Practical No. 4

Title: - Implementation of ETL transformation with Pentaho
Aim: - ETL Transformation with Pentaho.

Lab Objectives: -
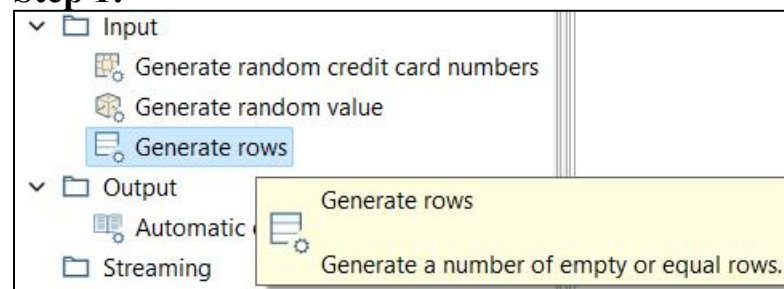**Students will understand following concepts:**
I. Copy data from Source (Table/Excel/ Oracle) and store it to Target (Table/Excel/ Oracle)
II. Adding sequence, Adding Calculator, Concatenation of two fields, Splitting of two fields
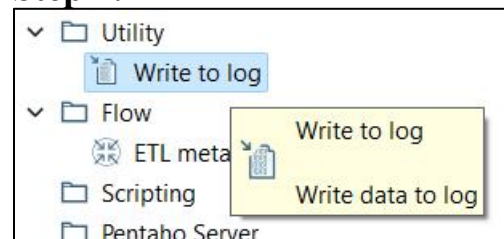**III. String Operations, Sorting data, Implement the merge join transformation on tables.**

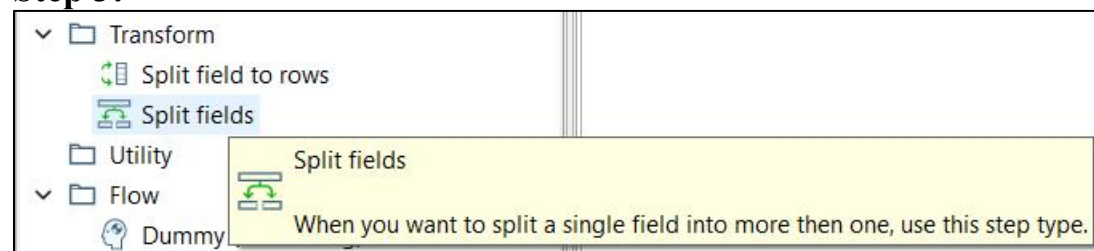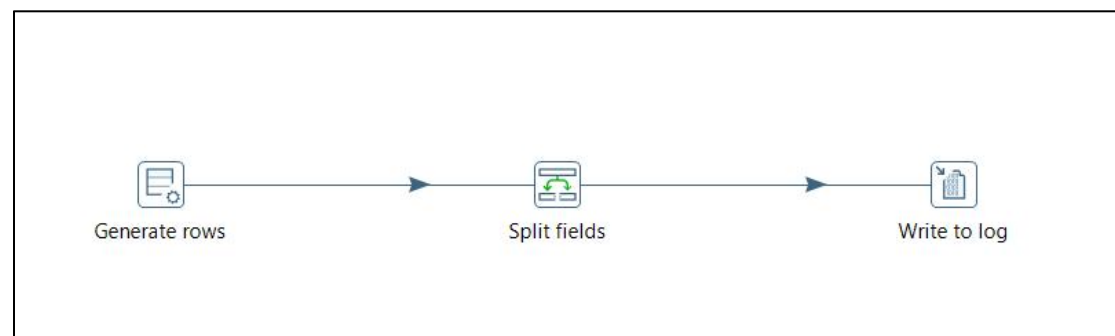**Transformation1: Fetch Data From Data Set**

Step 1:

Step 2:

Step 3:

Step 4:

Step 5:

Generate rows

Step name	Generate rows
Limit	1
Never stop generating rows	☐
Interval in ms (delay)	5000
Current row time field name	now
Previous row time field name	FiveSecondsAgo

Fields :

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Value	Set empty str
1	var	String		60					Onkar;Malawade	N
2	var	String		60					Omkar;Vele	N

Help   OK   Preview   Cancel

Step 6:

Split fields

Step name	Split fields
Field to split	var_1
Delimiter	;
Enclosure	

Fields

#	New field	ID	Remove ID?	Type	Length	Precision	Format	Group	Decimal	Currency	Nullif	Default	Trim type
1	var11		N	String									none
2	var12		N	String									none

Help   OK   Cancel

Step 7:

Write to log

Step name	Write to log
Log level	Basic
Print header	☑
Limit rows?	☐
Nr of rows to print	0
Write to log	

Fields

#	Field
1	var11
2	var12

Help   OK   Get Fields   Cancel

Step 8:

Generate rows → Split fields → Write to log
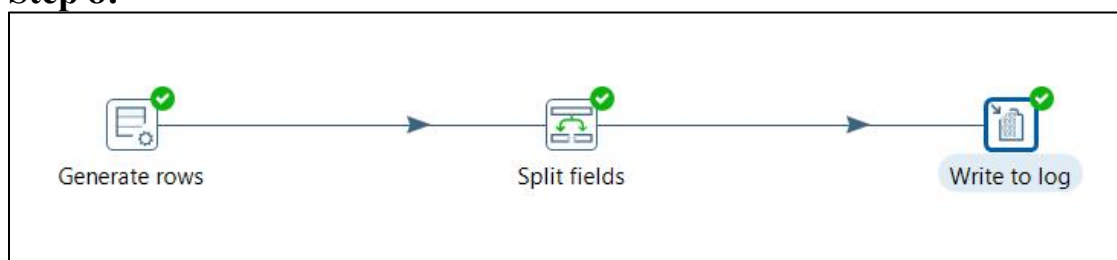
Step 9:

#	var	var11	var12
1	Onkar;Malawade	Omkar	Vele

Logging | Execution History | Step Metrics
● First rows ○ Last rows ○ Off

Step 10:

```
2023/12/04 18:31:25 - Write to log.0 - ------------> Linenr 1-----------------------------
2023/12/04 18:31:25 - Write to log.0 - var11 = Omkar
2023/12/04 18:31:25 - Write to log.0 - var12 = Vele
2023/12/04 18:31:25 - Write to log.0 -
2023/12/04 18:31:25 - Write to log.0 - ====================
```
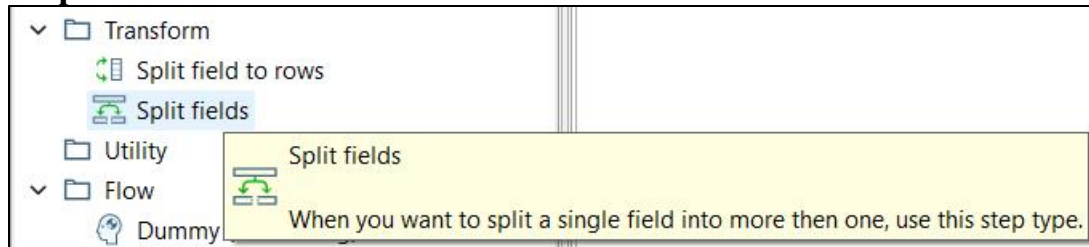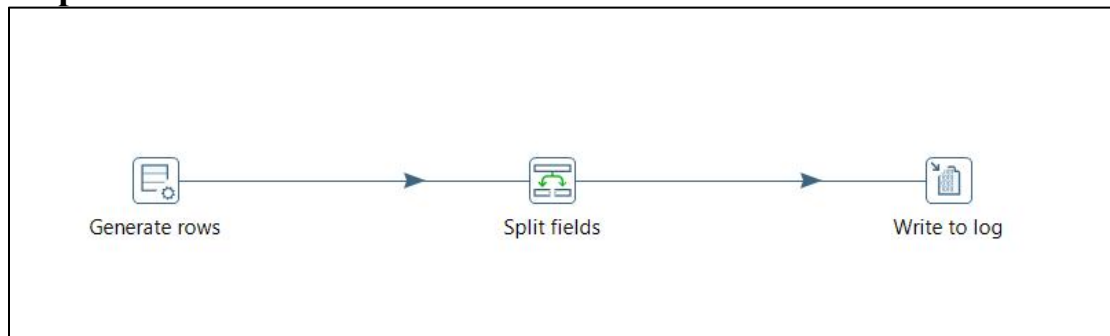
Transformation 2: String Concatenation

Step 1:

Step 2:

Step 3:

Step 4:

Step 5:

Step 6:

Split fields							— □ ✕	

Step name: Split fields
Field to split: StdMk
Delimiter: ;
Enclosure:

Fields

#	New field	ID	Remove ID?	Type	Length	Precision	Format	Group	Decimal	Currency	Nullif	Default	Trim type
1	num1		N	String									none
2	num12		N	String									none
3	num13		N	String									none

⑦ Help   OK   Cancel

Step 7:

```
2023/12/04 18:34:26 - Write to log.0 - ------------> Linenr 1----------------------------
2023/12/04 18:34:26 - Write to log.0 - num1 = 12
2023/12/04 18:34:26 - Write to log.0 - num12 = 34
2023/12/04 18:34:26 - Write to log.0 - num13 = 67
2023/12/04 18:34:26 - Write to log.0 -
2023/12/04 18:34:26 - Write to log.0 - =====================
```

Transformation 3: Merge Join
Step 1:

Step 2:

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Null if	Set empty strin...
1	name	String		23						N
2	empid	Integer								N
3	deptid	Integer								N
4	age	Integer								N
5	address	String		50						N

Step name: Employee — Meta / Data

OK · Preview · Cancel · Help

Step 3:

Step name: Employee — Meta / Data

#	name	empid	deptid	age	address
1	Onkar	101	10	23	Talere
2	Aditya	102	10	22	Dubai
3	Omkar	103	11	21	Pele
4					

OK · Preview · Cancel · Help

Step 4:

Step name: Department — Meta / Data

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Null if	Set empty string?
1	deptid	Integer								N
2	deptNm	String		23						N

OK · Preview · Cancel · Help

Step 5:

#	deptid	deptNm
1	10	Information
2	11	Police

Step name: Department

Meta | Data

Help | Preview | Cancel

Step 6:

Merge join

Field	Value
Step name	Merge join
First Step:	Sort rows
Second Step:	Sort rows 2
Join Type:	INNER

Keys for 1st step:

#	Key field
1	deptid

Keys for 2nd step:

#	Key field
1	deptid

Get key fields | Get key fields

Help | OK | Cancel

Step 7:

Employee → Sort rows

Department → Sort rows 2

→ Merge join
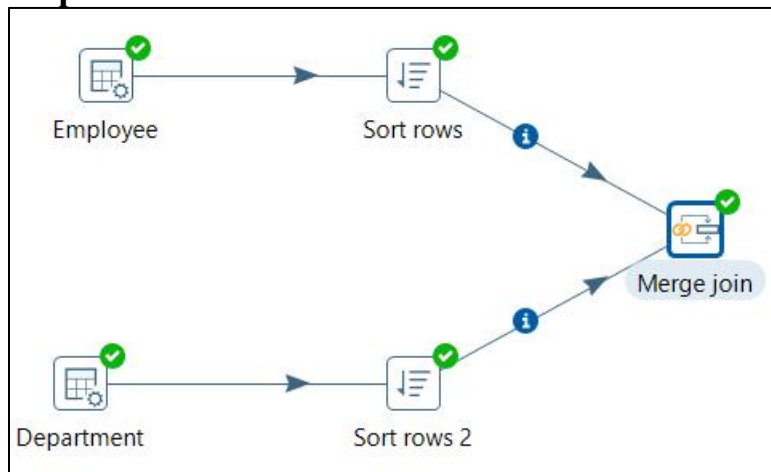
Step 8:

First rows ○ Last rows ○ Off

#	name	empid	deptid	age	address	deptid_1	deptNm
1	Onkar	101	10	23	Talere	10	Information
2	Aditya	102	10	22	Dubai	10	Information
3	Omkar	103	11	21	Pele	11	Police

Transformation Calculator

Step 1:

calculator → Sort rows → Calculator 2
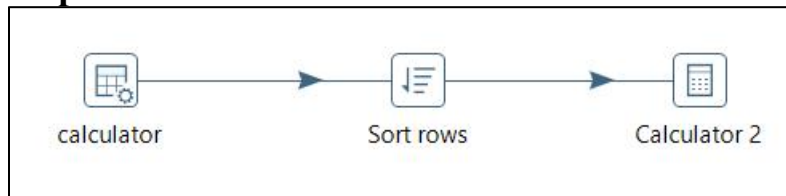
Step 2:

Data grid										— □ ×

Step name   calculator

Meta \ Data

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Null if	Set empty strin...
1	var1	Number								N
2	var2	Number								N

? Help   OK   Preview   Cancel

Step 3:

D	— □ ×

Step   calculator

Meta | Data

#	var1	var2
1	10	12
2	21	12
3	123	123
4	2	3
5	23	43
6	21	44
7	67	90

? Help   Preview   Cancel

Step 4:

Calculator	— □ ×

Step name

Calculator 2

☑ Throw an error on non existing files

Fields:

#	New field	Calculation	Field A	Field B	Field C	Value type
1	addition	A + B	var1	var2	addition	String
2	sub	A - B	var1	var2	sub	String
3	multi	A * B	var1	var2	multi	String
4	div	A / B	var1	var2	div	String
5	sqr	A * A	var1		sqr	String
6	sqrt	SQRT(A)	var1		sqrt	String

? Help   OK   Cancel

Step 5:

● First rows ○ Last rows ○ Off

#	var1	var2	addition	sub	multi	div	sqr	sqrt
1	2.0	3.0	5.0	-1.0	6.0	0.6666666667	4.0	1.4142135624
2	10.0	12.0	22.0	-2.0	120.0	0.8333333333	100.0	3.1622776602
3	21.0	12.0	33.0	9.0	252.0	1.75	441.0	4.582575695
4	21.0	44.0	65.0	-23.0	924.0	0.4772727273	441.0	4.582575695
5	23.0	43.0	66.0	-20.0	989.0	0.5348837209	529.0	4.7958315233
6	67.0	90.0	157.0	-23.0	6030.0	0.7444444444	4489.0	8.1853527719
7	123.0	123.0	246.0	0.0	15129.0	1.0	15129.0	11.0905365064

Transformation Data Grid Add Sequence

Step 1:

Step 2:

#	Name	Type	Format
1	stdid	Number	
2	sname	String	
3	dob	Date	dd-MM-yyyy
4	marks	Number	
5	percentage	Number	
6	areaOfInterest	String	

Step 3:

| #	stdid	sname	dob	marks	percentage	areaOfInterest
1	101	Onkar	30-05-2000	100	100	Coding
2	102	Aniket	10-02-2001	100	100	Cricket
3	103	Raju	10-12-1999	100	100	Hockey
4	104	Aditya	10-11-2002	130	100	Football
5	105	Yash	12-01-1989	50	12	InstagramReel

Step 4:

Step 5:

Execution Results

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

○ First rows ○ Last rows ○ Off

#	stdid	sname	dob	marks	percentage	areaOfInterest	addSequence
1	101.0	Onkar	30-05-2000	100.0	100.0	Coding	1
2	102.0	Aniket	10-02-2001	100.0	100.0	Cricket	2
3	103.0	Raju	10-12-1999	100.0	100.0	Hockey	3
4	104.0	Aditya	10-11-2002	130.0	100.0	Football	4
5	105.0	Yash	12-01-1989	50.0	12.0	InstagramReel	5

Step 6:

Data grid → Add sequence
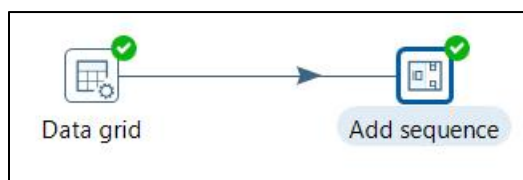
Transformation Csv File Operations on It

Step 1:
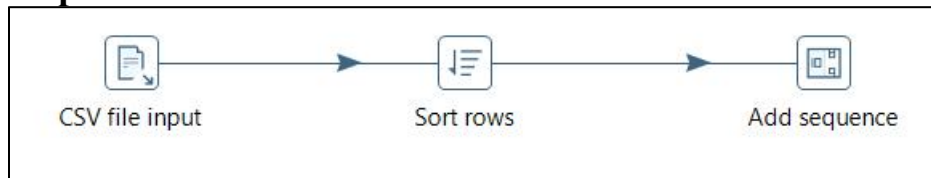
Step 2:

CSV file input

Field	Value
Step name	CSV file input
Filename	C:\Users\student\Downloads\custsales.csv.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?	☐
Format	mixed
File encoding	

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	id	String							none

Help   OK   Get Fields   Preview   Cancel

Step 3:

CSV file input → Sort rows → Add sequence
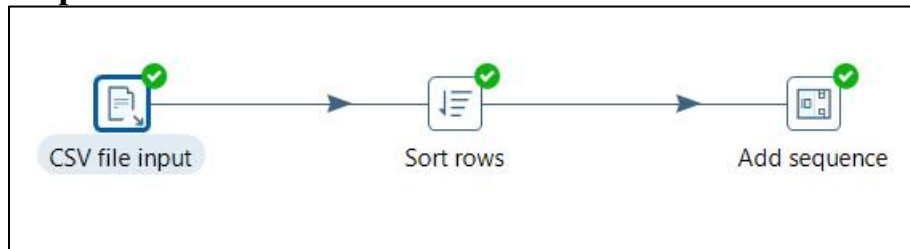
Step 4:

Execution Results

Logging | Execution History | Step Metrics

● First rows ○ Last rows ○ Off

#	id	firstname	lastname	sales
1	1	France	44	72000
2	2	Spain	27	48000
3	3	Germany	30	54000
4	4	Spain	38	61000
5	5	Germany	40	NA
6	6	France	35	58000
7	7	Spain	NA	52000
8	8	France	48	79000
9	9	Germany	50	83000
1..	10	France	37	67000