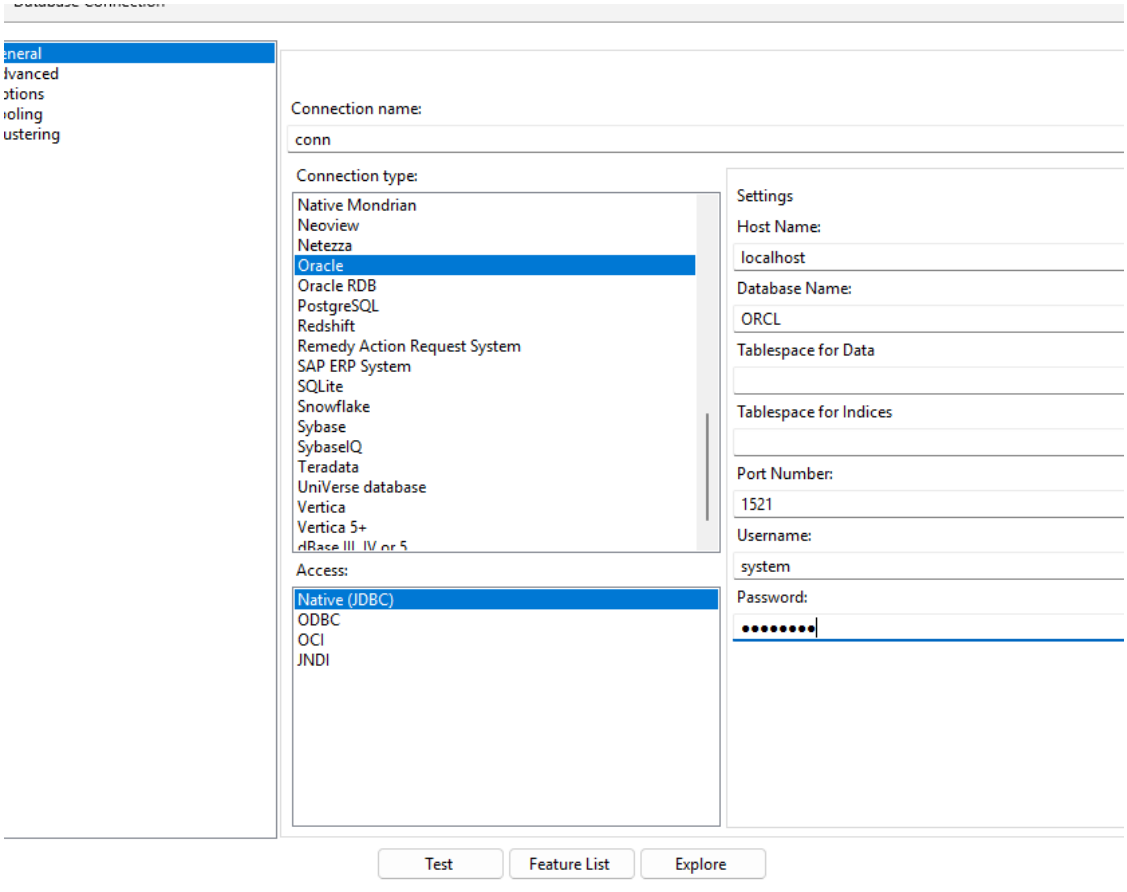
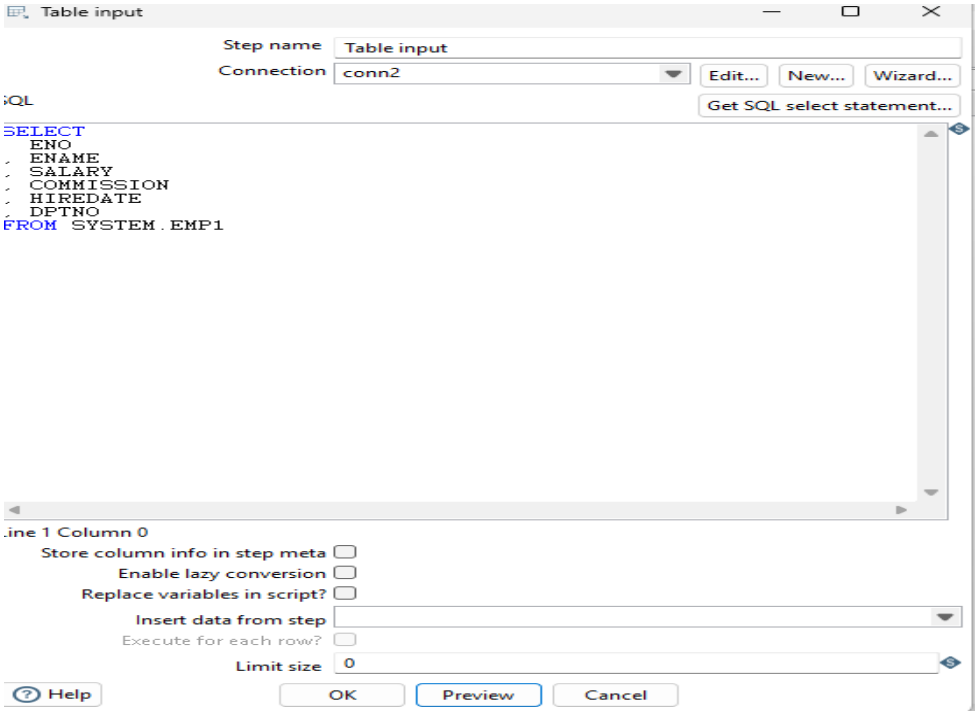
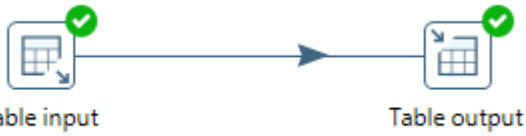
Assignment No	14
Title	**Pentaho, ETL(Extraction Transformation and Loading)**
Objective	1) Extracting data form SQL table and ETL(extraction transformation and loading) storing the data in sql 2) Sorting the data and add Sequence 3) Calculation Operations 4) Concatenation Operation 5) Split Operation
Roll No	MCA2565

1) Extracting data form SQL table and ETL(extraction transformation and loading) storing the data in sql







Examine preview data

Rows of step: Table output (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

2) Sorting the data and add Sequence

Sorting

Sort rows

Step name: Sort rows
Sort directory: %%java.io.tmpdir%% Browse...
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	SALARY	N	N	N	0	N

Examine preview data

Rows of step: Sort rows (10 rows)

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

Sequence add

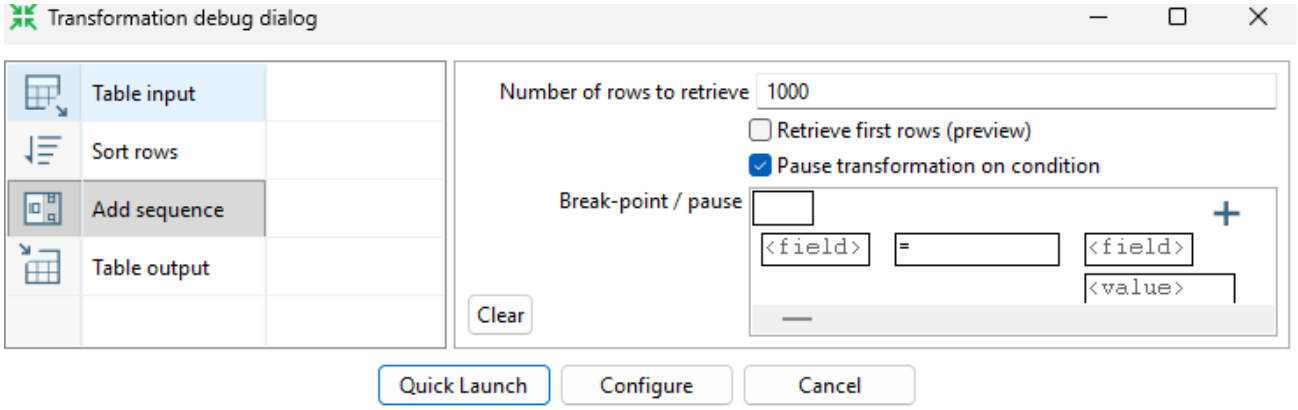
Add sequence

Step name: Add sequence
Name of value: valuename

Use a database to generate the sequence
Use DB to get sequence?
Connection: conn Edit... New... Wizard...
Schema name: Schemas...
Sequence name: SEQ_ Sequences...

Use a transformation counter to generate the sequence
Use counter to calculate sequence?
Counter name (optional):
Start at value: 1
Increment by: 5
Maximum value: 999999999

Help OK Cancel



Examine preview data

Rows of step: Add sequence (10 rows)

#	ENO	ENAME	SALARY	COMMISSION	HIREDATE	DPTNO	valuename
1	6.0	Banner	12000.0	<null>	2012/03/13 00:00:00.000000000	1.0	46
2	4.0	Harry	15000.0	3000	2005/11/12 00:00:00.000000000	6.0	41
3	5.0	Mary	18000.0	3000	1999/11/01 00:00:00.000000000	5.0	36
4	8.0	Sameer	20000.0	1000	1996/10/18 00:00:00.000000000	2.0	31
5	2.0	Alex	25000.0	5000	1996/01/12 00:00:00.000000000	5.0	26
6	3.0	LEX	30000.0	8000	1985/02/25 00:00:00.000000000	12.0	21
7	9.0	MAX	35000.0	9000	1986/05/05 00:00:00.000000000	10.0	16
8	10.0	Xavier	39000.0	10000	1990/07/30 00:00:00.000000000	9.0	11
9	7.0	Yadhnesh	50000.0	<null>	2001/04/12 00:00:00.000000000	6.0	6
10	1.0	Atharva	80000.0	50000	2000/12/02 00:00:00.000000000	5.0	1
11							

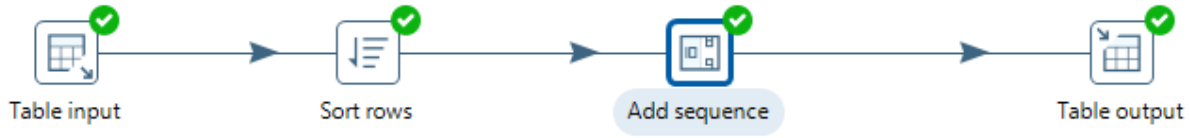


Table output with sort and sequence

Examine preview data

Rows of step: Table output (10 rows)

#	ENO	ENAME	SALARY	COMMISSION	HIREDATE	DPTNO	valuename
1	6.0	Banner	12000.0	<null>	2012/03/13 00:00:00.000000000	1.0	46
2	4.0	Harry	15000.0	3000	2005/11/12 00:00:00.000000000	6.0	41
3	5.0	Mary	18000.0	3000	1999/11/01 00:00:00.000000000	5.0	36
4	8.0	Sameer	20000.0	1000	1996/10/18 00:00:00.000000000	2.0	31
5	2.0	Alex	25000.0	5000	1996/01/12 00:00:00.000000000	5.0	26
6	3.0	LEX	30000.0	8000	1985/02/25 00:00:00.000000000	12.0	21
7	9.0	MAX	35000.0	9000	1986/05/05 00:00:00.000000000	10.0	16
8	10.0	Xavier	39000.0	10000	1990/07/30 00:00:00.000000000	9.0	11
9	7.0	Yadhnesh	50000.0	<null>	2001/04/12 00:00:00.000000000	6.0	6
10	1.0	Atharva	80000.0	50000	2000/12/02 00:00:00.000000000	5.0	1

3) CALCULATION Operation

Calculator

Step name

Calculator

Throw an error on non existing files

Fields:

#	New field	Calculation	Field A	Field B	Field C	Value type	Length	Precision	Remove
1	TOTAL	A + B	COMMI...	SALARY		Number			

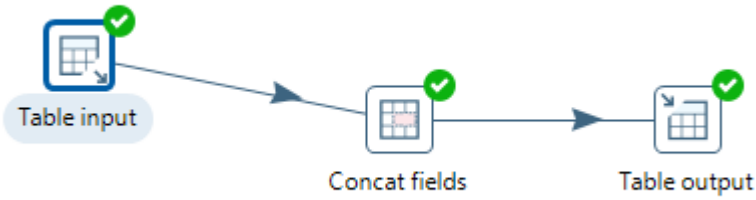
Help

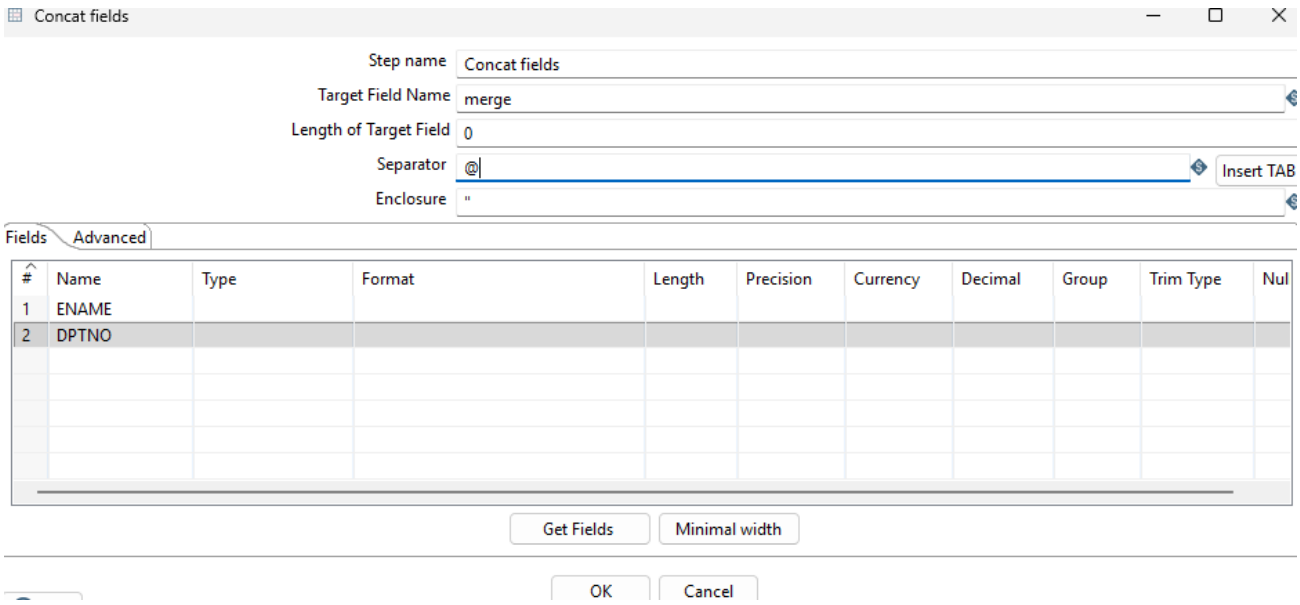
Examine preview data

Rows of step: Calculator (10 rows)

#	ENO	ENAME	SALARY	COMMISSION	HIREDATE	DPTNO	TOTAL
1	10.0	Xavier	39000.0	10000	1990/07/30 00:00:00.000000000	9.0	49000.0
2	9.0	MAX	35000.0	9000	1986/05/05 00:00:00.000000000	10.0	44000.0
3	8.0	Sameer	20000.0	1000	1996/10/18 00:00:00.000000000	2.0	21000.0
4	7.0	Yadhnesh	50000.0	<null>	2001/04/12 00:00:00.000000000	6.0	<null>
5	6.0	Banner	12000.0	<null>	2012/03/13 00:00:00.000000000	1.0	<null>
6	5.0	Mary	18000.0	3000	1999/11/01 00:00:00.000000000	5.0	21000.0
7	4.0	Harry	15000.0	3000	2005/11/12 00:00:00.000000000	6.0	18000.0
8	3.0	LEX	30000.0	8000	1985/02/25 00:00:00.000000000	12.0	38000.0
9	2.0	Alex	25000.0	5000	1996/01/12 00:00:00.000000000	5.0	30000.0
10	1.0	Atharva	80000.0	50000	2000/12/02 00:00:00.000000000	5.0	130000.0

4) Concatenation Operation





Examine preview data

Rows of step: Concat fields (10 rows)

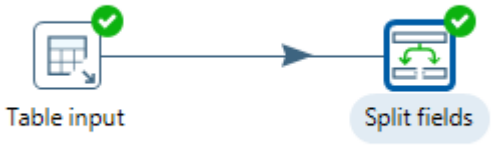
#	ENO	ENAME	SALARY	COMMISSION	HIREDATE	DPTNO	merge
1	10.0	Xavier	39000.0	10000	1990/07/30 00:00:00.000000000	9.0	Xavier--9.0
2	9.0	MAX	35000.0	9000	1986/05/05 00:00:00.000000000	10.0	MAX--10.0
3	8.0	Sameer	20000.0	1000	1996/10/18 00:00:00.000000000	2.0	Sameer--2.0
4	7.0	Yadhnesh	50000.0	<null>	2001/04/12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

Examine preview data

Rows of step: Table output (10 rows)

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

5) Split Operation – splitting the merge fields in 2 rows



Split fields

Step name: Split fields
Field to split: MERGE
Delimiter: --
Enclosure:

Fields

#	New field	ID	Remove ID?	Type	Length	Precision	Format	Group	Decimal	Currency	Nullif	Default	Trim type
1	new1			String	10								
2	new2			Number	10								

Help | OK | Cancel

Examine preview data

Rows of step: Split fields (10 rows)

#	ENO	ENAME	SALARY	COMMISSION	HIREDATE	DPTNO	new1	new2
1	10.0	Xavier	39000.0	10000	1990/07/30 00:00:00.000000000	9.0	Xavier	0000000009.00
2	9.0	MAX	35000.0	9000	1986/05/05 00:00:00.000000000	10.0	MAX	0000000010.00
3	8.0	Sameer	20000.0	1000	1996/10/18 00:00:00.000000000	2.0	Sameer	0000000002.00
4	7.0	Yadhnesh	50000.0	<null>	2001/04/12 00:00:00.000000000	6.0	Yadhnesh	0000000006.00
5	6.0	Banner	12000.0	<null>	2012/03/13 00:00:00.000000000	1.0	Banner	0000000001.00
6	5.0	Mary	18000.0	3000	1999/11/01 00:00:00.000000000	5.0	Mary	0000000005.00
7	4.0	Harry	15000.0	3000	2005/11/12 00:00:00.000000000	6.0	Harry	0000000006.00
8	3.0	LEX	30000.0	8000	1985/02/25 00:00:00.000000000	12.0	LEX	0000000012.00
9	2.0	Alex	25000.0	5000	1996/01/12 00:00:00.000000000	5.0	Alex	0000000005.00
10	1.0	Atharva	80000.0	50000	2000/12/02 00:00:00.000000000	5.0	Atharva	0000000005.00