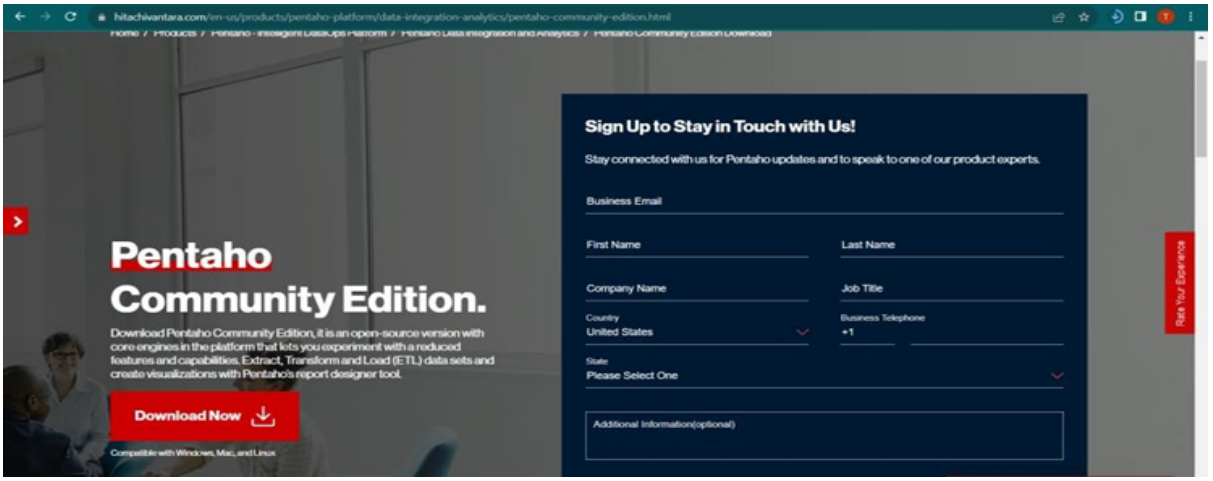
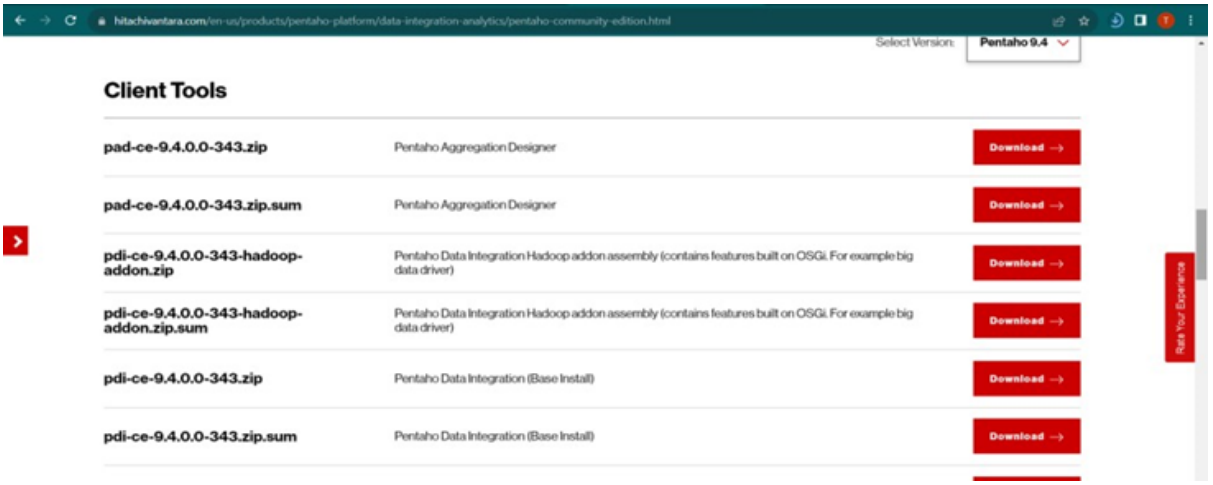
Practical no.4

INSTALLATION OF PENTAHO

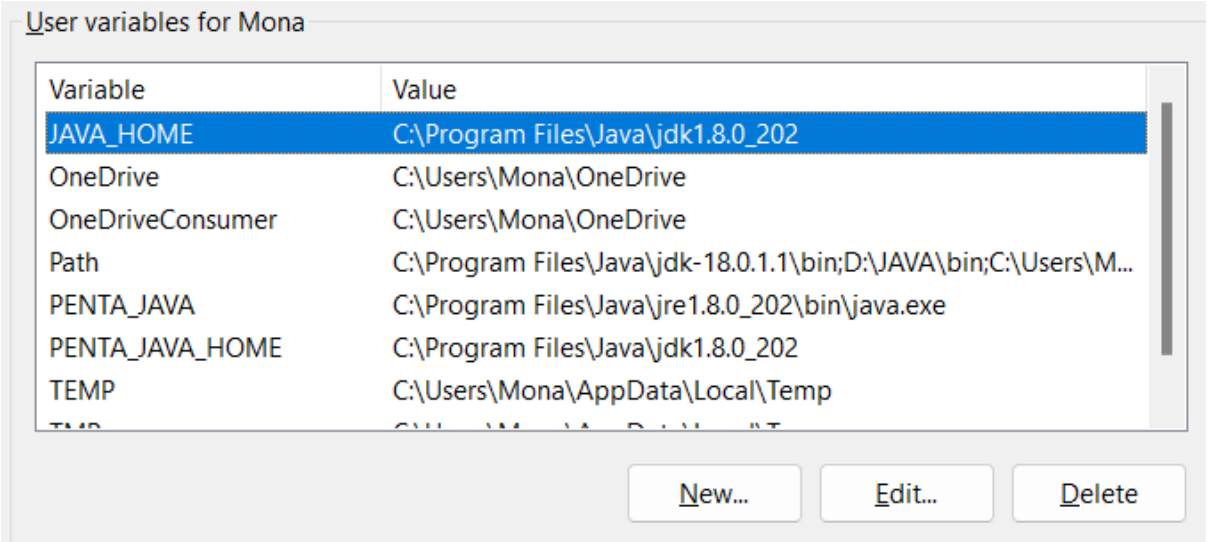
1.Go to https://www.hitachivantara.com/en-us/products/pentaho-platform/data-integrationanalytics/pentaho-community-edition.html site and click on download now.



1. Select and download pdi-ce-9.4.0.0.343.zip Pentaho Database Integration.



3. Change environmental variables of the system.



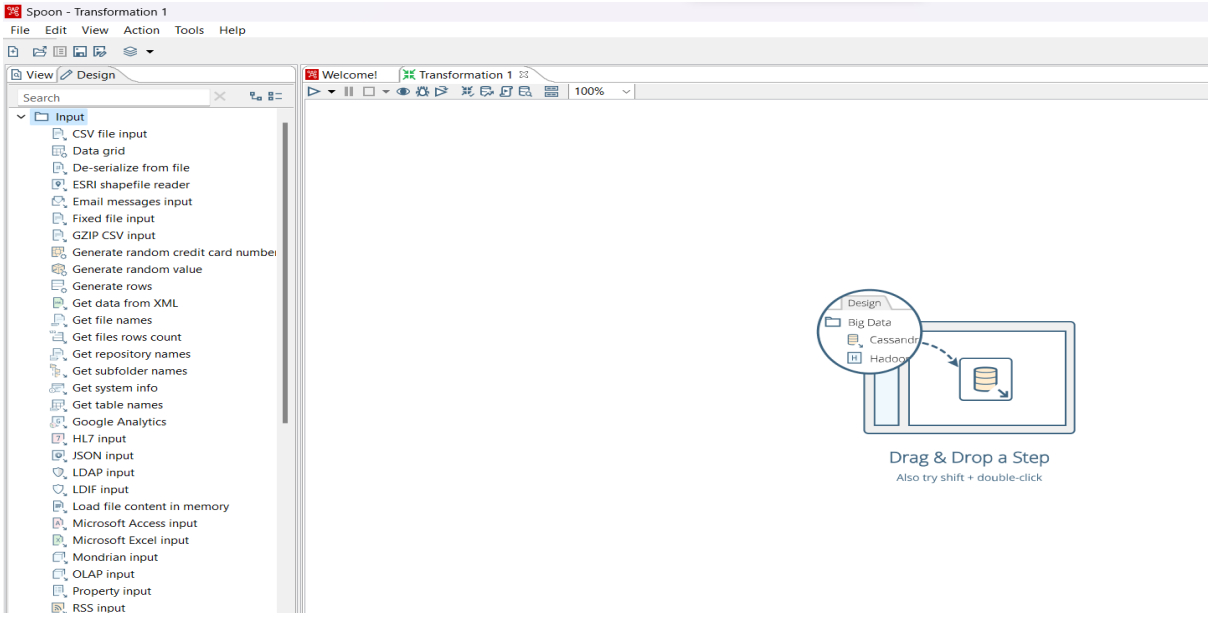
4. Run the spoon.bat (Window batch file) for opening the pentaho

Q1] Create a pentaho transformation to get data from mysql and push required data into an excel file.

Step 1: Create New Database penta and table prac1

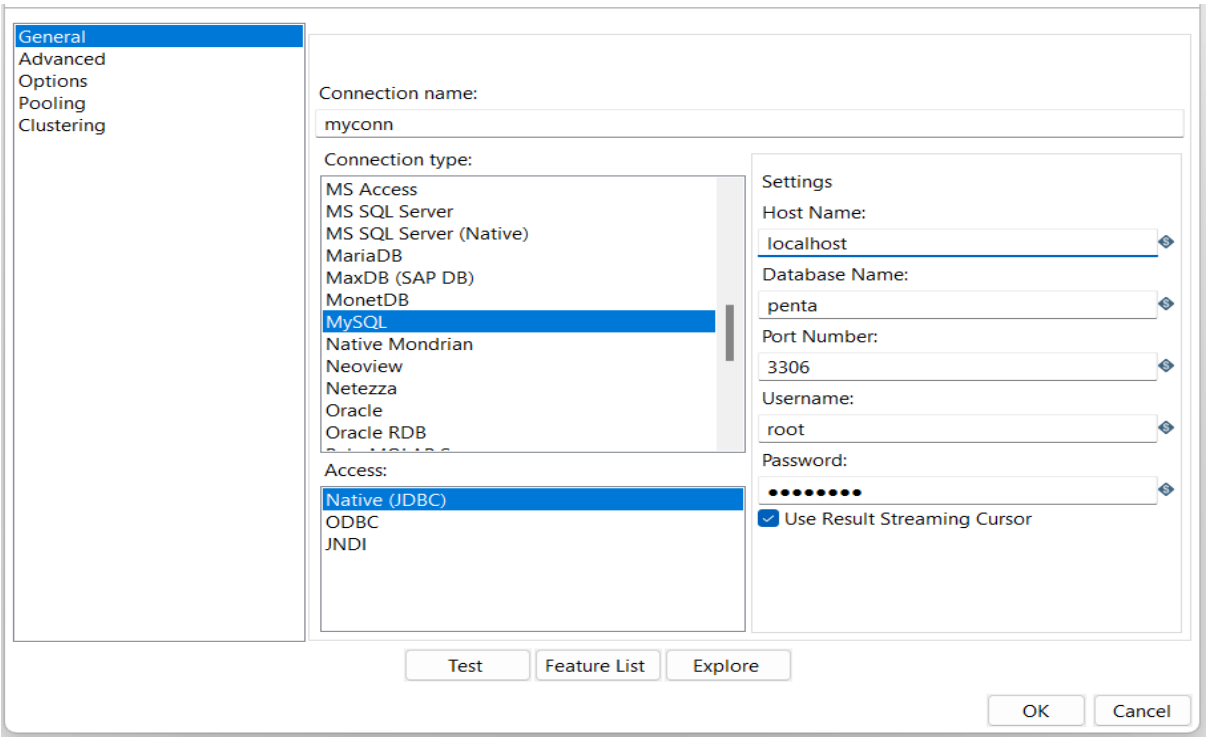
```
mysql> use penta;
Database changed
mysql> show tables;
+-----------------+
| Tables_in_penta |
+-----------------+
| prac1           |
+-----------------+
1 row in set (0.01 sec)

mysql> select * from prac1;
+------+------------+--------------+
| s_id | s_name     | s_department |
+------+------------+--------------+
|    1 | manoranjan | MCA          |
|    2 | monalisa   | MCA          |
|    3 | shriya     | MBA          |
|    4 | komal      | MBA          |
+------+------------+--------------+
4 rows in set (0.00 sec)

mysql>
```

Step 2: Open spoon batch file then click on new Transformation

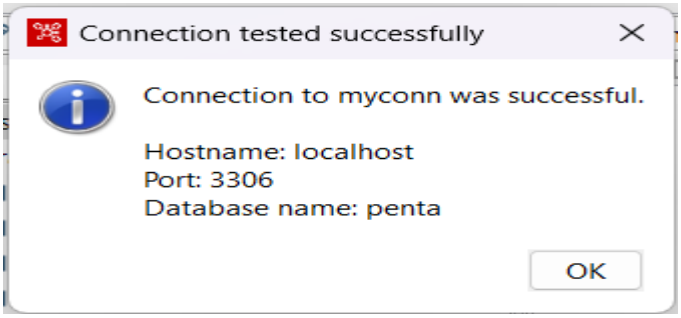


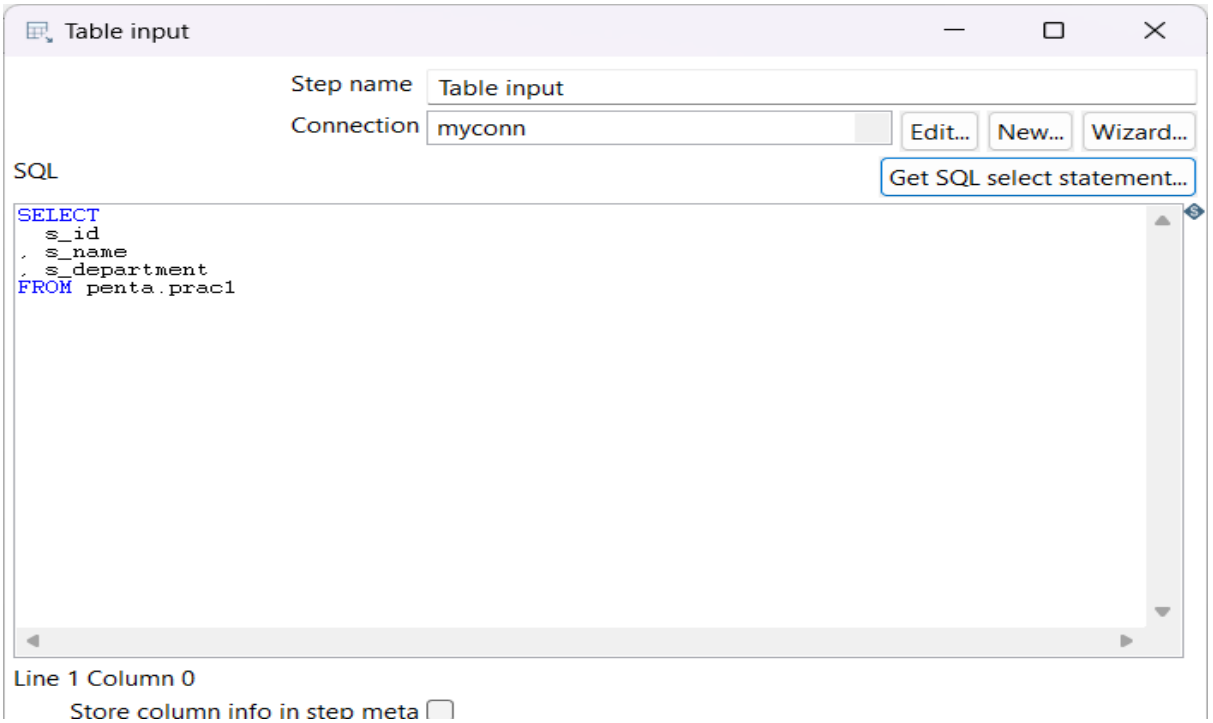
Step 3: Now drag and drop table input from input (in design tab) and excel output from the design tab and establish a connection between both by connecting the arrow.

Step 4: Double click on table input, new connection and select Get sql select statement. After this all the columns you have created will be shown in a sql statement. New Connection

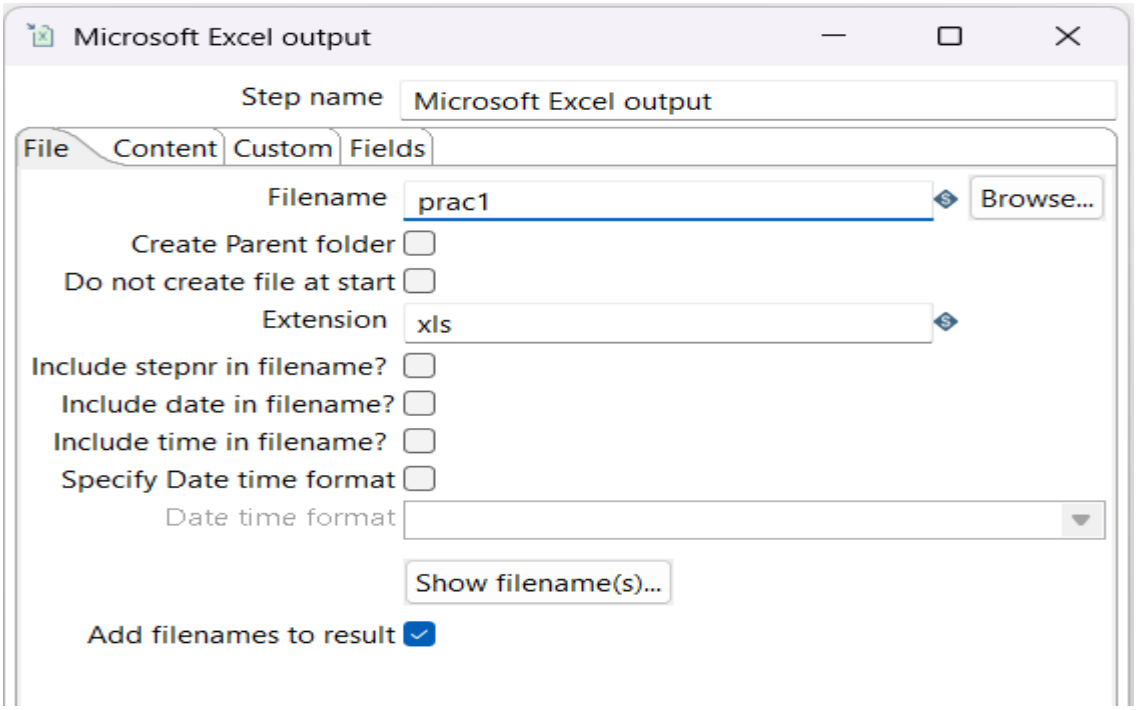


After Clicking Test

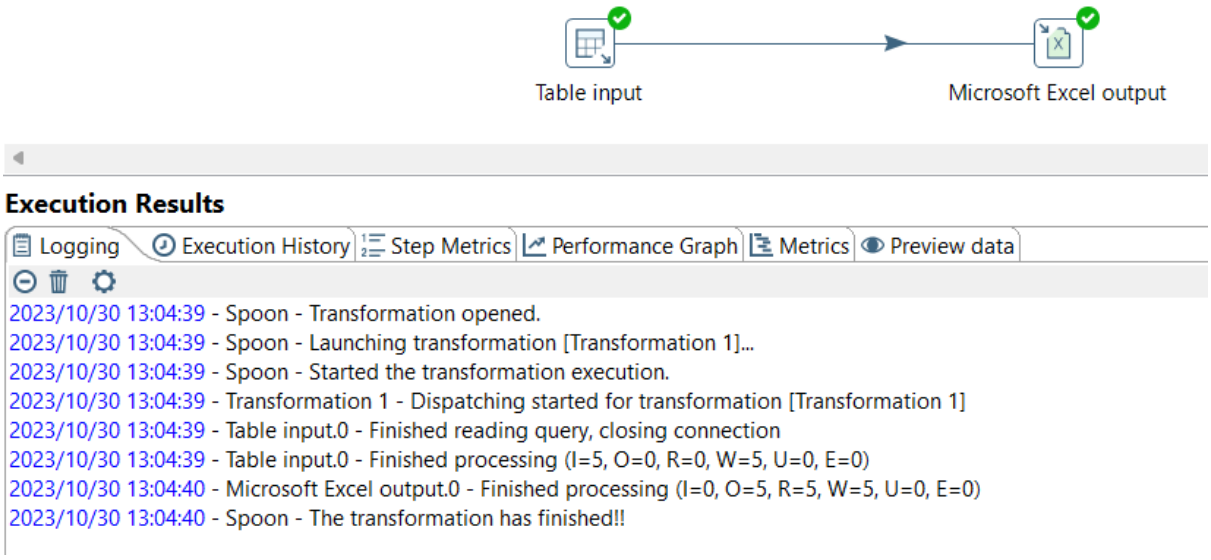




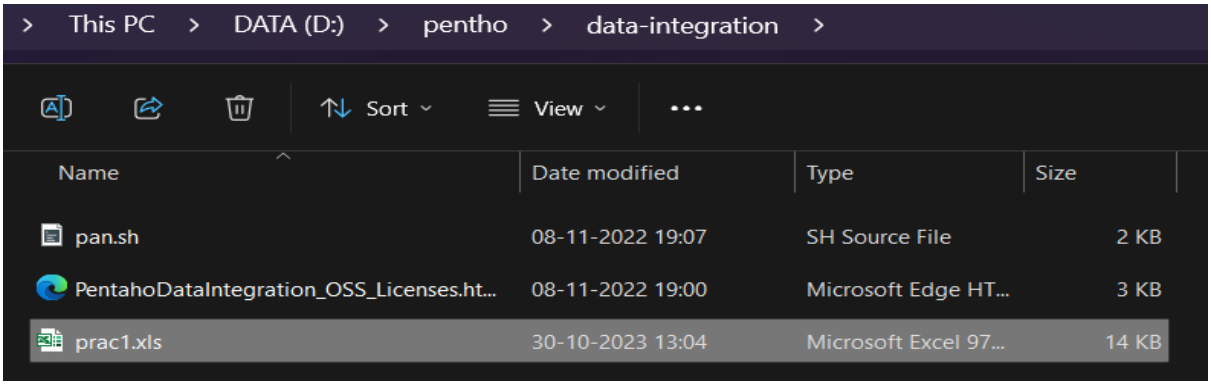
Step 5: then double click on Microsoft Excel output and inside the filename give the path of the file with the filename that you want to save.

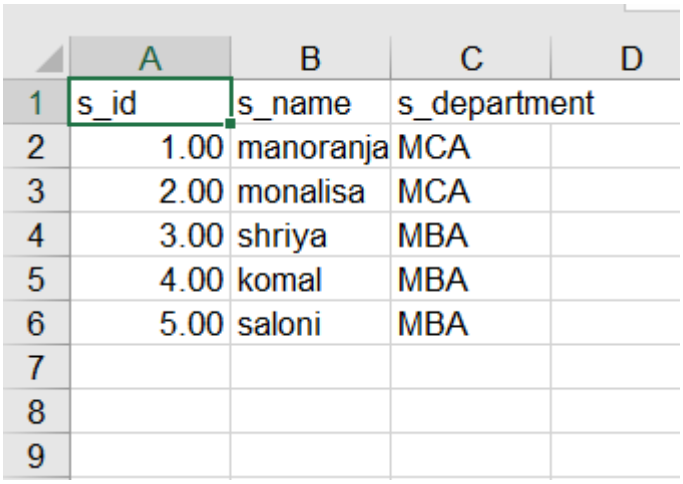


Step 6: Run the Transformation by clicking on ▶️ button on left corner



Step 7: Go to the Data Integration Folder, search for the name of the file that you have saved.





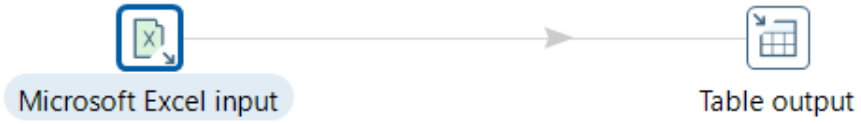
	A	B	C	D
1	s_id	s_name	s_department	
2	1.00	manoranja	MCA	
3	2.00	monalisa	MCA	
4	3.00	shriya	MBA	
5	4.00	komal	MBA	
6	5.00	saloni	MBA	
7				
8				
9				

Q2] Create a pentaho transformation to get required data from excel file and push it into mysql.

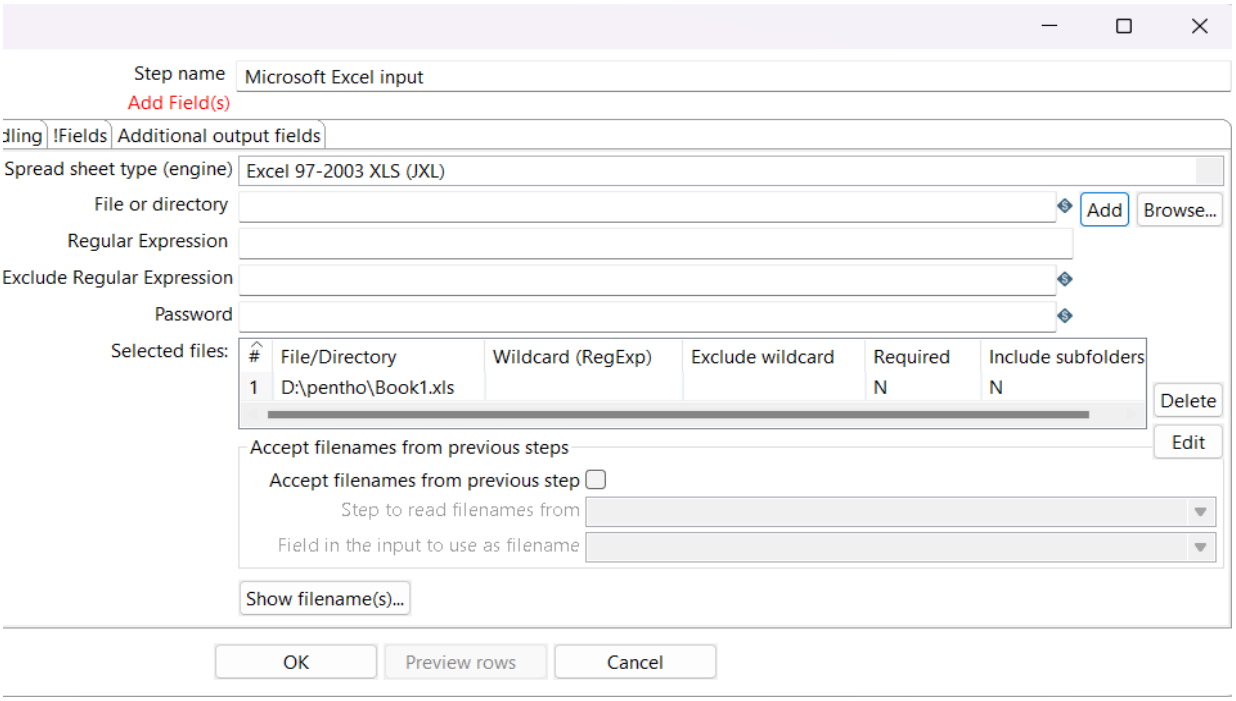
Step 1: Create a new Excel file and save it with .xls extension. Put data of your choice in the same.

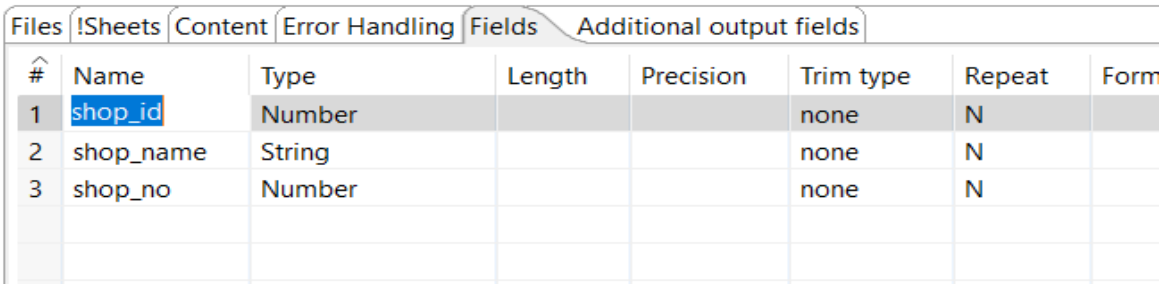
	A	B	C
1	shop_id	shop_name	shop_no
2	1	jiomart	101
3	2	dmert	102
4	3	tata	103
5	4	zudio	104
6			

Step2: Go to Pentaho -> New Transformation.
Under the Design Tab -> Input -> Select Microsoft Excel Input &
Under the Design Tab -> Output -> Select Table Output

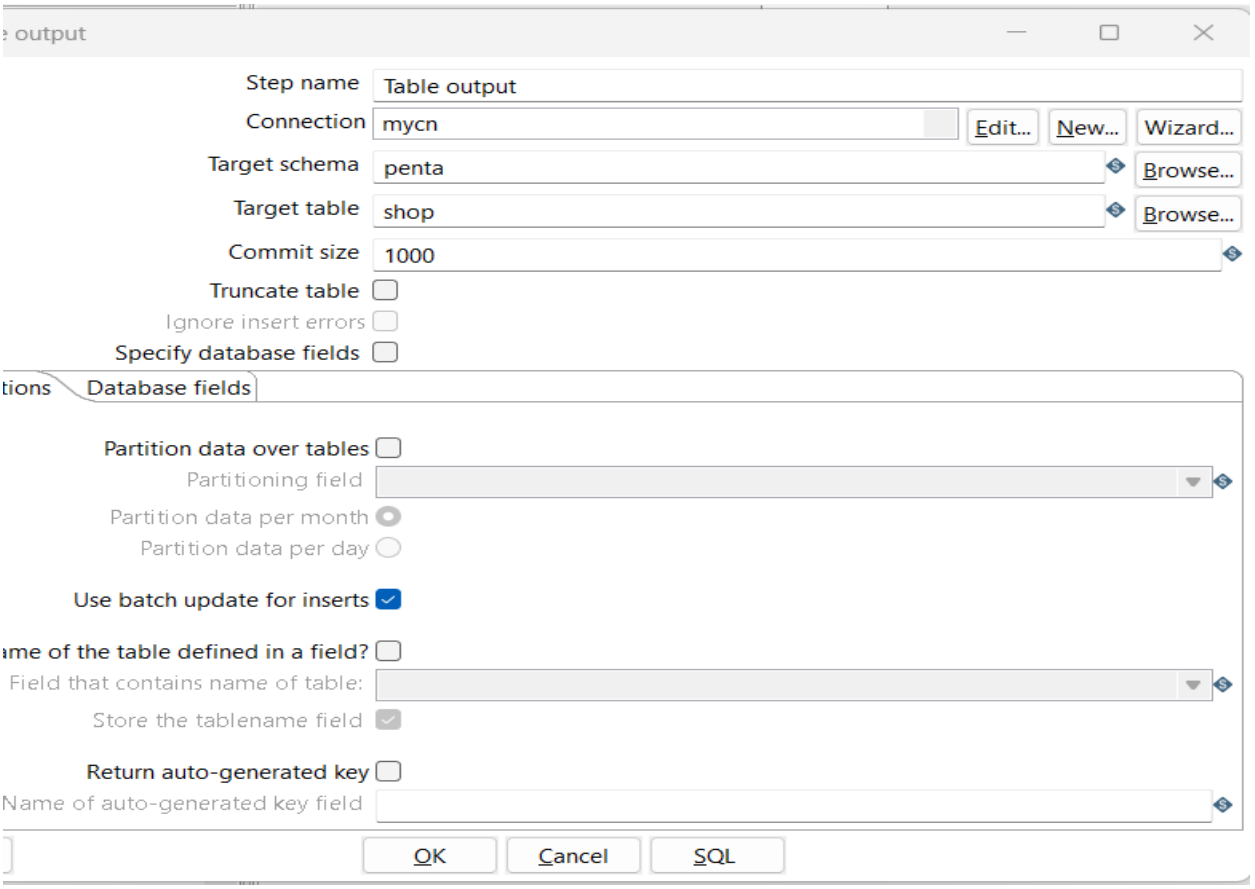


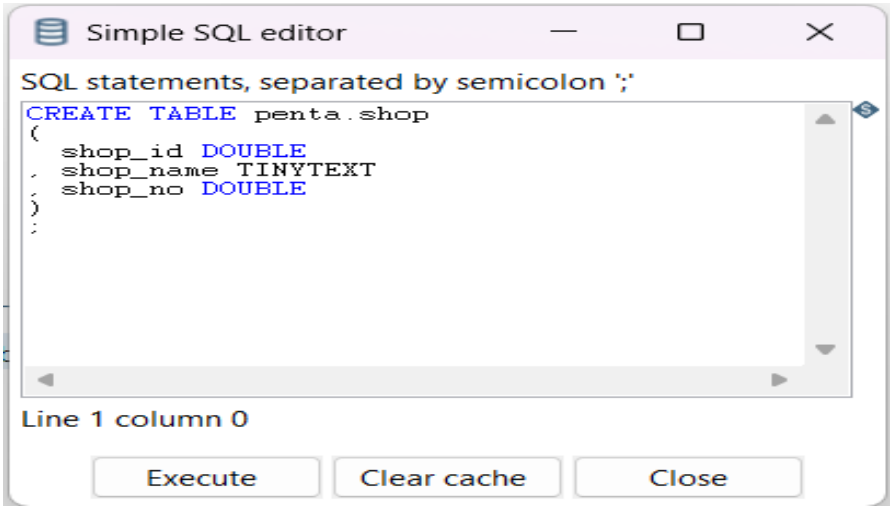
Step 4: Double Click on Microsoft Excel Input, Select the File or directory of the excel sheet created by clicking on Browse and then click on add. Under Fields tab, click on Get Fields from Header Row -> Click OK

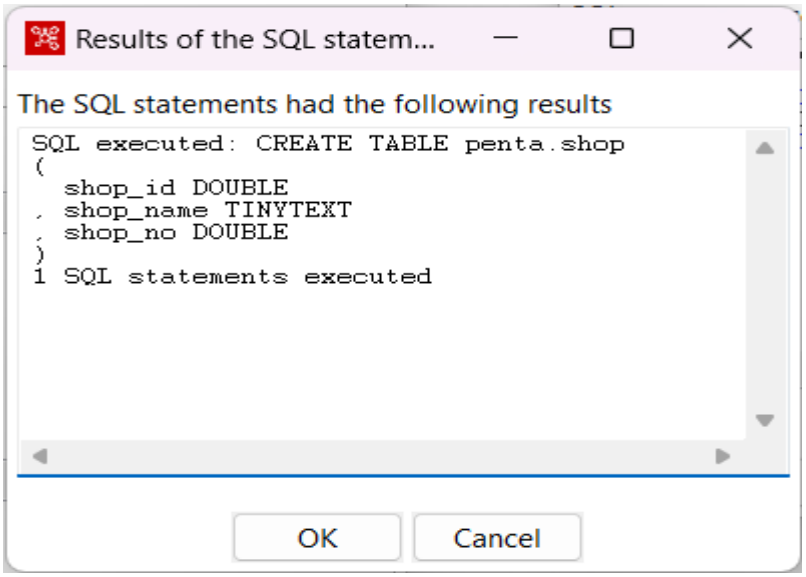




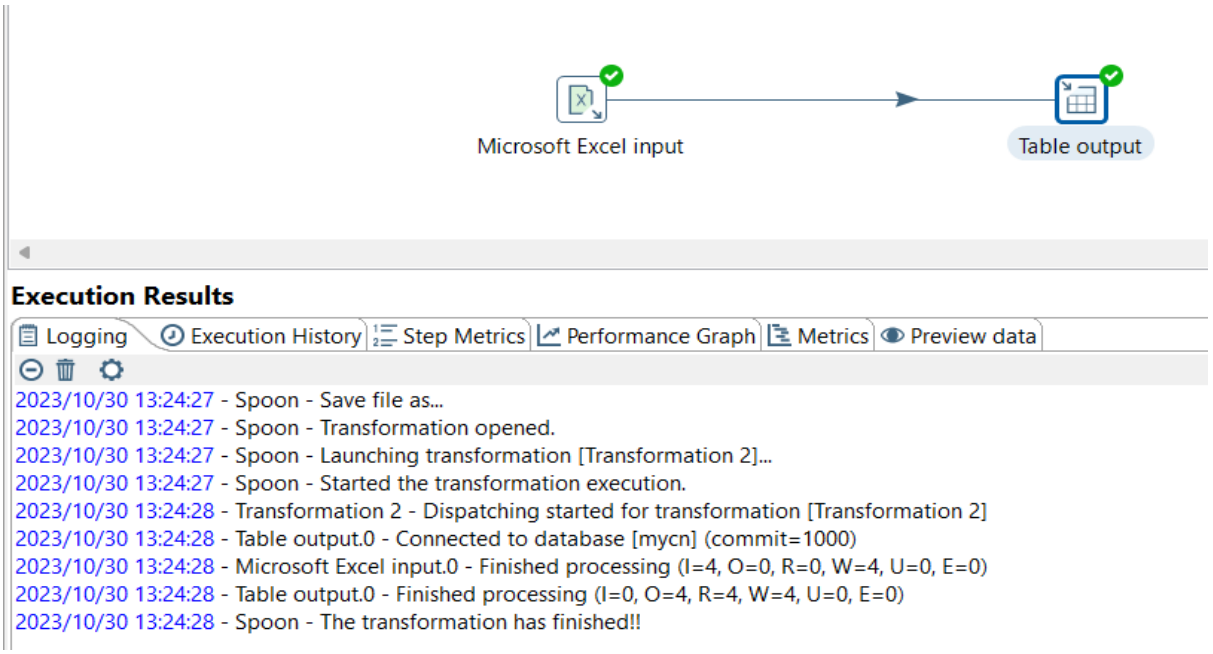
Step 5 : Double click on Table Output, Create New Connection ->Connection Type :MySQL -> Access: Native(JDBC) -> Enter Host Name, Database Name, Username and password -> Test >OK







Step 7: Run the Transformation By Clicking on ▶️ in the left corner of the window.



To Check the output: Go to MySQL and type the following Commands.

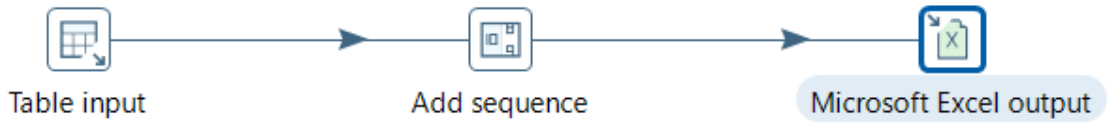
```
mysql> use penta;
Database changed
mysql> select * from shop;
+---------+-----------+---------+
| shop_id | shop_name | shop_no |
+---------+-----------+---------+
|       1 | jiomart   |     101 |
|       2 | dmert     |     102 |
|       3 | tata      |     103 |
|       4 | zudio     |     104 |
+---------+-----------+---------+
4 rows in set (0.00 sec)

mysql>
```

Q3] Create a pentaho transformation to get data from mysql table, add serial no to the data using pentaho sequences and push data into excel.

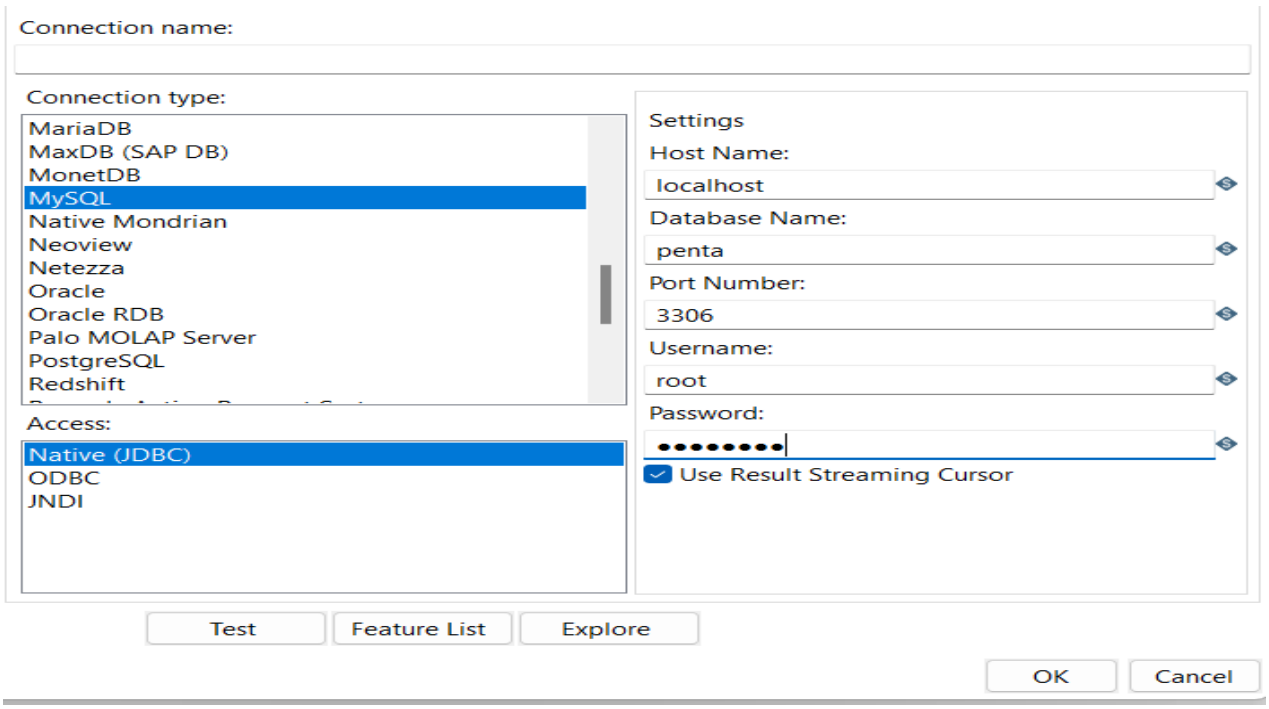
Step1: Create a table student and add some data.

```
mysql> use penta;
Database changed
mysql> select * from student;
+-------------+---------------+---------+
| s_name      | s_department  | gender  |
+-------------+---------------+---------+
| manoranjan  | MCA           | Male    |
| monalisa    | MCA           | Female  |
| shriya      | MCA           | Female  |
| Harshal     | MCA           | Male    |
+-------------+---------------+---------+
4 rows in set (0.01 sec)

mysql>
```

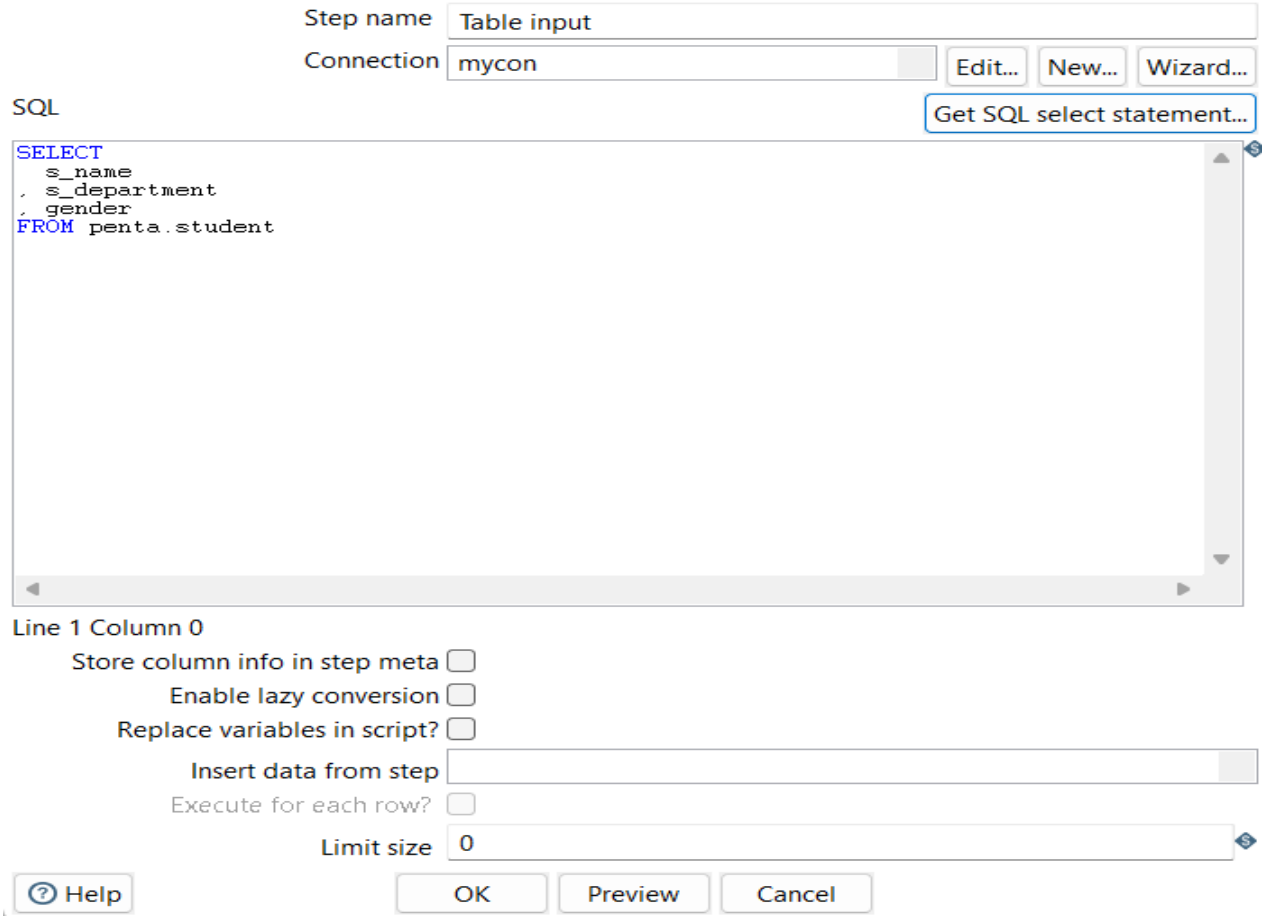
Step2: Open Pentaho -> New Transformation. Under Design Tab, Drag and drop Table Input(From Input), Add Sequence(From Transform) and Microsoft Excel Output(From output).



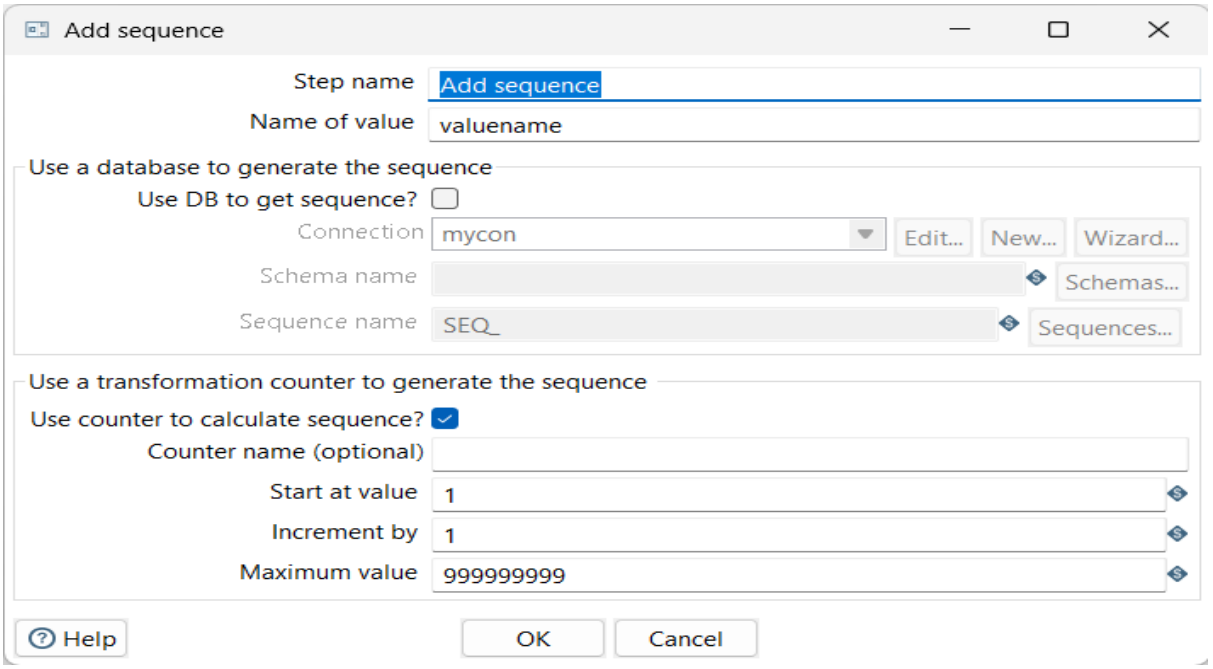
Step 3: Double Click On Table Input (my_sql_data) -> New Connection



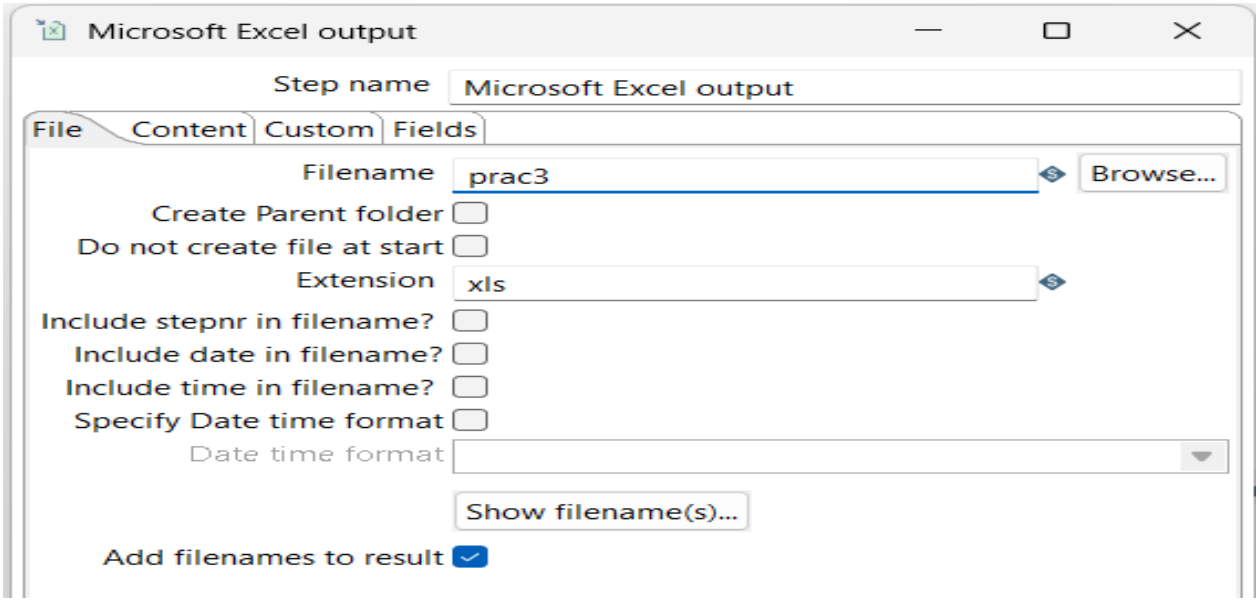
Step 4:After Creating New Connection Select Get SQL select Statement and select the table name -> Click OK.



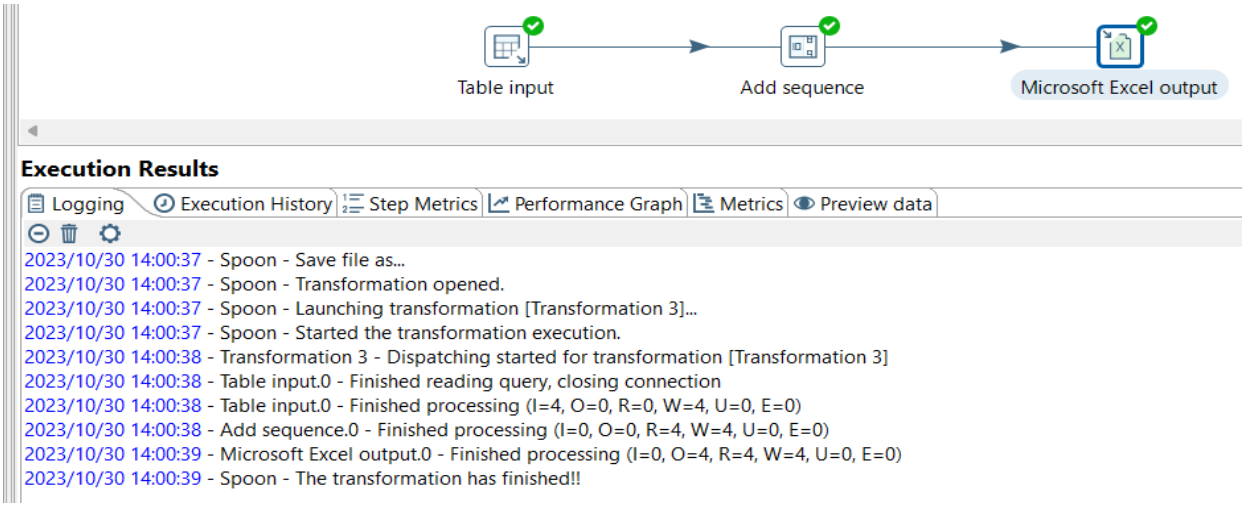
Step 5: Double Click On Add Sequence -> Give the Name of value -> OK



Step 6: Double Click on Microsoft Excel Output -> Give the File Name -> OK



Step 7: To Run the transformation click on the ▶ button on the left corner of the window.



Step 8: Check the Output from the Data Integration Folder by opening the file with the name saved by the user.

	A	B	C	D	E
1	s_name	s_departm	gender	srno	
2	manoranja	MCA	Male	1.00	
3	monalisa	MCA	Female	2.00	
4	shriya	MCA	Female	3.00	
5	Harshal	MCA	Male	4.00	
6					
7					
8					

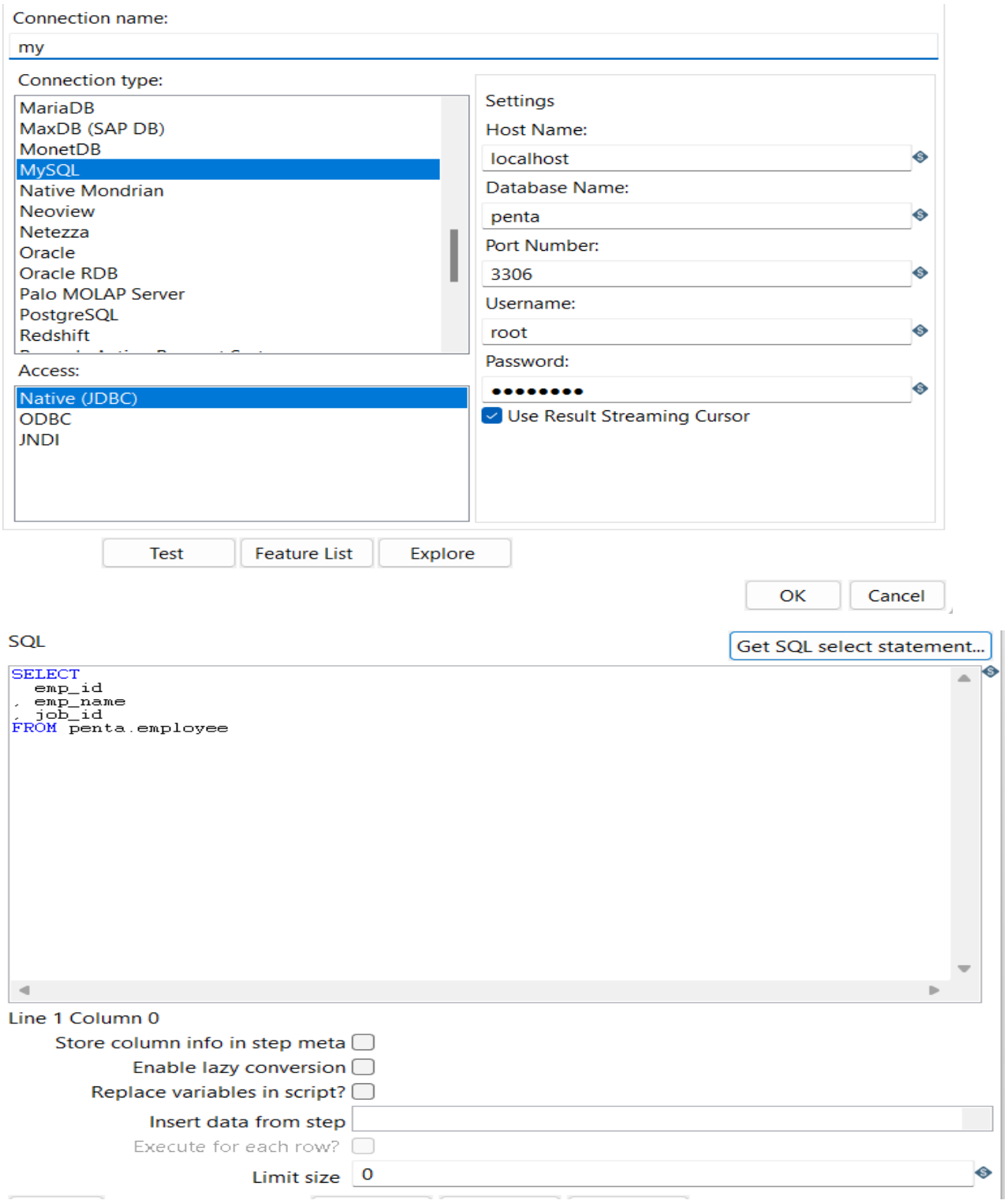
Q4] Create a pentaho transformation to get data from a mysql table named as employee have the following columns (emp_id,full name and job_id). Split the full name of the employee using pentaho split transformation and store the data into a new table in mysql name as emp1 having the following columns (emp_id,fname,mname,lname and job_id)

Step 1: Create a new Table employee in the database. Insert Values in it.

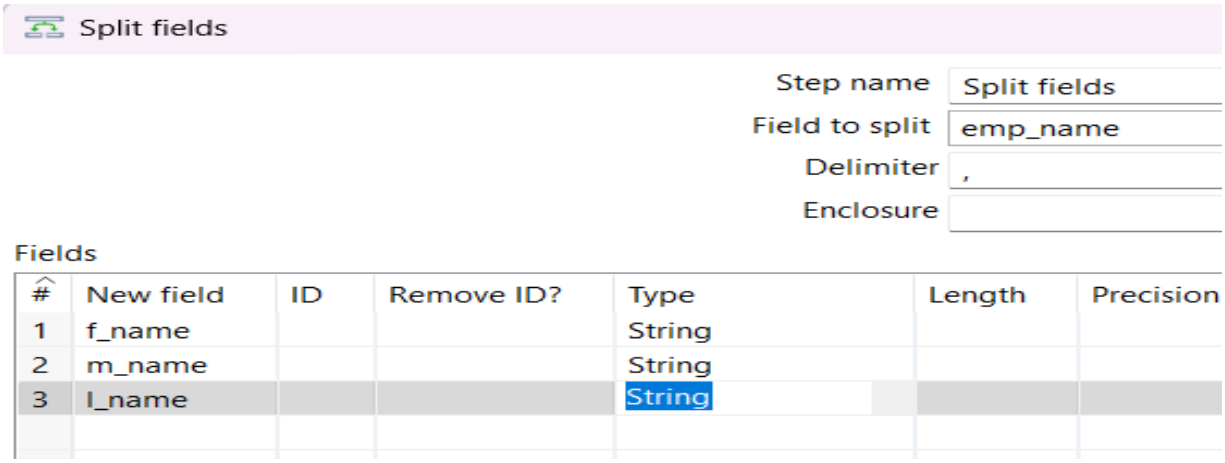
```
mysql> use penta;
Database changed
mysql> select * from employee;
+--------+---------------------------+--------+
| emp_id | emp_name                  | job_id |
+--------+---------------------------+--------+
|      1 | Manoranjan Mangaraj Baral |    101 |
|      2 | Monalisa Mangaraj Baral   |    102 |
|      3 | Shriya Narayan Rane       |    103 |
|      4 | Komal Vijay Bhamble       |    104 |
+--------+---------------------------+--------+
4 rows in set (0.00 sec)

mysql>
```

Step 2: Open Pentaho -> New Transformation -> Drag and Drop Table Input(From Input), Split fields and Table Output (From Output)

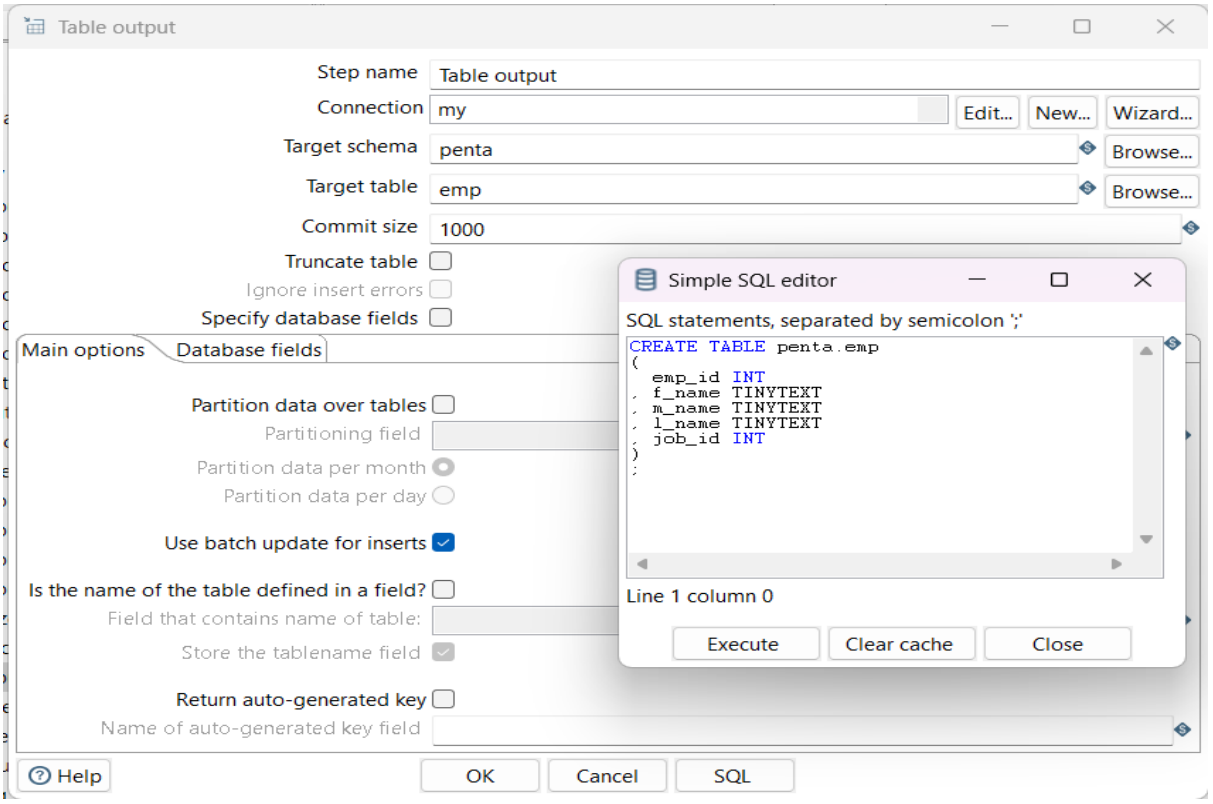
Step 3: Double Click on Table Input



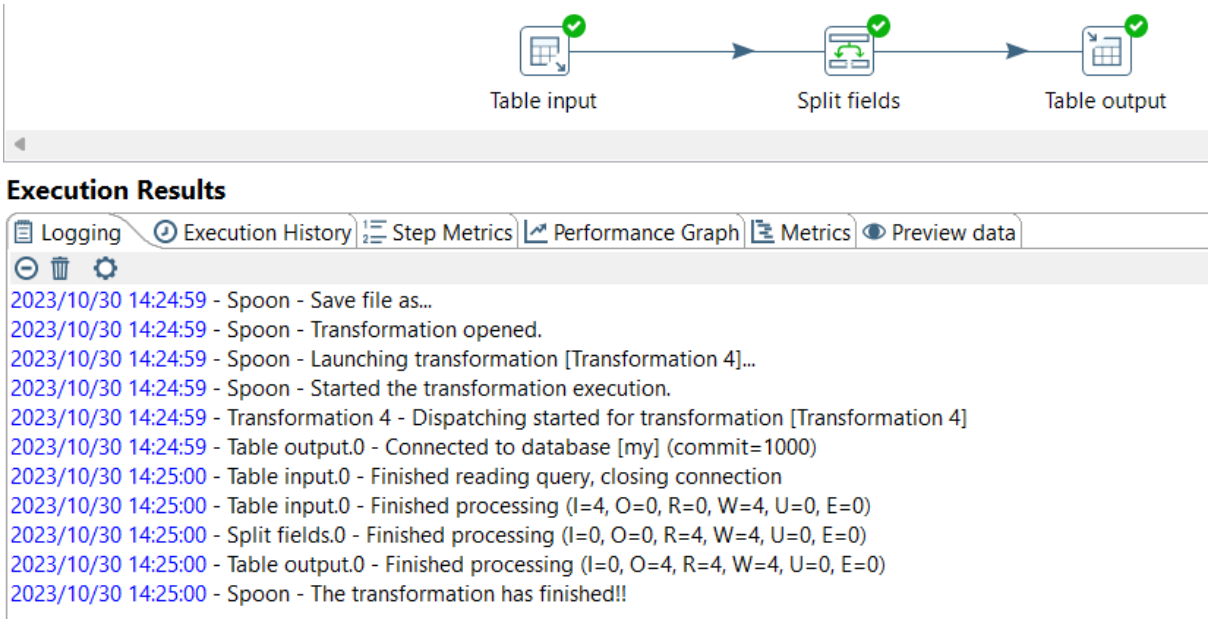
Step 4: Double click on split fields .inside Field to split write the name of the column you want to split i.e full_name then inside delimiter give single space (it will split after space).then add 3 field name f_name,m_name and l_name and specify its type as string . -->OK



Step 5: Double click on table output, give the connection as usual and execute the SQL queries .



Step 6: After completing all the above steps properly click on Run.

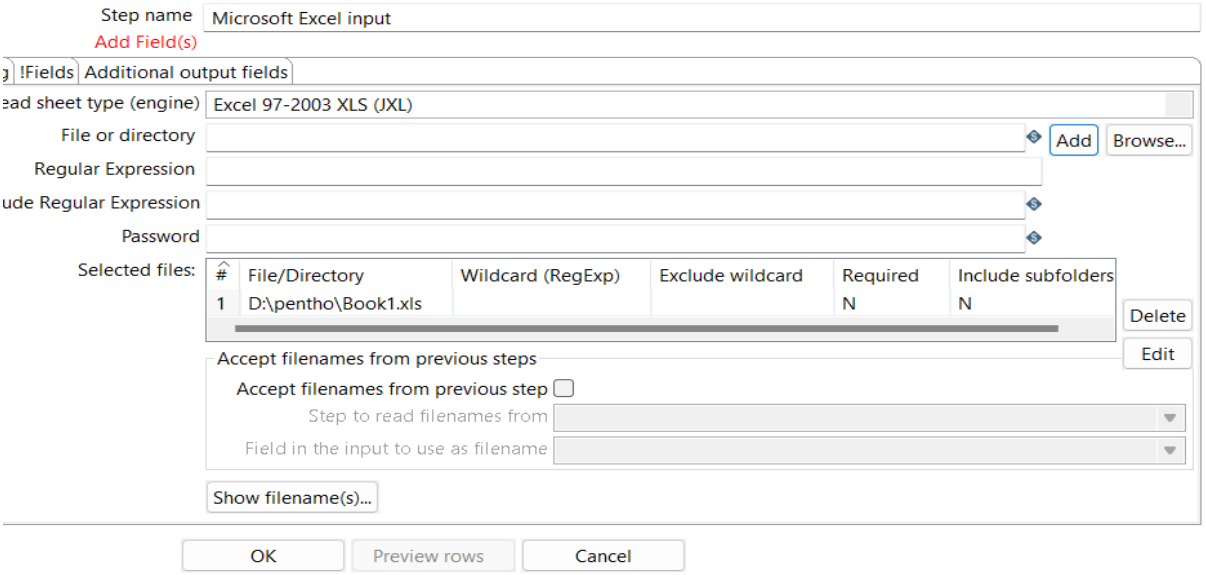


Step 7: Open MySQL and type the following Command.

```
mysql> select * from emp;
+--------+-----------+----------+---------+--------+
| emp_id | f_name    | m_name   | l_name  | job_id |
+--------+-----------+----------+---------+--------+
|      1 | Manoranjan | Mangaraj | Baral   |    101 |
|      2 | Monalisa  | Mangaraj | Baral   |    102 |
|      3 | Shriya    | Narayan  | Rane    |    103 |
|      4 | Komal     | Vijay    | Bhamble |    104 |
+--------+-----------+----------+---------+--------+
4 rows in set (0.00 sec)

mysql>
```

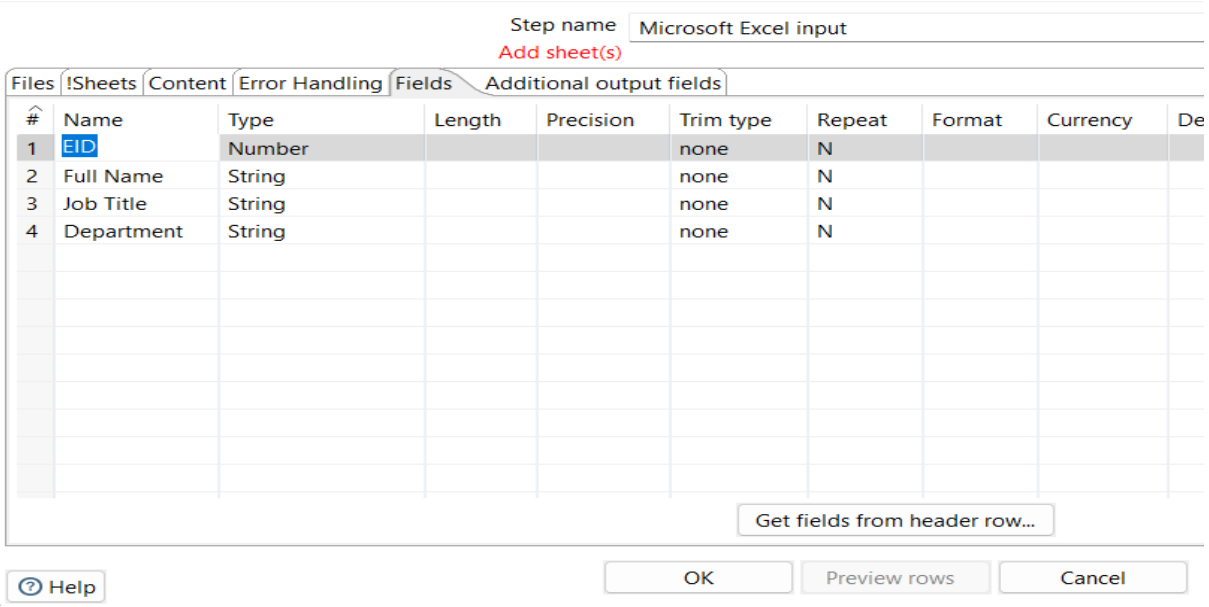
Q5] create two excel files with following columns (a) data 1: EID,Full Name,Job,Title,Department. data 2:EID,Business,Unit,Gender,Ethnicity,Age.EID in data 1 points to EID in data 2 . Make sure to have at least 10 to 15 rows in the file. sort the data using sort transformation of pentaho.
(b) Merge join the sorted in pentaho and remove repeated rows and save the data in mysql table.

Step 1: Open Spoon. Then take 2 input Microsoft excel and name it as data 1 and data 2.
Step 2: Double Click on input excel and browse the data 1 file and select it after selecting click on add.



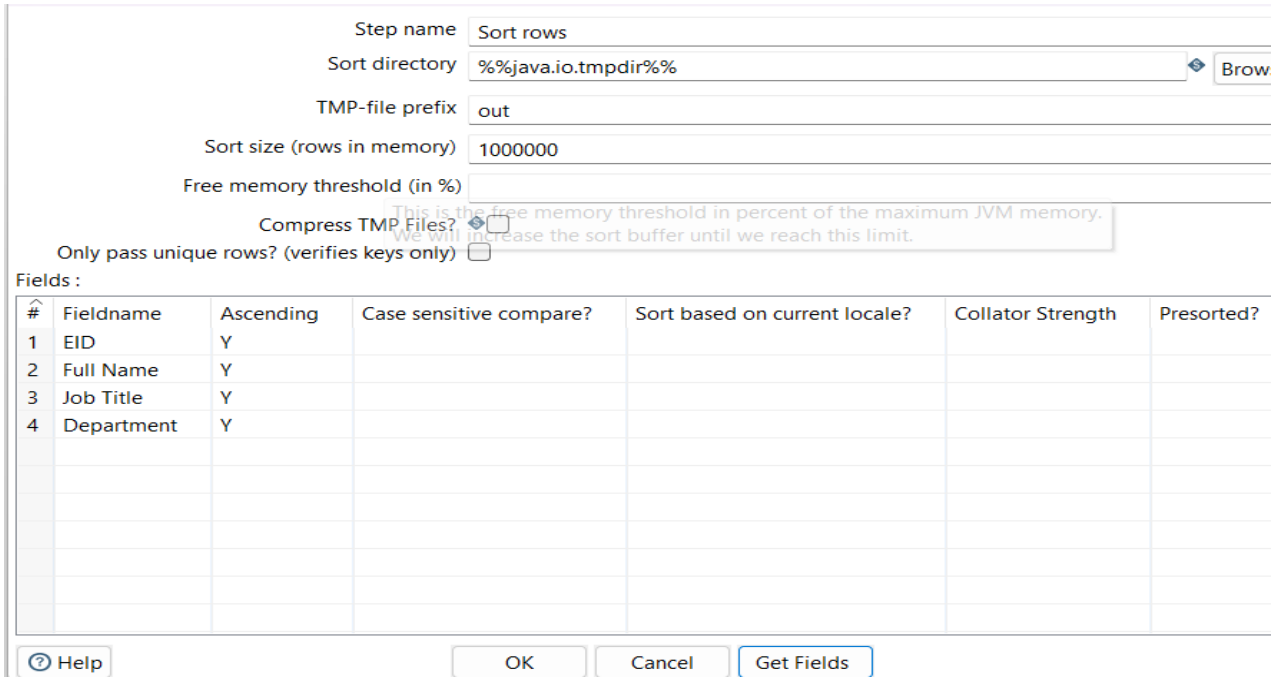
Step 3: Then click on field option and select Get field from header row button.

Do The same steps for Data 2 excel file.

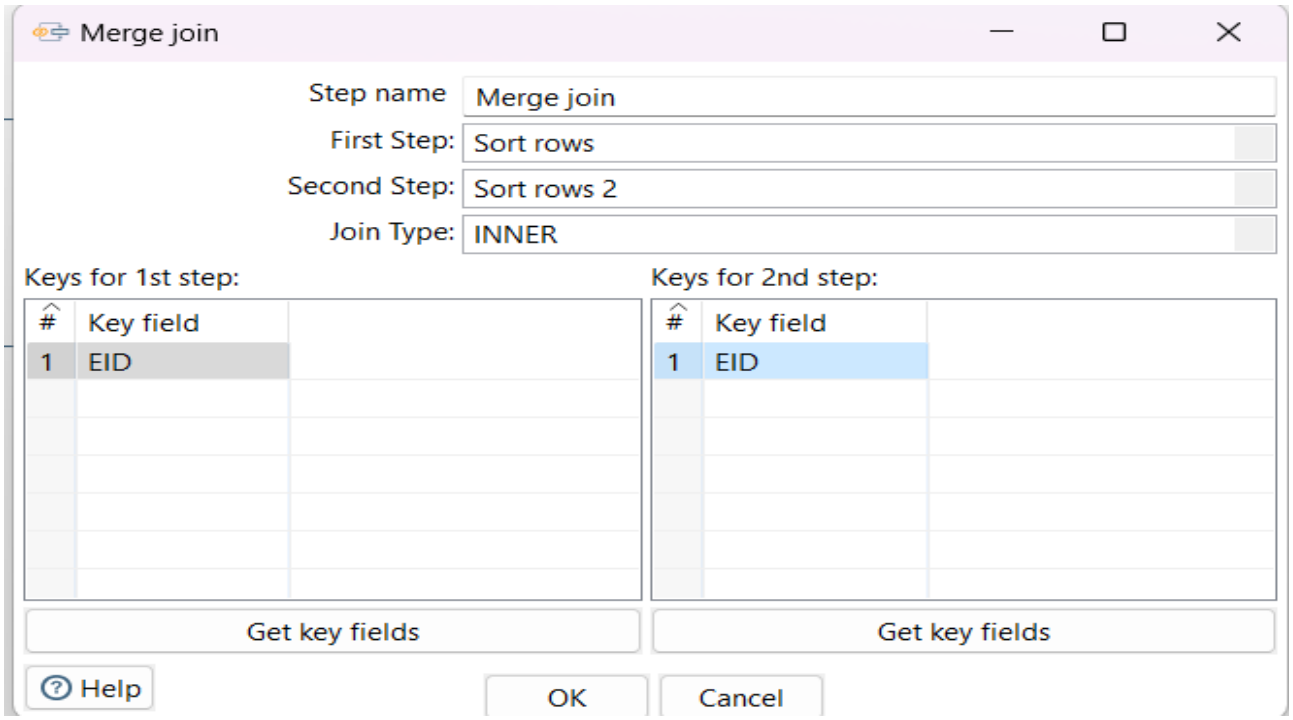


Step 4: Then select 2 sort rows from Transform and connect those to that excel input files.

Step 5: Double click on sort rows and click on Fields button. This will display the headers of the columns in that particular excel that you imported. Do the same step for sort row 2.



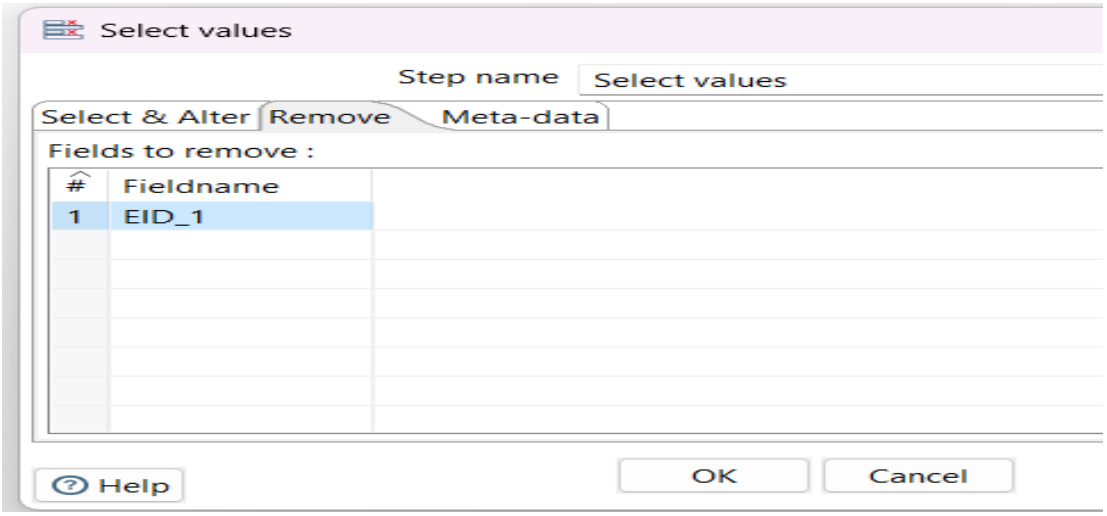
Step 6: From join drag merge join and connect both sort rows with merge join.
Step 7: Double click merge join. Then inside First step type sort rows and In second step select sort row 2 and click on Get key fields for both steps.



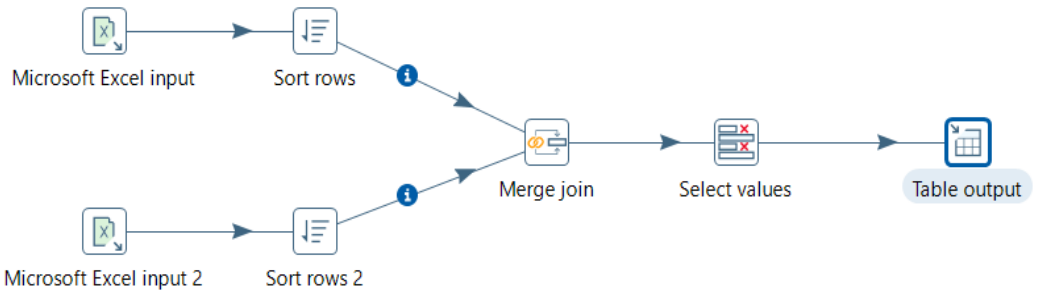
Then , delete all uncommon columns and keep the common column i.e EID.

Step 8: Then from transform drag select values and connect it to merge join.

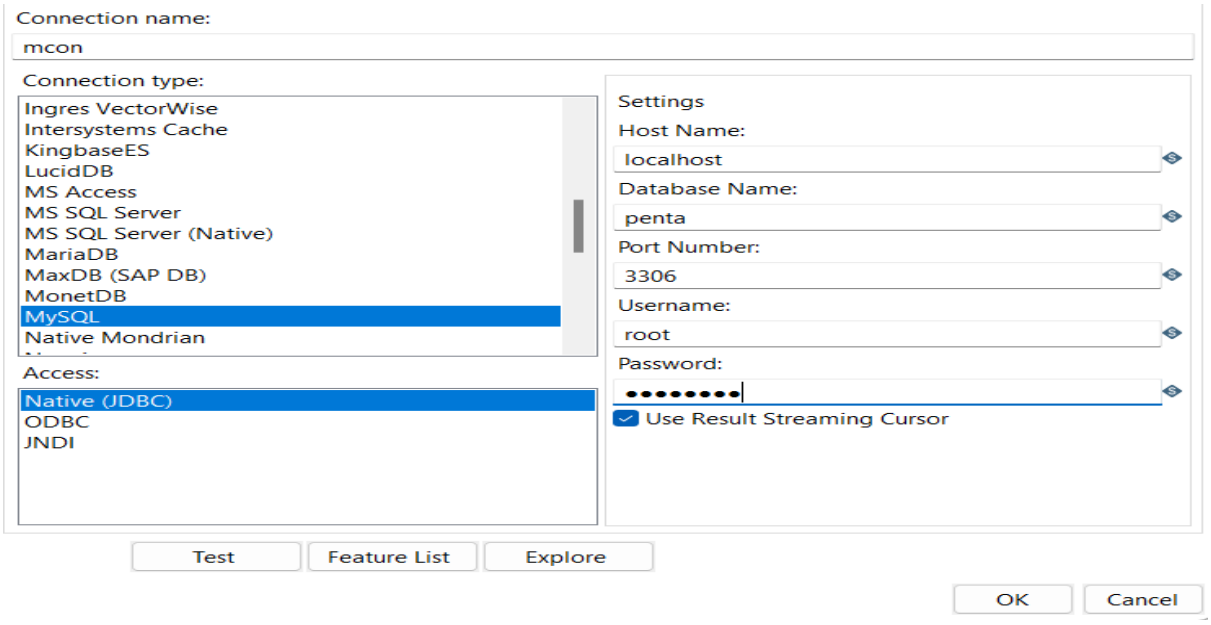
Step 9: Double click on select values and click on get fields to select . we will be able to see 2 EIDs. Then we have to delete EID_1 from the fields section and click on OK.



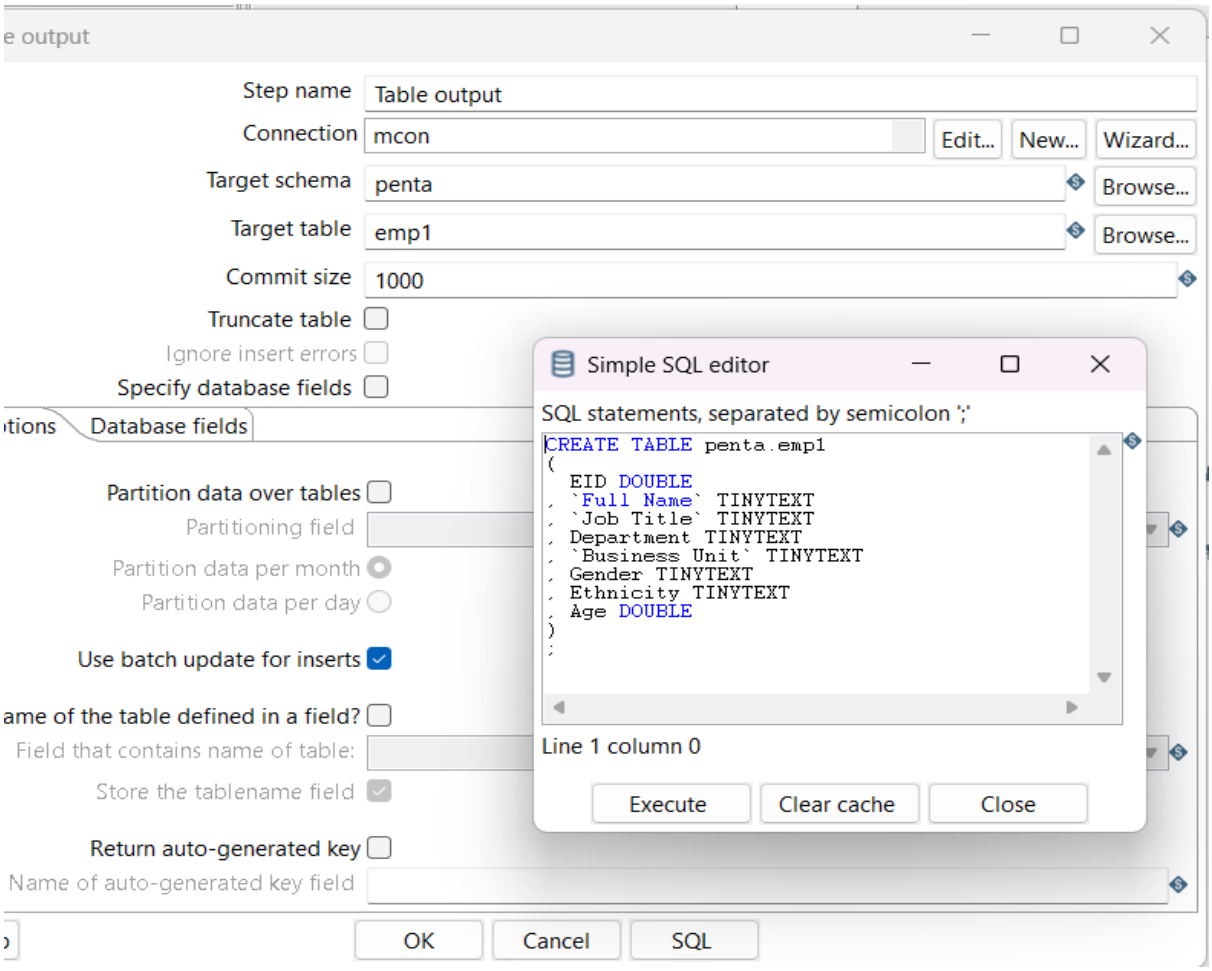
Step 10: Drag table output from output session as shown below.



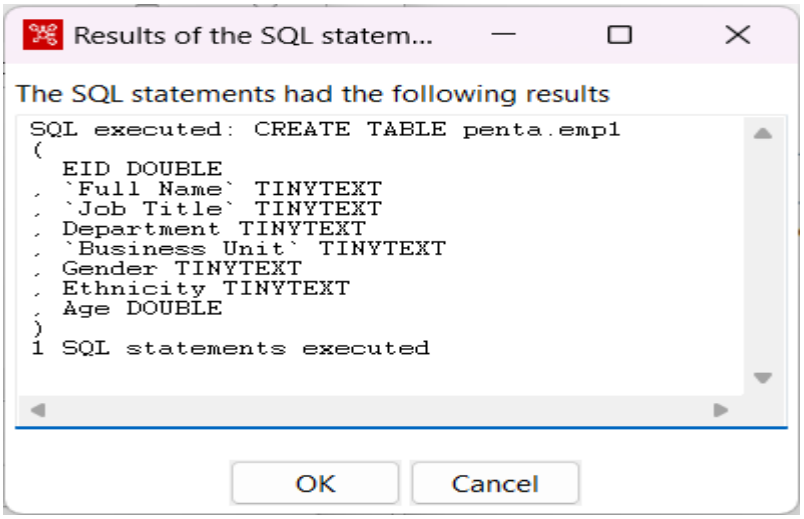
Step 11: Double click on Table Output -> New Connection -> OK



Step 12: Click on SQL

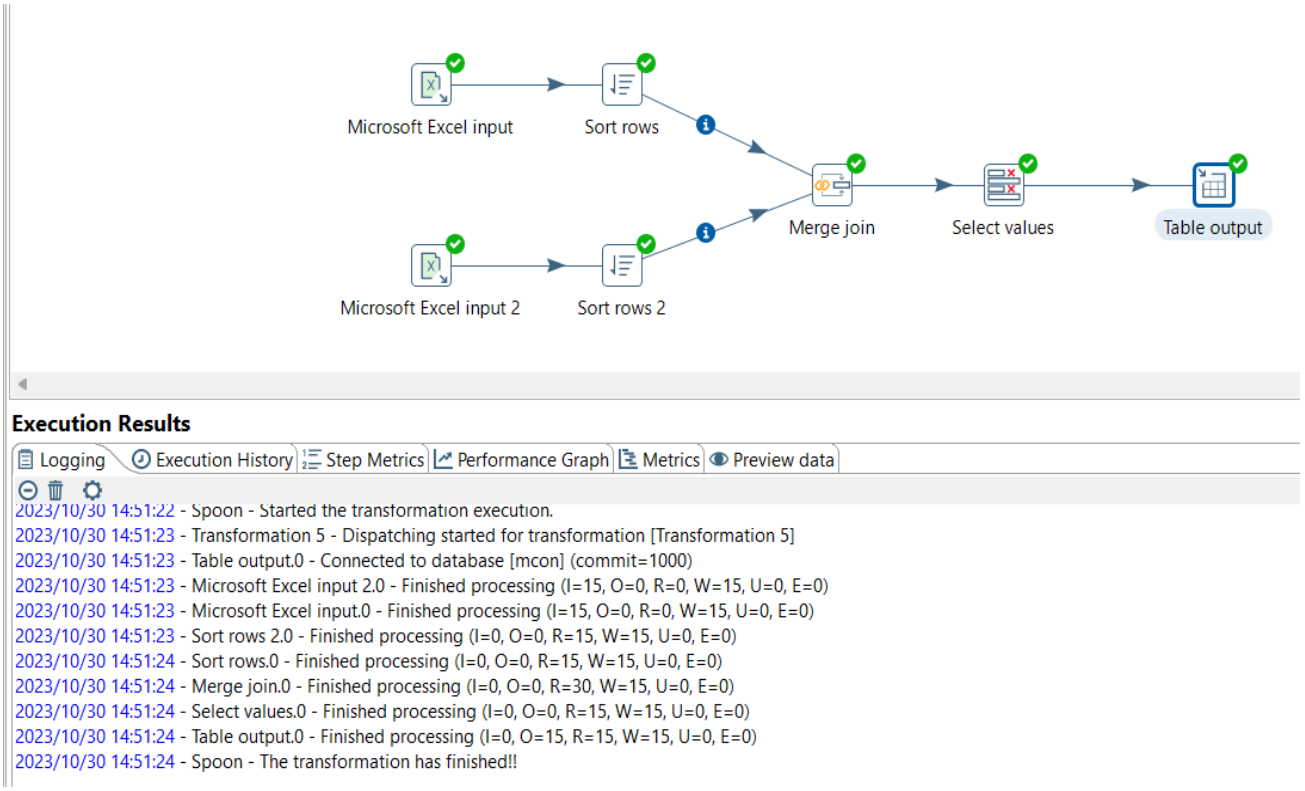


Click on Execute



Click OK

Step 13: To run the transformation click on ▶ button on the left corner of the window.

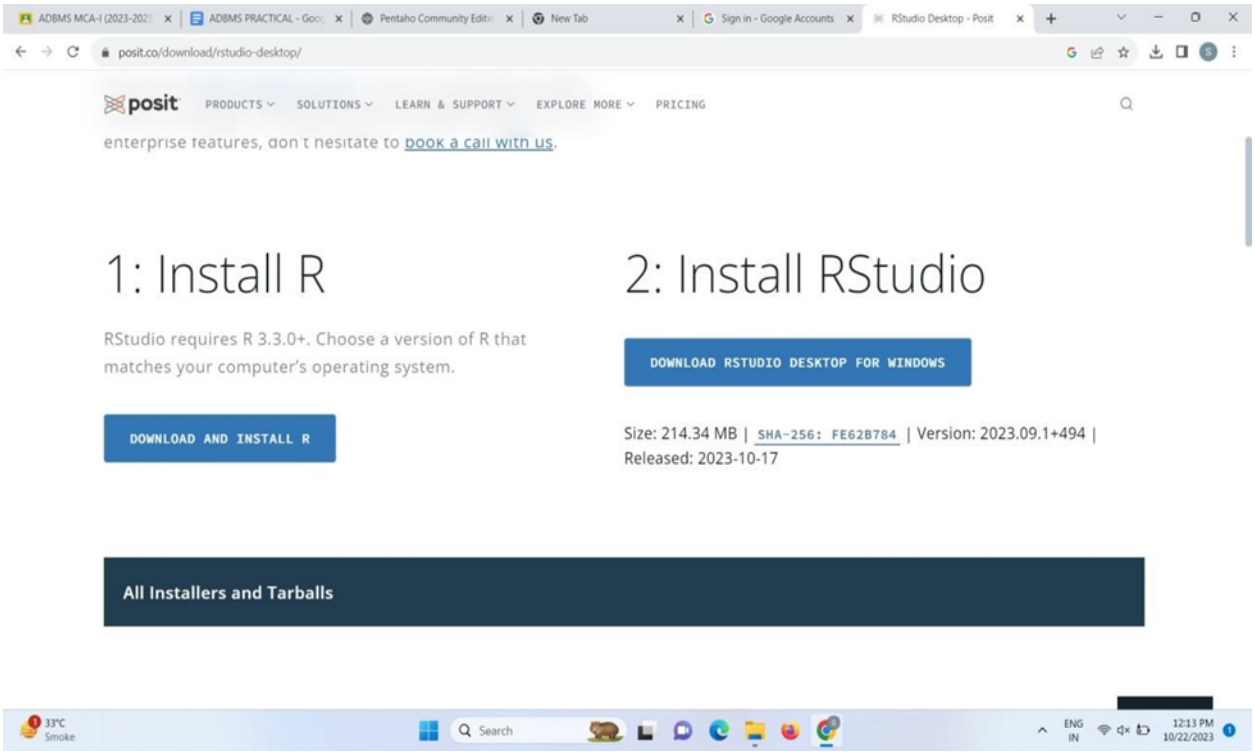


Step 14: Execute the following Command on MySQL to Get the Output.

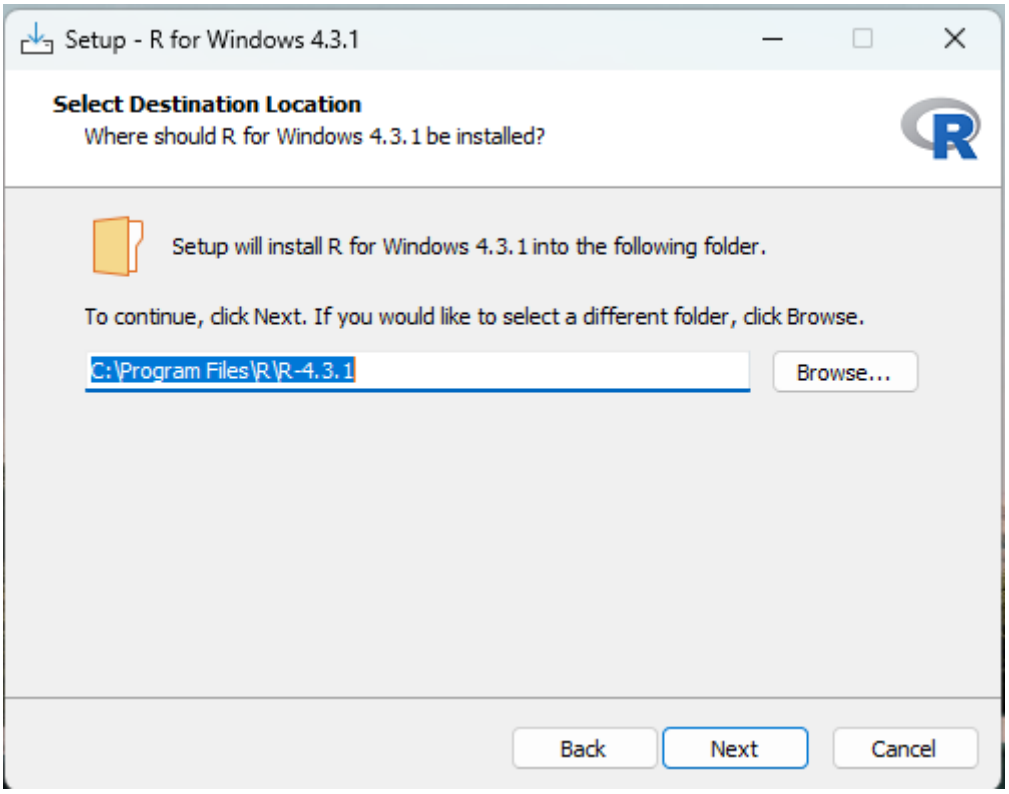
```
mysql> use penta;
Database changed
mysql> select * from emp1;
+------+----------------+-------------------------+----------------+-----------------------+--------+-----------+------+
| EID  | Full Name      | Job Title               | Department     | Business Unit         | Gender | Ethnicity | Age  |
+------+----------------+-------------------------+----------------+-----------------------+--------+-----------+------+
|    1 | Emily Davis    | Sr. Manger              | IT             | Research & Development | Female | Black     |   55 |
|    2 | Luna Sanders   | Director                | Finance        | Speciality Products   | Female | Caucasian |   50 |
|    3 | Penelope Jordan | Computer Systems Manager | IT            | Manufacturing         | Female | Caucasian |   26 |
|    4 | Theodore Dinh  | Technical Architect     | IT             | Manufacturing         | Male   | Asian     |   59 |
|    5 | Austin Vo      | Sr. Analyst             | Finance        | Manufacturing         | Male   | Asian     |   55 |
|    6 | Ruby Barnes    | Manager                 | IT             | Corporate             | Female | Caucasian |   27 |
|    7 | Joshua Gupta   | Account Representative  | Sales          | Corporate             | Male   | Asian     |   57 |
|    8 | Luke Martin    | Analyst                 | Finance        | Manufacturing         | Male   | Black     |   25 |
|    9 | Easton Bailey  | Manager                 | Accounting     | Manufacturing         | Male   | Caucasian |   29 |
|   10 | Savannah Ali   | Sr. Manger              | Human Resources | Manufacturing        | Female | Asian     |   36 |
|   11 | Madeline Walker | Sr. Analyst             | Finance        | Speciality Products   | Female | Caucasian |   34 |
|   12 | Eli Jones      | Manager                 | Human Resources | Manufacturing        | Male   | Caucasian |   59 |
|   13 | Camila Rogers  | Controls Engineer       | Engineering    | Speciality Products   | Female | Caucasian |   27 |
|   14 | Robert Yang    | Sr. Analyst             | Accounting     | Speciality Products   | Male   | Asian     |   31 |
|   15 | Everleigh Ng   | Sr. Manger              | Finance        | Research & Development | Female | Asian     |   51 |
+------+----------------+-------------------------+----------------+-----------------------+--------+-----------+------+
15 rows in set (0.00 sec)

mysql>
```

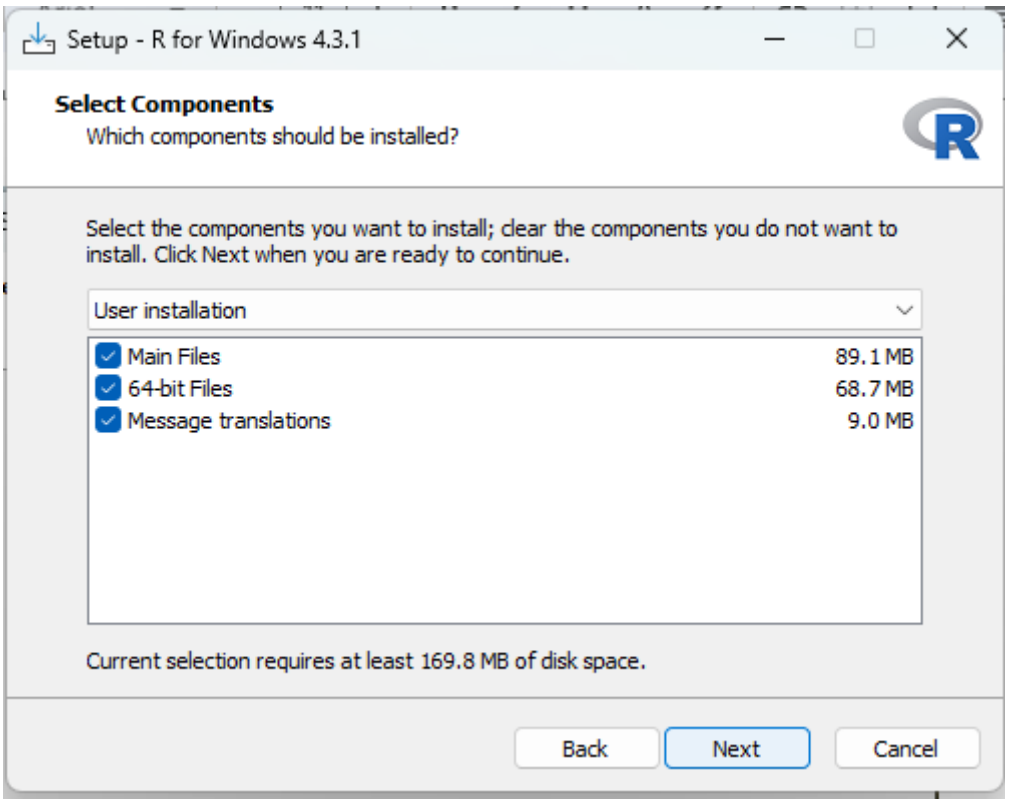
Practical no.5

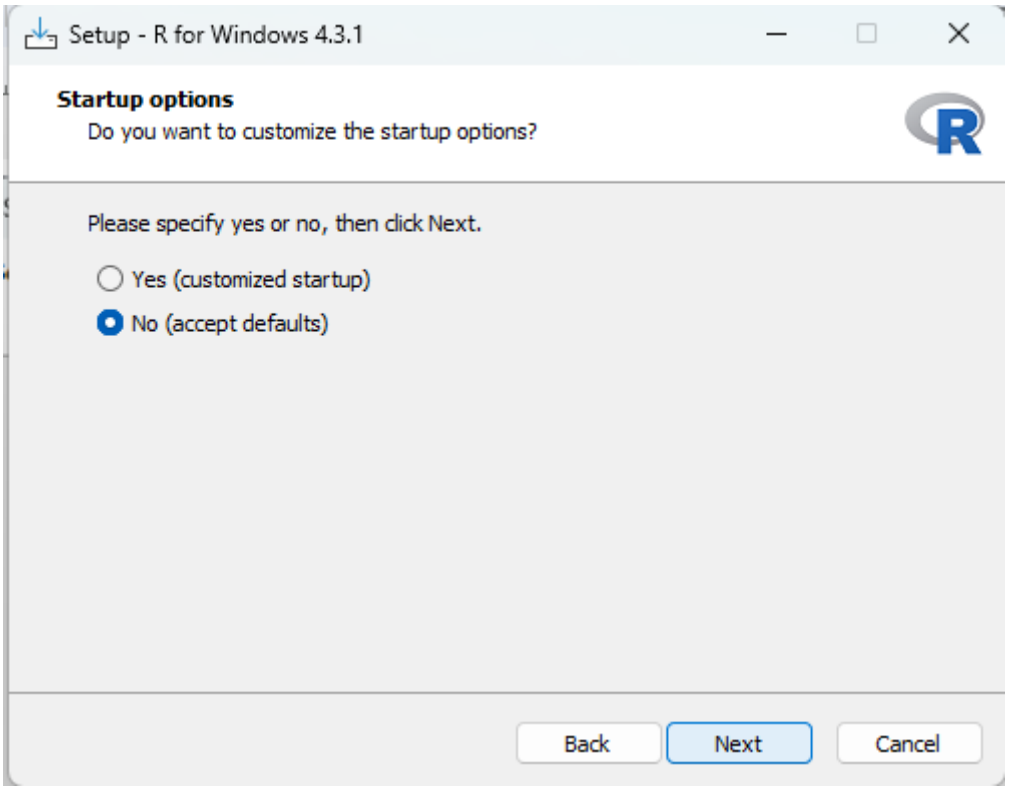
R and RStudio Installation

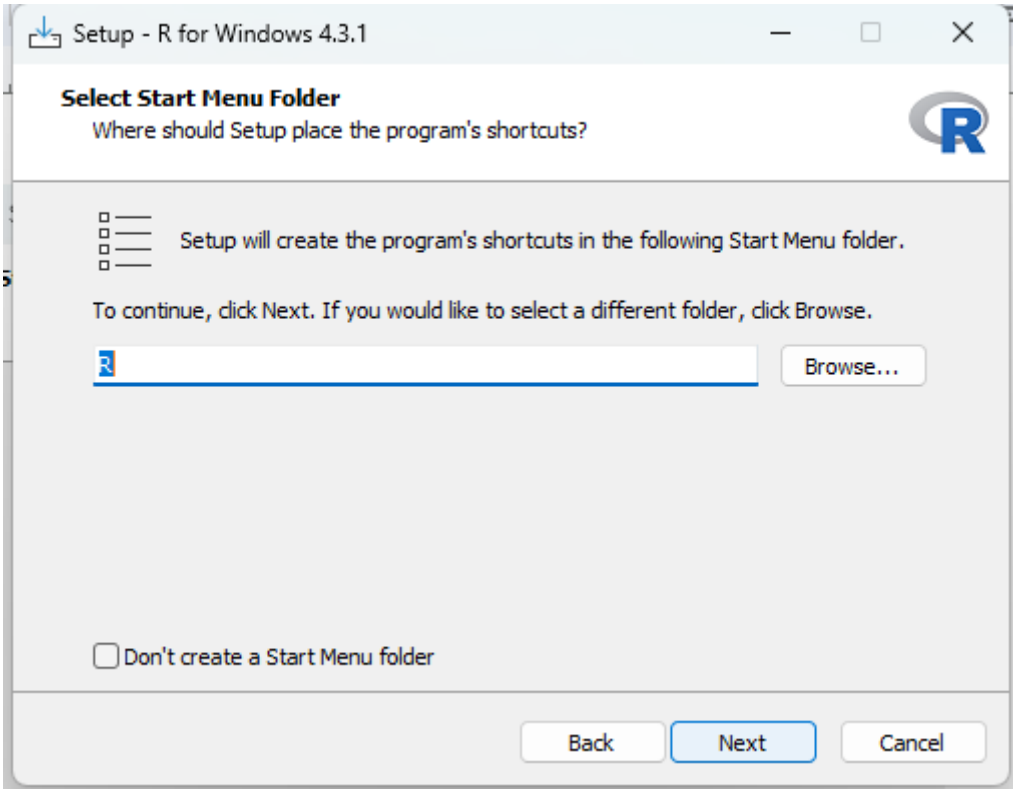


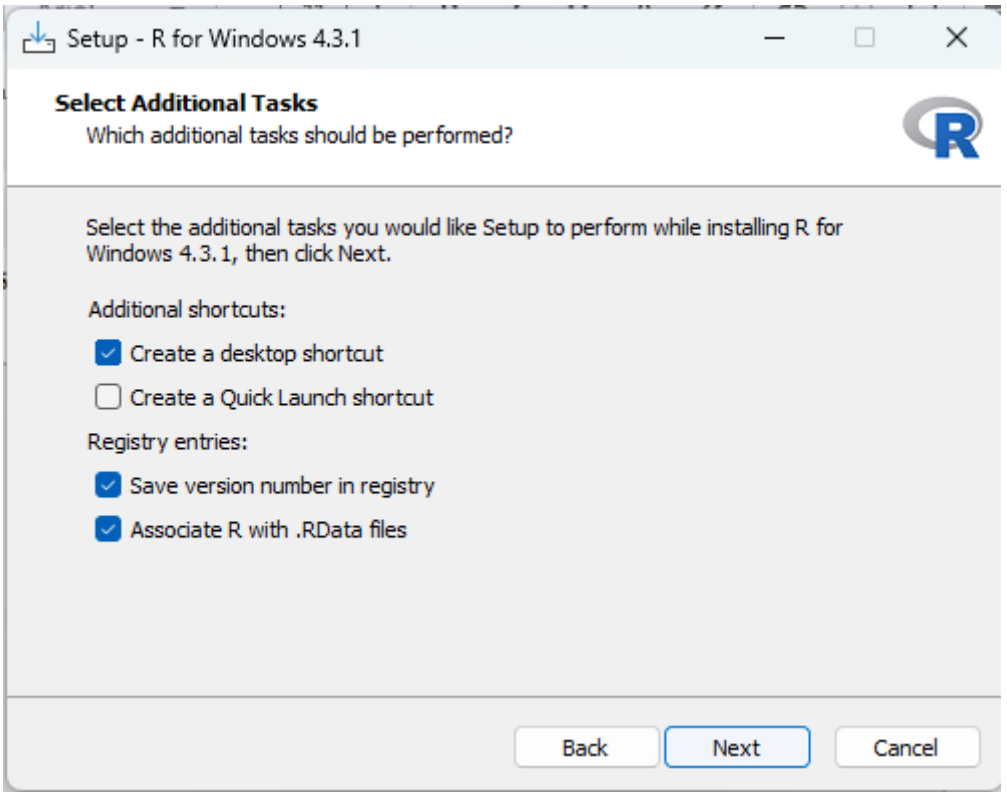
Rstudio installation

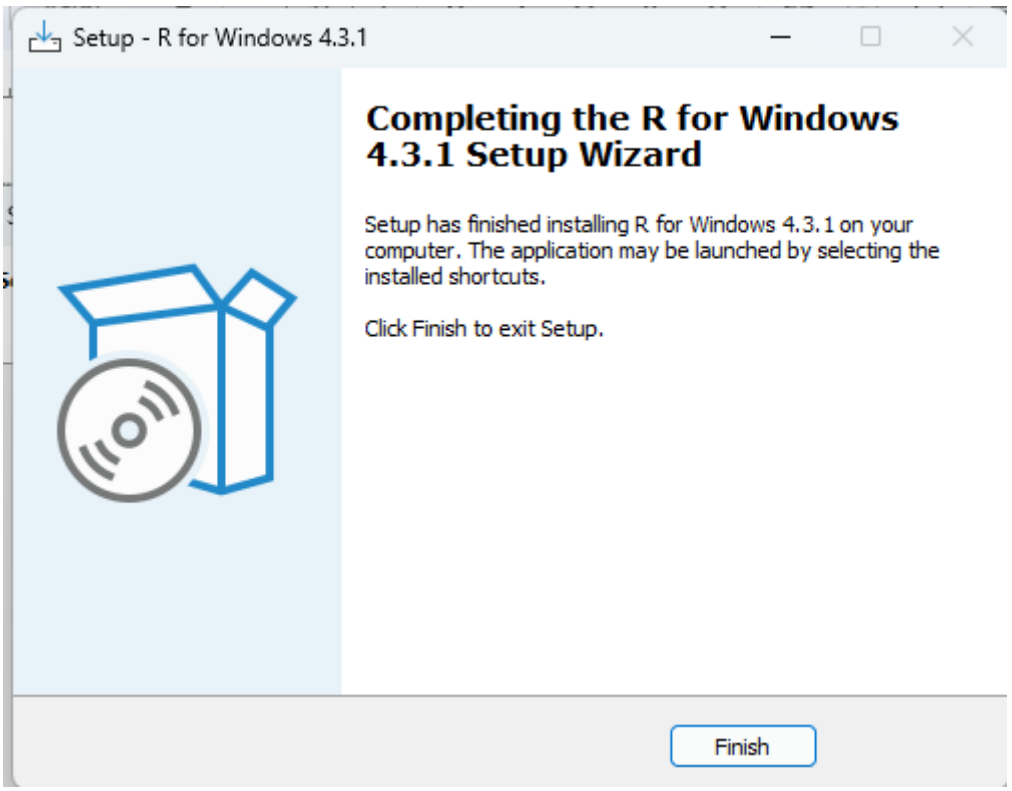


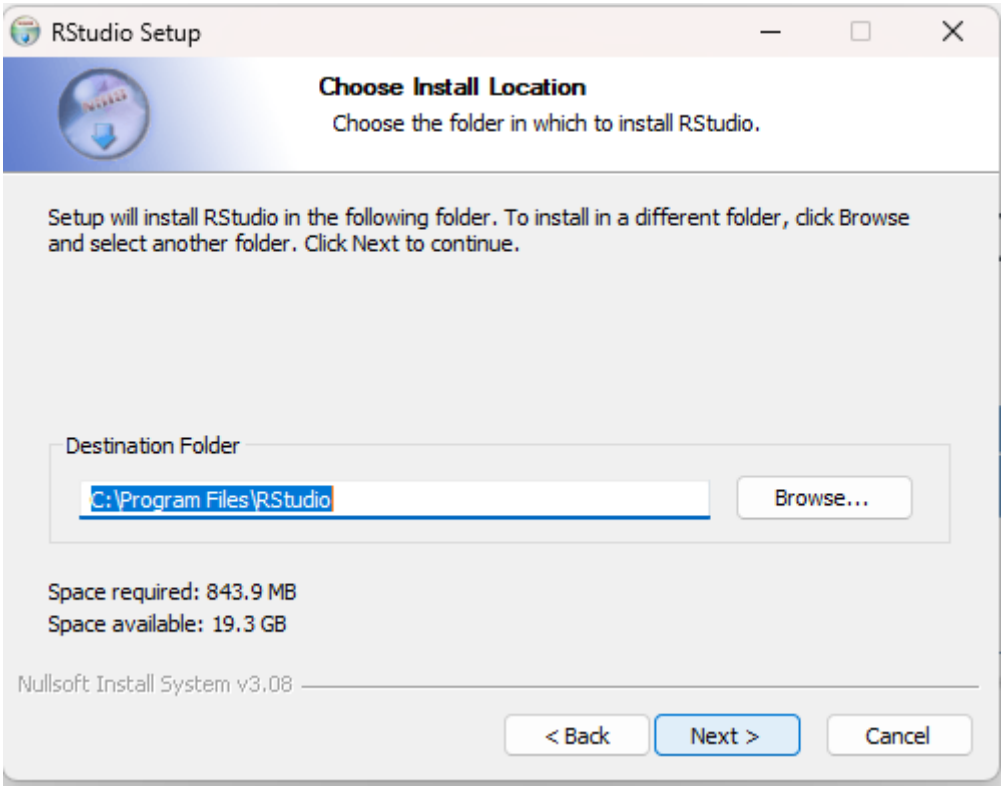

Setup - R for Windows 4.3.1
Select Components
Which components should be installed?
Select the components you want to install; clear the components you do not want to install. Click Next when you are ready to continue.
User installation
Main Files 89.1 MB
64-bit Files 68.7 MB
Message translations 9.0 MB
Current selection requires at least 169.8 MB of disk space.
Back
Next
Cancel

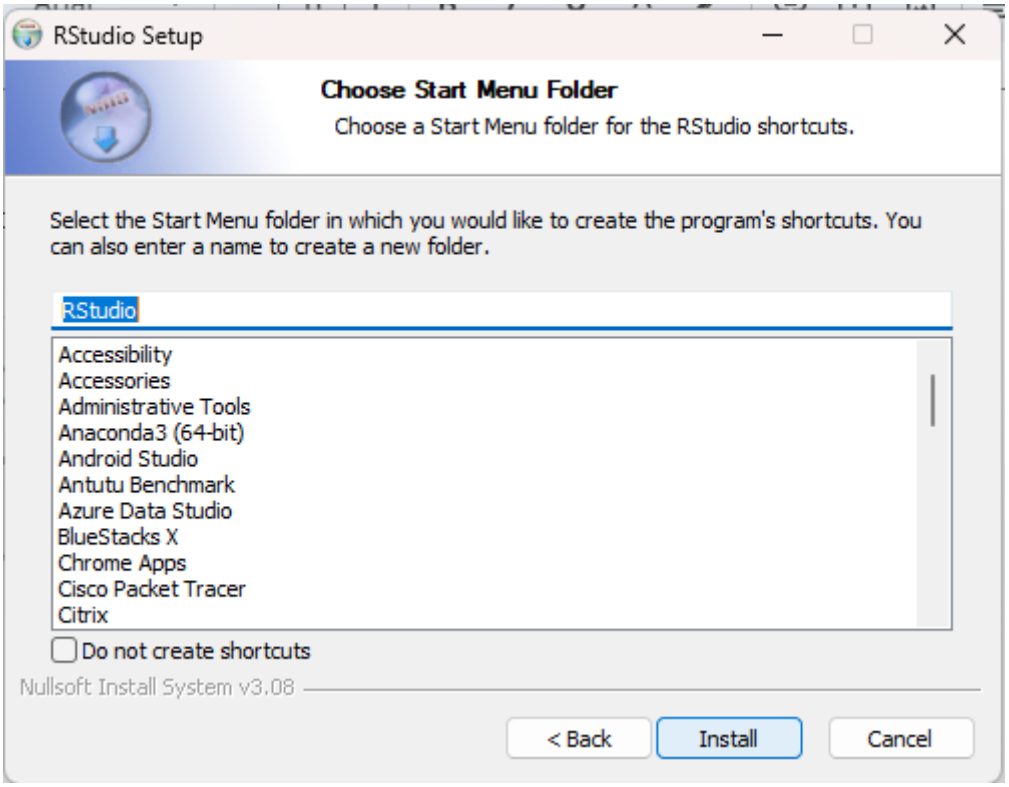

Setup - R for Windows 4.3.1
Startup options
Do you want to customize the startup options?
Please specify yes or no, then click Next.
Yes (customized startup)
No (accept defaults)
Back
Next
Cancel

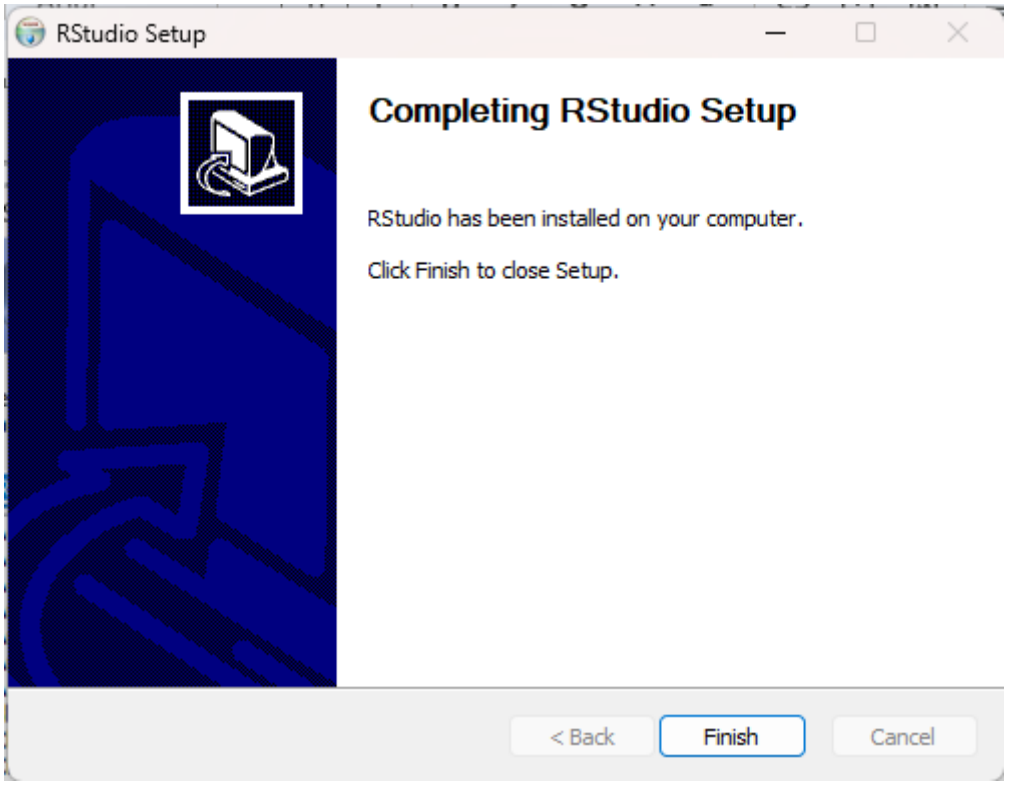

Setup - R for Windows 4.3.1
Select Start Menu Folder
Where should Setup place the program's shortcuts?
Setup will create the program's shortcuts in the following Start Menu folder.
To continue, click Next. If you would like to select a different folder, click Browse.
R
Browse...
Don't create a Start Menu folder
Back
Next
Cancel


Setup - R for Windows 4.3.1
Select Additional Tasks
Which additional tasks should be performed?
Select the additional tasks you would like Setup to perform while installing R for Windows 4.3.1, then click Next.
Additional shortcuts:
Create a desktop shortcut
Create a Quick Launch shortcut
Registry entries:
Save version number in registry
Associate R with .RData files
Back
Next
Cancel


Setup - R for Windows 4.3.1
Completing the R for Windows 4.3.1 Setup Wizard
Setup has finished installing R for Windows 4.3.1 on your computer. The application may be launched by selecting the installed shortcuts.
Click Finish to exit Setup.
Finish


RStudio Setup
Choose Install Location
Choose the folder in which to install RStudio.
Setup will install RStudio in the following folder. To install in a different folder, click Browse and select another folder. Click Next to continue.
Destination Folder
C:\Program Files\RStudio
Browse...
Space required: 843.9 MB
Space available: 19.3 GB
Nullsoft Install System v3.08
< Back
Next >
Cancel


RStudio Setup
Choose Start Menu Folder
Choose a Start Menu folder for the RStudio shortcuts.
Select the Start Menu folder in which you would like to create the program's shortcuts. You can also enter a name to create a new folder.
RStudio
Accessibility
Accessories
Administrative Tools
Anaconda3 (64-bit)
Android Studio
Antutu Benchmark
Azure Data Studio
BlueStacks X
Chrome Apps
Cisco Packet Tracer
Citrix
Do not create shortcuts
Nullsoft Install System v3.08
< Back
Install
Cancel


RStudio Setup
Completing RStudio Setup
RStudio has been installed on your computer.
Click Finish to close Setup.
< Back
Finish
Cancel

Q1.

```
> x = 5.6
> print(class(x))
[1] "numeric"
> print(typeof(x))
[1] "double"
> print (x)
[1] 5.6
>
```

Q2.

```
> y = 5
> print(class(y))
[1] "numeric"
> print(typeof(y))
[1] "double"
>
```

Q3.

```
> x = as.integer(5)
> print(class(x))
[1] "integer"
> print(typeof(x))
[1] "integer"
> y = 5L
> print(class(y))
[1] "integer"
> print(typeof(y))
[1] "integer"
>
```

Q4.

```
> x = 4
> y = 3
> z = x > y
> print(z)
[1] TRUE
> print(class(z))
[1] "logical"
> print(typeof(z))
[1] "logical"
>
```

Q5.

```
> x = 4 + 3i
> print(class(x))
[1] "complex"
> print(typeof(x))
[1] "complex"
>
```

Q6.

```
> char = "Mumbai University"
> print(class(char))
[1] "character"
> print(typeof(char))
[1] "character"
>
```

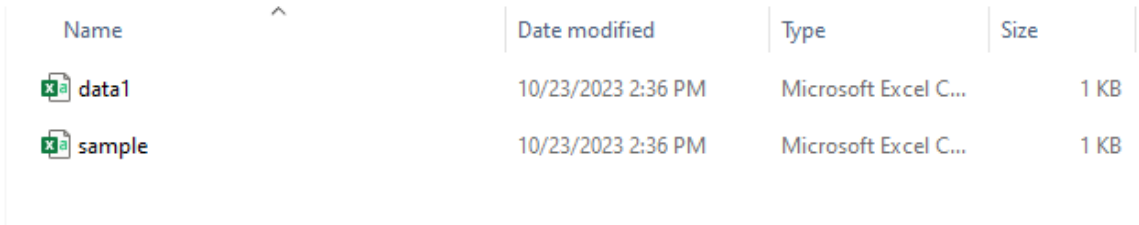
Q7

```
> record_data <- read.table("D:/MCA/New folder/new.txt")
Warning message:
In read.table("D:/MCA/New folder/new.txt") :
  incomplete final line found by readTableHeader on 'D:/MCA/New folder/new.txt'
> head(record_data)
      V1      V2  V3  V4 V5         V6      V7  V8      V9  V10     V11     V12  V13
1 RStudio Desktop Pro Get  a commercial license and support with RStudio Desktop Pro.
```

```
> record_data <- read.csv("D:/MCA/New folder/data.csv")
> head(record_data)
  roll.no.        name surname course
1        1 manoraqnjan   baral    mca
2        2       komal bhamble    mca
3        3      shriya    rane    mca
4        4    monalisa   baral    mca
5        5       tanvi   parab    mca
>
```

```
> library(readxl)
Warning message:
package 'readxl' was built under R version 4.2.3
> data <- read_excel("D:/MCA/New folder/data.xls", sheet = 1)
> print(data)
# A tibble: 5 x 4
  `roll no.` name        surname course
       <dbl> <chr>       <chr>   <chr>
1          1 manoraqnjan baral   mca
2          2 komal       bhamble mca
3          3 shriya      rane    mca
4          4 monalisa    baral   mca
5          5 tanvi       parab   mca
>
```

```
> setwd("D:/MCA/New folder")
> getwd()
[1] "D:/MCA/New folder"
> x <- data.frame (name = c("manoranjan", "Monalisa", "shriya"), department = c("Sales", "Marketing","Finance"))
> write.table(x, file ="data1.csv", sep = ",")
> z <- data.frame(a = 10, b = 40, c = pi)
> write.csv(z, file = "sample.csv")
>
```



Name	Date modified	Type	Size
data1	10/23/2023 2:36 PM	Microsoft Excel C...	1 KB
sample	10/23/2023 2:36 PM	Microsoft Excel C...	1 KB

data1

	A	B	C	D
1	name	department		
2	1	manoranjan	Sales	
3	2	Monalisa	Marketing	
4	3	shriya	Finance	
5				

sample

	A	B	C	D
1		a	b	c
2	1	10	40	3.141593

```
mysql> use penta
Database changed
mysql> Create database prac5;
Query OK, 1 row affected (0.03 sec)

mysql> Create table emp1(name varchar(50),age int,city varchar(50));
ERROR 1050 (42S01): Table 'emp1' already exists
mysql> Create table emp6(name varchar(50),age int,city varchar(50));
Query OK, 0 rows affected (0.04 sec)

mysql>
```

```
> library(RMySQL)
> con <- dbConnect(MySQL(), user='root', password='password', dbname='prac5', host='localhost',port=3306)
> data <- data.frame( name = c("manoranjan", "Monalisa", "shriya"),age = c(30, 25, 35),city = c("chembur","chembur", "parel"))
> insert_query <- paste0("INSERT INTO emp6 (name, age, city) VALUES ")
> values <- paste0("('", data$name, "', ", data$age, ", '", data$city, "')")
> insert_query <- paste0(insert_query, paste(values, collapse = ", "))
> dbSendQuery(con, insert_query)
<MySQLResult:-65713072,1,0>
> dbDisconnect(con)
[1] TRUE
```

```
mysql> select * from emp6;
+------------+------+---------+
| name       | age  | city    |
+------------+------+---------+
| manoranjan |   30 | chembur |
| Monalisa   |   25 | chembur |
| shriya     |   35 | parel   |
+------------+------+---------+
3 rows in set (0.00 sec)
```

Practical 6:

```
> dataframe <- read.csv("D:/MCA/prac5.csv")
> print(dataframe)
       Name Age Salary Occupation Gender
1 Manoranjan  21  90000  Developer   Male
2   Monalisa  22  80000     Doctor Female
3      Komal  26  75000   Engineer Female
4     shriya  22 100000   Designer Female
```

```
> names(dataframe)[names(dataframe) == "Name"] <- "FirstName"
> print(dataframe)
   FirstName Age Salary Occupation Gender
1 Manoranjan  21  90000  Developer   Male
2   Monalisa  22  80000     Doctor Female
3      Komal  26  75000   Engineer Female
4     shriya  22 100000   Designer Female
```

```
> dataframe$Email <- c('Mano@gmail.com', 'Mona@gmail.com', 'komal@gm
arva@gmail.com')
> print(dataframe)
   FirstName Age Salary Occupation Gender             Email
1 Manoranjan  21  90000  Developer   Male    Mano@gmail.com
2   Monalisa  NA  80000     Doctor Female    Mona@gmail.com
3      Komal  26     NA   Engineer Female   komal@gmail.com
4     shriya  22 100000   Designer Female  shriya@gmail.com
5    harshal  23  95000         Hr   Male harshal2gmial.com
6     chirag  NA  80000     Writer   Male chrirag@gmail.com
7    atharva  24  75000   Business   Male atharva@gmail.com
```

```
> dataframe <- dataframe[complete.cases(dataframe), ]
> print(dataframe)
   FirstName Age Salary Occupation Gender             Email
1 Manoranjan  21  90000  Developer   Male    Mano@gmail.com
4     shriya  22 100000   Designer Female  shriya@gmail.com
5    harshal  23  95000         Hr   Male harshal2gmial.com
7    atharva  24  75000   Business   Male atharva@gmail.com
```

```
> dataframe$Age <- ifelse(is.na(dataframe$Age),ave(dataframe$Age,
> print(dataframe)
       Name  Age Salary Occupation Gender
1 Manoranjan 21.0  90000  Developer   Male
2   Monalisa 23.2  80000     Doctor Female
3      Komal 26.0     NA   Engineer Female
4     shriya 22.0 100000   Designer Female
5    harshal 23.0  95000         Hr   Male
6     chirag 23.2  80000     Writer   Male
7    atharva 24.0  75000   Business   Male
```

```
> dataframe$Salary <- ifelse(is.na(dataframe$Salary), ave(dataf
> print(dataframe)
       Name  Age    Salary Occupation Gender
1 Manoranjan 21.0  90000.00  Developer   Male
2   Monalisa 23.2  80000.00     Doctor Female
3      Komal 26.0  86666.67   Engineer Female
4     shriya 22.0 100000.00   Designer Female
5    harshal 23.0  95000.00         Hr   Male
6     chirag 23.2  80000.00     Writer   Male
7    atharva 24.0  75000.00   Business   Male
```

```
> dataframe$Gender <- as.numeric(factor(dataframe$Gender))
> print(dataframe)
       Name  Age    Salary Occupation Gender
1 Manoranjan 21.0  90000.00  Developer      2
2   Monalisa 23.2  80000.00     Doctor      1
3      Komal 26.0  86666.67   Engineer      1
4     shriya 22.0 100000.00   Designer      1
5    harshal 23.0  95000.00         Hr      2
6     chirag 23.2  80000.00     Writer      2
7    atharva 24.0  75000.00   Business      2
```

```
> subset_data <- dataframe[dataframe$Age > 23, ]
> print(subset_data)
      Name  Age    Salary Occupation Gender
2 Monalisa 23.2 80000.00     Doctor      1
3    Komal 26.0 86666.67   Engineer      1
6   chirag 23.2 80000.00     Writer      2
7  atharva 24.0 75000.00   Business      2
```

```
> sample_data <- dataframe[sample(nrow(dataframe), sample_size, replace = FALSE), ]
> print(sample_data)
       Name Age Salary Occupation Gender
1 Manoranjan  21  90000  Developer   Male
3     Komal  26     NA   Engineer Female
7   atharva  24  75000   Business   Male
2  Monalisa  NA  80000     Doctor Female
6    chirag  NA  80000     Writer   Male
```