

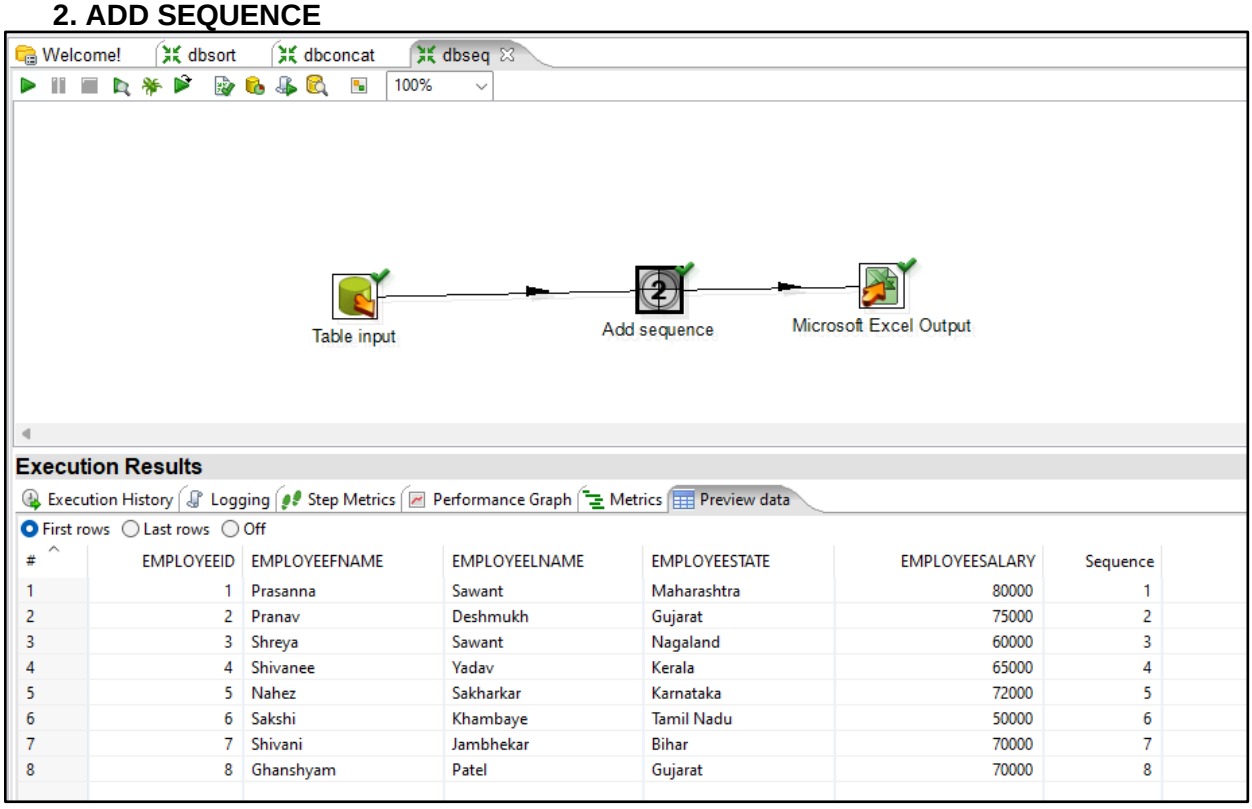
TRANSFORMATIONS:

1. SORT ROWS



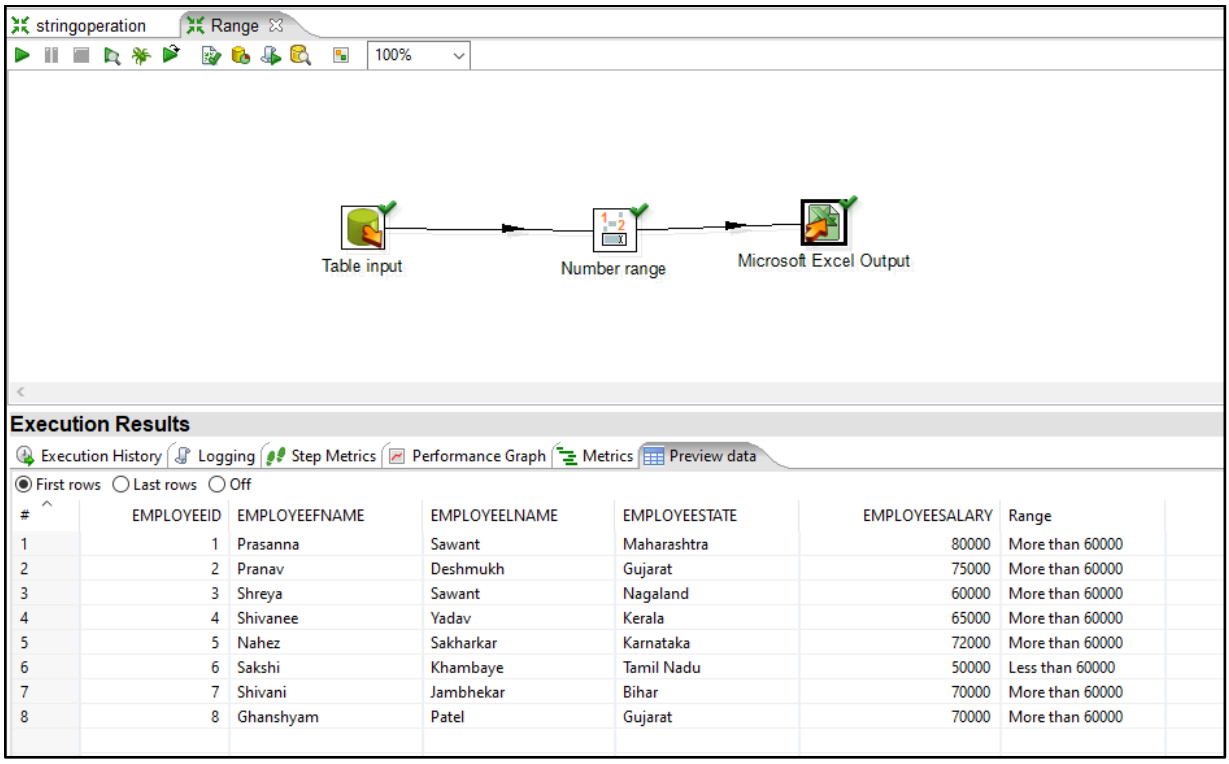
#	EMPLOYEEID	EMPLOYEEFNAME	EMPLOYEELNAME	EMPLOYEESTATE	EMPLOYEESALARY
1	8	Ghanshyam	Patel	Gujarat	70000
2	5	Nahez	Sakharkar	Karnataka	72000
3	2	Pranav	Deshmukh	Gujarat	75000
4	1	Prasanna	Sawant	Maharashtra	80000
5	6	Sakshi	Khambaye	Tamil Nadu	50000
6	4	Shivanee	Yadav	Kerala	65000
7	7	Shivani	Jambhekar	Bihar	70000
8	3	Shreya	Sawant	Nagaland	60000

2. ADD SEQUENCE



#	EMPLOYEEID	EMPLOYEEFNAME	EMPLOYEELNAME	EMPLOYEESTATE	EMPLOYEESALARY	Sequence
1	1	Prasanna	Sawant	Maharashtra	80000	1
2	2	Pranav	Deshmukh	Gujarat	75000	2
3	3	Shreya	Sawant	Nagaland	60000	3
4	4	Shivanee	Yadav	Kerala	65000	4
5	5	Nahez	Sakharkar	Karnataka	72000	5
6	6	Sakshi	Khambaye	Tamil Nadu	50000	6
7	7	Shivani	Jambhekar	Bihar	70000	7
8	8	Ghanshyam	Patel	Gujarat	70000	8

3. NUMBER RANGE



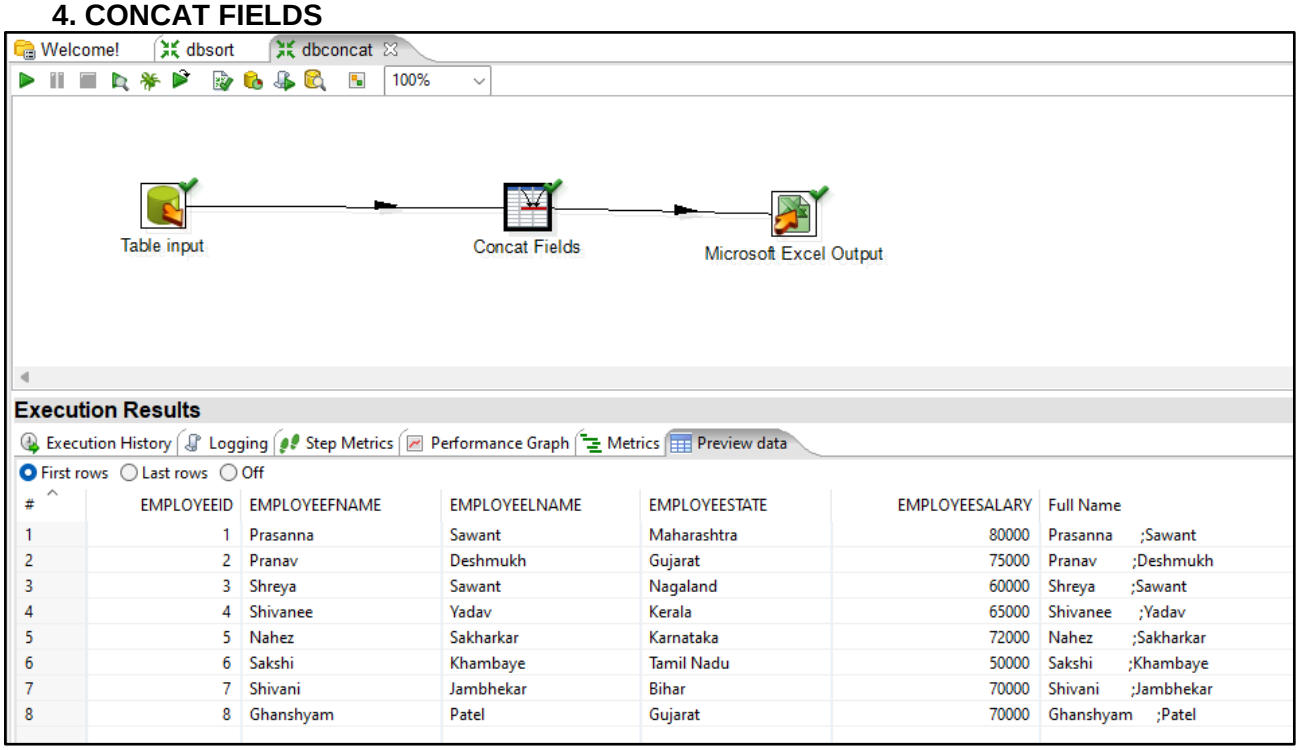
Execution Results

Execution History | Logging | Step Metrics | Performance Graph | Metrics | Preview data

First rows | Last rows | Off

#	EMPLOYEEID	EMPLOYEEFNAME	EMPLOYEELNAME	EMPLOYEESTATE	EMPLOYEESALARY	Range
1	1	Prasanna	Sawant	Maharashtra	80000	More than 60000
2	2	Pranav	Deshmukh	Gujarat	75000	More than 60000
3	3	Shreya	Sawant	Nagaland	60000	More than 60000
4	4	Shivanee	Yadav	Kerala	65000	More than 60000
5	5	Nahez	Sakharkar	Karnataka	72000	More than 60000
6	6	Sakshi	Khambaye	Tamil Nadu	50000	Less than 60000
7	7	Shivani	Jambhekar	Bihar	70000	More than 60000
8	8	Ghanshyam	Patel	Gujarat	70000	More than 60000

4. CONCAT FIELDS



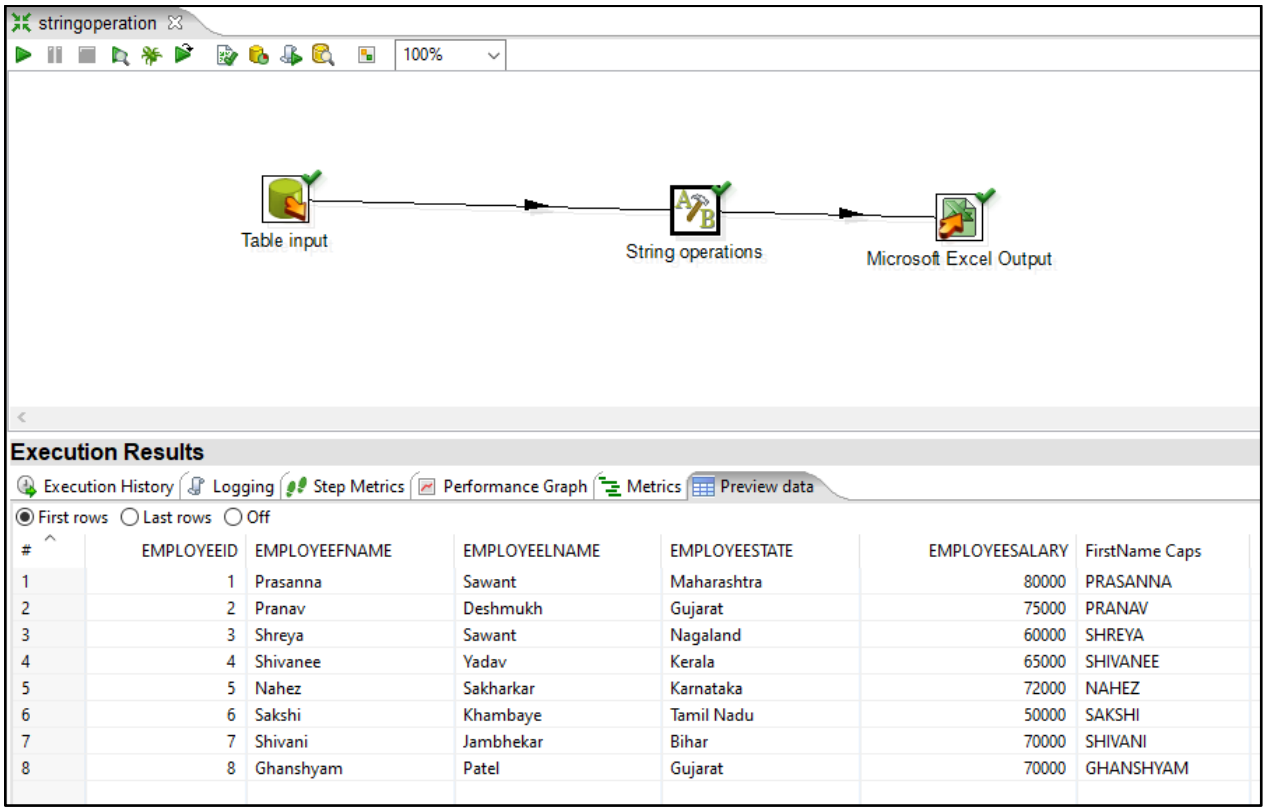
Execution Results

Execution History | Logging | Step Metrics | Performance Graph | Metrics | Preview data

First rows | Last rows | Off

#	EMPLOYEEID	EMPLOYEEFNAME	EMPLOYEELNAME	EMPLOYEESTATE	EMPLOYEESALARY	Full Name
1	1	Prasanna	Sawant	Maharashtra	80000	Prasanna ;Sawant
2	2	Pranav	Deshmukh	Gujarat	75000	Pranav ;Deshmukh
3	3	Shreya	Sawant	Nagaland	60000	Shreya ;Sawant
4	4	Shivanee	Yadav	Kerala	65000	Shivanee ;Yadav
5	5	Nahez	Sakharkar	Karnataka	72000	Nahez ;Sakharkar
6	6	Sakshi	Khambaye	Tamil Nadu	50000	Sakshi ;Khambaye
7	7	Shivani	Jambhekar	Bihar	70000	Shivani ;Jambhekar
8	8	Ghanshyam	Patel	Gujarat	70000	Ghanshyam ;Patel

5. STRING OPERATIONS



Execution Results

Execution History | Logging | Step Metrics | Performance Graph | Metrics | Preview data

First rows | Last rows | Off

#	EMPLOYEEID	EMPLOYEEFNAME	EMPLOYEELNAME	EMPLOYEESTATE	EMPLOYEESALARY	FirstName Caps
1	1	Prasanna	Sawant	Maharashtra	80000	PRASANNA
2	2	Pranav	Deshmukh	Gujarat	75000	PRANAV
3	3	Shreya	Sawant	Nagaland	60000	SHREYA
4	4	Shivanee	Yadav	Kerala	65000	SHIVANEE
5	5	Nahez	Sakharkar	Karnataka	72000	NAHEZ
6	6	Sakshi	Khambaye	Tamil Nadu	50000	SAKSHI
7	7	Shivani	Jambhekar	Bihar	70000	SHIVANI
8	8	Ghanshyam	Patel	Gujarat	70000	GHANSHYAM

CONCLUSION: I understood the concept of ETL using Pentaho and how to apply various transformations.