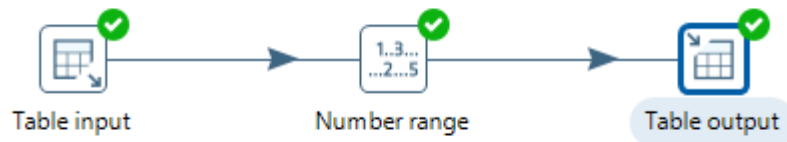| Assignment No	15
Title	PENTAHO transformation
Objective	1) Number range 2) String Operations 3) CSV File to target Output 4) SQL join operation data grid PENTAHO MERGE JOIN 5) DATA GRID TO DATA VALIDATOR
Roll No	MCA2565

1) Number Range

Table output

Step name	Table output
Connection	test
Target schema	SYSTEM
Target table	EMP1
Commit size	1000
Truncate table	☑
Ignore insert errors	☐
Specify database fields	☑

Main options | **Database fields**

Fields to insert:

#	Table field	Stream field
1	ENO	ENO
2	ENAME	ENAME
3	SALARY	SALARY
4	COMMISSI...	COMMISSION
5	HIREDATE	HIREDATE
6	DPTNO	DPTNO
7	range	range

Get fields

Enter field mapping

Help   OK   Cancel   SQL

Examine preview data

Rows of step: Table output (10 rows)

#	ENO	ENAME	SALARY	COMMISSION	HIREDATE	DPTNO	range
1	10.0	***************Xavier	39000.0	10000	1990/07/30 00:00:00.000000000	9.0	More than 10
2	9.0	******************Max	35000.0	9000	1986/05/05 00:00:00.000000000	10.0	5-10
3	8.0	***************Sameer	20000.0	1000	1996/10/18 00:00:00.000000000	2.0	5-10
4	7.0	*************Yadhnesh	50000.0	<null>	2001/04/12 00:00:00.000000000	6.0	5-10
5	6.0	***************Banner	12000.0	<null>	2012/03/13 00:00:00.000000000	1.0	5-10
6	5.0	*****************Mary	18000.0	3000	1999/11/01 00:00:00.000000000	5.0	5-10
7	4.0	****************Harry	15000.0	3000	2005/11/12 00:00:00.000000000	6.0	Less than 5
8	3.0	******************Lex	30000.0	8000	1985/02/25 00:00:00.000000000	12.0	Less than 5
9	2.0	*****************Alex	25000.0	5000	1996/01/12 00:00:00.000000000	5.0	Less than 5
10	1.0	**************Atharva	80000.0	50000	2000/12/02 00:00:00.000000000	5.0	Less than 5

2) String Operations
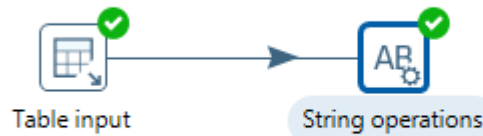
String operations — □ ✕

Step name String operations

The fields to process:

#	In stream field	Out stream field	Trim type	Lower/Upper	Padding	Pad char	Pad Length	InitCap	Escape	Digits	Remove Special character
1	ENAME		left		left	*	20	Y			

⑦ Help   OK   Get fields   Cancel

Examine preview data

Rows of step: String operations (10 rows)

#	ENO	ENAME	SALARY	COMMISSION	HIREDATE	DPTNO
1	10.0	**************Xavier	39000.0	10000	1990/07/30 00:00:00.000000000	9.0
2	9.0	******************Max	35000.0	9000	1986/05/05 00:00:00.000000000	10.0
3	8.0	**************Sameer	20000.0	1000	1996/10/18 00:00:00.000000000	2.0
4	7.0	************Yadhnesh	50000.0	<null>	2001/04/12 00:00:00.000000000	6.0
5	6.0	**************Banner	12000.0	<null>	2012/03/13 00:00:00.000000000	1.0
6	5.0	****************Mary	18000.0	3000	1999/11/01 00:00:00.000000000	5.0
7	4.0	***************Harry	15000.0	3000	2005/11/12 00:00:00.000000000	6.0
8	3.0	******************Lex	30000.0	8000	1985/02/25 00:00:00.000000000	12.0
9	2.0	****************Alex	25000.0	5000	1996/01/12 00:00:00.000000000	5.0
10	1.0	**************Atharva	80000.0	50000	2000/12/02 00:00:00.000000000	5.0
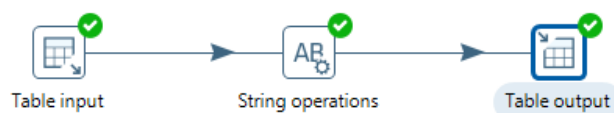
Table output — □ ×

Step name	Table output	
Connection	conn	Edit... New... Wizard...
Target schema		Browse...
Target table	EMP1	Browse...
Commit size	1000	
Truncate table	☑	
Ignore insert errors	☐	
Specify database fields	☑	

Main options | Database fields

Fields to insert:

#	Table field	Stream field
1	ENO	ENO
2	ENAME	ENAME
3	SALARY	SALARY
4	COMMISSI...	COMMISSION
5	HIREDATE	HIREDATE
6	DPTNO	DPTNO

Get fields

Enter field mapping

Execution Results

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

◉ First rows ○ Last rows ○ Off

#	ENO	ENAME	SALARY	COMMISSION	HIREDATE	DPTNO
1	1.0	*************Atharva	80000.0	50000	2000/12/02 00:00:00.000000000	5.0
2	2.0	*****************Alex	25000.0	5000	1996/01/12 00:00:00.000000000	5.0
3	3.0	******************Lex	30000.0	8000	1985/02/25 00:00:00.000000000	12.0
4	4.0	****************Harry	15000.0	3000	2005/11/12 00:00:00.000000000	6.0
5	5.0	*****************Mary	18000.0	3000	1999/11/01 00:00:00.000000000	5.0
6	6.0	***************Banner	12000.0	<null>	2012/03/13 00:00:00.000000000	1.0
7	7.0	*************Yadhnesh	50000.0	<null>	2001/04/12 00:00:00.000000000	6.0
8	8.0	***************Sameer	20000.0	1000	1996/10/18 00:00:00.000000000	2.0
9	9.0	*******************Max	35000.0	9000	1986/05/05 00:00:00.000000000	10.0
10	10.0	***************Xavier	39000.0	10000	1990/07/30 00:00:00.000000000	9.0

3) CSV File to target output

CSV file input   Table output

CSV file input

Step name	CSV file input
Filename	A:\data_apriori.csv Browse...
Delimiter	, Insert TAB
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 ty
1	Customer_Id	Integer	#	15	0	Rs.	.	,	none

⑦ Help   OK   Get Fields   Preview   Cancel

Examine preview data

Rows of step: CSV file input (50 rows)

#	Customer_Id	Products
1	1	laptop, mouse , headphones, pendrive, speakers
2	2	laptop, headphones
3	3	laptop, mouse, pendrive
4	4	mouse, speakers
5	5	laptop, pendrive
6	6	laptop, mouse , headphones, pendrive, speakers
7	7	laptop, headphones
8	8	laptop, mouse, pendrive
9	9	mouse, speakers
10	10	laptop, pendrive
11	11	laptop, mouse , headphones, pendrive, speakers
12	12	laptop, headphones
13	13	laptop, mouse, pendrive
14	14	laptop, mouse , headphones, pendrive, speakers
15	15	laptop, headphones
16	16	laptop, mouse, pendrive
17	17	mouse, speakers
18	18	laptop, pendrive
19	19	laptop, mouse , headphones, pendrive, speakers
20	20	laptop, headphones

Table output

Step name	Table output
Connection	CONN Edit... New... Wizard...
Target schema	Browse...
Target table	csvoutput Browse...
Commit size	1000
Truncate table	☑
Ignore insert errors	☐
Specify database fields	☑

Main options | Database fields

Fields to insert:

#	Table field	Stream field
1	Customer_Id	Customer_Id
2	Products	Products

Get fields

Enter field mapping

Transformation debug dialog — □

CSV file input	Number of rows to retrieve 1000
Table output	☐ Retrieve first rows (preview)
	☑ Pause transformation on condition
	Break-point / pause [ ] +
	`<field>` = `<field>`
	`<value>`
	Clear —

Quick Launch   Configure   Cancel

Examine preview data

Rows of step: Table output (50 rows)

#	Customer_Id	Products
1	50	laptop, mouse , headphones, pendrive, speakers
2	49	laptop, pendrive
3	48	mouse, speakers
4	47	laptop, mouse, pendrive
5	46	laptop, headphones
6	45	laptop, mouse , headphones, pendrive, speakers
7	44	laptop, pendrive
8	43	mouse, speakers
9	42	laptop, mouse, pendrive
10	41	laptop, headphones
11	40	laptop, mouse , headphones, pendrive, speakers
12	39	laptop, mouse, pendrive
13	38	laptop, headphones
14	37	laptop, mouse , headphones, pendrive, speakers
15	36	laptop, pendrive
16	35	mouse, speakers
17	34	laptop, mouse, pendrive
18	33	laptop, headphones
19	32	laptop, mouse , headphones, pendrive, speakers
20	31	laptop, pendrive

4) SQL join operation data grid PENTAHO   MERGE JOIN

Data grid

Data grid —

Step name employee

Meta \ Data

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Null if	Set em
1	name	String								N
2	book	String								
3	depno	Integer								
4	dept	String								
5	salary	Number								
6										

OK   Preview   Cancel

Help

Data grid

Step name: employee

Meta | Data

#	name	book	depno	dept	salary
1	puja	dbms	10	ics	4200
2	chetna	ds	10	ics	1000
3	bhushan	spm	10	ics	2000
4	om	java	20	ics	5000
5	asdasd	asdasd	30	mms	3454
6	grx	tyjy	30	iom	4536

DATA Grid 2

Data grid 2

Data grid

Step name: Department

Meta | Data

#	Name	Type	Format
1	ics	String	
2	depno	Number	
3	subjects	String	

Data grid

Step name: Department

Meta | Data

#	ics	depno	subjects
1	qwdqwd	10	qwe
2	awdw	20	olhj
3	fhtrdtg	10	jyn
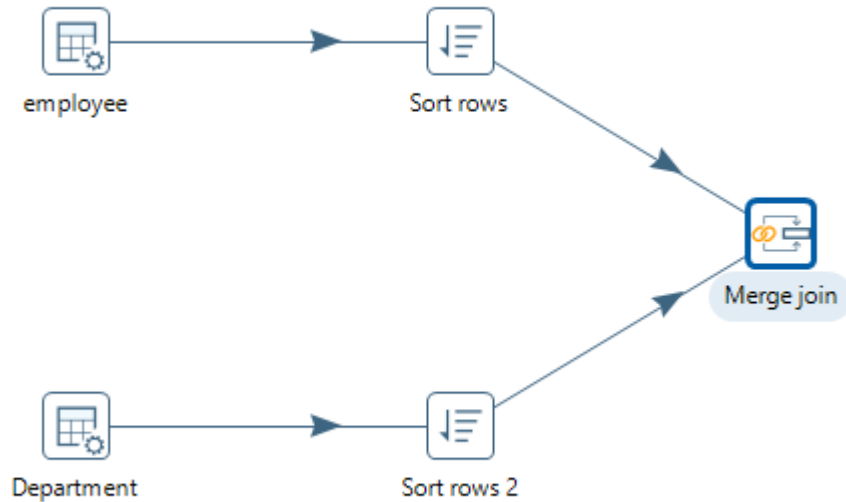
Sort rows

Step name	Sort rows
Sort directory	%%java.io.tmpdir%%
TMP-file prefix	out
Sort size (rows in memory)	1000000
Free memory threshold (in %)	
Compress TMP Files?	☐
Only pass unique rows? (verifies keys only)	☐

Fields :

#	Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Collator Strength	Presorted?
1	depno	Y	N	N	0	N

Sort rows

Step name	Sort rows 2
Sort directory	%%java.io.tmpdir%%
TMP-file prefix	out
Sort size (rows in memory)	1000000
Free memory threshold (in %)	
Compress TMP Files?	☐
Only pass unique rows? (verifies keys only)	☐

Fields :

| #	Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Collator Strength	Presorted?
1	depno	Y	N	N	0	N
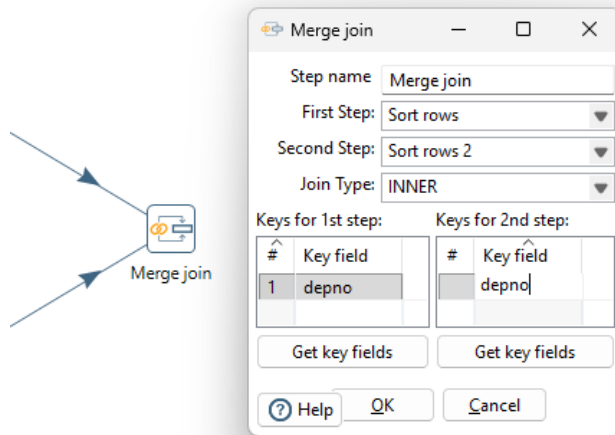
Examine preview data

Rows of step: Merge join (7 rows)

#	name	book	depno	dept	salary	ics	depno_1	subjects
1	om	java	20	ics	5000.0	awdw	20.0	olhj
2	bhushan	spm	10	ics	2000.0	fhtrdtg	10.0	jyn
3	bhushan	spm	10	ics	2000.0	qwdqwd	10.0	qwe
4	chetna	ds	10	ics	1000.0	fhtrdtg	10.0	jyn
5	chetna	ds	10	ics	1000.0	qwdqwd	10.0	qwe
6	puja	dbms	10	ics	4200.0	fhtrdtg	10.0	jyn
7	puja	dbms	10	ics	4200.0	qwdqwd	10.0	qwe

5) DATA GRID TO DATA VALIDATOR

Data grid

Step name product

Meta \ Data
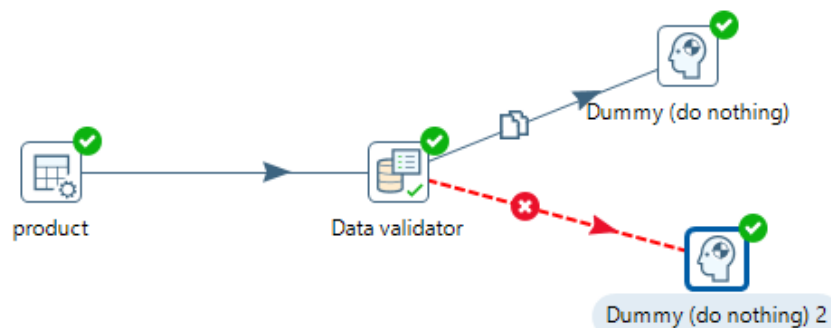
#	Name	Type	Format	Length	Precision	Currency
1	ID	Number				
2	PRICE	Number				
3	NAME	String				
4	DISPATCH	String				

Data grid

Step name product

Meta [Data]

#	ID	PRICE	NAME	DISPATCH
1	1	75000	LAPTOP	SHIPPED
2	2	1564	KEYBOARD	NO
3	3	6874	TV	SHIPPED
4	4	454102	SDF	WHY

Enter validation name ✕

Enter a unique name for the validation rule you want to add

STATUS

[OK]   [Cancel]

Only null values allowed?

Data validator — ☐ ✕

Stepname : Data validator

Select a validation to edit :

STATUS

☐ Report all errors, not only the first

☐ Output one row, concatenate errors with separator : | ◈

Validation description	STATUS
Name of field to validate	DISPATCH ▼
Error code	◈
Error description	◈

Type

Data validator

Stepname : Data validator

Select a validation to edit :

STATUS

☐ Report all errors, not only the first

☐ Output one row, concatenate errors with separator : |

Grouping Symbol

Data

Null allowed?	☑
Only null values allowed?	☐
Only numeric data expected	☐
Max string length	
Min string length	
Maximum value	
Minimum value	
Expected start string	
Expected end string	
Not allowed start string	
Not allowed end string	
Regular expression expected to match	
Regular expression not allowed to	

Allowed values

SHIPPED
NO

Add

Remove

Read allowed values from another step? ☐

The step to read from

The field to read from

Examine preview data

Rows of step: Dummy (do nothing) (3 rows)

#	ID	PRICE	NAME	DISPATCH
1	3.0	6874.0	TV	SHIPPED
2	2.0	1564.0	KEYBOARD	NO
3	1.0	75000.0	LAPTOP	SHIPPED

Examine preview data

Rows of step: Dummy (do nothing) 2 (1 rows)

#	ID	PRICE	NAME	DISPATCH
1	4.0	454102.0	SDF	WHY