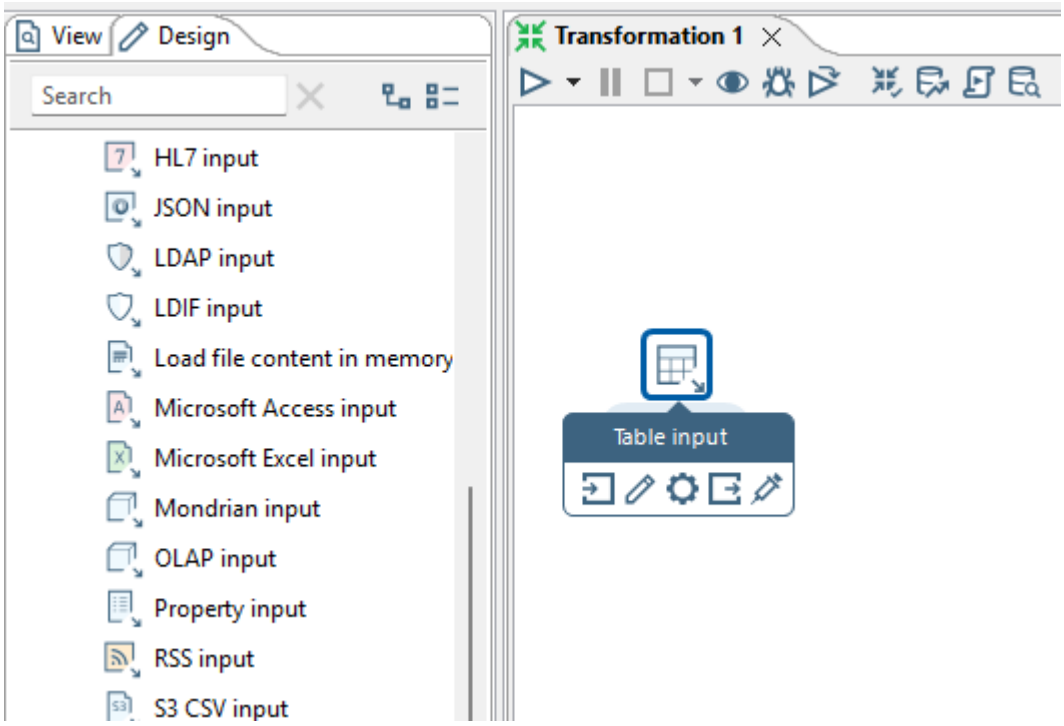
Assignment No	13
Title	Pentaho Operations
Objective	1. Concatenation Operations 2. Splitting Operations 3. Number Range Operations 4. String Operations 5. Importing CSV File 6. Merge Joint 7. Data Validation
Roll No	MCA2516

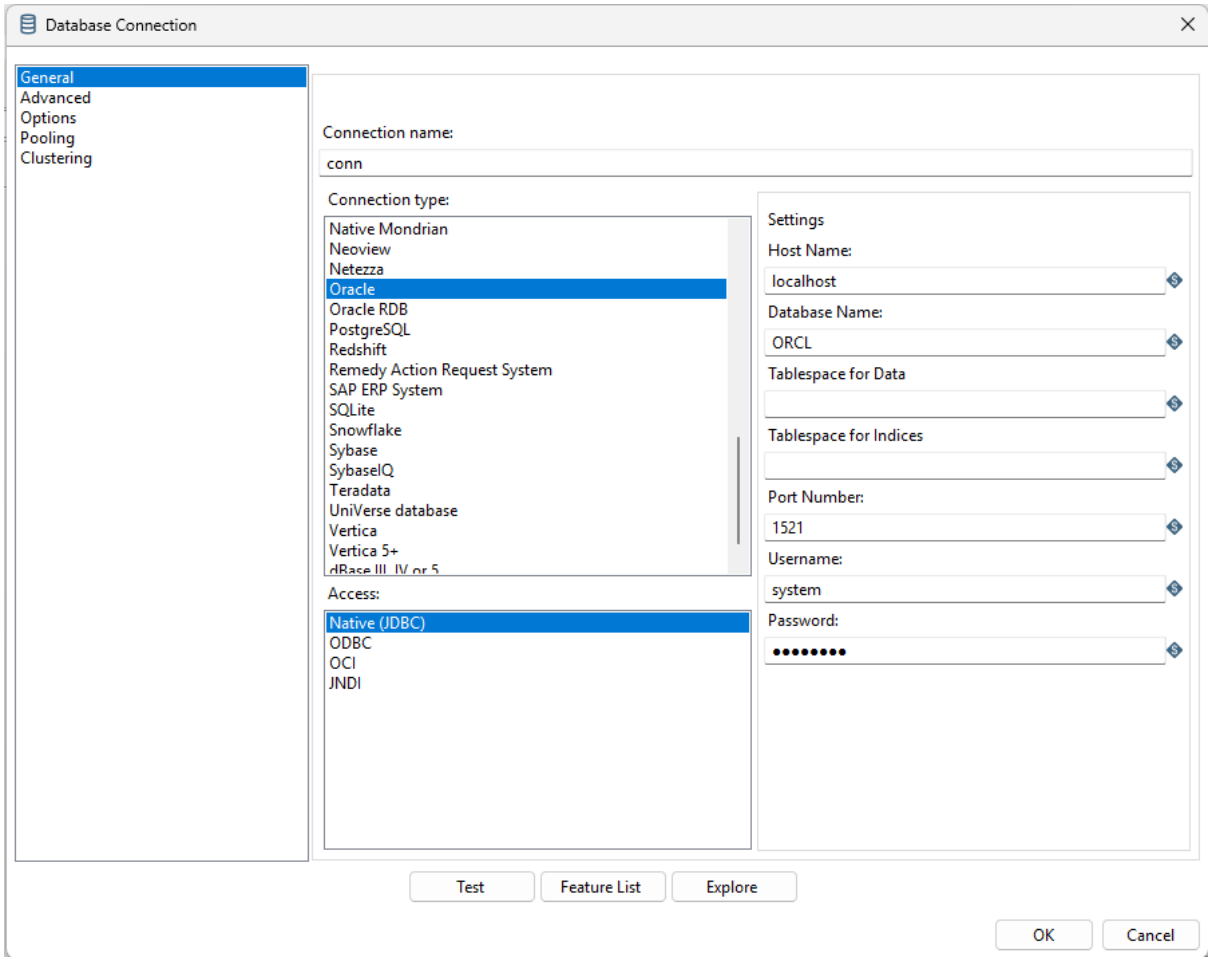
Program 1: Concatenation Operations

Step 1: File -> New -> Transformation

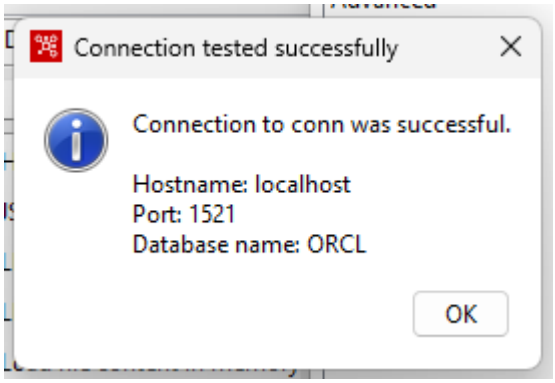
Step 2: Design -> Input -> Table Input



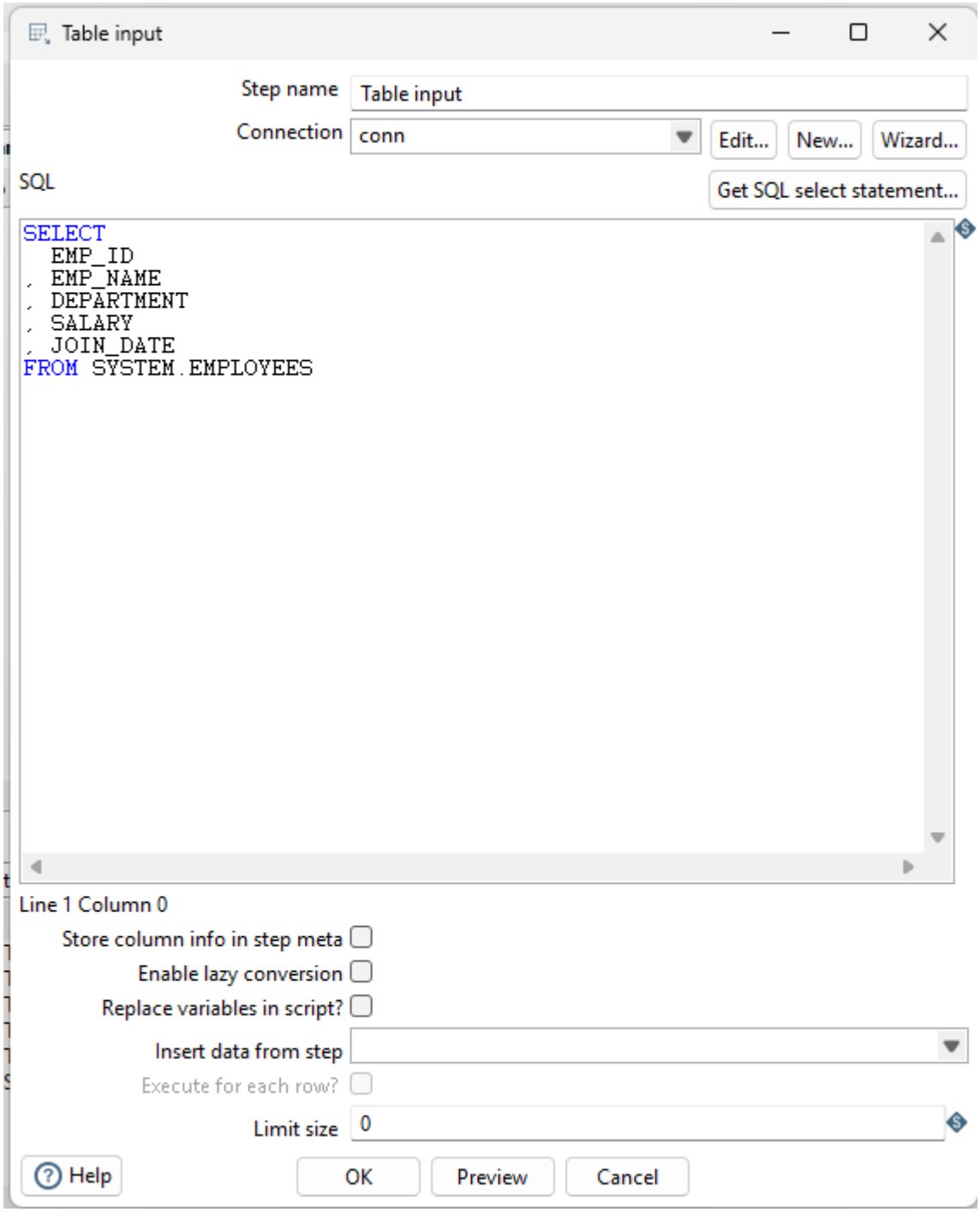
Step 3 : Database Connection configuration



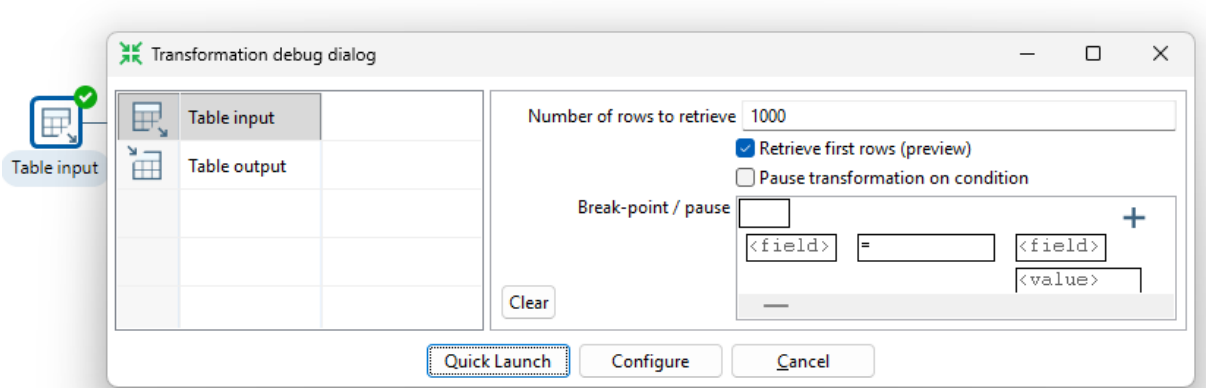
Step 4: Test for connection



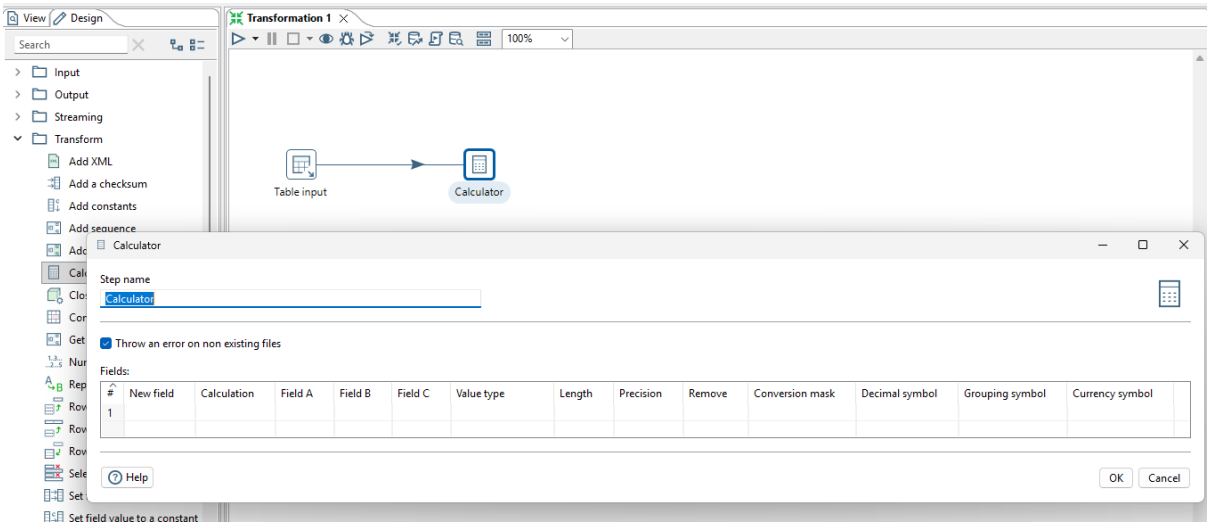
Step 5: Get SQL Statement

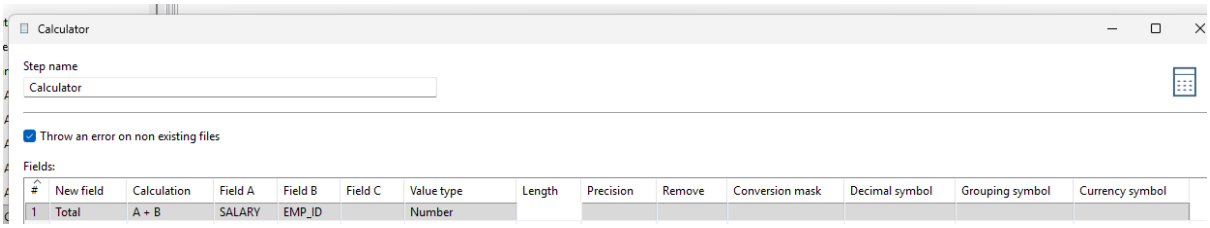


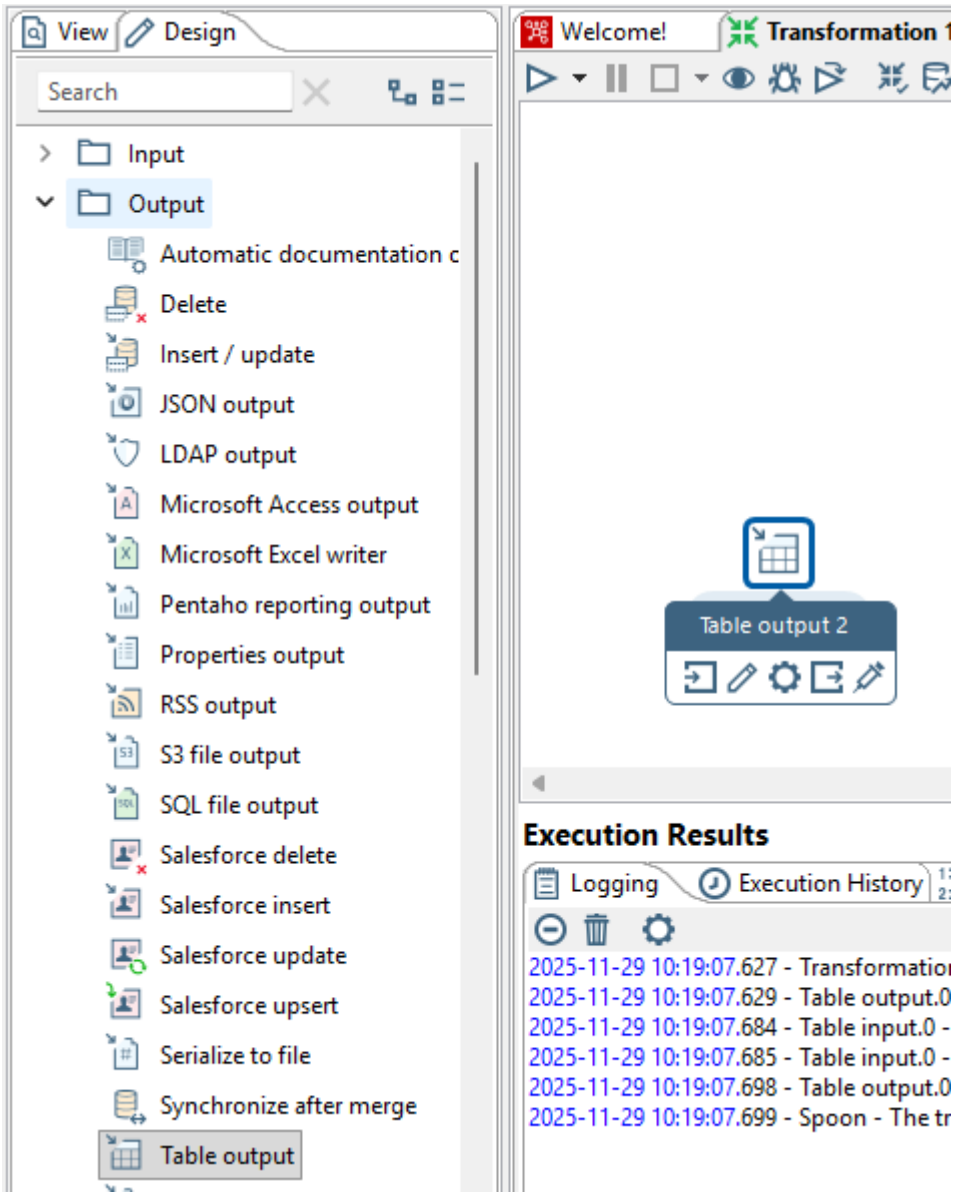
Step 6: Right Click -> Preview -> Quick Launch



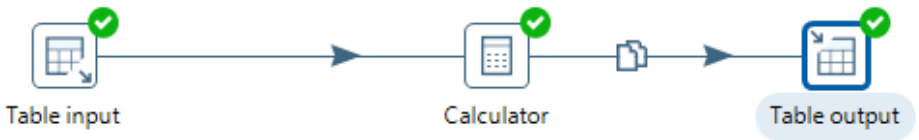
Step 7 : Design -> Transform -> Calculator



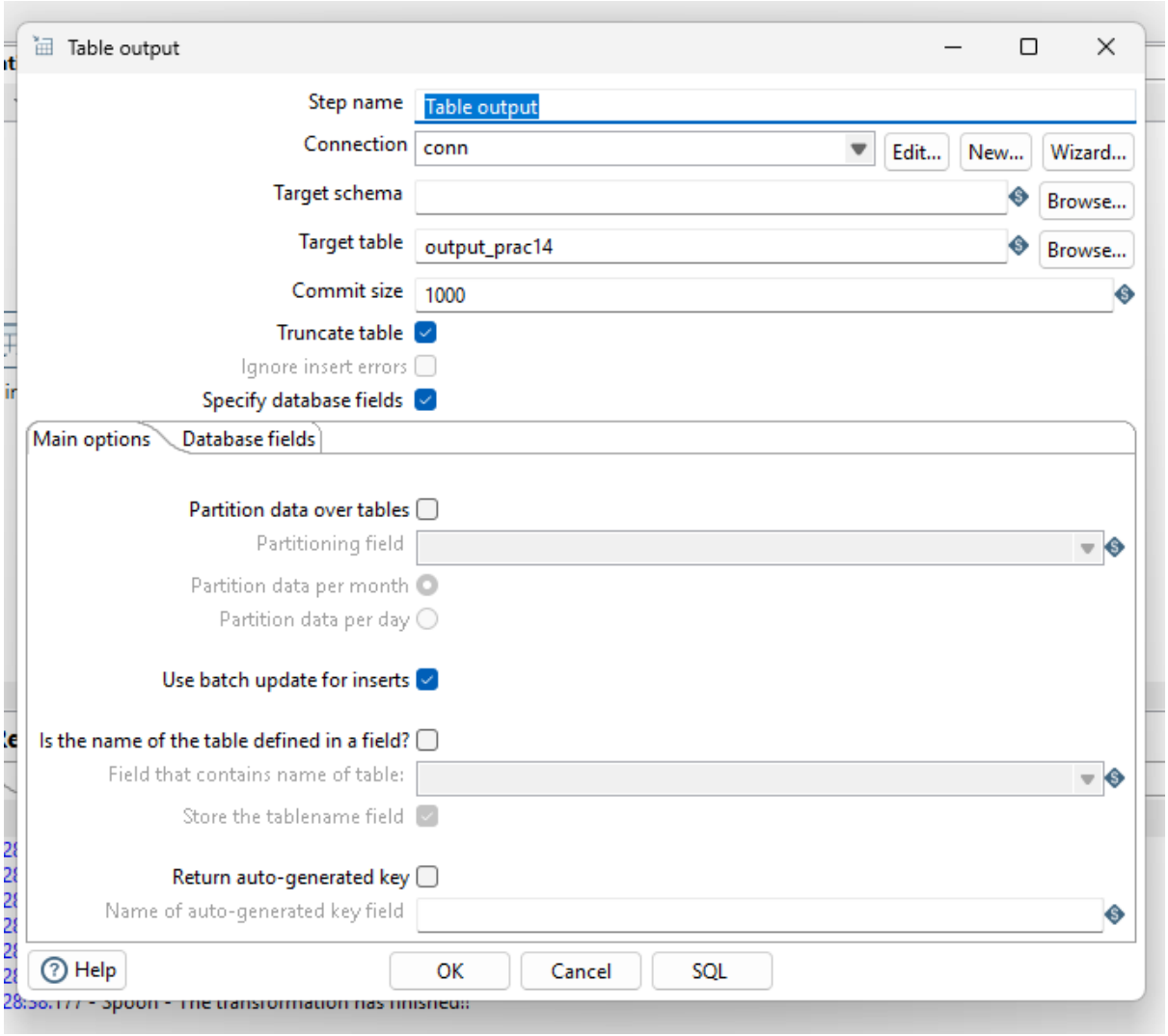




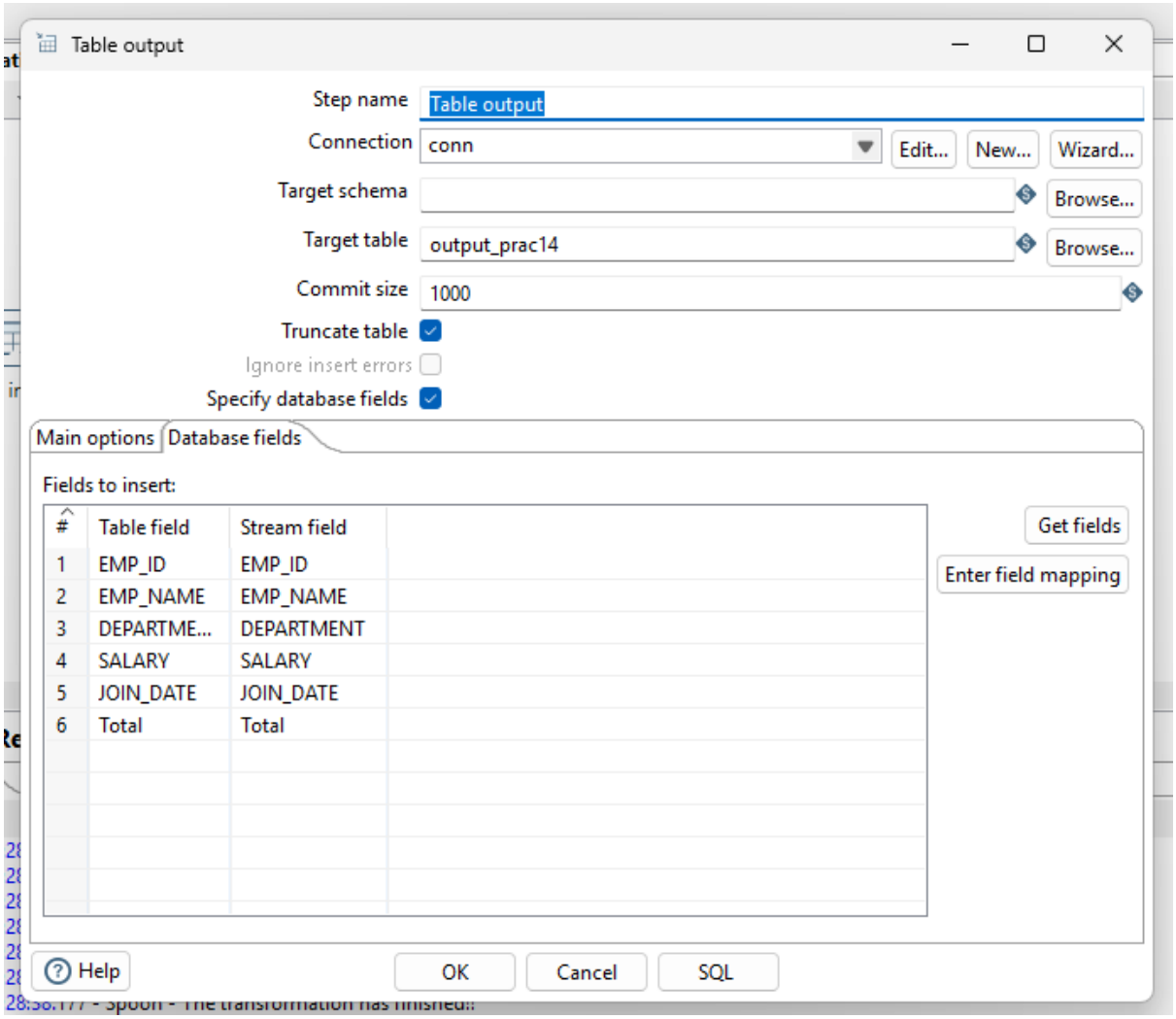
Step 9: Connect with table output



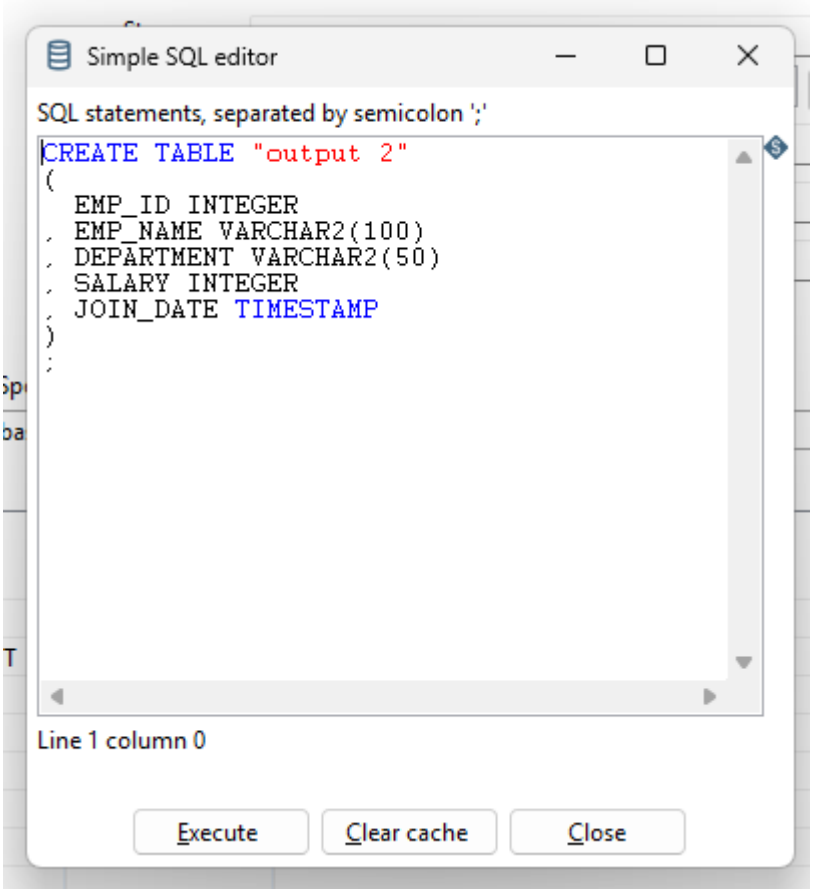
Step 10: Double Click Table output



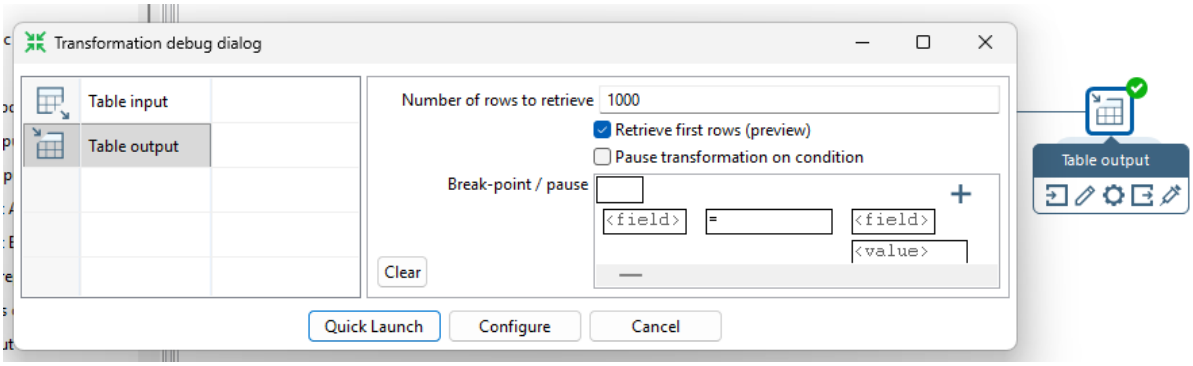
Step 11: Data field base -> get fields



Step 12: Click on SQL and execute



Step 14:



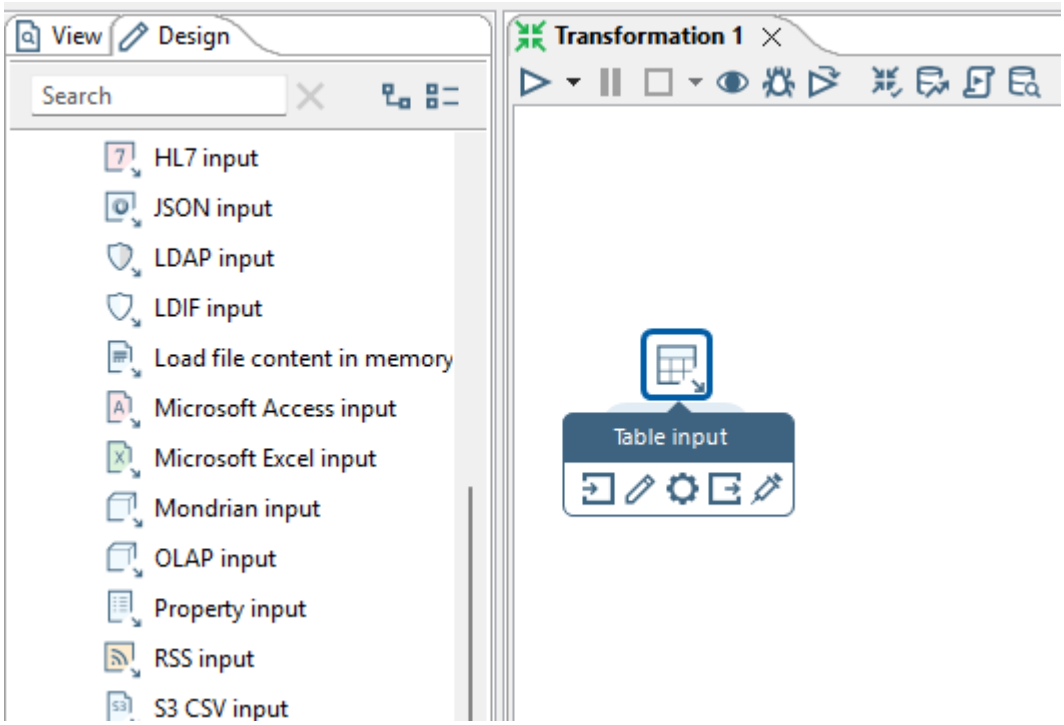
SQL Plus Output

```
SQL> select * from output_prac14;

    EMP_ID
----------
EMP_NAME
--------------------------------------------------------------------------------
DEPARTMENT                                           SALARY
-------------------------------------------------- ----------
JOIN_DATE
---------------------------------------------------------------------------
     TOTAL
----------
         1
Amit Sharma
IT                                                    55000

    EMP_ID
----------
EMP_NAME
--------------------------------------------------------------------------------
DEPARTMENT                                           SALARY
-------------------------------------------------- ----------
JOIN_DATE
---------------------------------------------------------------------------
     TOTAL
----------
10-JAN-23 12.00.00.000000 AM
     55001
```

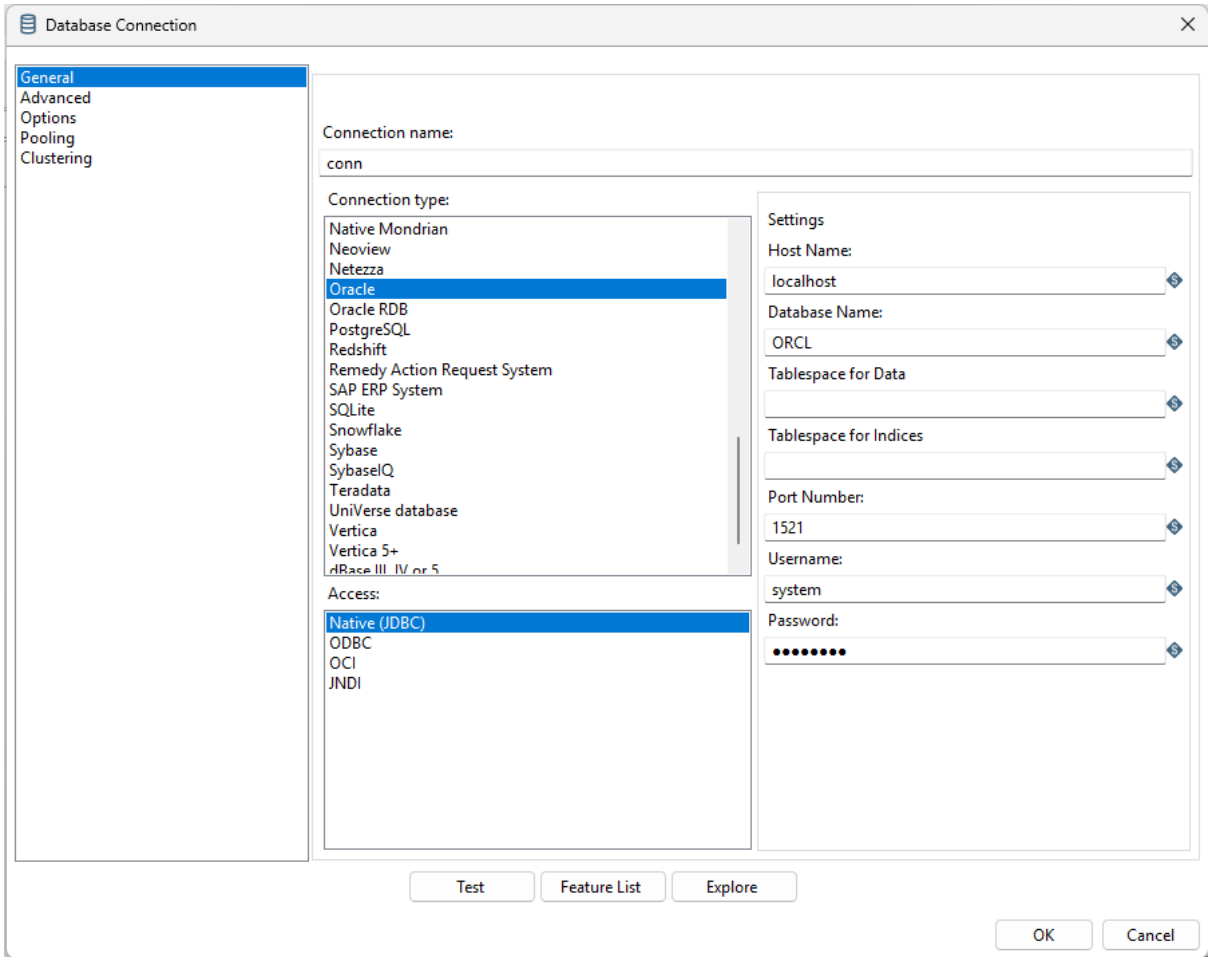
Program 1: Concatenation Operations

Step 1: File -> New -> Transformation

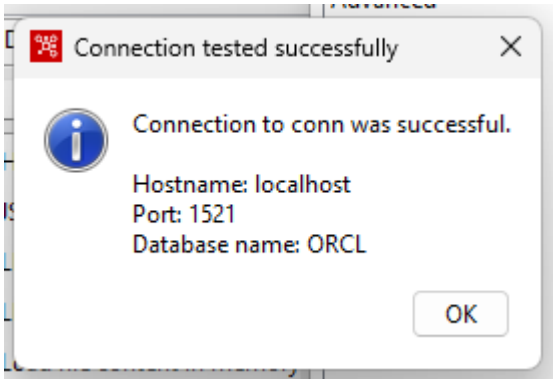
Step 2: Design -> Input -> Table Input



Step 3 : Database Connection configuration



Step 4: Test for connection



Step 5: Get SQL Statement

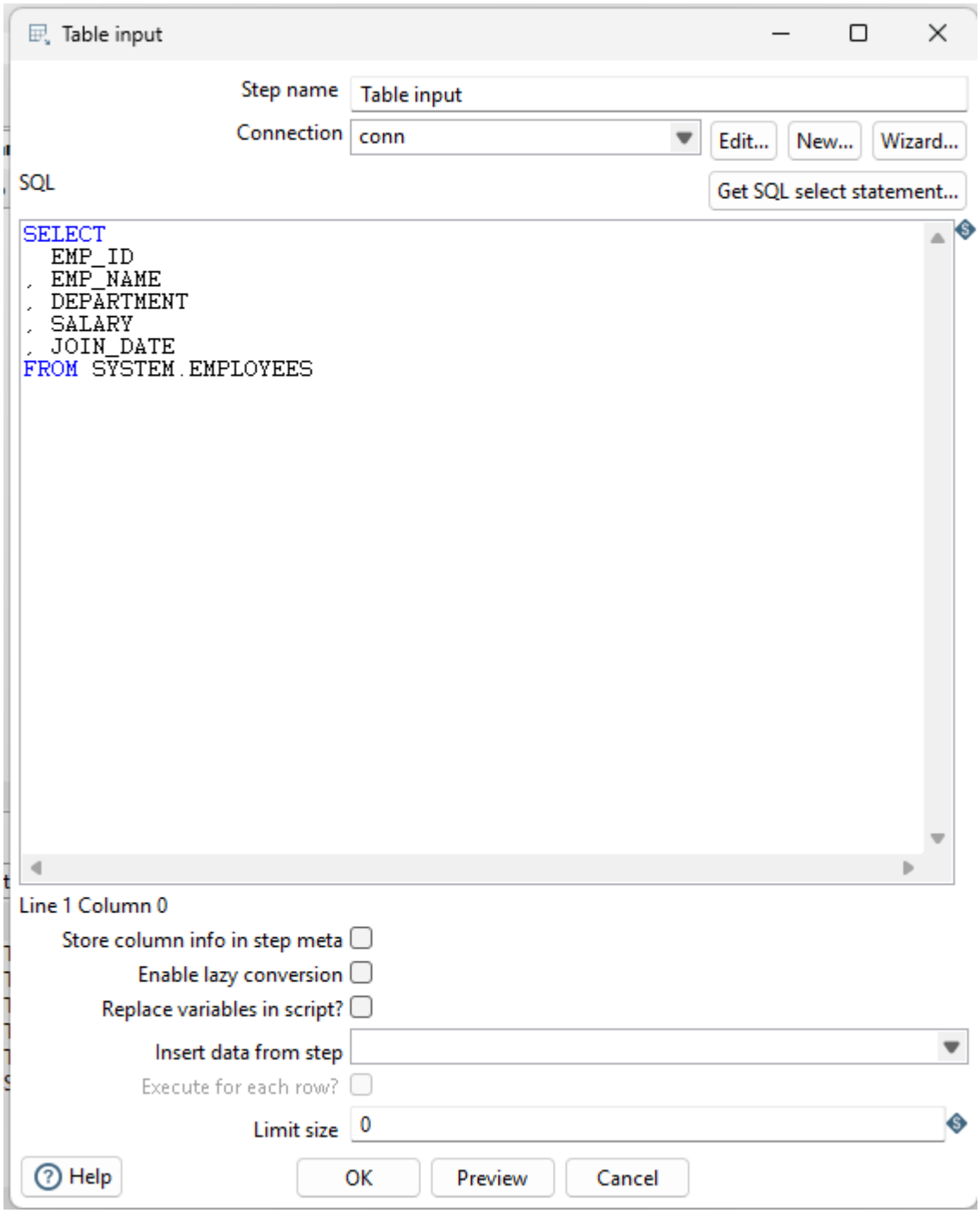
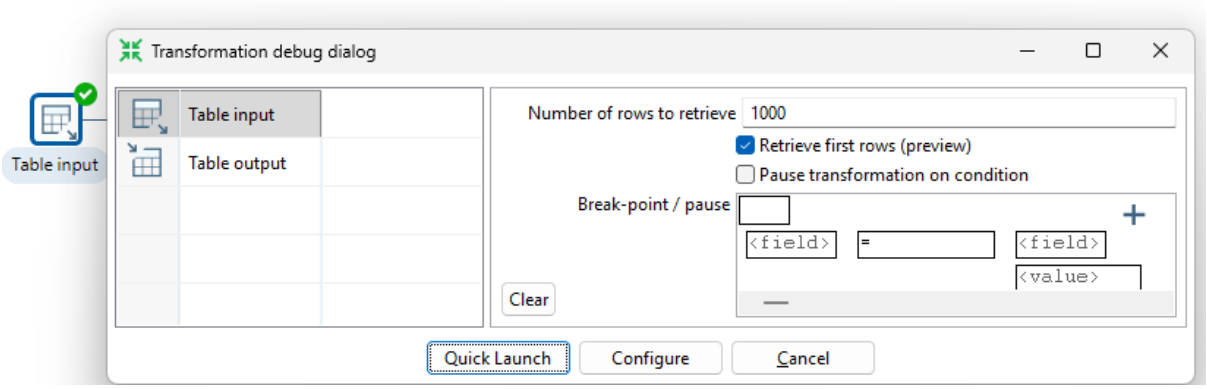
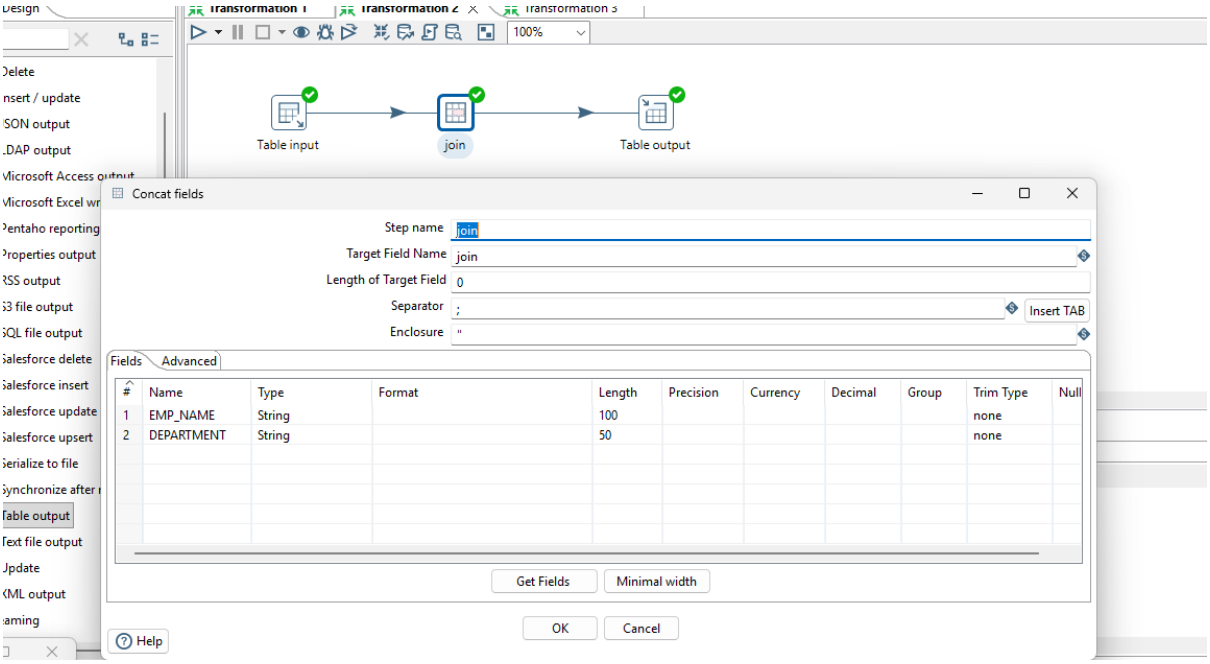


Table input
Step name
Table input
Connection
conn
Edit...
New...
Wizard...
SQL
Get SQL select statement...
SELECT
EMP_ID
, EMP_NAME
, DEPARTMENT
, SALARY
, JOIN_DATE
FROM SYSTEM.EMPLOYEES
Line 1 Column 0
Store column info in step meta
Enable lazy conversion
Replace variables in script?
Insert data from step
Execute for each row?
Limit size
0
Help
OK
Preview
Cancel

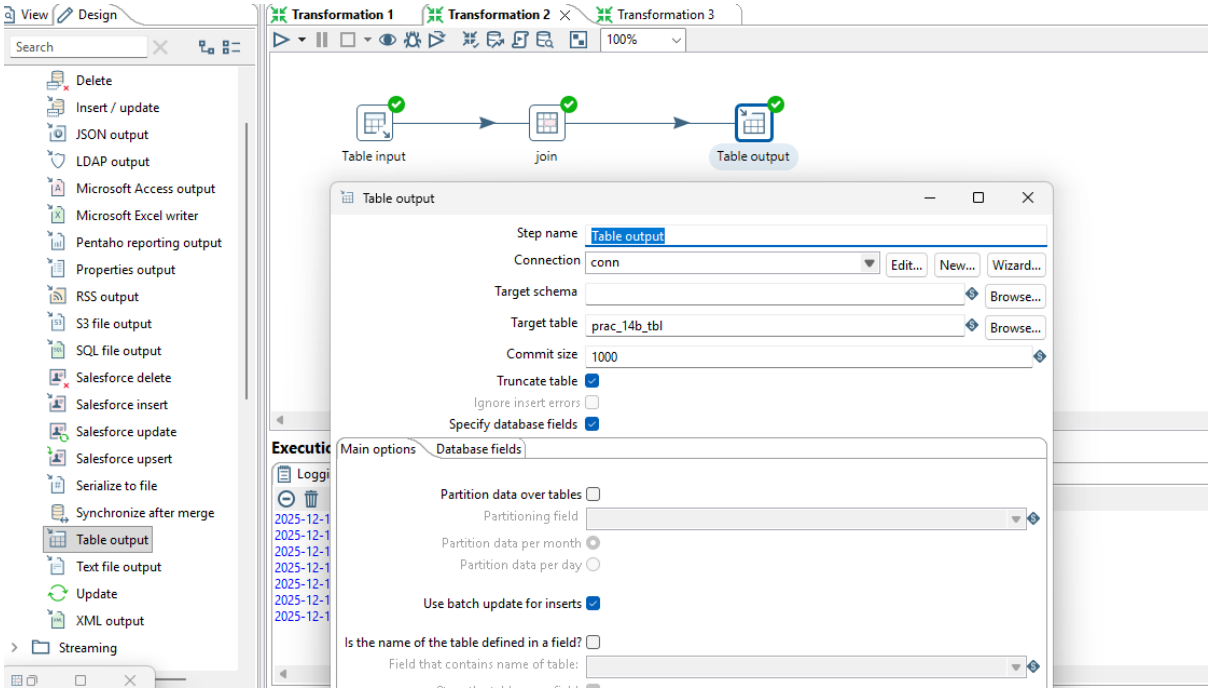
Step 6: Right Click -> Preview -> Quick Launch

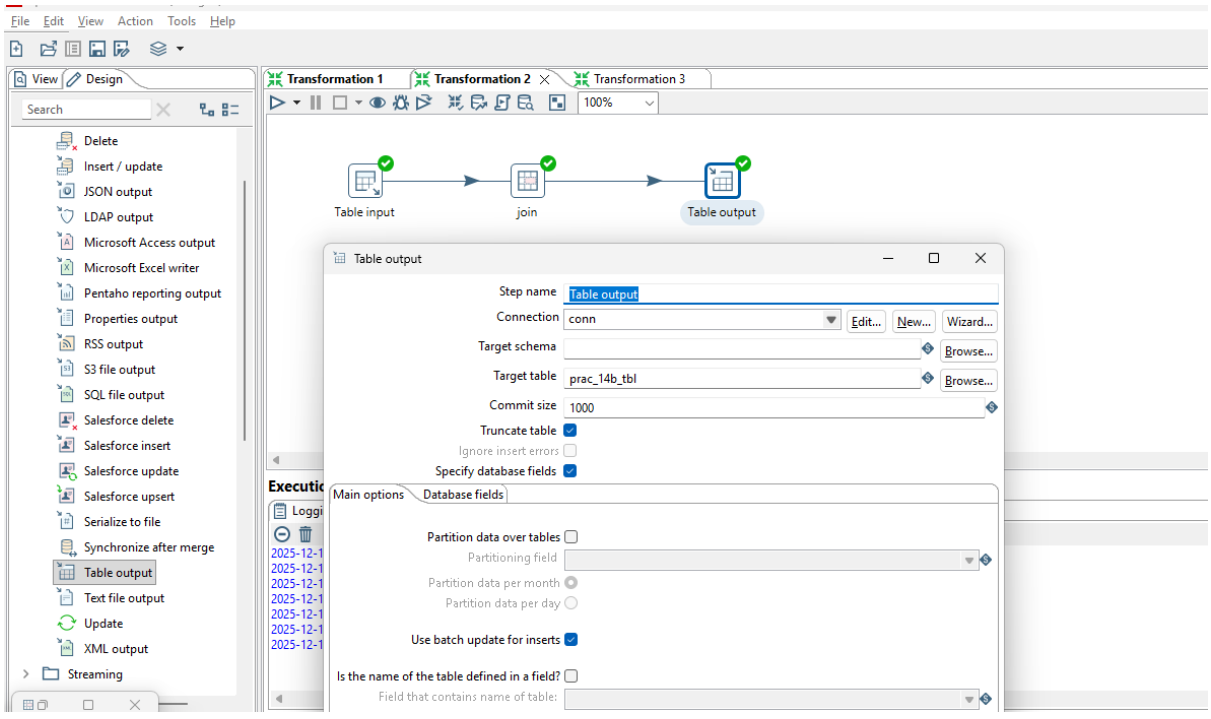


Step 7: Design -> Tranformation ->concat string

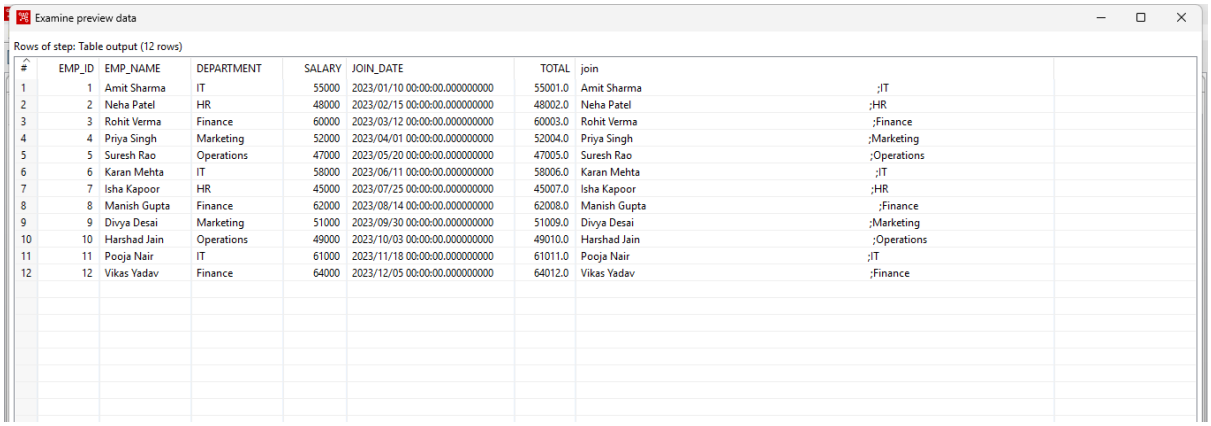


Step 9: Connect the table input with table output


View
Design
Search
Delete
Insert / update
JSON output
LDAP output
Microsoft Access output
Microsoft Excel writer
Pentaho reporting output
Properties output
RSS output
S3 file output
SQL file output
Salesforce delete
Salesforce insert
Salesforce update
Salesforce upsert
Serialize to file
Synchronize after merge
Table output
Text file output
Update
XML output
Streaming
Transformation 1
Transformation 2
Transformation 3
100%
Table input
join
Table output
Table output
Step name
Table output
Connection
conn
Edit...
New...
Wizard...
Target schema
Browse...
Target table
prac_14b_tbl
Browse...
Commit size
1000
Truncate table
Ignore insert errors
Specify database fields
Main options
Database fields
Partition data over tables
Partitioning field
Partition data per month
Partition data per day
Use batch update for inserts
Is the name of the table defined in a field?
Field that contains name of table:


File
Edit
View
Action
Tools
Help
View
Design
Search
Delete
Insert / update
JSON output
LDAP output
Microsoft Access output
Microsoft Excel writer
Pentaho reporting output
Properties output
RSS output
S3 file output
SQL file output
Salesforce delete
Salesforce insert
Salesforce update
Salesforce upsert
Serialize to file
Synchronize after merge
Table output
Text file output
Update
XML output
Streaming
Transformation 1
Transformation 2
Transformation 3
100%
Table input
join
Table output
Table output
Step name
Table output
Connection
conn
Edit...
New...
Wizard...
Target schema
Browse...
Target table
prac_14b_tbl
Browse...
Commit size
1000
Truncate table
Ignore insert errors
Specify database fields
Main options
Database fields
Partition data over tables
Partitioning field
Partition data per month
Partition data per day
Use batch update for inserts
Is the name of the table defined in a field?
Field that contains name of table:

Output:



Examine preview data

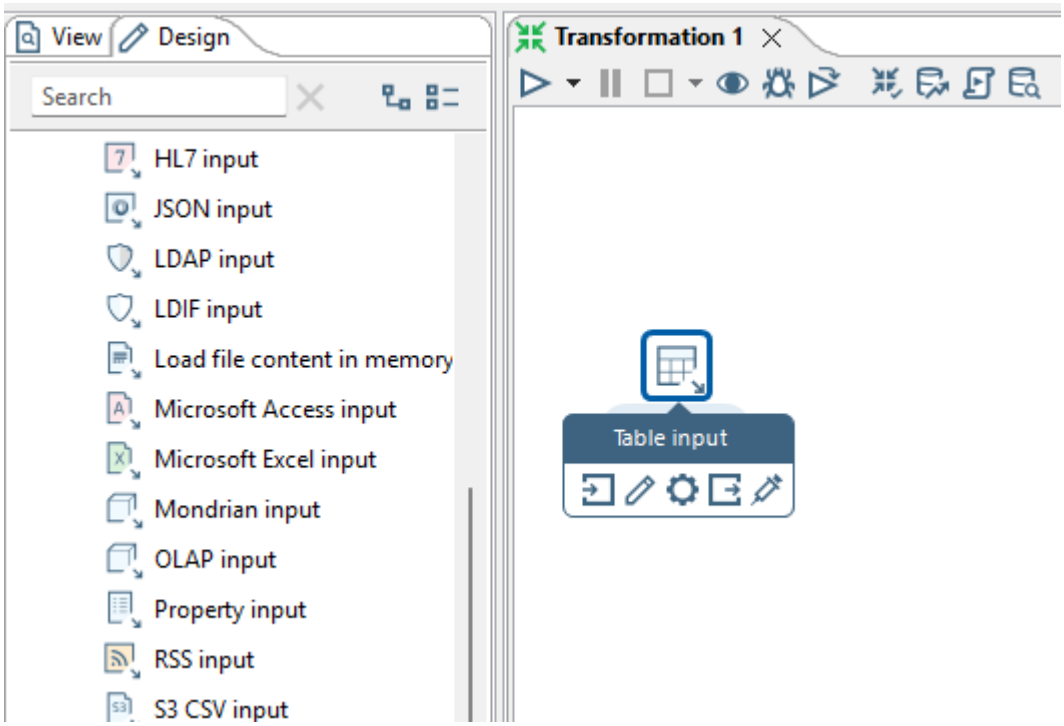
Rows of step: Table output (12 rows)

#	EMP_ID	EMP_NAME	DEPARTMENT	SALARY	JOIN_DATE	TOTAL	join
1	1	Amit Sharma	IT	55000	2023/01/10 00:00:00.000000000	55001.0	Amit Sharma ;IT
2	2	Neha Patel	HR	48000	2023/02/15 00:00:00.000000000	48002.0	Neha Patel ;HR
3	3	Rohit Verma	Finance	60000	2023/03/12 00:00:00.000000000	60003.0	Rohit Verma ;Finance
4	4	Priya Singh	Marketing	52000	2023/04/01 00:00:00.000000000	52004.0	Priya Singh ;Marketing
5	5	Suresh Rao	Operations	47000	2023/05/20 00:00:00.000000000	47005.0	Suresh Rao ;Operations
6	6	Karan Mehta	IT	58000	2023/06/11 00:00:00.000000000	58006.0	Karan Mehta ;IT
7	7	Isha Kapoor	HR	45000	2023/07/25 00:00:00.000000000	45007.0	Isha Kapoor ;HR
8	8	Manish Gupta	Finance	62000	2023/08/14 00:00:00.000000000	62008.0	Manish Gupta ;Finance
9	9	Divya Desai	Marketing	51000	2023/09/30 00:00:00.000000000	51009.0	Divya Desai ;Marketing
10	10	Harshad Jain	Operations	49000	2023/10/03 00:00:00.000000000	49010.0	Harshad Jain ;Operations
11	11	Pooja Nair	IT	61000	2023/11/18 00:00:00.000000000	61011.0	Pooja Nair ;IT
12	12	Vikas Yadav	Finance	64000	2023/12/05 00:00:00.000000000	64012.0	Vikas Yadav ;Finance

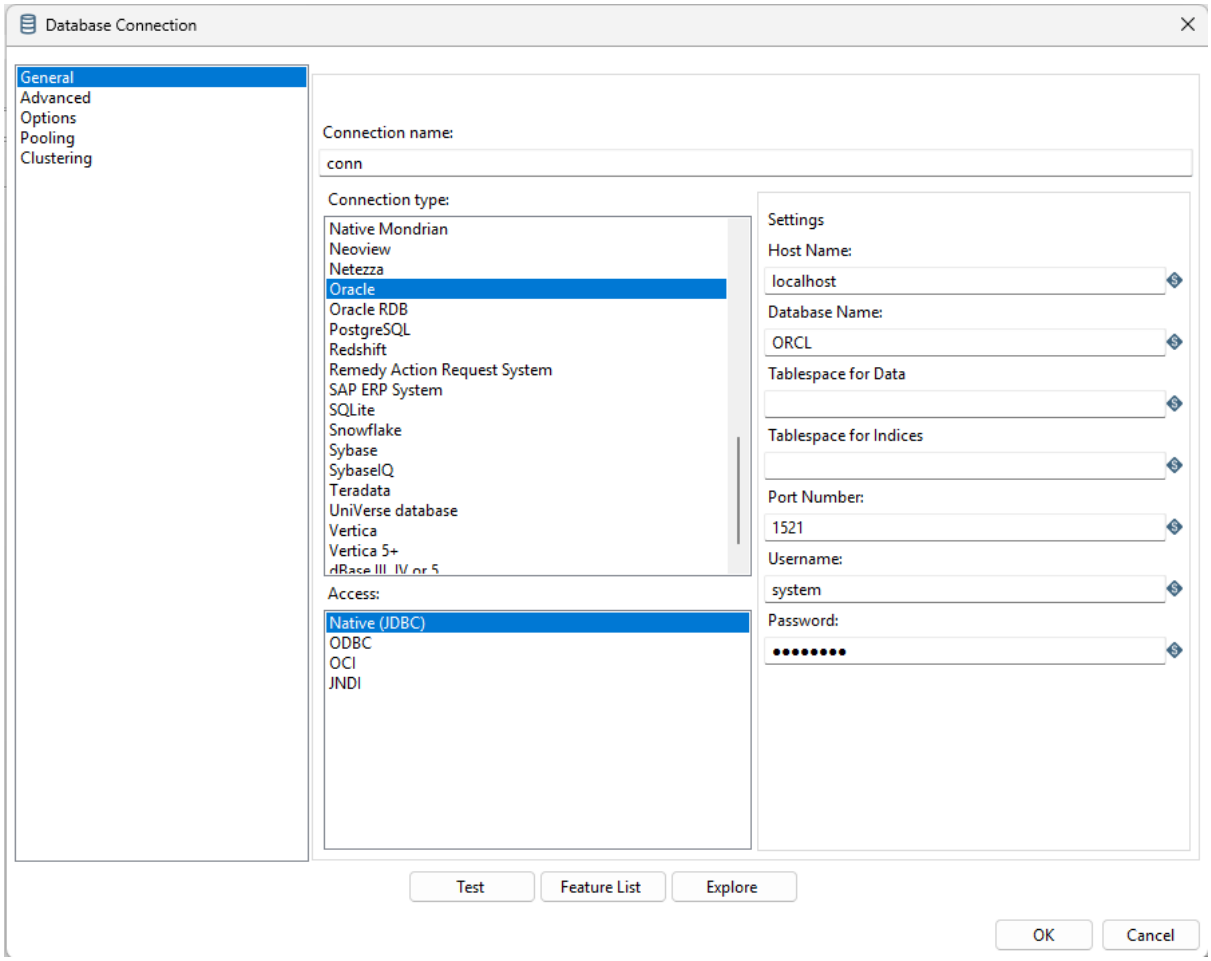
Program 2: Splitting Operations

Step 1: File -> New -> Transformation

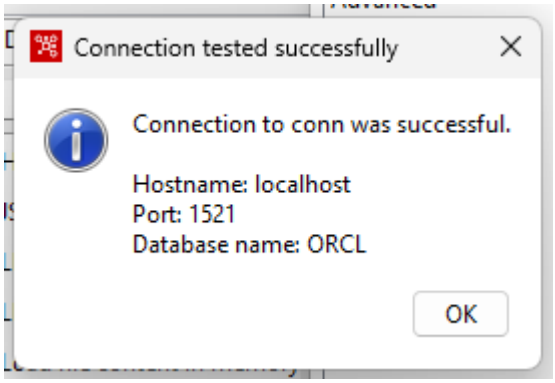
Step 2: Design -> Input -> Table Input



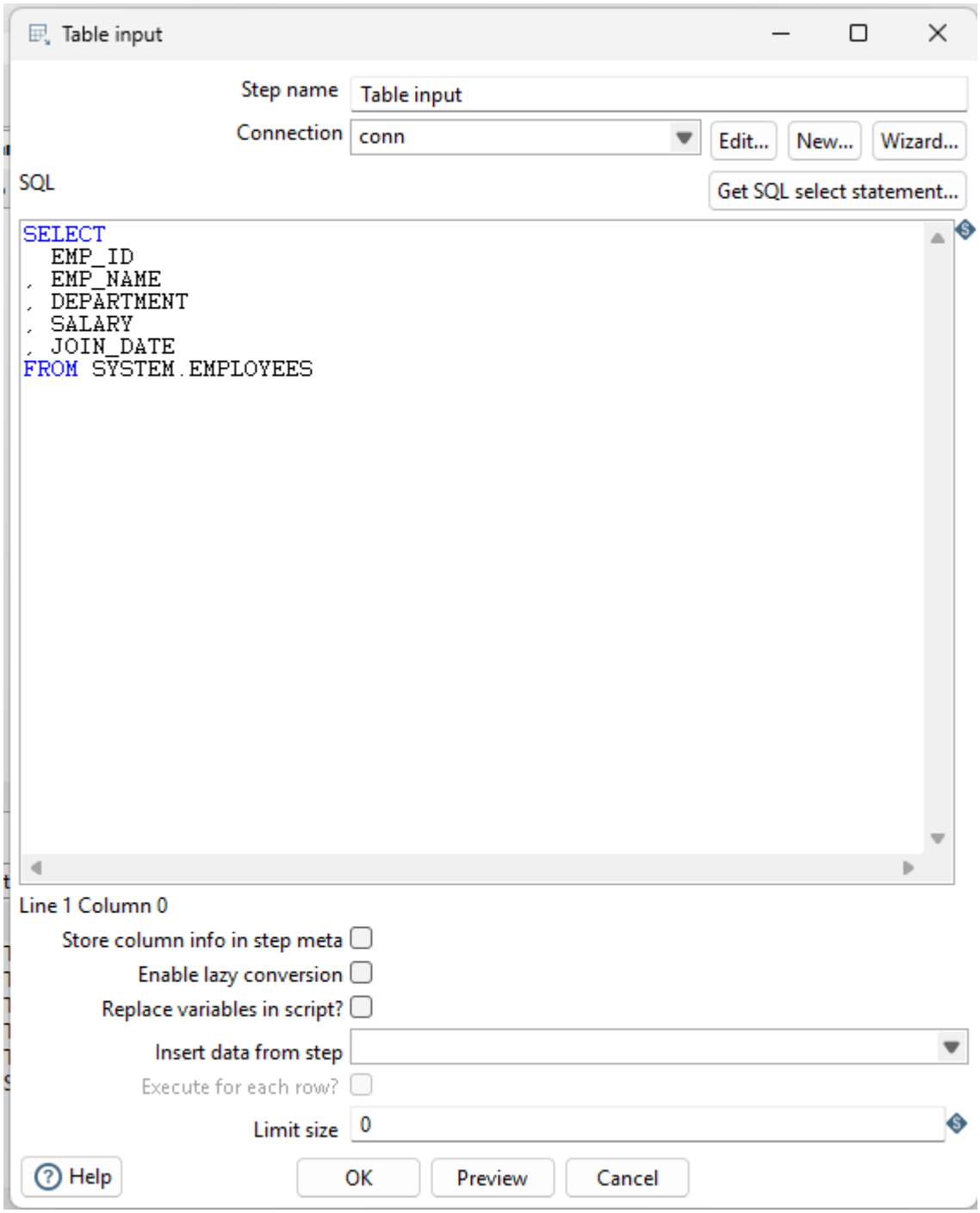
Step 3 : Database Connection configuration

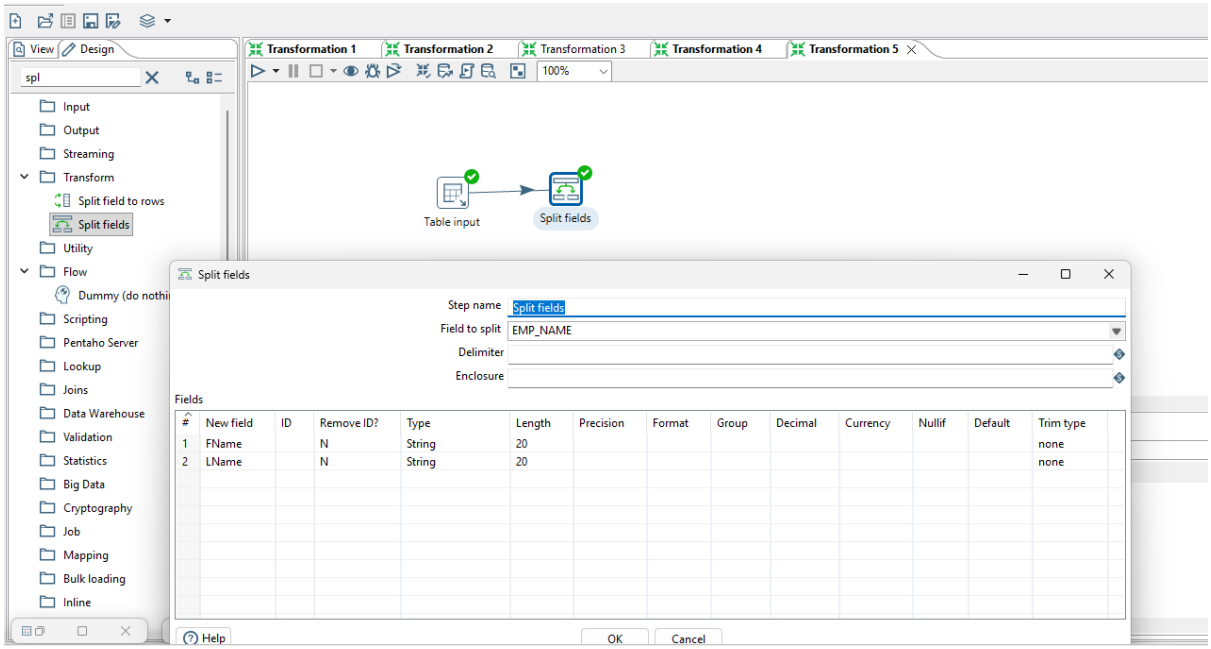


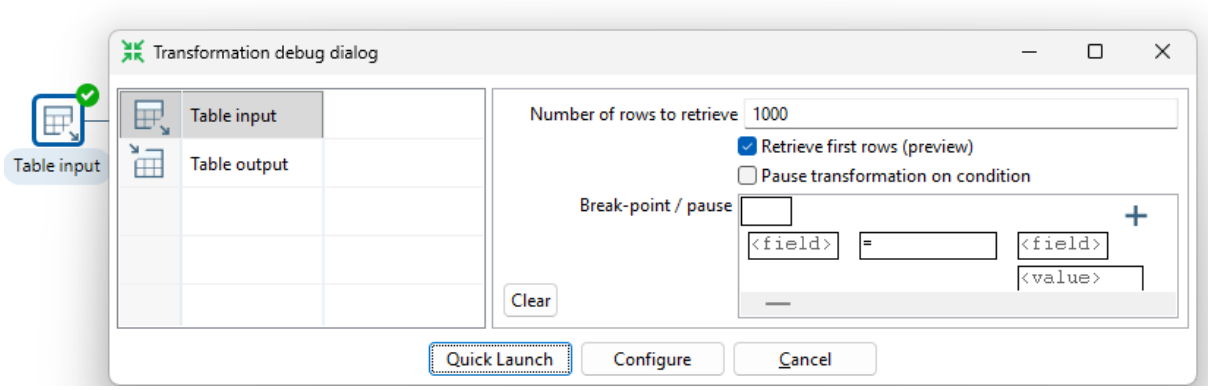
Step 4: Test for connection

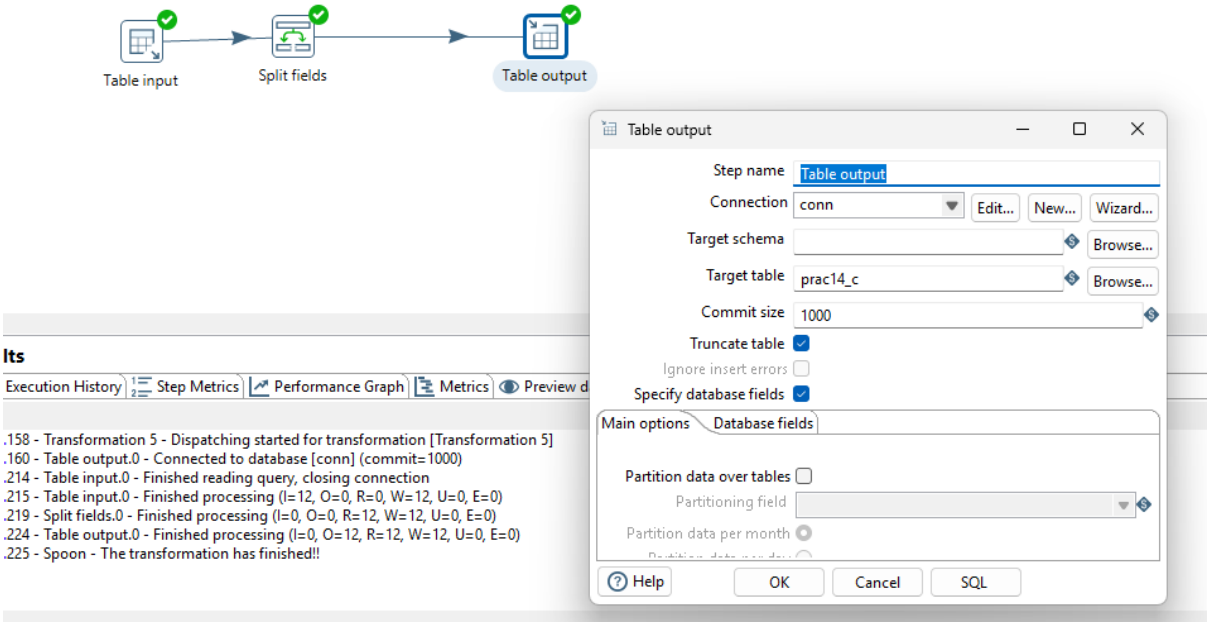


Step 5: Get SQL Statement



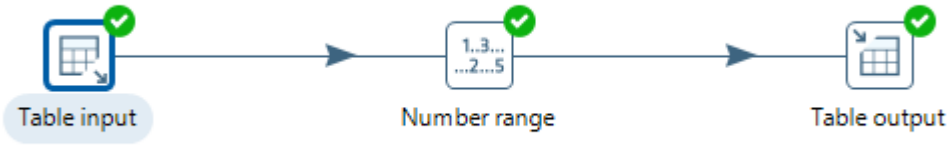

Transformation 1
Transformation 2
Transformation 3
Transformation 4
Transformation 5
Input
Output
Streaming
Transform
Split field to rows
Split fields
Utility
Flow
Scripting
Pentaho Server
Lookup
Joins
Data Warehouse
Validation
Statistics
Big Data
Cryptography
Job
Mapping
Bulk loading
Inline
Table input
Split fields
Step name
Split fields
Field to split
EMP_NAME
Delimiter
Enclosure
Fields
New field
ID
Remove ID?
Type
Length
Precision
Format
Group
Decimal
Currency
Nullif
Default
Trim type
FName
LName
String
none
Help
OK
Cancel

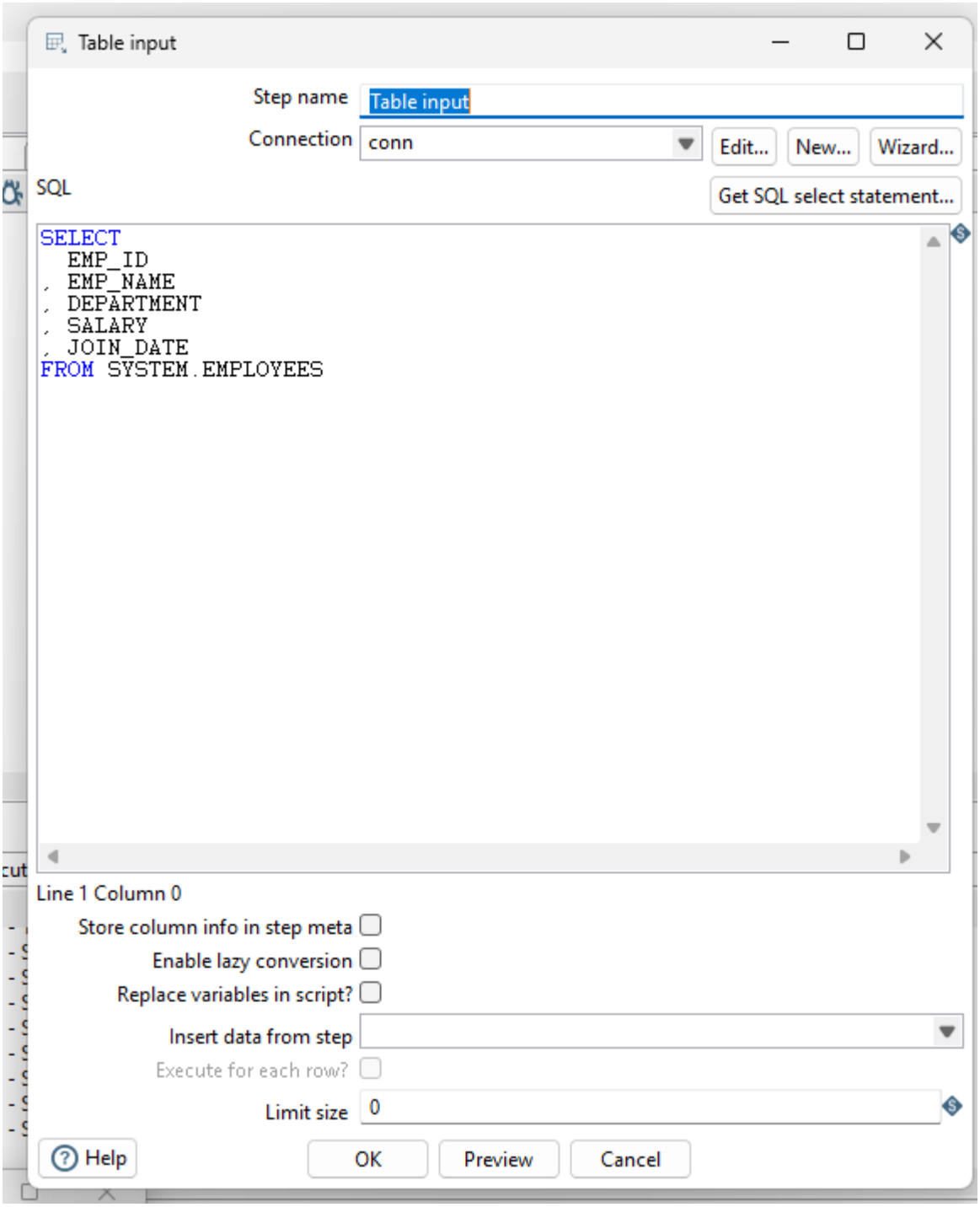

Transformation debug dialog
Table input
Table output
Number of rows to retrieve
1000
Retrieve first rows (preview)
Pause transformation on condition
Break-point / pause
<field>
=
<field>
<value>
Clear
Quick Launch
Configure
Cancel

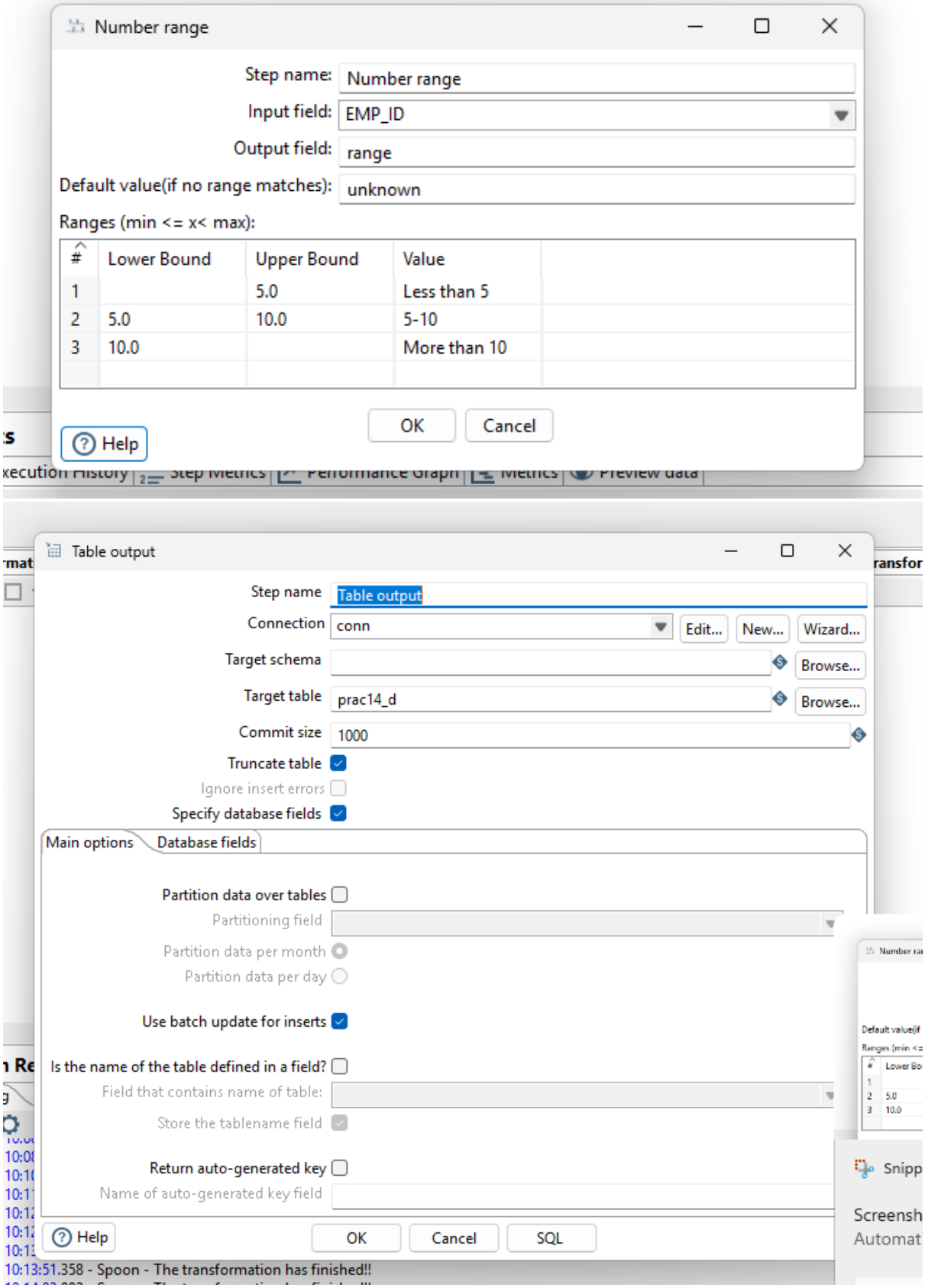



```
SQL> select * from prac14_c;

    EMP_ID FNAME                LNAME
---------- -------------------- --------------------
DEPARTMENT                                            SALARY
----------------------------------------------------- ----------
JOIN_DATE
---------------------------------------------------------------------------
         1 Amit                 Sharma
IT                                                     55000
10-JAN-23 12.00.00.000000 AM

         2 Neha                 Patel
HR                                                     48000
15-FEB-23 12.00.00.000000 AM

    EMP_ID FNAME                LNAME
---------- -------------------- --------------------
DEPARTMENT                                            SALARY
----------------------------------------------------- ----------
JOIN_DATE
---------------------------------------------------------------------------

         3 Rohit                Verma
Finance                                                60000
12-MAR-23 12.00.00.000000 AM

         4 Priya                Singh
Marketing                                              52000
```

Program 3: Number Range Operations

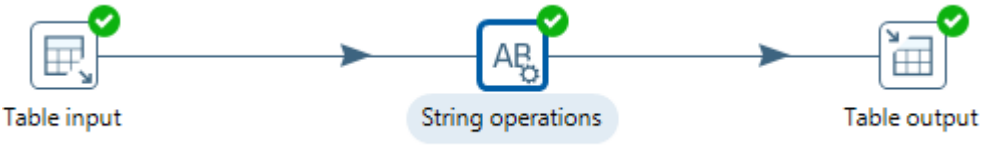


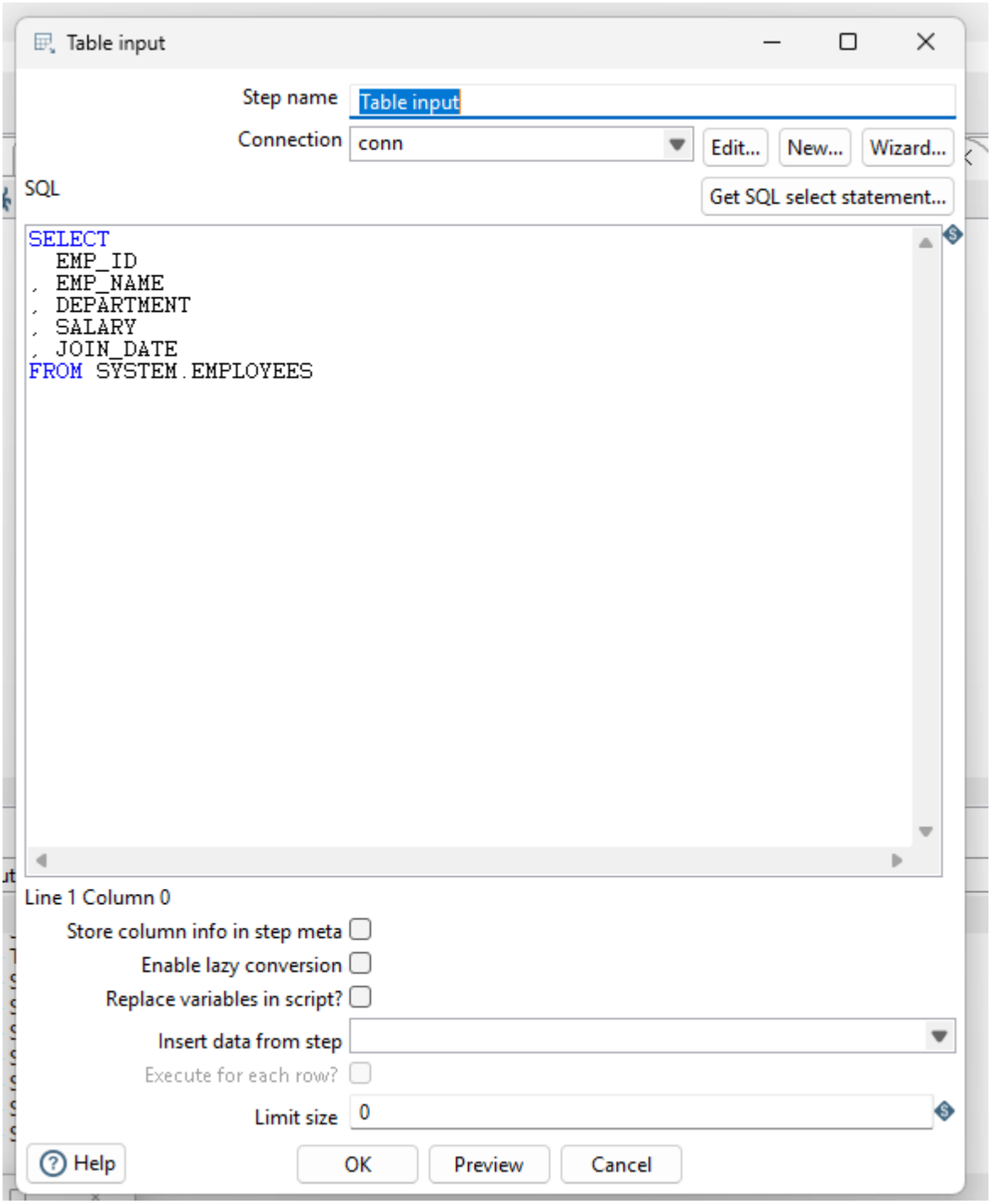


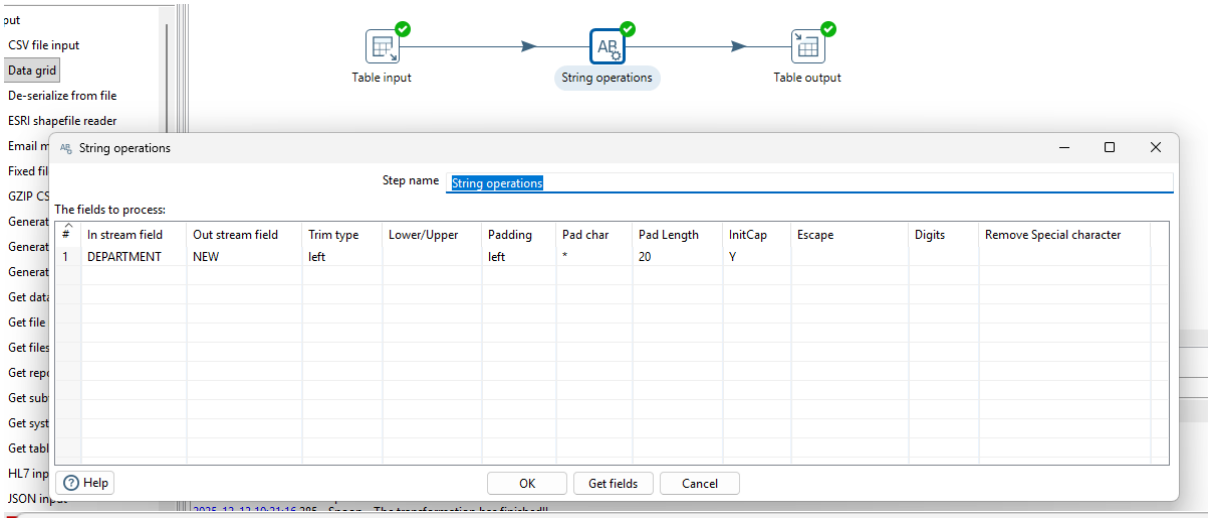

Number range
Step name: Number range
Input field: EMP_ID
Output field: range
Default value(if no range matches): unknown
Ranges (min <= x< max):
Lower Bound Upper Bound Value
1 5.0 Less than 5
2 5.0 10.0 5-10
3 10.0 More than 10
Help
OK
Cancel
Table output
Step name Table output
Connection conn
Edit...
New...
Wizard...
Target schema
Browse...
Target table prac14_d
Browse...
Commit size 1000
Truncate table
Ignore insert errors
Specify database fields
Main options
Database fields
Partition data over tables
Partitioning field
Partition data per month
Partition data per day
Use batch update for inserts
Is the name of the table defined in a field?
Field that contains name of table:
Store the tablename field
Return auto-generated key
Name of auto-generated key field
Help
OK
Cancel
SQL
10:13:51.358 - Spoon - The transformation has finished!!

```
SQL> select * from number_range;

    EMP_ID
----------
EMP_NAME
--------------------------------------------------------------------------------
DEPARTMENT                                                 SALARY
-------------------------------------------------- ----------
JOIN_DATE
---------------------------------------------------------------------------
RANGE
--------------------------------------------------------------------------------
         1
Amit Sharma
IT                                                          55000

    EMP_ID
----------
EMP_NAME
--------------------------------------------------------------------------------
DEPARTMENT                                                 SALARY
-------------------------------------------------- ----------
JOIN_DATE
---------------------------------------------------------------------------
RANGE
--------------------------------------------------------------------------------
10-JAN-23 12.00.00.000000 AM
Less than 5
```

Program 4: String Operations





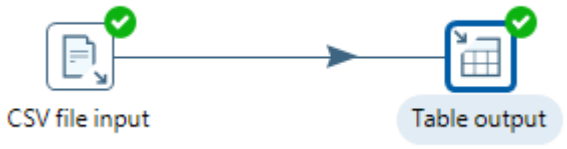


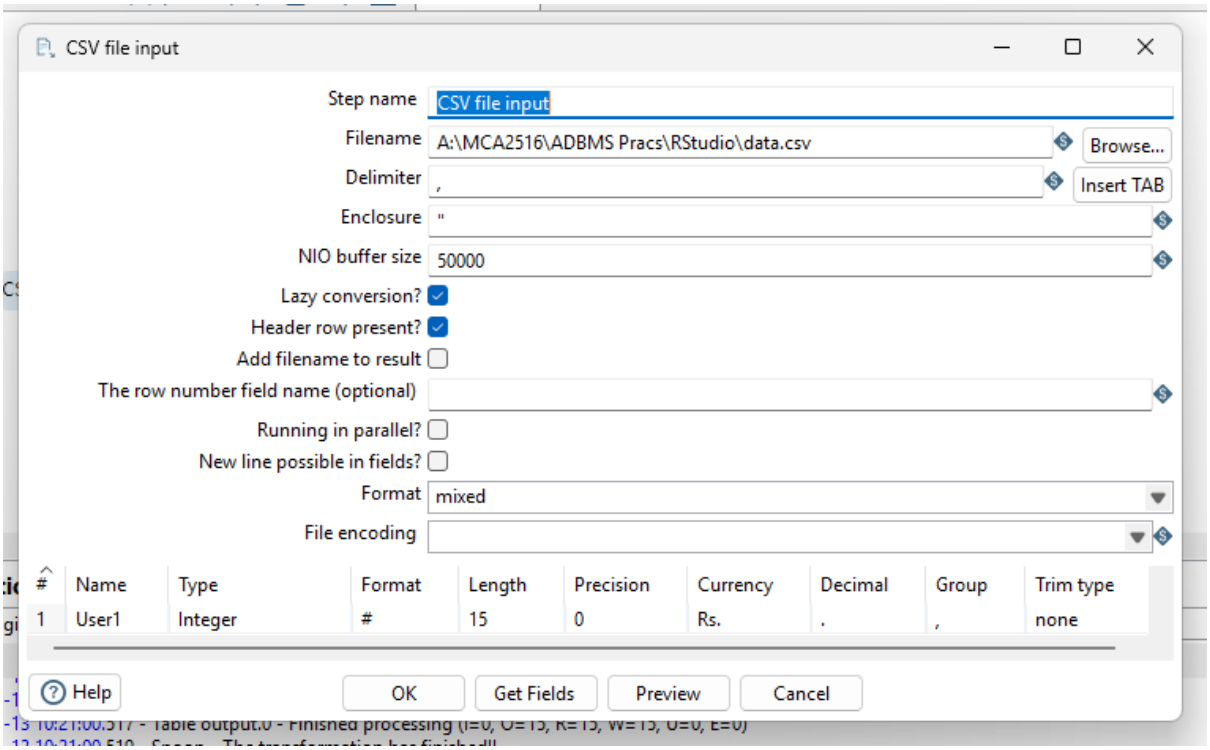
Examine preview data

Rows of step: Table output (12 rows)

#	EMP_ID	EMP_NAME	DEPARTMENT	SALARY	JOIN_DATE	NEW
1	12	Vikas Yadav	Finance	64000	2023/12/05 00:00:00.000000000	**********Finance
2	11	Pooja Nair	IT	61000	2023/11/18 00:00:00.000000000	**********It
3	10	Harshad Jain	Operations	49000	2023/10/03 00:00:00.000000000	**********Operations
4	9	Divya Desai	Marketing	51000	2023/09/30 00:00:00.000000000	**********Marketing
5	8	Manish Gupta	Finance	62000	2023/08/14 00:00:00.000000000	**********Finance
6	7	Isha Kapoor	HR	45000	2023/07/25 00:00:00.000000000	**********Hr
7	6	Karan Mehta	IT	58000	2023/06/11 00:00:00.000000000	**********It
8	5	Suresh Rao	Operations	47000	2023/05/20 00:00:00.000000000	**********Operations
9	4	Priya Singh	Marketing	52000	2023/04/01 00:00:00.000000000	**********Marketing
10	3	Rohit Verma	Finance	60000	2023/03/12 00:00:00.000000000	**********Finance
11	2	Neha Patel	HR	48000	2023/02/15 00:00:00.000000000	**********Hr
12	1	Amit Sharma	IT	55000	2023/01/10 00:00:00.000000000	**********It

Program 5: Import CSV





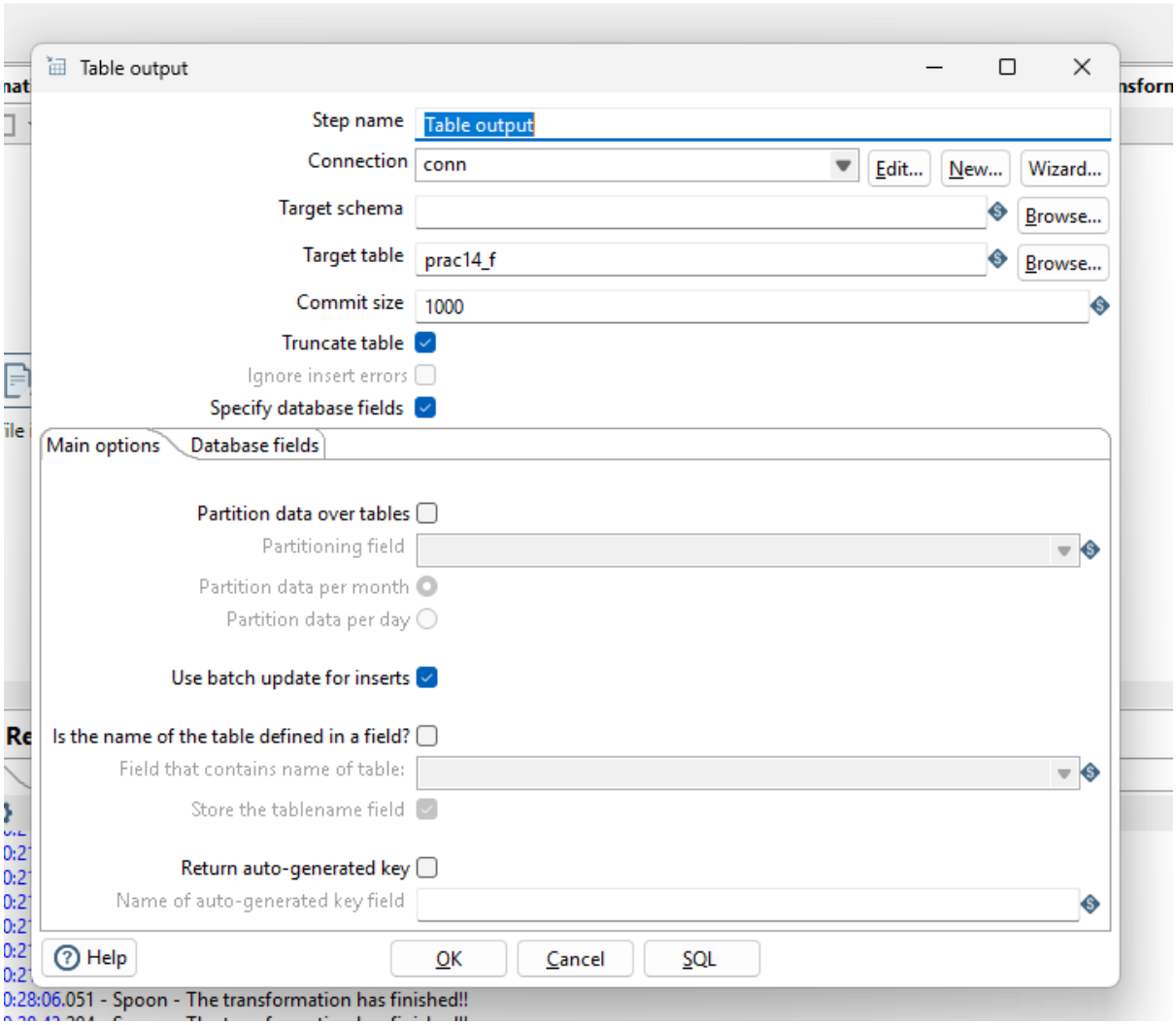


Table output
Step name
Table output
Connection
conn
Edit...
New...
Wizard...
Target schema
Browse...
Target table
prac14_f
Browse...
Commit size
1000
Truncate table
Ignore insert errors
Specify database fields
Main options
Database fields
Partition data over tables
Partitioning field
Partition data per month
Partition data per day
Use batch update for inserts
Is the name of the table defined in a field?
Field that contains name of table:
Store the tablename field
Return auto-generated key
Name of auto-generated key field
Help
OK
Cancel
SQL
0:28:06.051 - Spoon - The transformation has finished!!

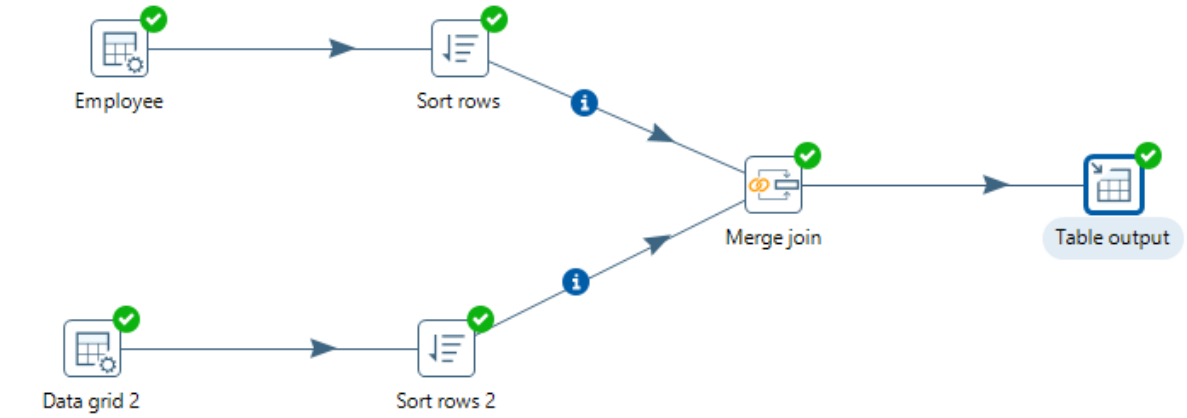
Examine preview data

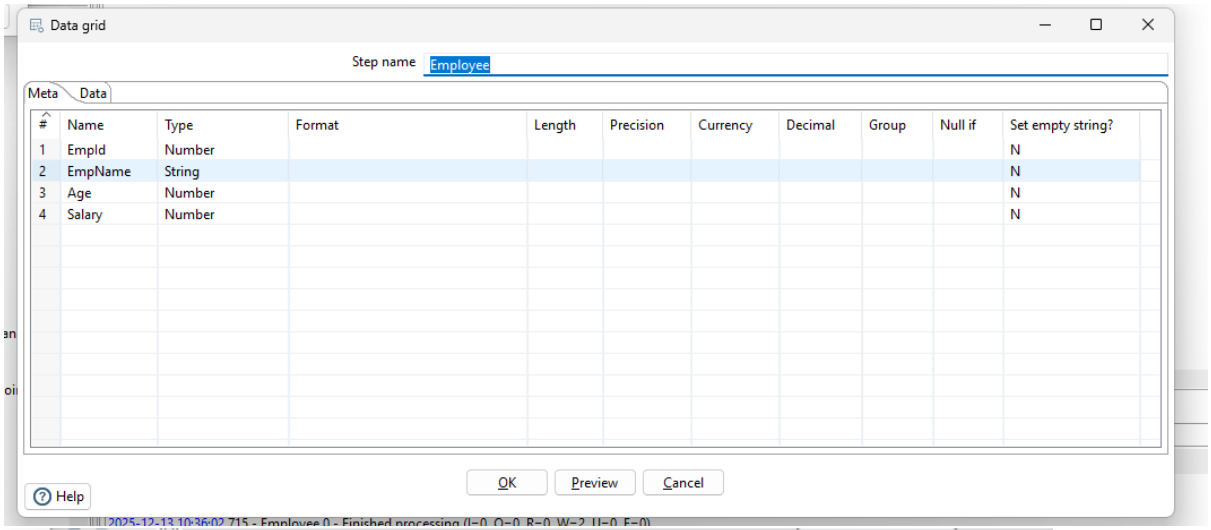
Rows of step: Table output (15 rows)

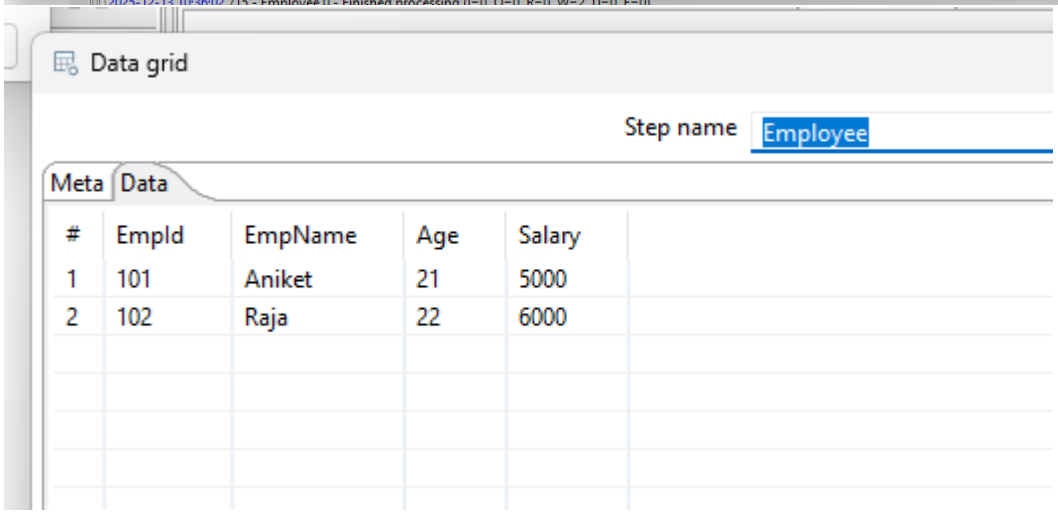
#	User1	User2	User3	User4	User5	User6	User7	User8	User9	User10
1	1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5	<null>
6	6	6	6	6	6	6	6	6	6	<null>
7	7	7	7	7	<null>	7	7	7	7	<null>
8	8	8	8	8	<null>	8	8	8	8	<null>
9	9	9	9	9	<null>	9	9	9	9	<null>
10	10	10	<null>	10	<null>	10	10	10	10	<null>
11	11	11	<null>	11	<null>	11	11	11	11	<null>
12	12	12	<null>	12	<null>	12	<null>	12	12	<null>
13	13	13	<null>	13	<null>	13	<null>	13	<null>	<null>
14	14	14	<null>	14	<null>	14	<null>	14	<null>	<null>
15	15	15	<null>	15	<null>	15	<null>	15	<null>	<null>

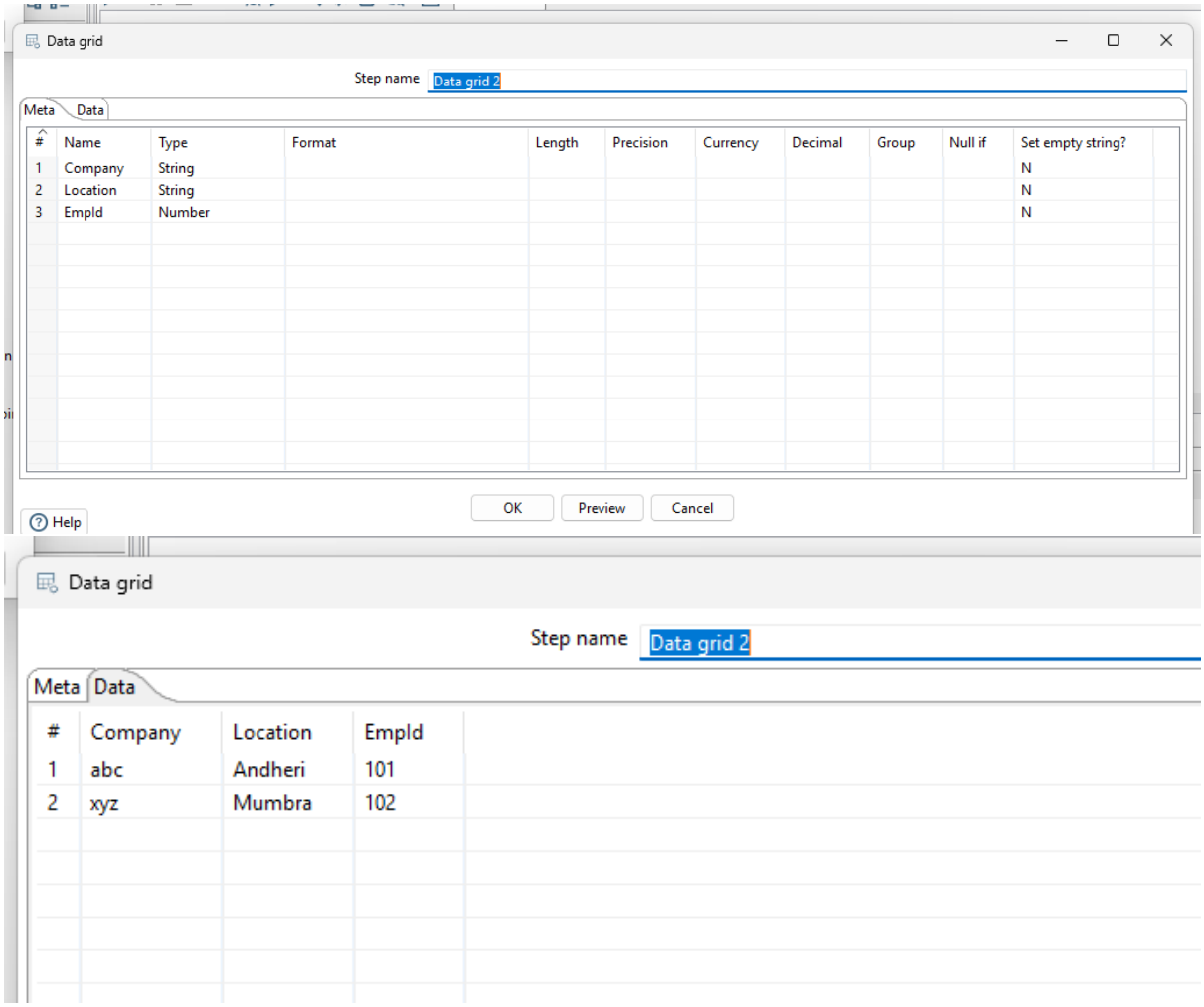
```
SQL> select * from prac14_f;

     USER1      USER2      USER3      USER4      USER5      USER6      USER7
---------- ---------- ---------- ---------- ---------- ---------- ----------
     USER8      USER9     USER10
---------- ---------- ----------
         1          1          1          1          1          1          1
         1          1          1

         2          2          2          2          2          2          2
         2          2          2

         3          3          3          3          3          3          3
         3          3          3


     USER1      USER2      USER3      USER4      USER5      USER6      USER7
---------- ---------- ---------- ---------- ---------- ---------- ----------
     USER8      USER9     USER10
---------- ---------- ----------
         4          4          4          4          4          4          4
         4          4          4

         5          5          5          5          5          5          5
         5          5
```

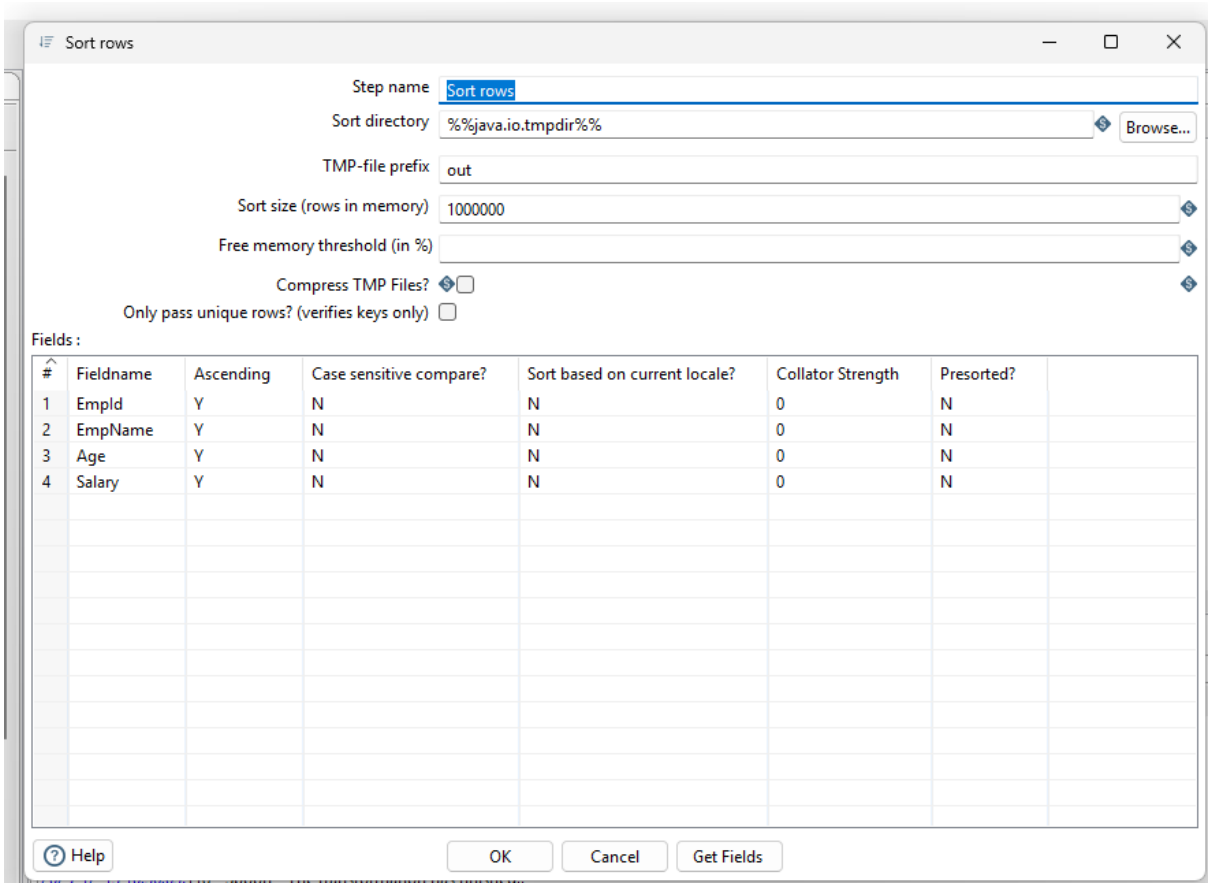
Program 6: Merge Join

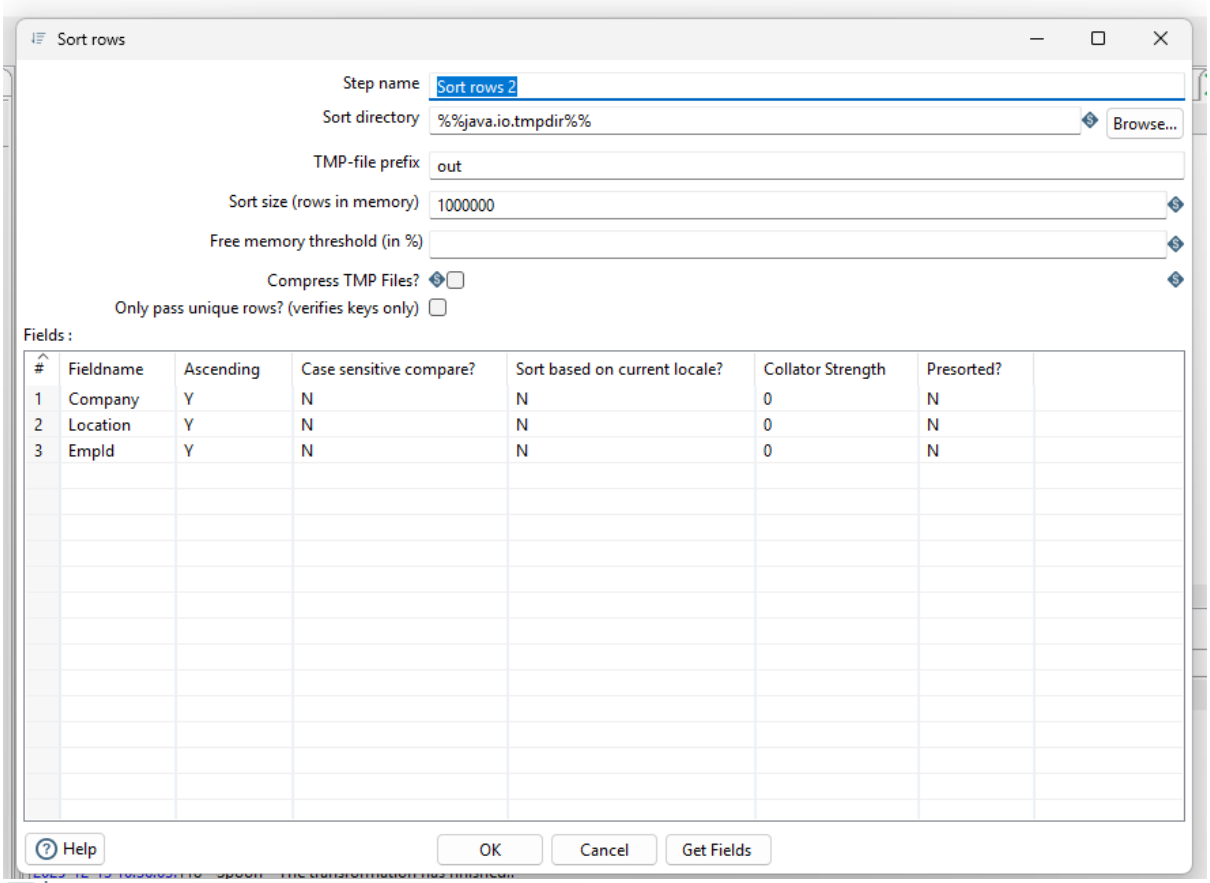


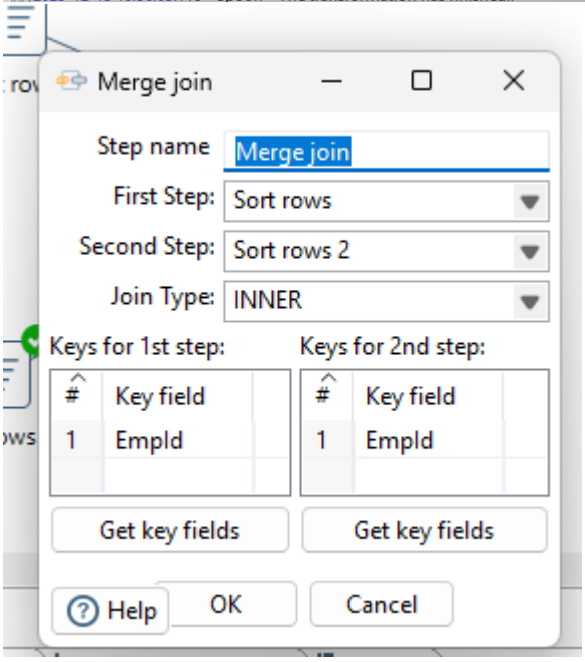


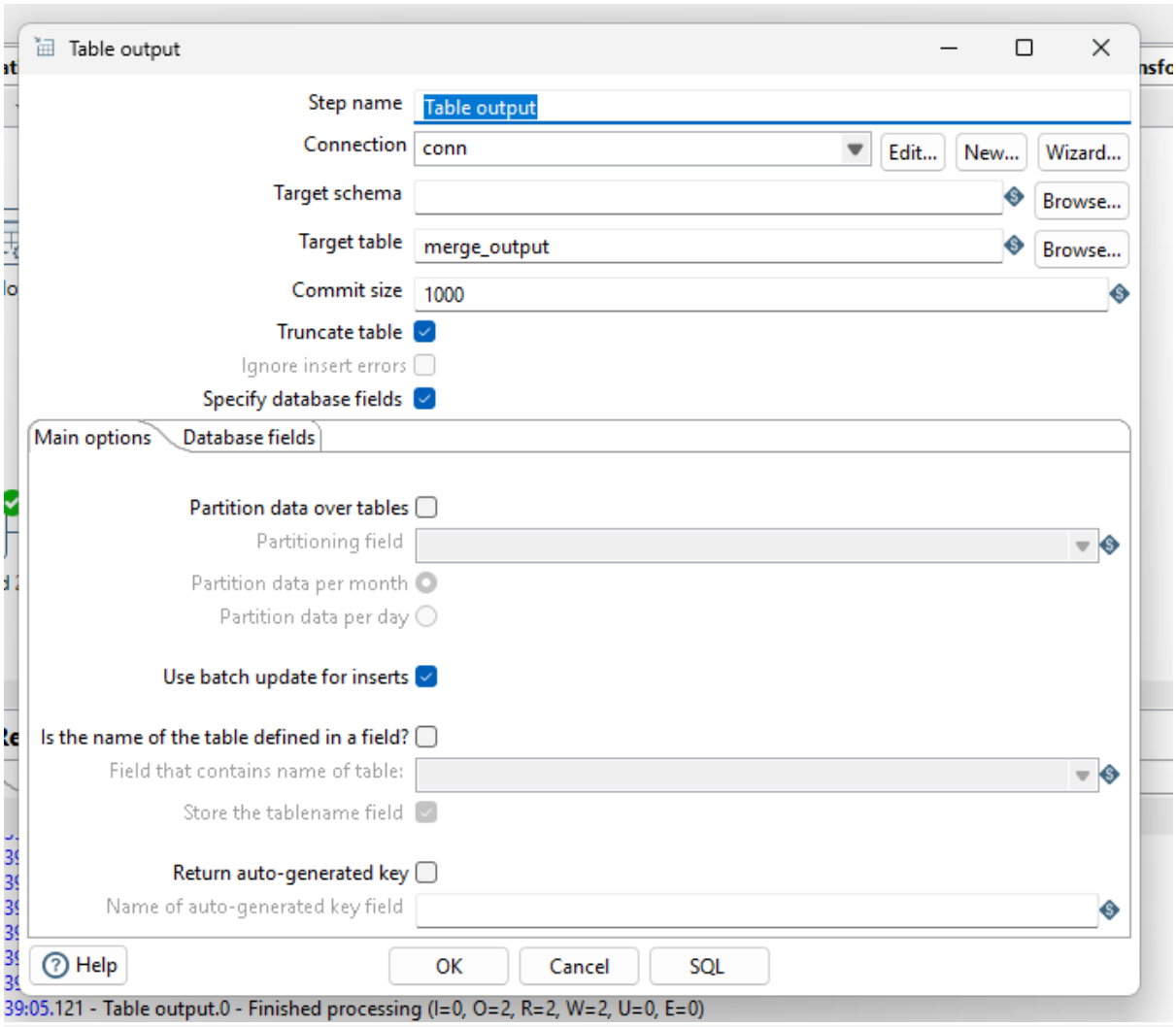



Data grid
Step name
Data grid 2
Meta
Data
Name Type Format Length Precision Currency Decimal Group Null if Set empty string?
1 Company String N
2 Location String N
3 EmpId Number N
OK
Preview
Cancel
Help
Data grid
Step name
Data grid 2
Meta
Data
Company Location EmpId
1 abc Andheri 101
2 xyz Mumbra 102


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 EmpId Y N N 0 N
2 EmpName Y N N 0 N
3 Age Y N N 0 N
4 Salary Y N N 0 N
Help
OK
Cancel
Get Fields


Sort rows
Step name
Sort rows 2
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 Company Y N N 0 N
2 Location Y N N 0 N
3 EmpId Y N N 0 N
Help
OK
Cancel
Get Fields

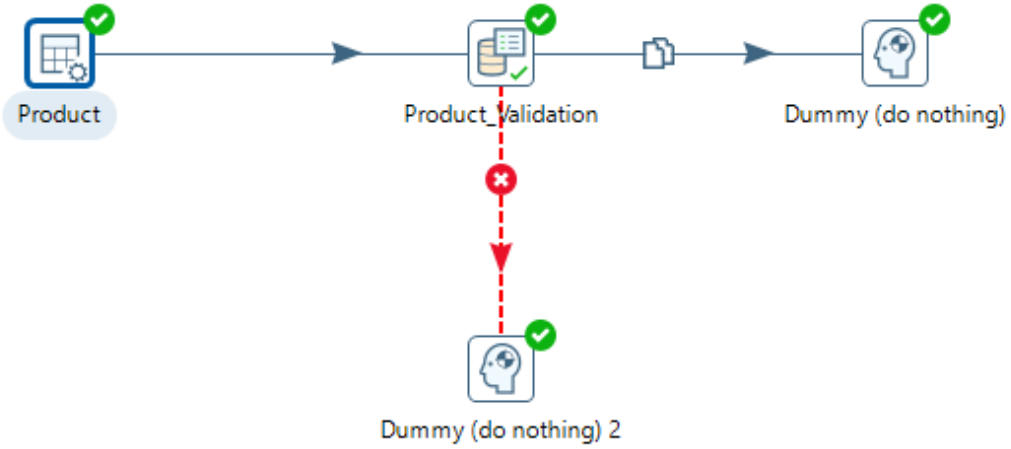

Merge join
Step name
Merge join
First Step:
Sort rows
Second Step:
Sort rows 2
Join Type:
INNER
Keys for 1st step:
Keys for 2nd step:
Key field
1 EmpId
Key field
1 EmpId
Get key fields
Get key fields
Help
OK
Cancel

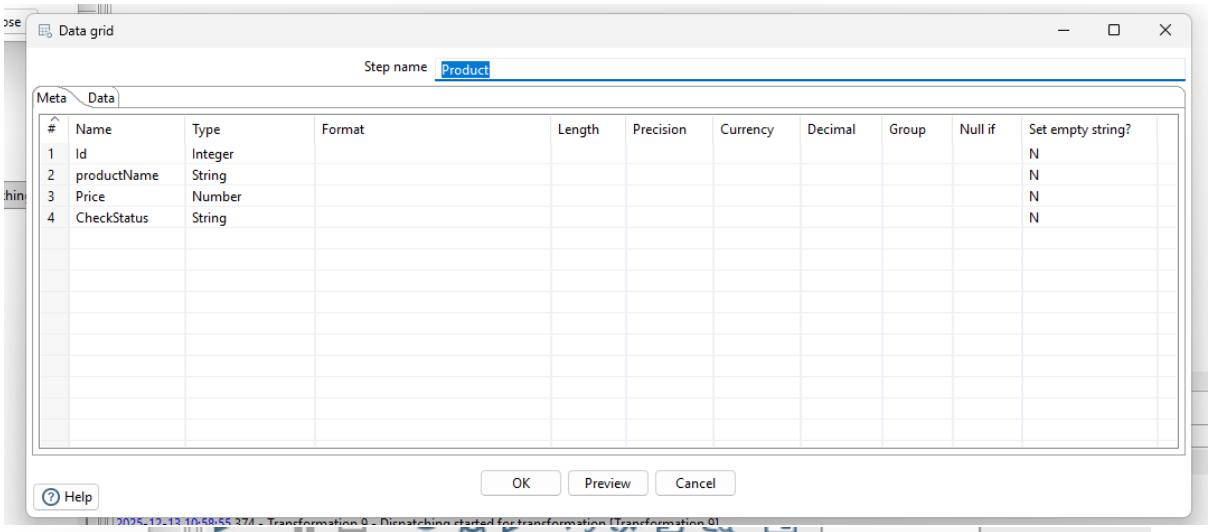


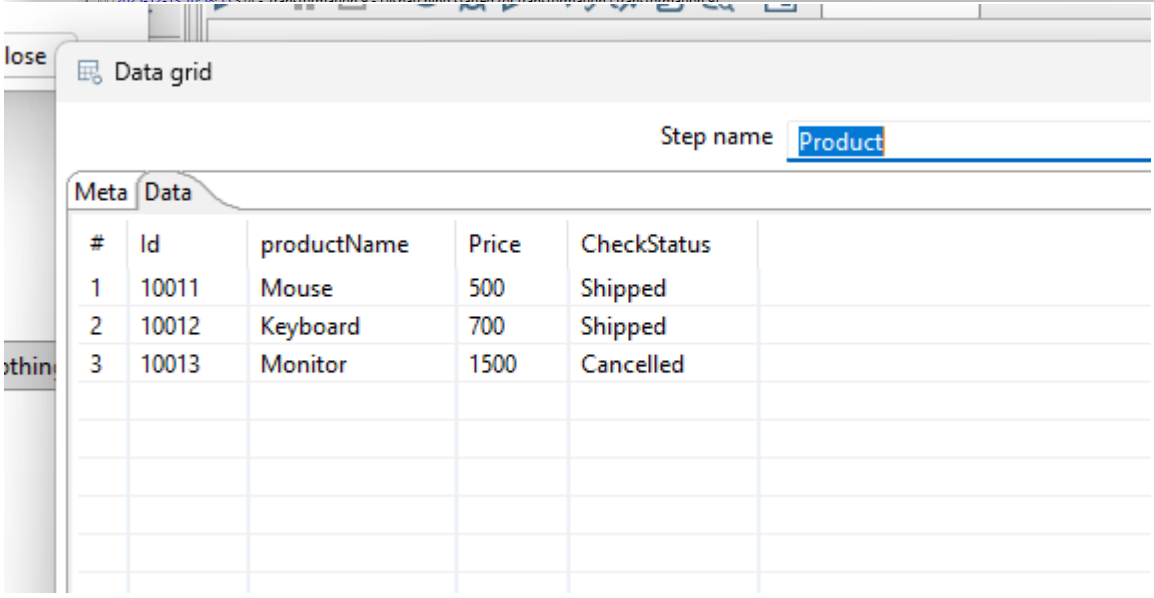
Rows of step: Table output (2 rows)

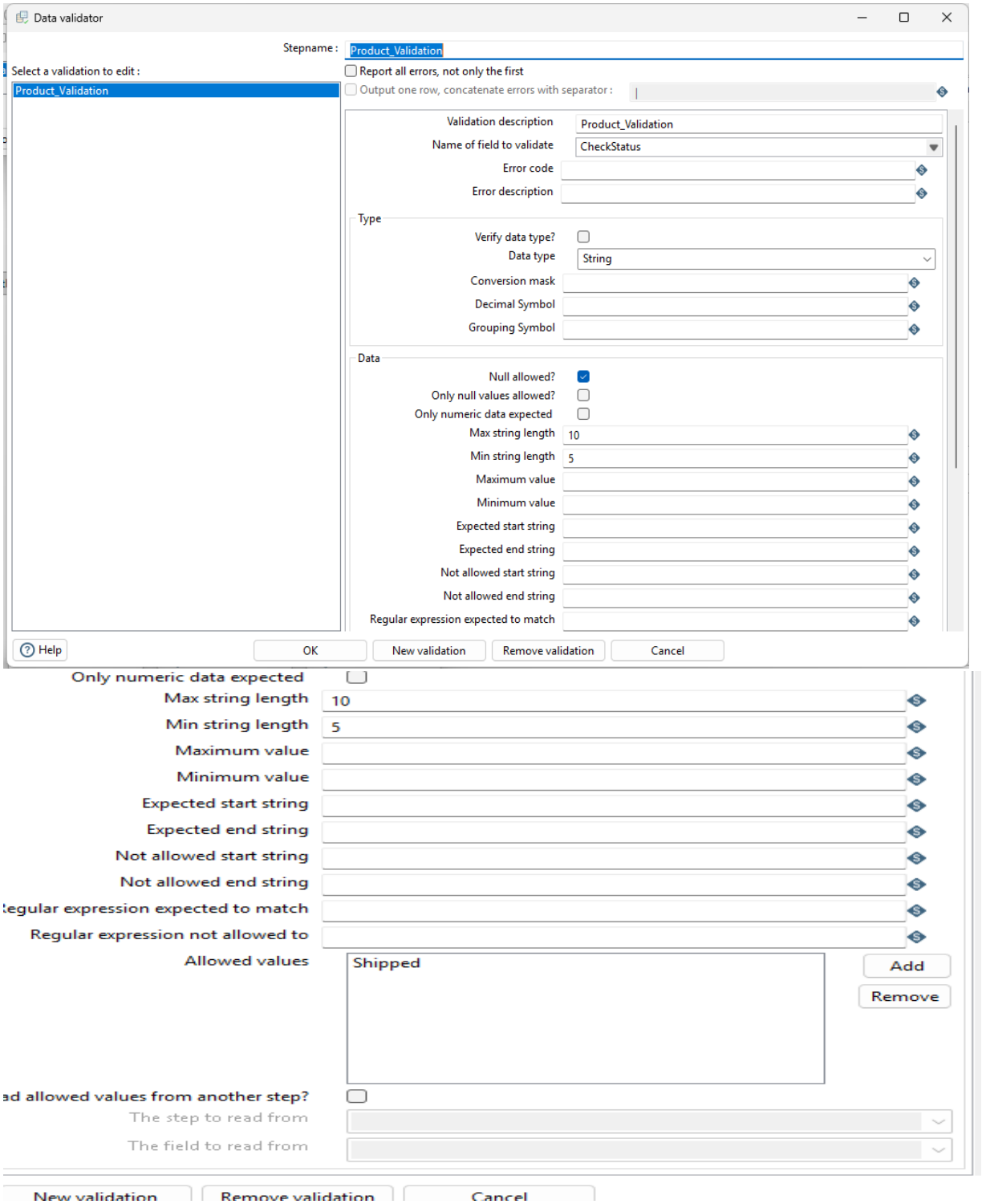
#	EmpId	EmpName	Age	Salary	Company	Location	EmpId_1
1	101.0	Aniket	21.0	5000.0	abc	Andheri	101.0
2	102.0	Raja	22.0	6000.0	xyz	Mumbra	102.0

Program 7: Data Validation








Data validator
Stepname : Product_Validation
Select a validation to edit :
Product_Validation
Report all errors, not only the first
Output one row, concatenate errors with separator :
Validation description
Product_Validation
Name of field to validate
CheckStatus
Error code
Error description
Type
Verify data type?
Data type
String
Conversion mask
Decimal Symbol
Grouping Symbol
Data
Null allowed?
Only null values allowed?
Only numeric data expected
Max string length
10
Min string length
5
Maximum value
Minimum value
Expected start string
Expected end string
Not allowed start string
Not allowed end string
Regular expression expected to match
Help
OK
New validation
Remove validation
Cancel
Only numeric data expected
Max string length
10
Min string length
5
Maximum value
Minimum value
Expected start string
Expected end string
Not allowed start string
Not allowed end string
egular expression expected to match
Regular expression not allowed to
Allowed values
Shipped
Add
Remove
ad allowed values from another step?
The step to read from
The field to read from
New validation
Remove validation
Cancel

Examine preview data

Rows of step: Product_Validation (2 rows)

#	Id	productName	Price	CheckStatus
1	10012	Keyboard	700.0	Shipped
2	10011	Mouse	500.0	Shipped

Examine preview data

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

#	Id	productName	Price	CheckStatus
1	10011	Mouse	500.0	Shipped
2	10012	Keyboard	700.0	Shipped

Examine preview data

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

#	Id	productName	Price	CheckStatus
1	10013	Monitor	1500.0	Cancelled