**Assignment-2**

**Pentaho Data Integration**

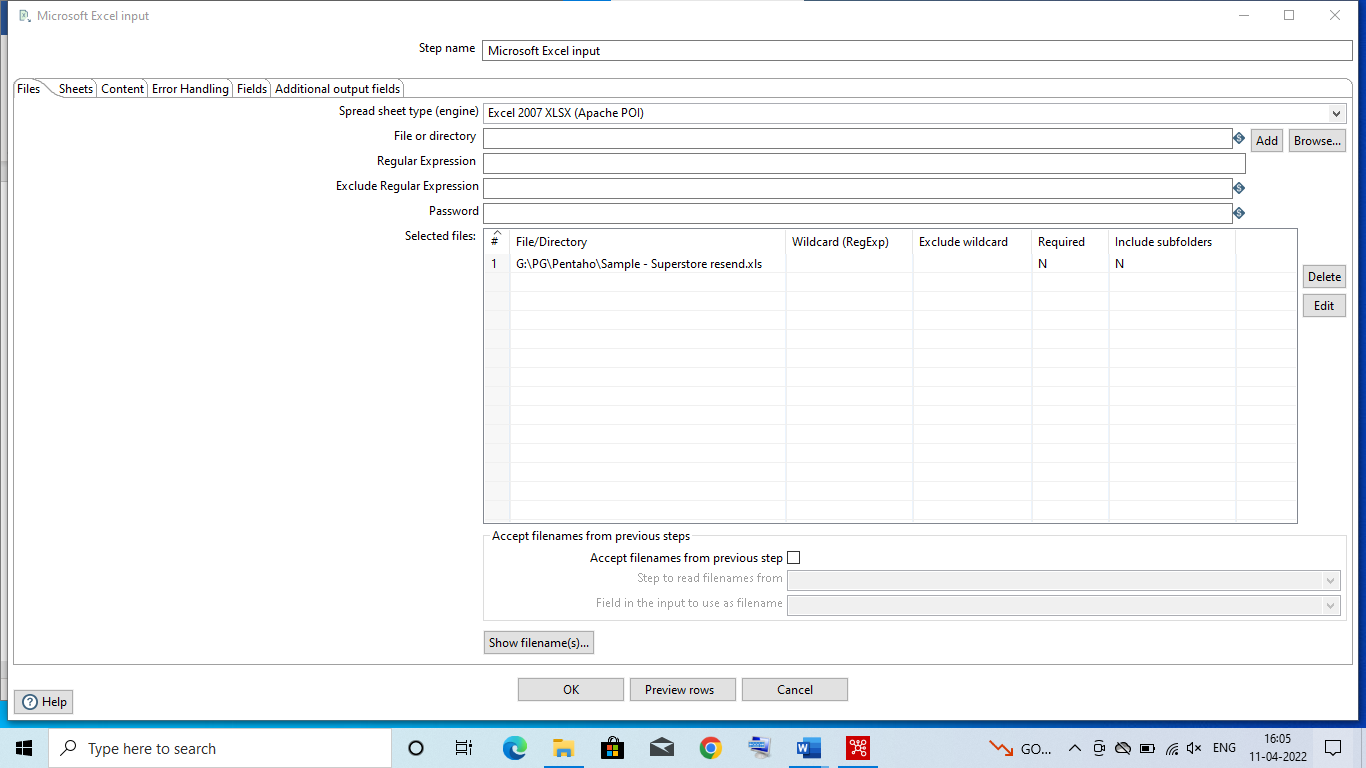
1. Migrate a table data from one database to another.

First Method : To get all fields from Excel input to Excel output.

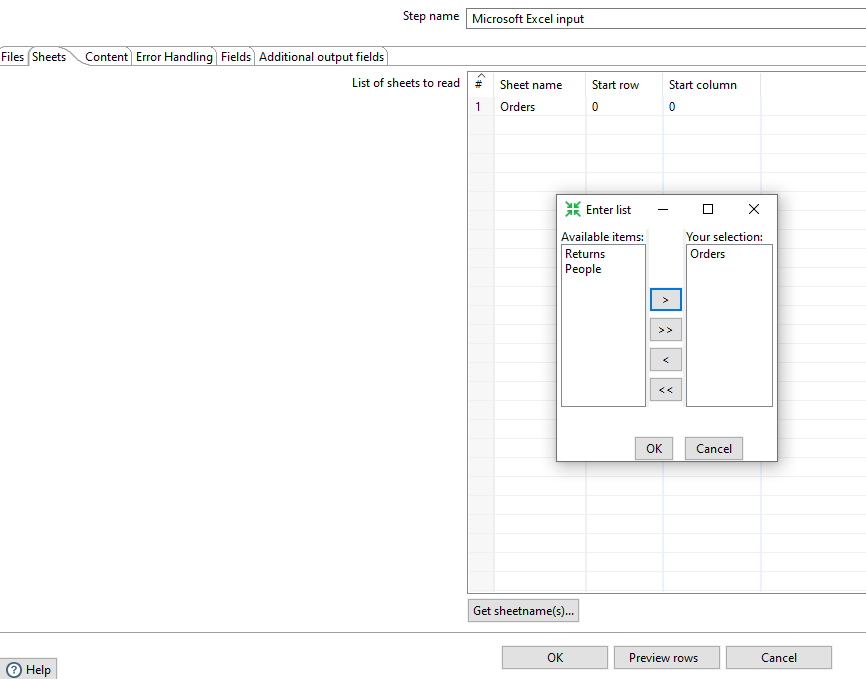
**Excel Input :**

**To change the Excel 2007 XLSX(Apache POI)**

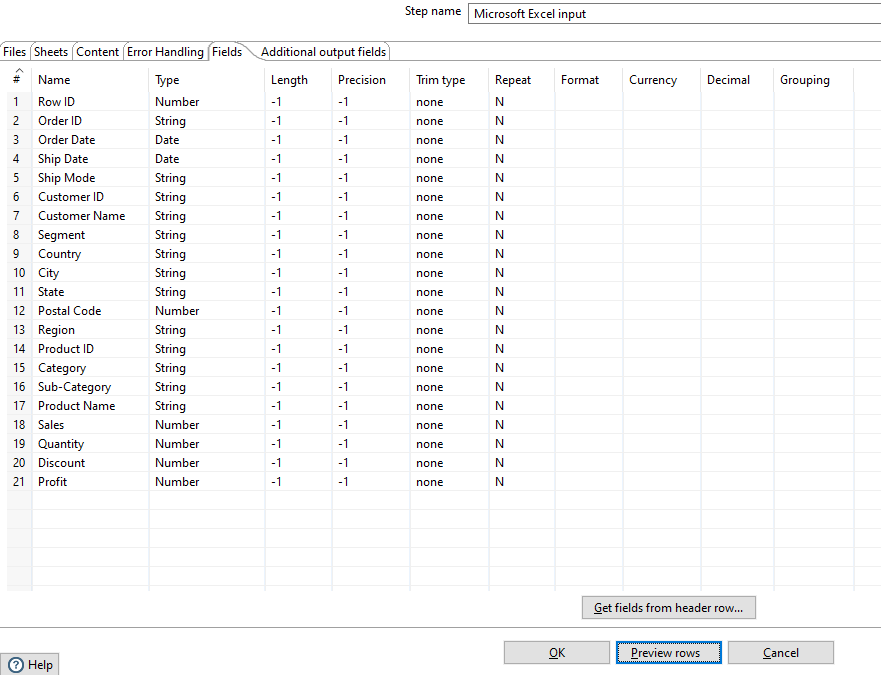
**Open the sample superstore file**



**To select the order items only**



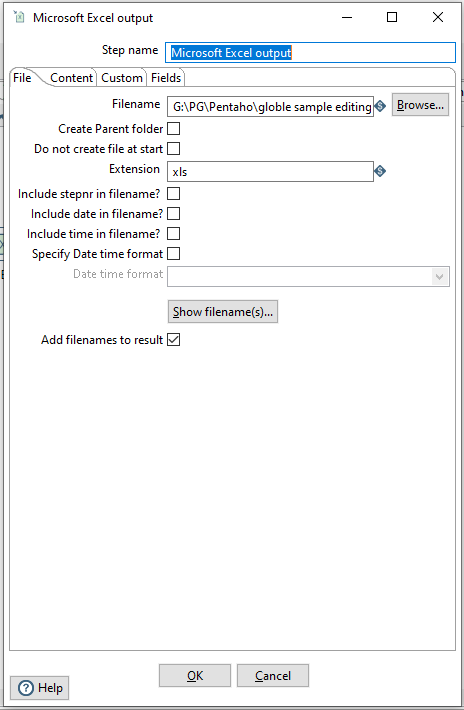
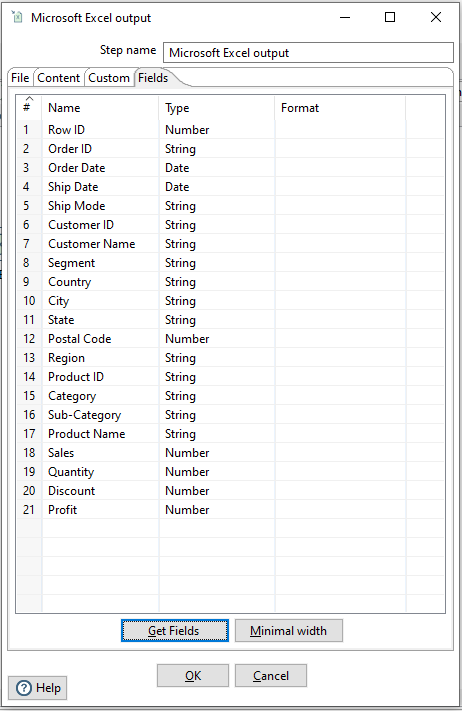
**Go to fields Tab Click the Get fields from header row**



**Excel Output :**

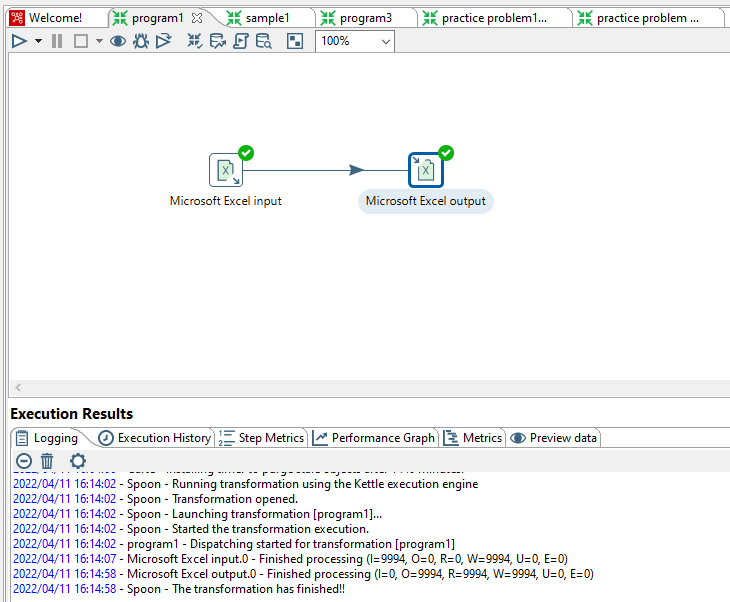
**Open the Microsoft excel output go to fields tab and click get fields**

**to browse the filename**

Finally to click the Run button after asking to save this program

Then successfully run this program shown green Thicks

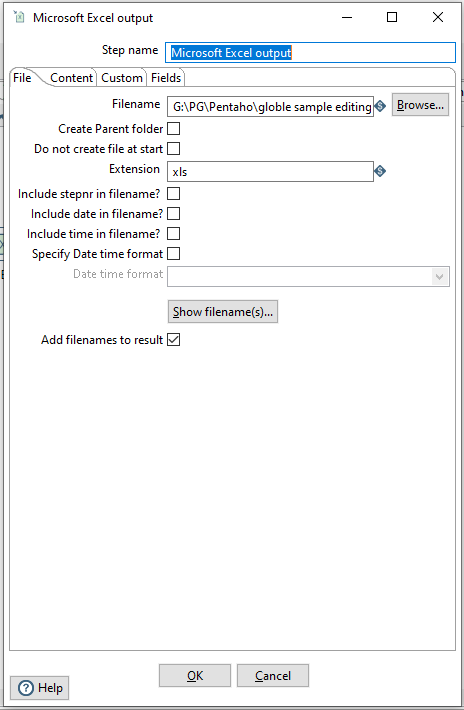
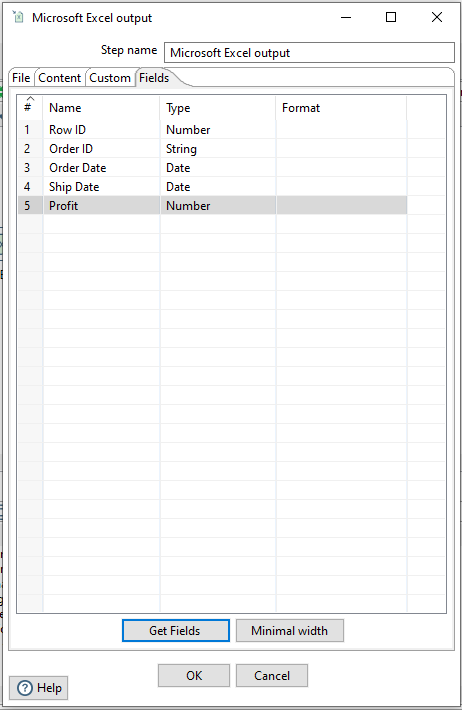


Method 2 : Excel to Excel specified fields only

Excel input :

Same procedure for Method 1 inputs so do follow this program Excel input

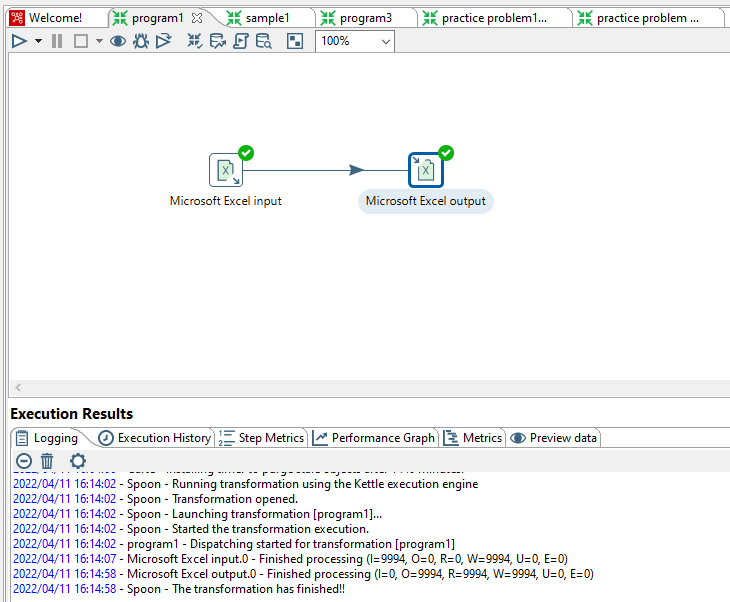
Excel Output :

**Open the Microsoft excel output to browse the filename go to fields tab and click get fields**  ****

Finally to click the Run button after asking to save this program

Then successfully run this program shown green Thicks

To check the preview its shown only 5 features.



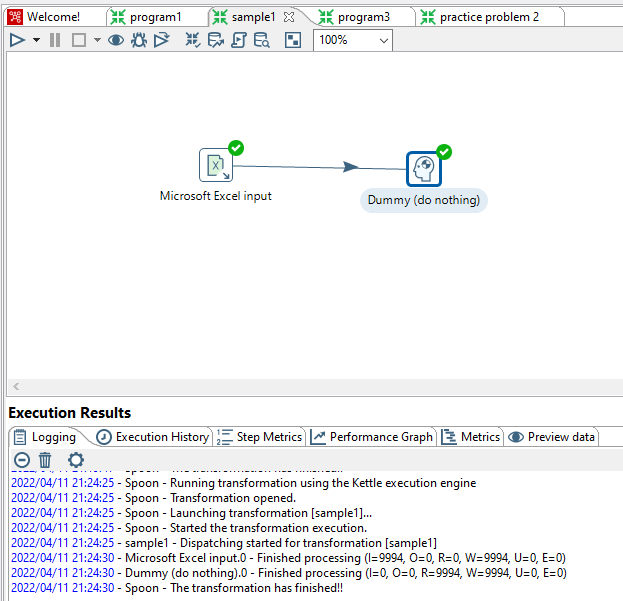
Method 3 : Excel to Dummy

Excel input :

Same procedure for Method 1 inputs so do follow this program Excel input

Dummys Output :

To taking a dummies after run this program



Method 4 : Excel to MySQL

Excel Input :

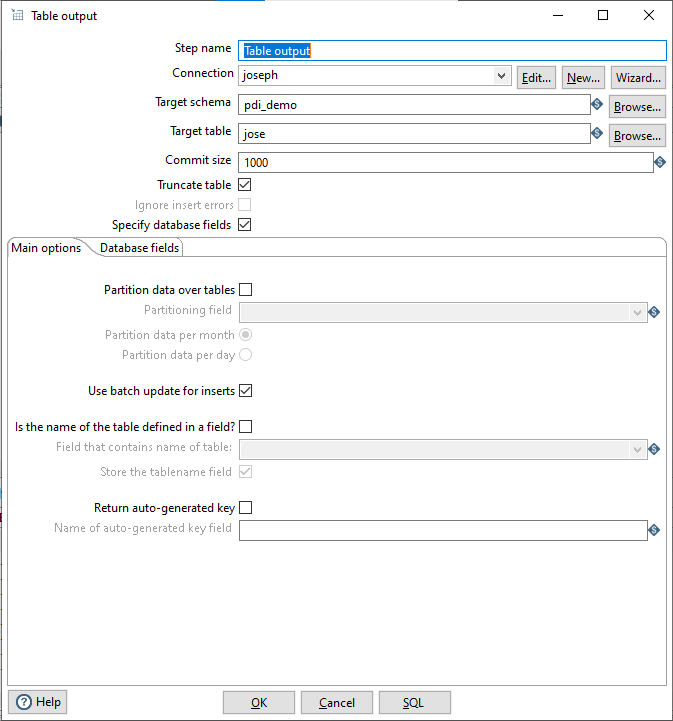
Same procedure for Method 1 inputs so do follow this program Excel input

MySQL Table Output :

Create a Table output and gave the Connection Name

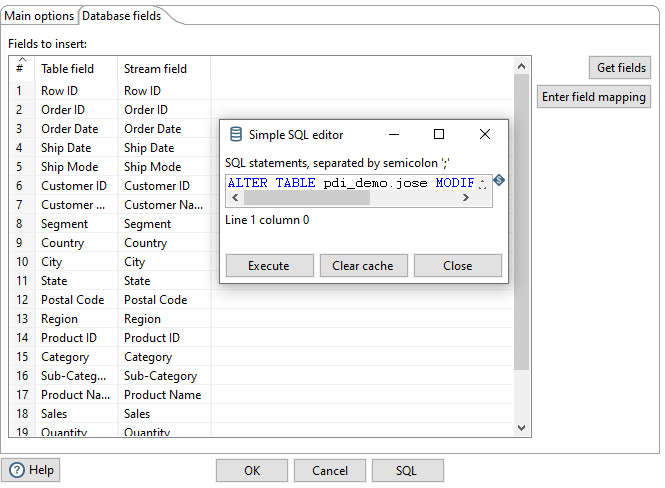
To gave a database name and Target table name

Enable the Two check box Truncate table and Specify database fields

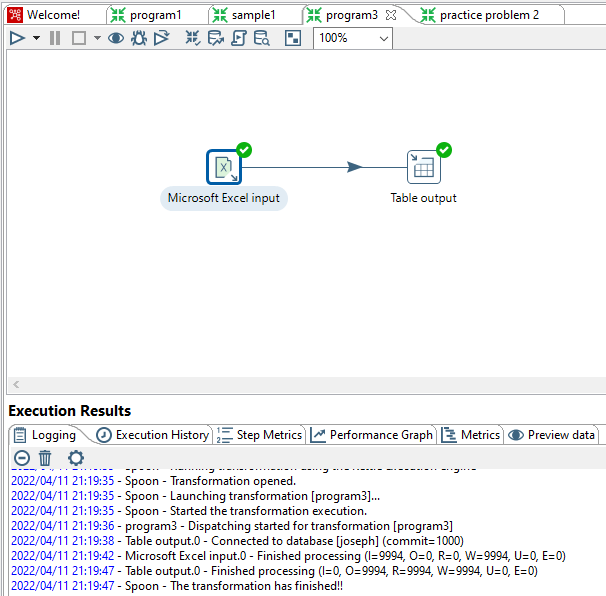


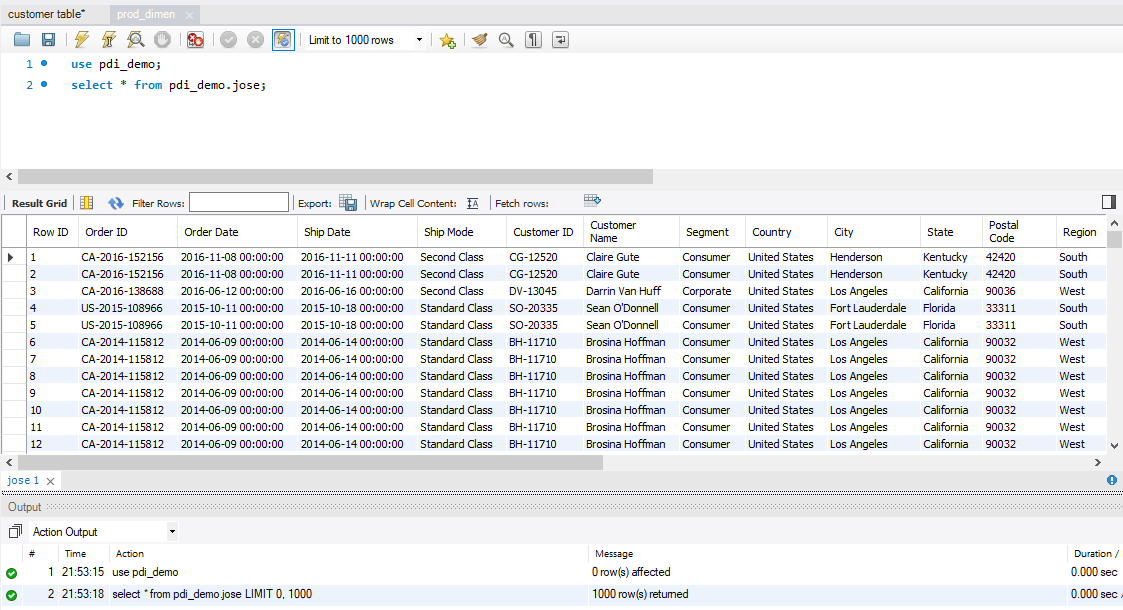
To change the tab Main options to Database fields

And click the SQL button opening the popup box and click the excute button



Run this mysql table output program



To open the MySql workbench and checking the output

2. Extract data from CSV file and load it into a database table.

Graphical user interface, application, table

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Text

Description automatically generated

Output on dummy

Graphical user interface, text, application

Description automatically generated

1. Discuss the following:-

**A, Need of DataWarehouse :**

\*Need for Data Warehouse to deliver business reports, analysis, data to BI tools, and perform ad-hoc queries becomes an essential part of efficient processing.

\* Data Warehouse gives the ability to quickly run analysis on huge volumes of datasets.

1. **Need of ETL :**

\* ETL platforms extract, transform, and load data from a source to a destination.

\*The ETL process can pull information from multiple databases.

\*The ETL process can search databases to find specific types of information.

\*Transforming data often includes formatting, duplicating, integrating, splitting, and other actions that make information more useful.

1. Explain about PDI components.

* **Spoon** - a data modelling and development tool for ETL developers. Use it to create transformations (elementary data flows) and jobs (execution sequences of transformations and other jobs)
* **Pan** - executes transformations modelled in Spoon
* **Kitchen** - executes jobs designed in Spoon
* **Carte** - a simple web server used for running and monitoring data integration tasks.