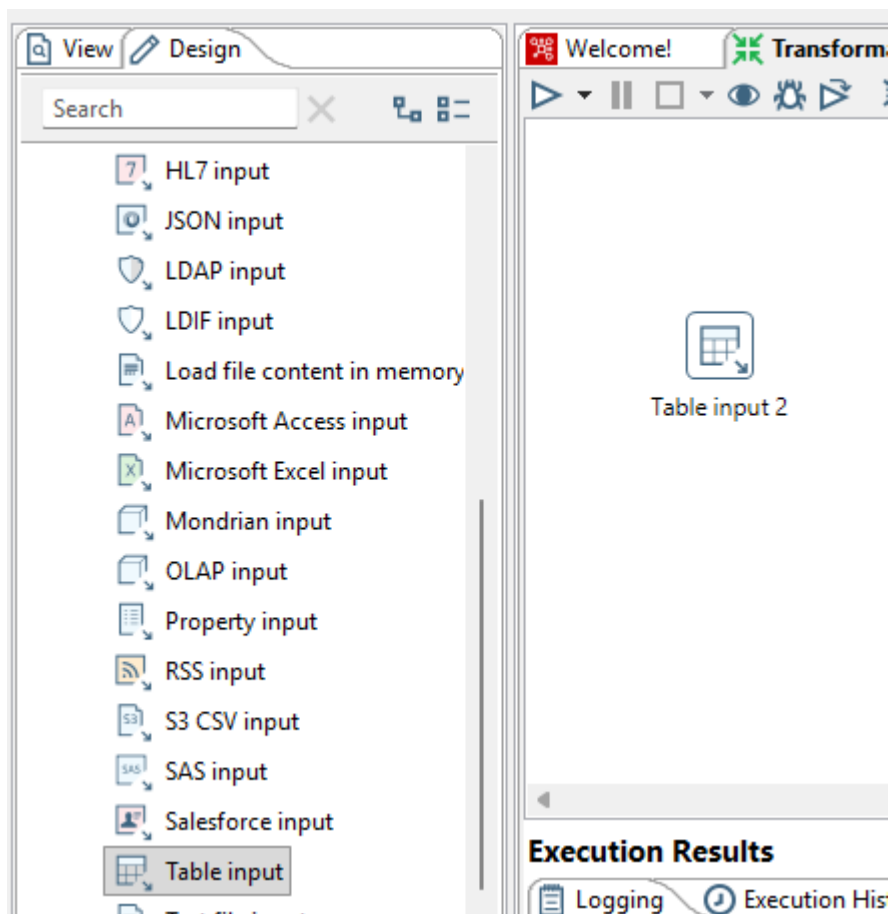


Assignment No	13
Title	Pentaho
Objective	<ol style="list-style-type: none"><li>1. Extraction of data from SQL and storing data in SQL</li><li>2. Sort Rows, Add Sequence</li></ol>
Roll No	MCA2516

## Program 1: Extraction of data from SQL and storing data in SQL

Step 1: File -> New -> Transformation

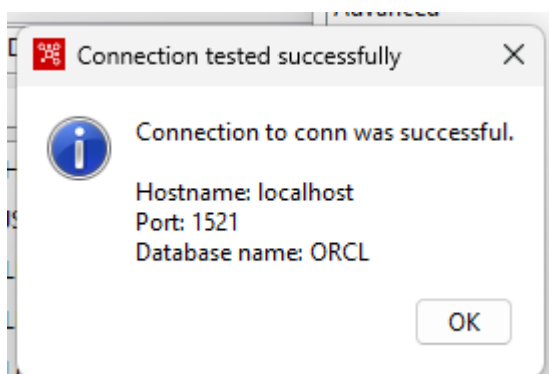
Step 2: Design -> Input -> Table Input



### Step 3 : Database Connection configuration

The screenshot shows the 'Database Connection' dialog box. On the left, a sidebar lists 'General', 'Advanced', 'Options', 'Pooling', and 'Clustering', with 'General' selected. The main area is divided into three sections: 'Connection name' (text field with 'conn'), 'Connection type' (list box with 'Oracle' selected), and 'Access' (list box with 'Native (JDBC)' selected). To the right, the 'Settings' section contains fields for 'Host Name' (localhost), 'Database Name' (ORCL), 'Tablespace for Data' (empty), 'Tablespace for Indices' (empty), 'Port Number' (1521), 'Username' (system), and 'Password' (masked with dots). At the bottom, there are buttons for 'Test', 'Feature List', 'Explore', 'OK', and 'Cancel'.

### Step 4: Test for connection



Step 5: Get SQL Statement

Table input

Step name: Table input

Connection: conn

Buttons: Edit..., New..., Wizard..., Get SQL select statement...

SQL

```
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? ☐

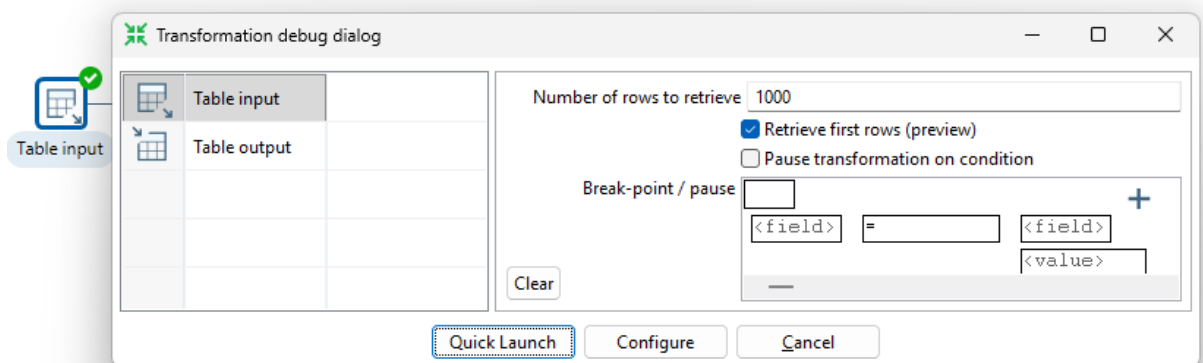
Insert data from step

Execute for each row? ☐

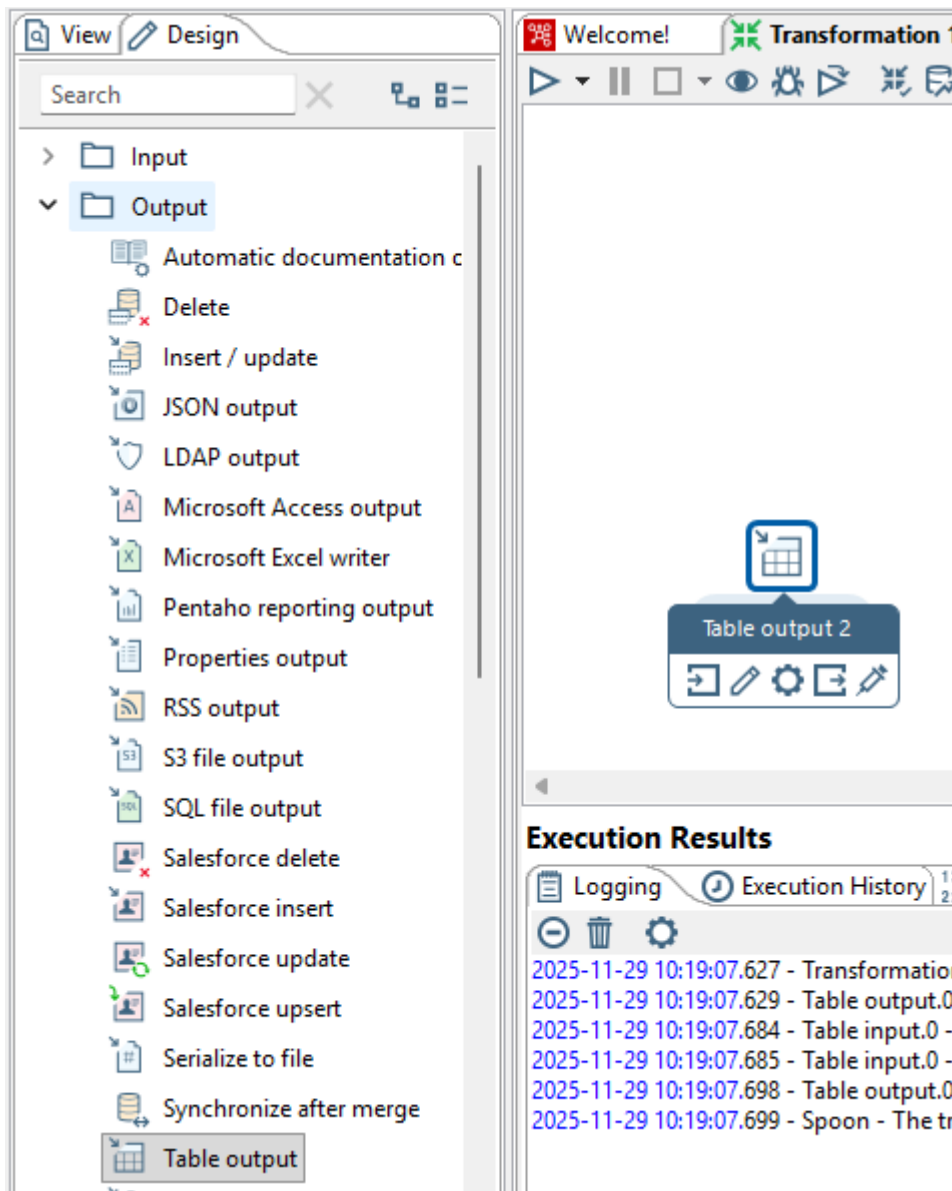
Limit size: 0

Buttons: Help, OK, Preview, Cancel

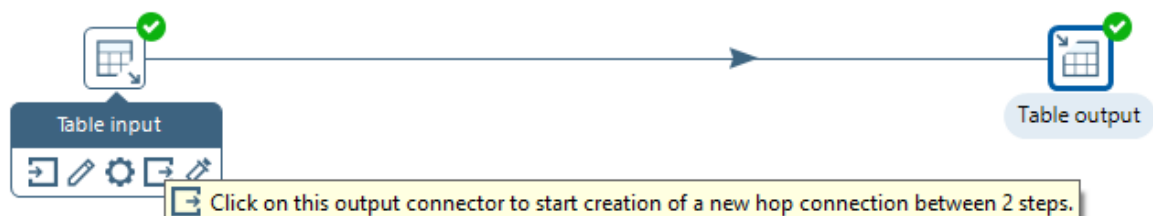
Step 6: Right Click -> Preview -> Quick Launch



Step 7 : Design -> Output -> Table Output



Step 9: Connect the table input with table output



Step 10: Double Click Table output

Table output

Step name:

Connection:

Target schema:

Target table:

Commit size:

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: ☒

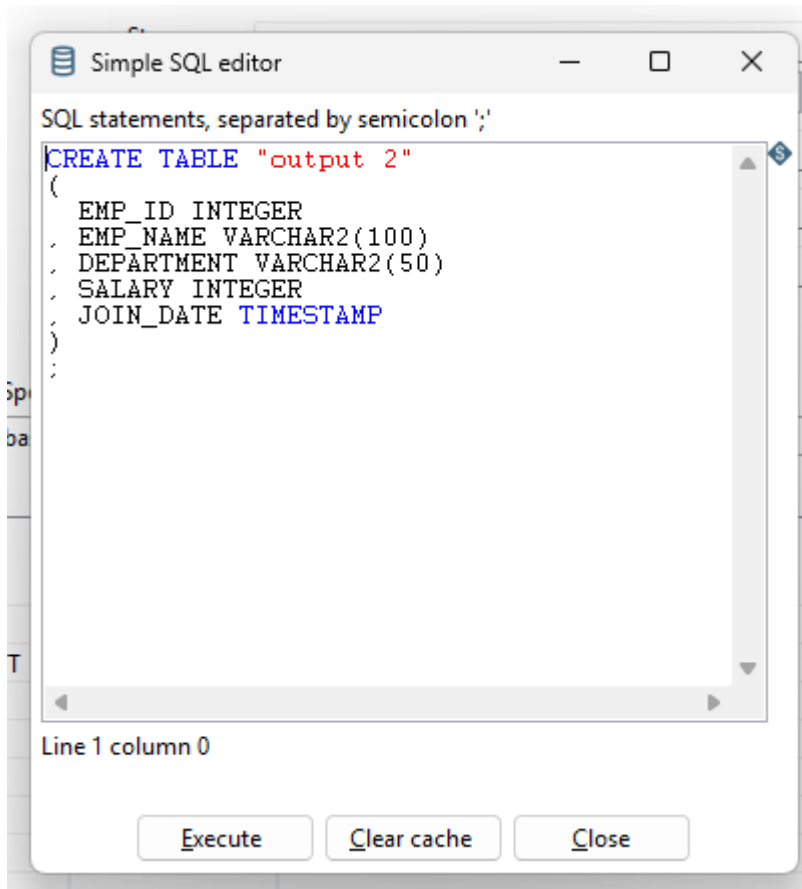
Return auto-generated key: ☐

Name of auto-generated key field:

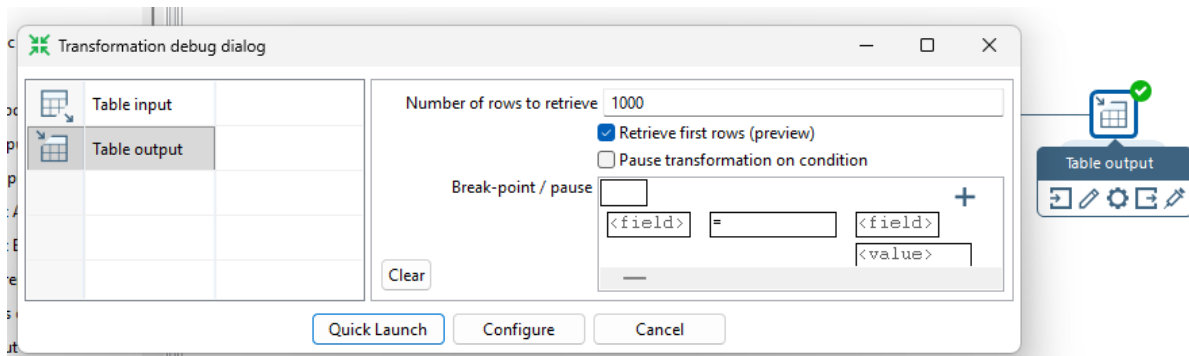
Step 11: Data field base -> get fields

[illegible]

Step 12: Click on SQL and execute



Step 14:





SQL Plus Output

```
SQL> select * from output;
```

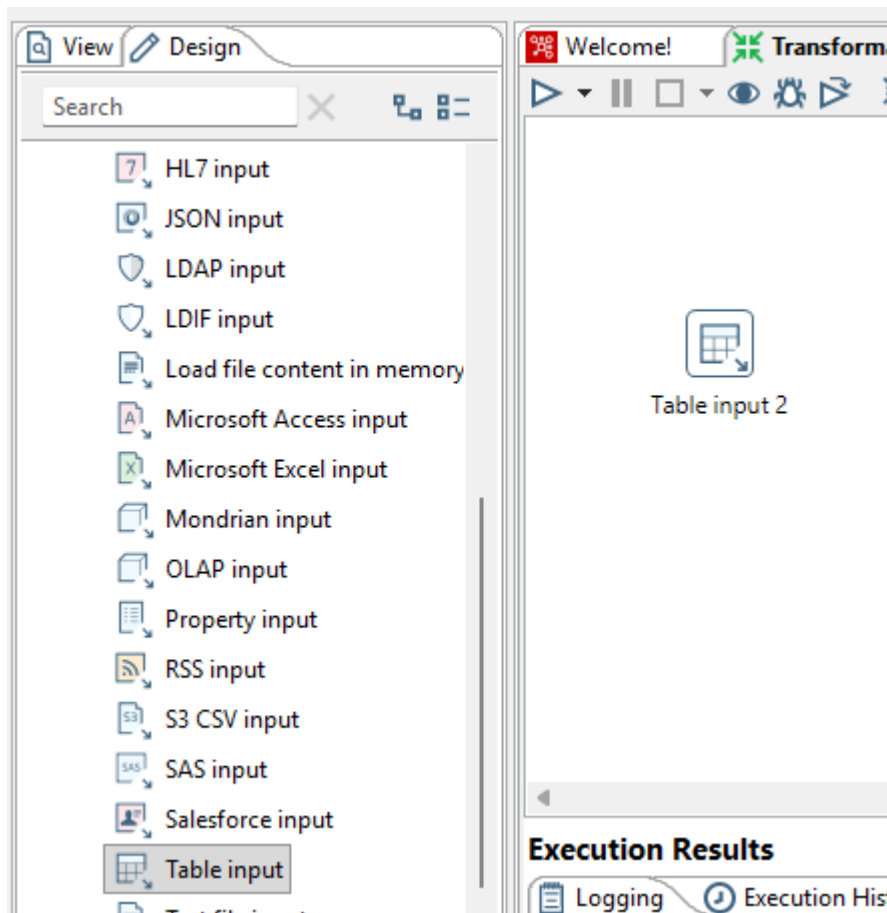
EMP_ID	EMP_NAME	DEPARTMENT	SALARY	JOIN_DATE
1	Amit Sharma	IT	55000	10-JAN-23 12.00.00.000000 AM

EMP_ID	EMP_NAME	DEPARTMENT	SALARY	JOIN_DATE
2	Neha Patel	HR	48000	15-FEB-23 12.00.00.000000 AM

## Program 2: Sort Rows, Add Sequence

Step 1: File -> New -> Transformation

Step 2: Design -> Input -> Table Input



Step 3 : Database Connection configuration

Database Connection

General  
Advanced  
Options  
Pooling  
Clustering

Connection name:  
conn

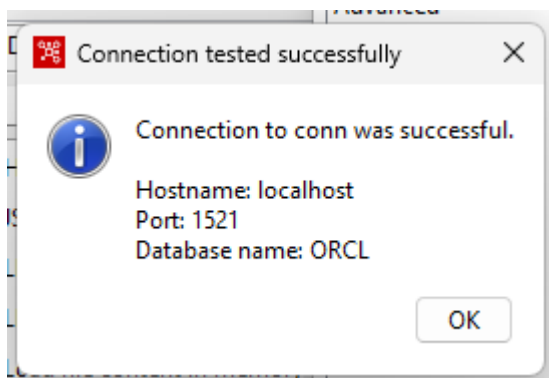
Connection type:  
Native Mondrian  
Neoview  
Netezza  
**Oracle**  
Oracle RDB  
PostgreSQL  
Redshift  
Remedy Action Request System  
SAP ERP System  
SQLite  
Snowflake  
Sybase  
SybaseIQ  
Teradata  
UniVerse database  
Vertica  
Vertica 5+  
dBase III, IV or 5

Access:  
**Native (JDBC)**  
ODBC  
OCI  
JNDI

Settings  
Host Name:  
localhost  
Database Name:  
ORCL  
Tablespace for Data  
Tablespace for Indices  
Port Number:  
1521  
Username:  
system  
Password:  
.....

Test Feature List Explore OK Cancel

#### 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

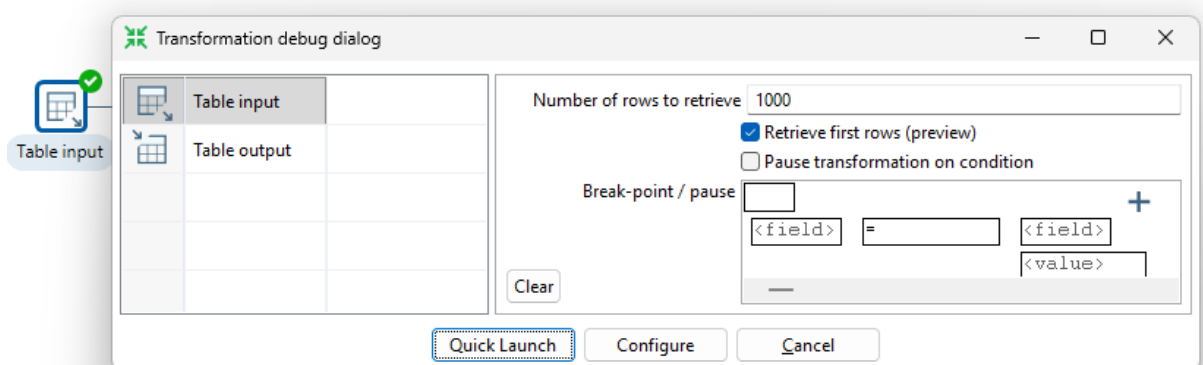
Help

OK

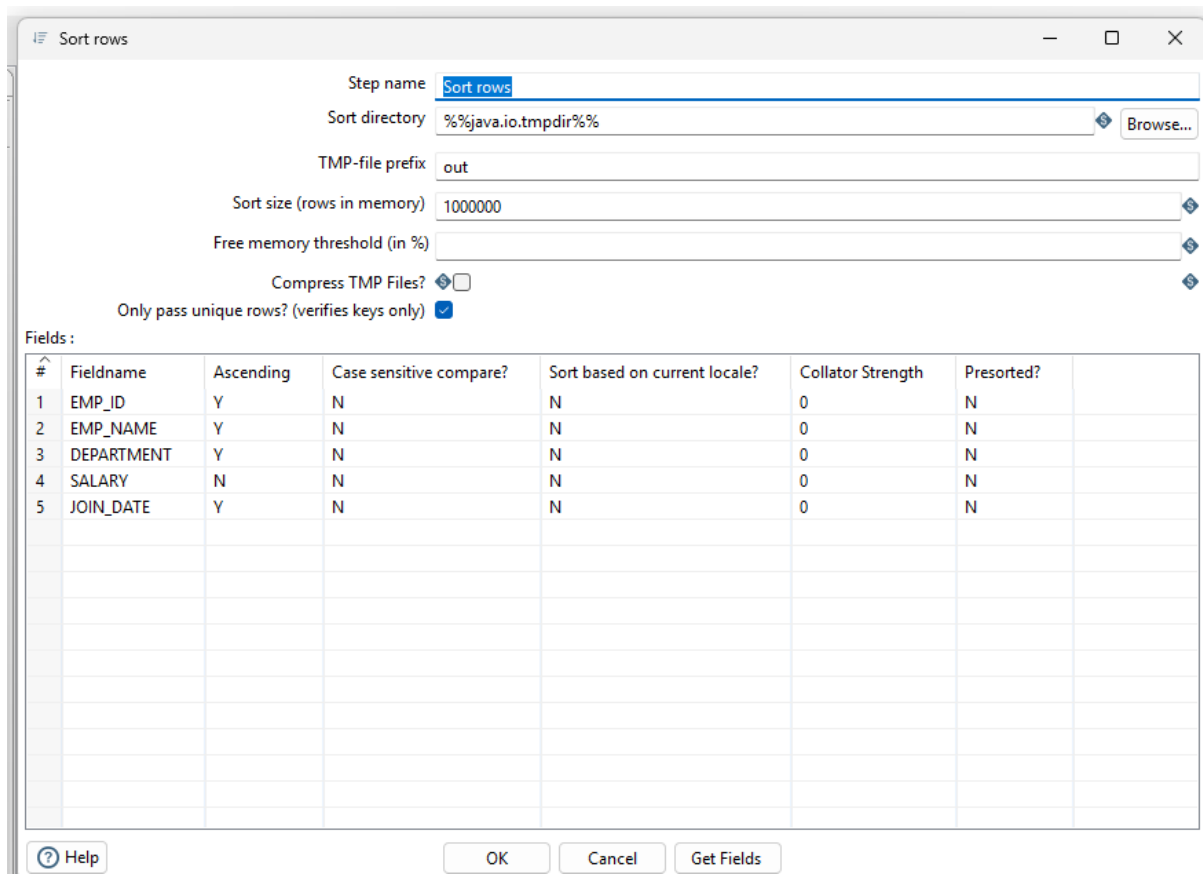
Preview

Cancel

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



Step 7: Design -> Transformation -> sort Row -> and add the connection



Step 8: Design -> Transformation -> Add Sequence



**Add sequence**

Step name: Add sequence

Name of value: new column

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):

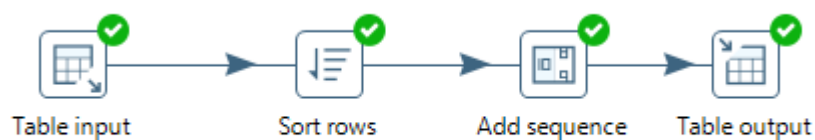
Start at value: 1

Increment by: 1

Maximum value: 50

Help OK Cancel

Step 9: Connect the table input with table output



Step 10: Double Click Table output

**Table output**

Step name: **Table output**

Connection: **conn** [Edit...] [New...] [Wizard...]

Target schema: [Browse...]

Target table: **sortedoutput** [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]

Step 11: Data field base -> get fields

Table output

Step name: Table output

Connection: conn [Edit...] [New...] [Wizard...]

Target schema: [Browse...]

Target table: output [Browse...]

Commit size: 1000

Truncate table: ☒

Ignore insert errors: ☐

Specify database fields: ☒

Main options Database fields

Fields to insert:

#	Table field	Stream field
1	EMP_ID	EMP_ID
2	EMP_NAME	EMP_NAME
3	DEPARTME...	DEPARTMENT
4	SALARY	SALARY
5	JOIN_DATE	JOIN_DATE

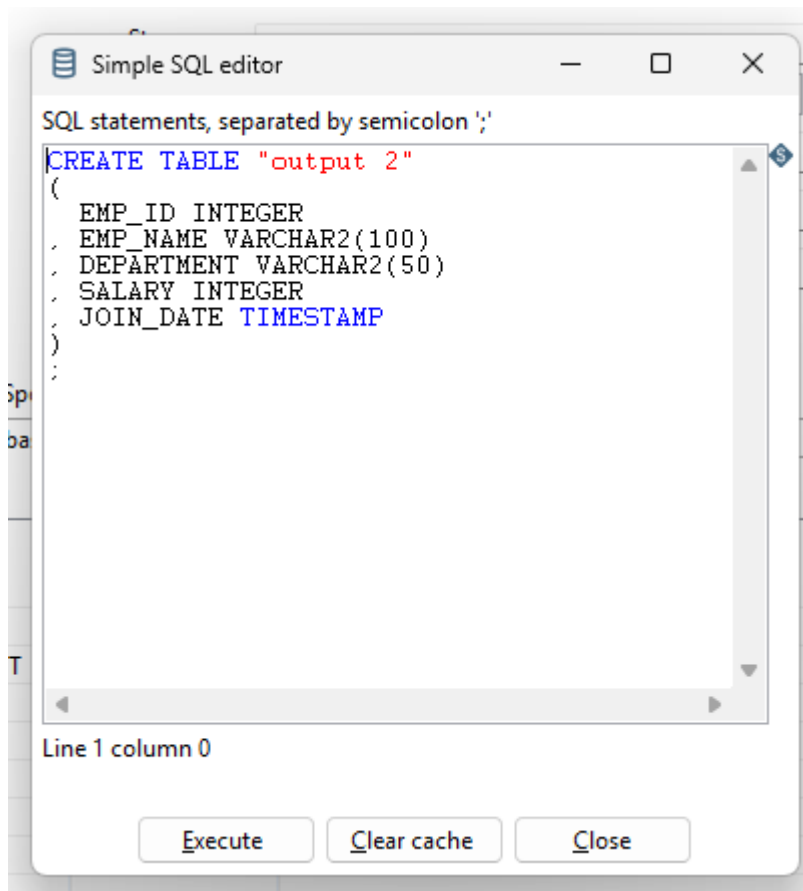
[Get fields]

Enter field mapping

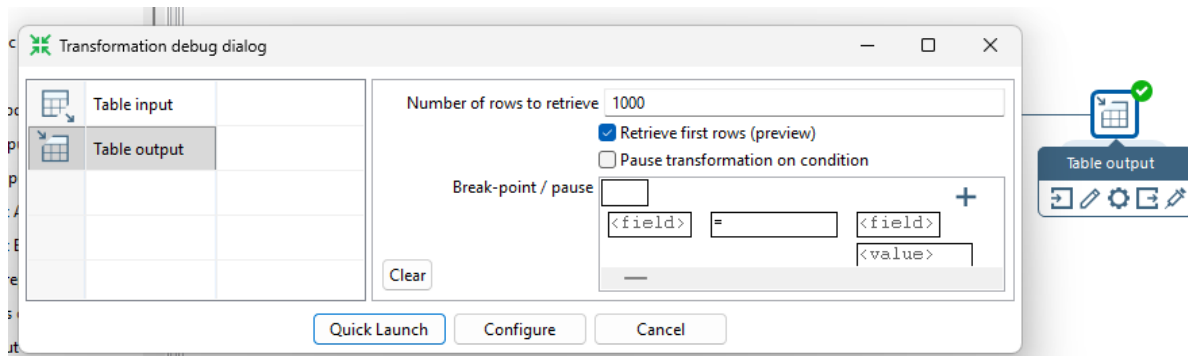
[?] Help [OK] [Cancel] [SQL]

Step 12: Click on SQL and execute





Step 14:



Output:

Examine preview data

Rows of step: Table output (12 rows)

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

SQL Output:

SQL Plus

```
SQL> select * from sortedoutput;
```

EMP_ID			
EMP_NAME			
DEPARTMENT		SALARY	
JOIN_DATE			
new column			
1			
Amit Sharma			
IT		55000	

EMP_ID			
EMP_NAME			
DEPARTMENT		SALARY	
JOIN_DATE			
new column			
10-JAN-23 12.00.00.000000 AM			
1			