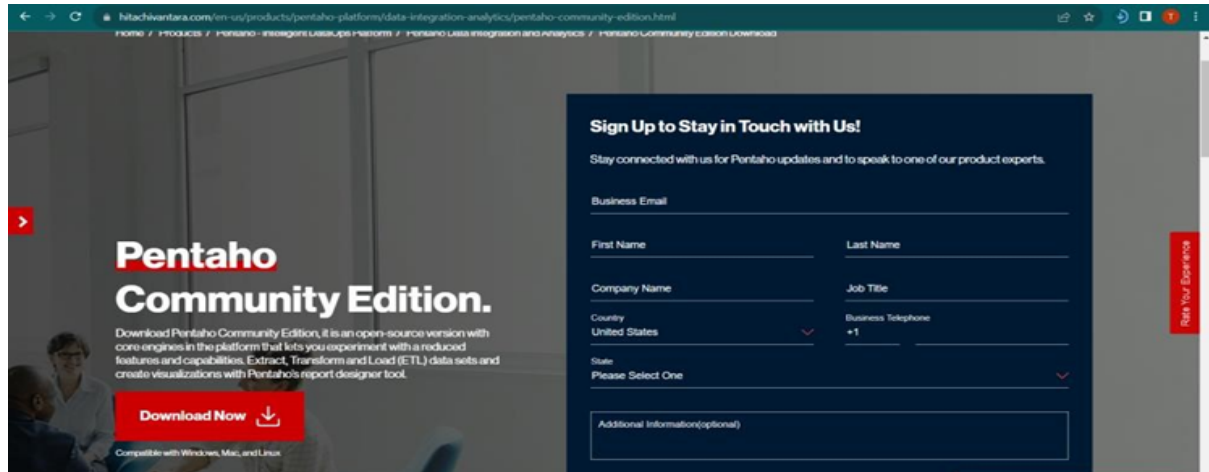


Practical no.4

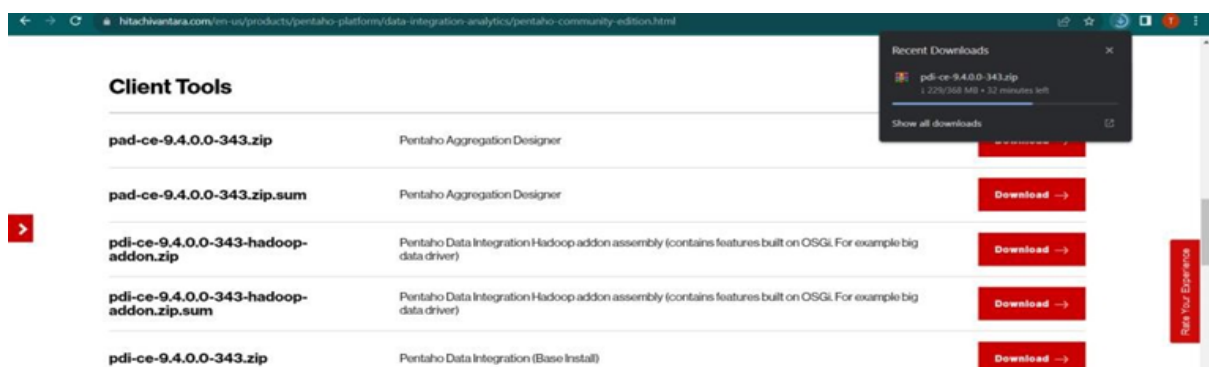
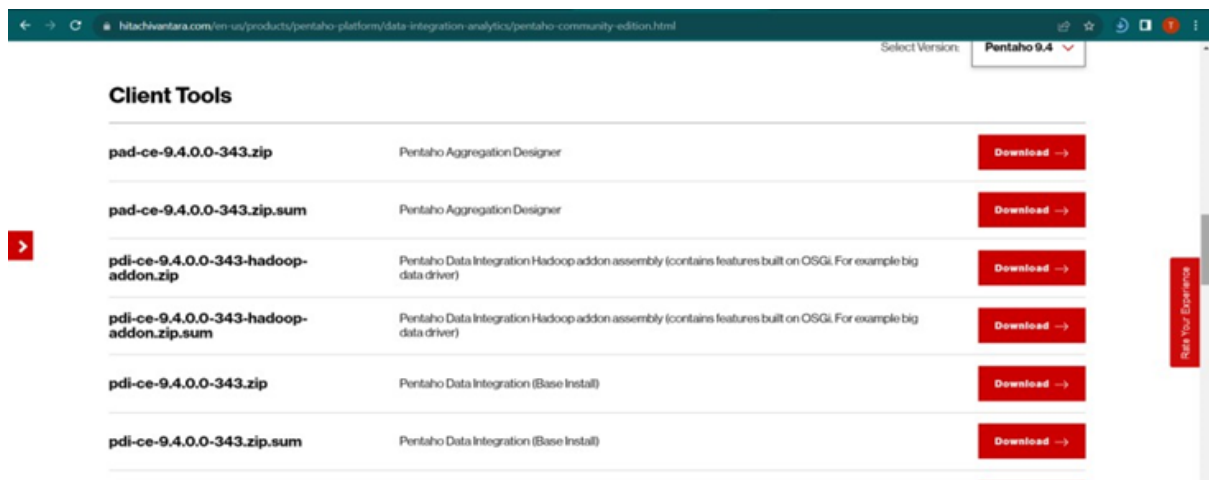
INSTALLATION OF PENTAHO

1. Go to

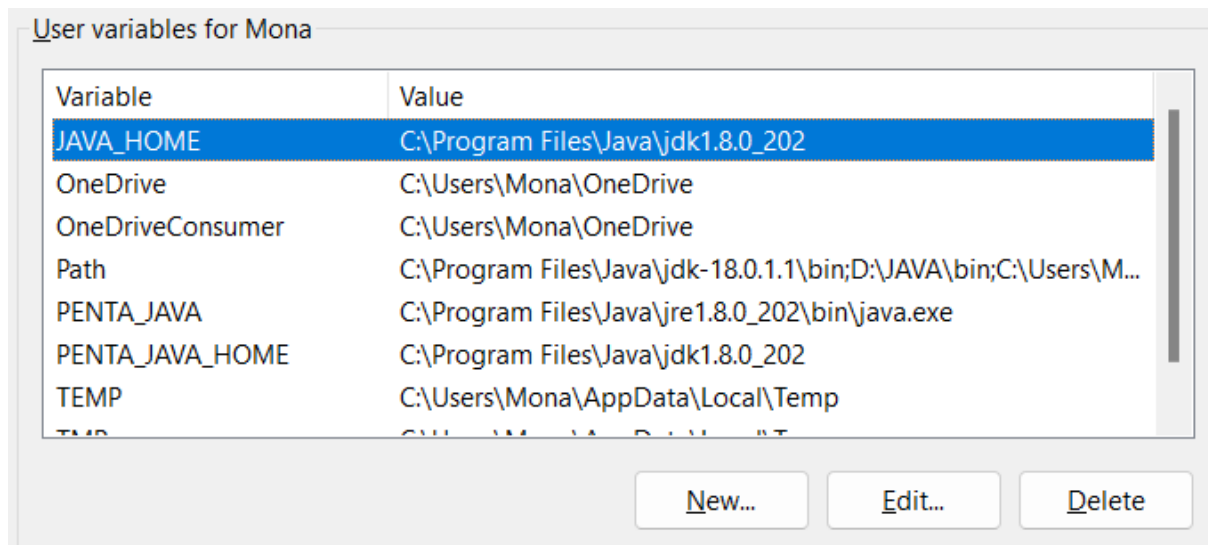
<https://www.hitachivantara.com/en-us/products/pentaho-platform/data-integrationanalytics/pentaho-community-edition.html> site and click on download now.






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

	set-pentaho-env.sh	08-11-2022 19:07	SH Source File	5 KB
	Spoon.bat	08-11-2022 19:07	Windows Batch File	6 KB
	spoon.command	08-11-2022 19:07	COMMAND File	2 KB

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

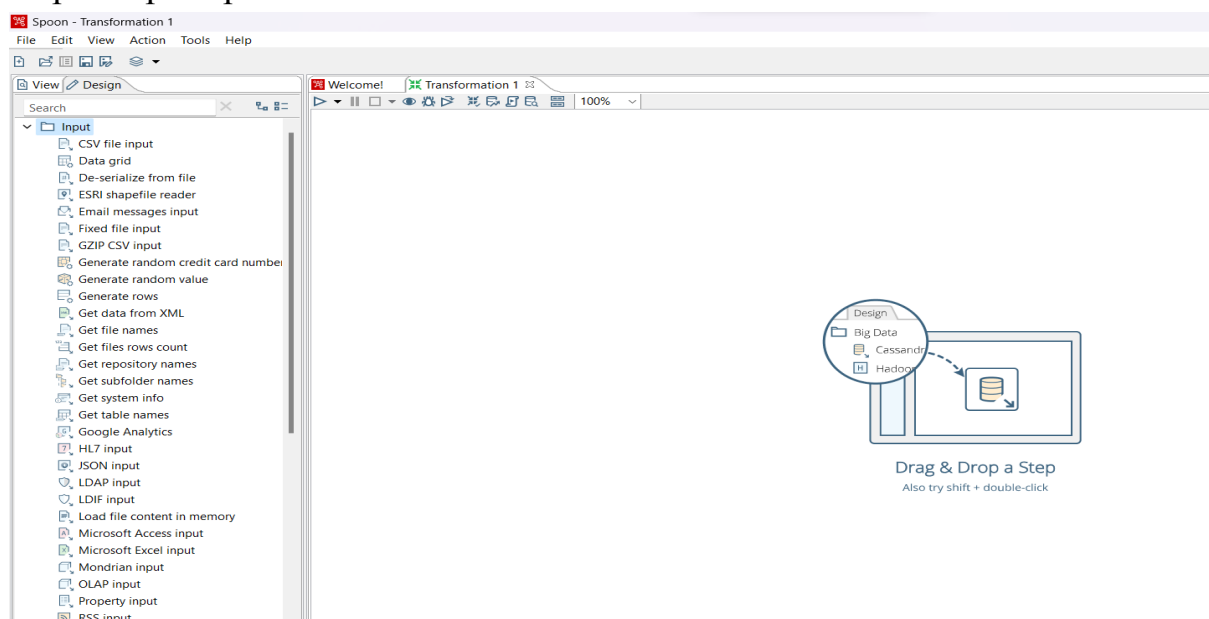
Step 1: Create New Database penta and table prac1

```
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    | manoranjana | MCA          |
| 2    | monalisa    | MCA          |
| 3    | shriya      | MBA          |
| 4    | komal       | MBA          |
+-----+-----+-----+
4 rows in set (0.00 sec)

mysql> |
```

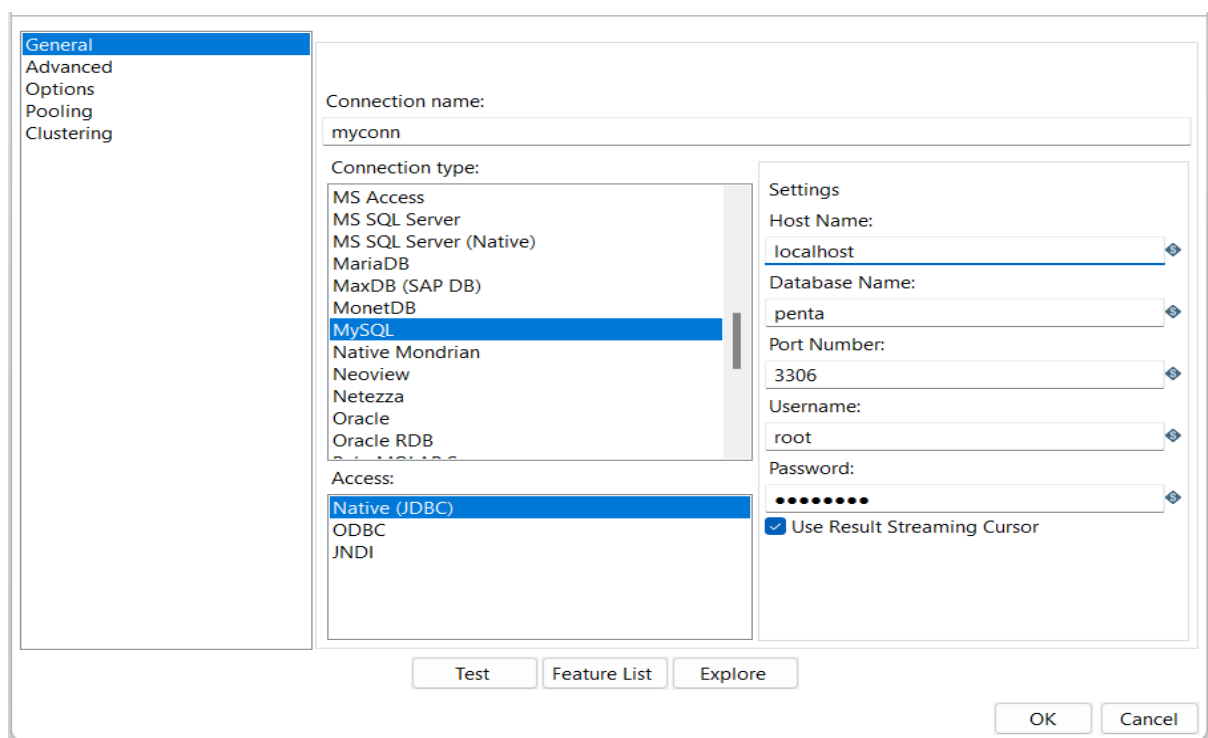
Step 2: Open spoon batch file then click on new Transformation



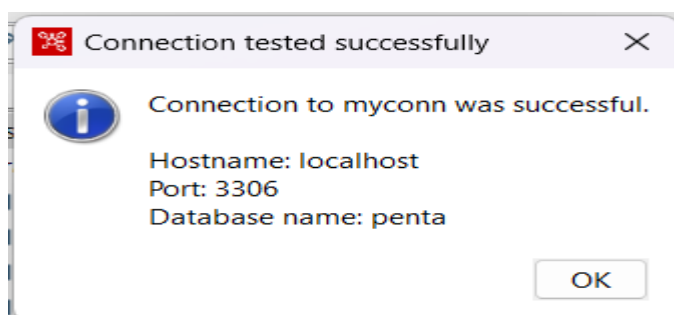
Step 3: Now drag and drop table input from input (in design tab) and excel output from the design tab and establish a connection between both by connecting the arrow.

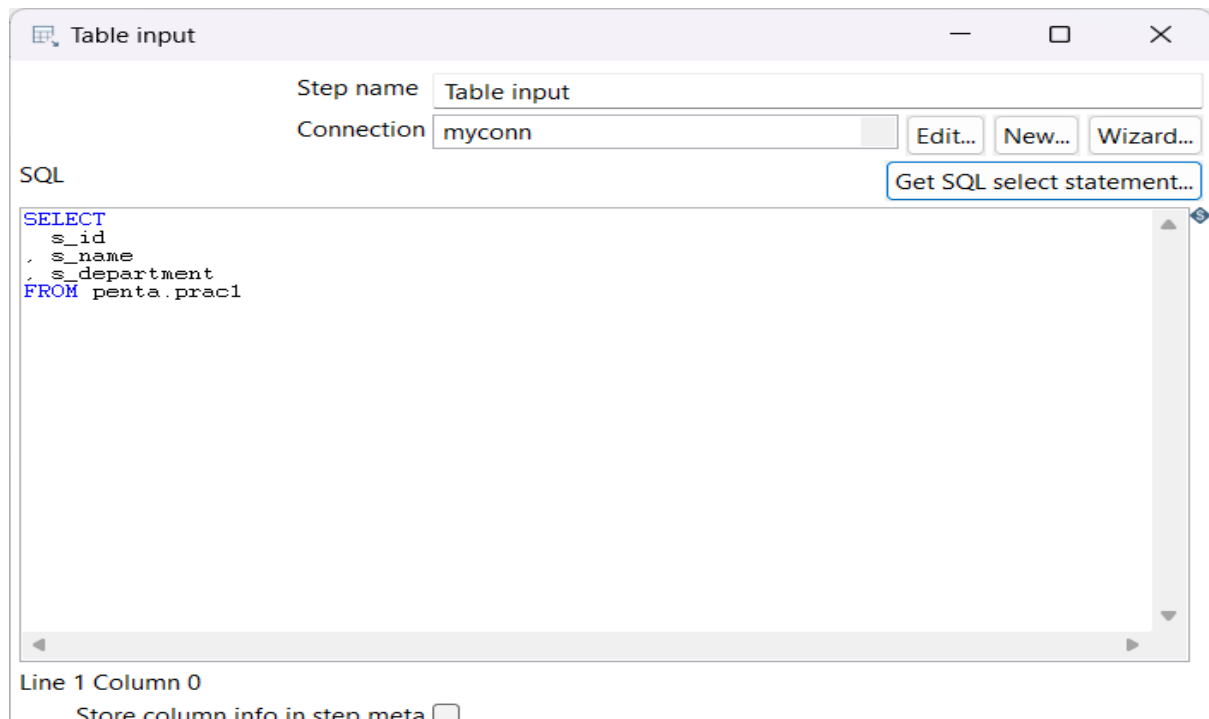


Step 4: Double click on table input, new connection and select Get sql select statement. After this all the columns you have created will be shown in a sql statement. New Connection

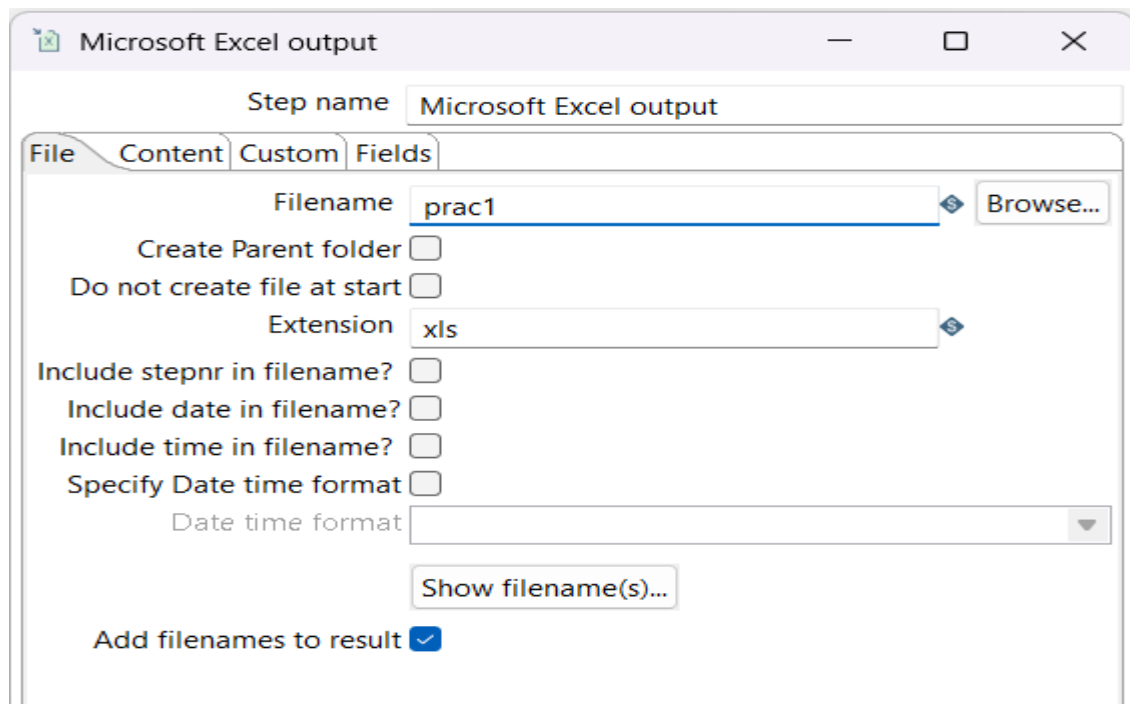



After Clicking Test



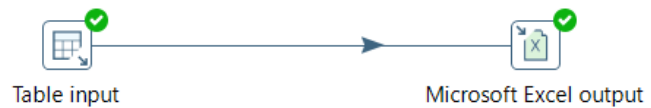


Step 5: then double click on Microsoft Excel output and inside the filename give the path of the file with the filename that you want to save.



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





Execution Results

Logging	Execution History	Step Metrics	Performance Graph	Metrics	Preview data
<div> </div> <p>2023/10/30 13:04:39 - Spoon - Transformation opened.</p> <p>2023/10/30 13:04:39 - Spoon - Launching transformation [Transformation 1]...</p> <p>2023/10/30 13:04:39 - Spoon - Started the transformation execution.</p> <p>2023/10/30 13:04:39 - Transformation 1 - Dispatching started for transformation [Transformation 1]</p> <p>2023/10/30 13:04:39 - Table input.0 - Finished reading query, closing connection</p> <p>2023/10/30 13:04:39 - Table input.0 - Finished processing (I=5, O=0, R=0, W=5, U=0, E=0)</p> <p>2023/10/30 13:04:40 - Microsoft Excel output.0 - Finished processing (I=0, O=5, R=5, W=5, U=0, E=0)</p> <p>2023/10/30 13:04:40 - Spoon - The transformation has finished!!</p>					

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

> This PC > DATA (D:) > pentho > data-integration >				
<div> </div> <div> Sort > View > ... </div>				
Name	Date modified	Type	Size	
pan.sh	08-11-2022 19:07	SH Source File	2 KB	
PentahoDataIntegration_OSS_Licenses.ht...	08-11-2022 19:00	Microsoft Edge HT...	3 KB	
prac1.xls	30-10-2023 13:04	Microsoft Excel 97...	14 KB	

	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 name: Microsoft Excel input

Add Field(s)

Fields

Additional output fields

Spread sheet type (engine): Excel 97-2003 XLS (JXL)

File or directory: Add Browse...

Regular Expression:

Exclude Regular Expression:

Password:

Selected files:

#	File/Directory	Wildcard (RegExp)	Exclude wildcard	Required	Include subfolders
1	D:\pentho\Book1.xls			N	N

Delete Edit

Accept filenames from previous steps

Accept filenames from previous step ☐

Step to read filenames from:

Field in the input to use as filename:

Show filename(s)...

OK Preview rows Cancel

Files ISheets Content Error Handling Fields Additional output fields							
#	Name	Type	Length	Precision	Trim type	Repeat	Form
1	shop_id	Number			none	N	
2	shop_name	String			none	N	
3	shop_no	Number			none	N	

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

Table output

Step name: Table output

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

Target schema: penta [Browse...]

Target table: shop [Browse...]

Commit size: 1000

Truncate table: ☐

Ignore insert errors: ☐

Specify database fields: ☐

Database fields

Partition data over tables: ☐

Partitioning field: []

Partition data per month: ☒

Partition data per day: ☐

Use batch update for inserts: ☒

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: []

[OK] [Cancel] [SQL]

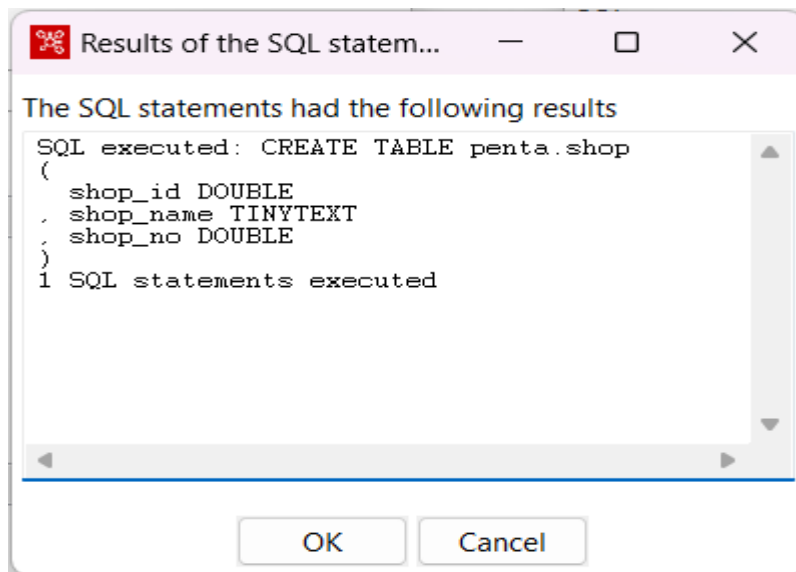
Simple SQL editor


SQL statements, separated by semicolon ';' :

```
CREATE TABLE penta.shop
(
  shop_id DOUBLE
  shop_name TINYTEXT
  shop_no DOUBLE
)
```

Line 1 column 0

[Execute] [Clear cache] [Close]



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

Execution Results

- Logging
- Execution History
- Step Metrics
- Performance Graph
- Metrics
- Preview data

```

2023/10/30 13:24:27 - Spoon - Save file as...
2023/10/30 13:24:27 - Spoon - Transformation opened.
2023/10/30 13:24:27 - Spoon - Launching transformation [Transformation 2]...
2023/10/30 13:24:27 - Spoon - Started the transformation execution.
2023/10/30 13:24:28 - Transformation 2 - Dispatching started for transformation [Transformation 2]
2023/10/30 13:24:28 - Table output.0 - Connected to database [mycn] (commit=1000)
2023/10/30 13:24:28 - Microsoft Excel input.0 - Finished processing (I=4, O=0, R=0, W=4, U=0, E=0)
2023/10/30 13:24:28 - Table output.0 - Finished processing (I=0, O=4, R=4, W=4, U=0, E=0)
2023/10/30 13:24:28 - Spoon - The transformation has finished!!
  
```

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> |
  
```

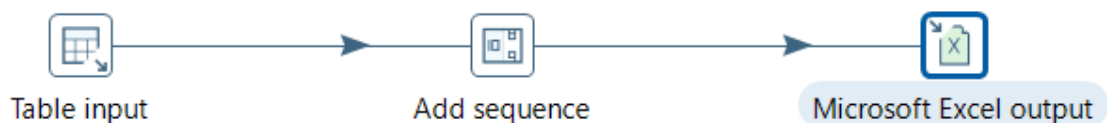
Q3] Create a pentaho transformation to get data from mysql table, add serial no to the data using pentaho sequences and push data into excel.

Step1: Create a table student and add some data.

```
mysql> use penta;
Database changed
mysql> select * from student;
+-----+-----+-----+
| s_name      | s_department | gender |
+-----+-----+-----+
| manoranjana | 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

Connection name:

Connection type:

- MariaDB
- MaxDB (SAP DB)
- MonetDB
- MySQL**
- Native Mondrian
- Neoview
- Netezza
- Oracle
- Oracle RDB
- Palo MOLAP Server
- PostgreSQL
- Redshift

Access:

- Native (JDBC)**
- ODBC
- JNDI

Settings

Host Name: localhost

Database Name: penta

Port Number: 3306

Username: root

Password: root

☒ Use Result Streaming Cursor

Test Feature List Explore OK Cancel

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

Step name: Table input

Connection: mycon

SQL:

```
SELECT
  s_name
, s_department
, gender
FROM penta.student
```

Get SQL select statement...

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

Add sequence

Step name: Add sequence

Name of value: valuename

Use a database to generate the sequence

Use DB to get sequence? ☐

Connection: mycon

Schema name

Sequence name: SEQ_

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: 99999999

Help OK Cancel

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

Step name: Microsoft Excel output

File | Content | Custom | Fields

Filename:

Create Parent folder: ☐

Do not create file at start: ☐

Extension:

Include stepnr in filename?: ☐

Include date in filename?: ☐

Include time in filename?: ☐

Specify Date time format: ☐

Date time format:

Add filenames to result: ☒

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

Workflow: Table input → Add sequence → Microsoft Excel output

Execution Results

Logging | Execution History | Step Metrics | Performance Graph | Metrics | Preview data

2023/10/30 14:00:37 - Spoon - Save file as...

2023/10/30 14:00:37 - Spoon - Transformation opened.

2023/10/30 14:00:37 - Spoon - Launching transformation [Transformation 3]...

2023/10/30 14:00:37 - Spoon - Started the transformation execution.

2023/10/30 14:00:38 - Transformation 3 - Dispatching started for transformation [Transformation 3]

2023/10/30 14:00:38 - Table input.0 - Finished reading query, closing connection

2023/10/30 14:00:38 - Table input.0 - Finished processing (I=4, O=0, R=0, W=4, U=0, E=0)

2023/10/30 14:00:38 - Add sequence.0 - Finished processing (I=0, O=0, R=4, W=4, U=0, E=0)

2023/10/30 14:00:39 - Microsoft Excel output.0 - Finished processing (I=0, O=4, R=4, W=4, U=0, E=0)

2023/10/30 14:00:39 - Spoon - The transformation has finished!!

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					

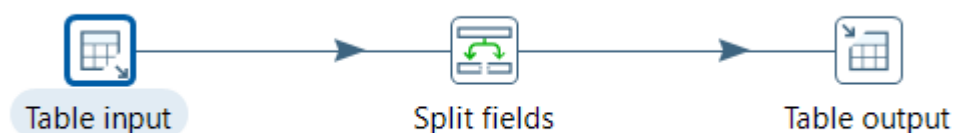
Q4] Create a pentaho transformation to get data from a mysql table named as employee have the following columns (emp_id,full name and job_id). Split the full name of the employee using pentaho split transformation and store the data into a new table in mysql name as emp1 having the following columns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

Connection name: my

Connection type:

- MariaDB
- MaxDB (SAP DB)
- MonetDB
- MySQL
- Native Mondrian
- Neoview
- Netezza
- Oracle
- Oracle RDB
- Palo MOLAP Server
- PostgreSQL
- Redshift

Access:

- Native (JDBC)
- ODBC
- JNDI

Settings

Host Name: localhost

Database Name: penta

Port Number: 3306

Username: root

Password:

☒ Use Result Streaming Cursor

Test Feature List Explore

OK Cancel

SQL

Get SQL select statement...

```
SELECT
  emp_id
, emp_name
, job_id
FROM penta.employee
```

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

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

 Split fields

Step name

Field to split


Delimiter

Enclosure

Fields

#	New field	ID	Remove ID?	Type	Length	Precision
1	f_name			String		
2	m_name			String		
3	l_name			String		

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

 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

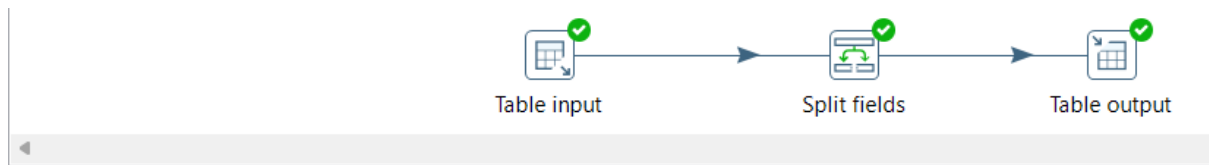
 Simple SQL editor

SQL statements, separated by semicolon ";"

```
CREATE TABLE penta.emp
(
  emp_id INT
  , f_name TINYTEXT
  , m_name TINYTEXT
  , l_name TINYTEXT
  , job_id INT
)
```

Line 1 column 0

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



Execution Results

Logging	Execution History	Step Metrics	Performance Graph	Metrics	Preview data
2023/10/30 14:24:59	Spoon - Save file as...				
2023/10/30 14:24:59	Spoon - Transformation opened.				
2023/10/30 14:24:59	Spoon - Launching transformation [Transformation 4]...				
2023/10/30 14:24:59	Spoon - Started the transformation execution.				
2023/10/30 14:24:59	Transformation 4 - Dispatching started for transformation [Transformation 4]				
2023/10/30 14:24:59	Table output.0 - Connected to database [my] (commit=1000)				
2023/10/30 14:25:00	Table input.0 - Finished reading query, closing connection				
2023/10/30 14:25:00	Table input.0 - Finished processing (I=4, O=0, R=0, W=4, U=0, E=0)				
2023/10/30 14:25:00	Split fields.0 - Finished processing (I=0, O=0, R=4, W=4, U=0, E=0)				
2023/10/30 14:25:00	Table output.0 - Finished processing (I=0, O=4, R=4, W=4, U=0, E=0)				
2023/10/30 14:25:00	Spoon - The transformation has finished!!				

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 name: Microsoft Excel input

Add Field(s)

!Fields | Additional output fields

Read sheet type (engine): Excel 97-2003 XLS (JXL)

File or directory: Add Browse...

Regular Expression:

Include Regular Expression:

Password:

Selected files:

#	File/Directory	Wildcard (RegExp)	Exclude wildcard	Required	Include subfolders
1	D:\pentho\Book1.xls			N	N

Delete Edit

Accept filenames from previous steps

Accept filenames from previous step ☐

Step to read filenames from:

Field in the input to use as filename:

Show filename(s)...

OK Preview rows Cancel

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 name: Microsoft Excel input

Add sheet(s)


Files | !Sheets | Content | Error Handling | Fields | Additional output fields

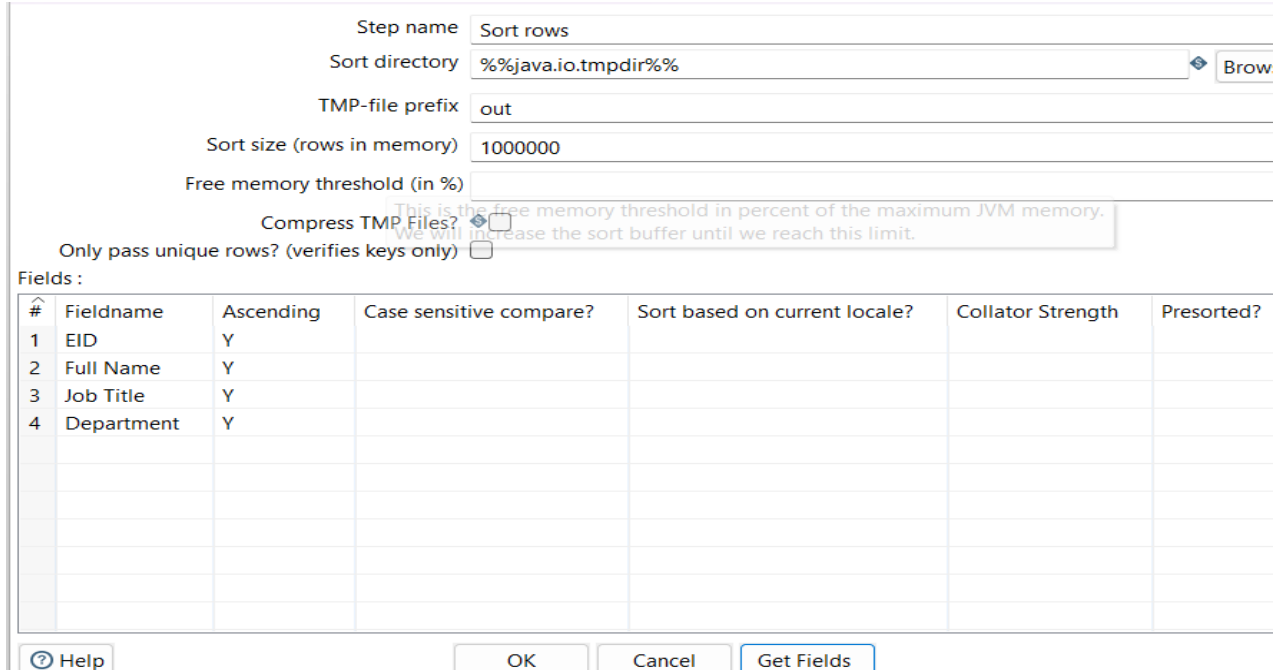
#	Name	Type	Length	Precision	Trim type	Repeat	Format	Currency	De
1	EID	Number			none	N			
2	Full Name	String			none	N			
3	Job Title	String			none	N			
4	Department	String			none	N			

Get fields from header row...


Help OK Preview rows Cancel

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 name: Sort rows

Sort directory: %%java.io.tmpdir%%  **Brow**

TMP-file prefix: out

Sort size (rows in memory): 1000000


Free memory threshold (in %):  This is the free memory threshold in percent of the maximum JVM memory. We will increase the sort buffer until we reach this limit.

Compress TMP Files? ☐


Only pass unique rows? (verifies keys only) ☐

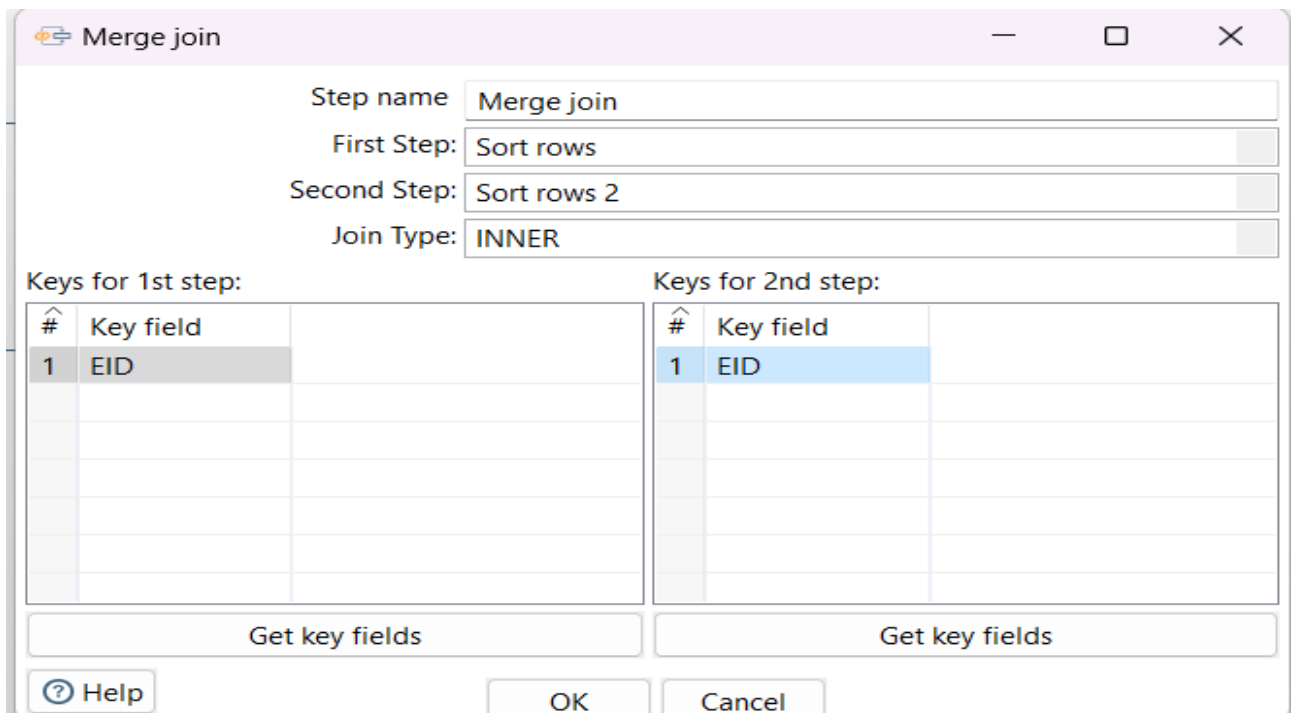
Fields:

#	Fieldname	Ascending	Case sensitive compare?	Sort based on current locale?	Collator Strength	Presorted?
1	EID	Y				
2	Full Name	Y				
3	Job Title	Y				
4	Department	Y				

 **Help** **OK** **Cancel** **Get Fields**

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.



Step name: Merge join

First Step: Sort rows

Second Step: Sort rows 2

Join Type: INNER


Keys for 1st step:

#	Key field
1	EID

Keys for 2nd step:

#	Key field
1	EID

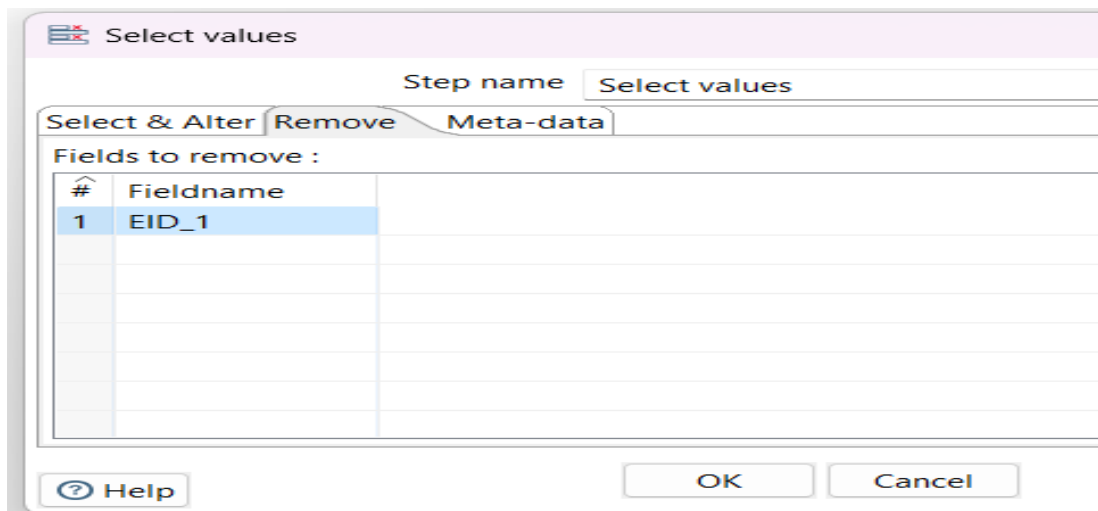
Get key fields **Get key fields**

 **Help** **OK** **Cancel**

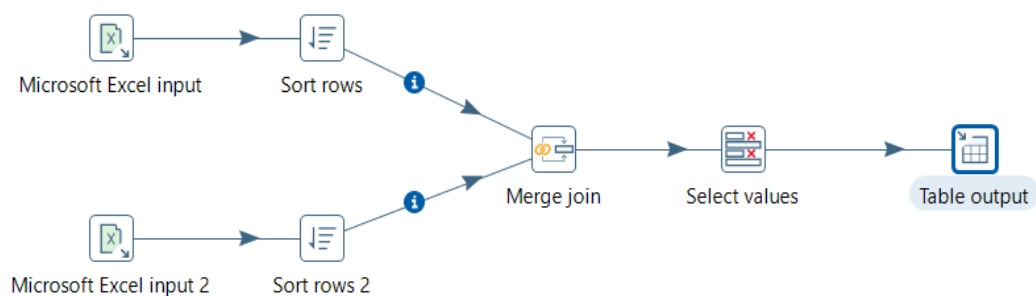
Then , delete all uncommon columns and keep the common column i.e EID.

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

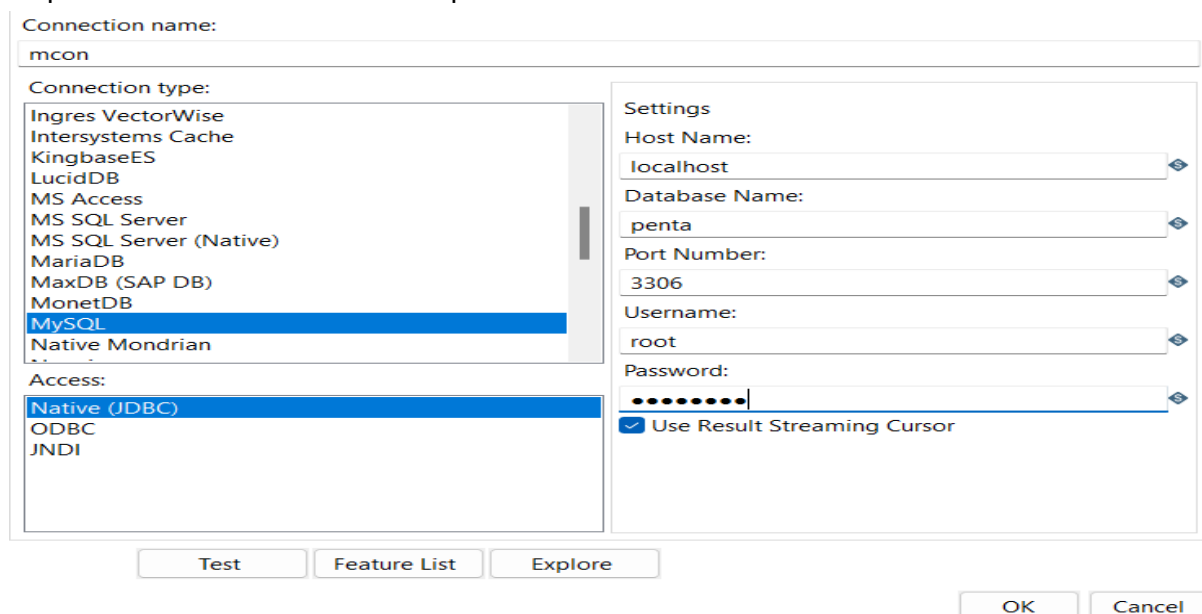
Step 9: Double click on select values and click on "get fields" to select . we will be able to see 2 EIDs. Then we have to delete EID_1 from the fields section and click on "OK".



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

Table output

Step name: Table output

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

Target schema: penta [Browse...]

Target table: emp1 [Browse...]

Commit size: 1000

Truncate table: ☐

Ignore insert errors: ☐

Specify database fields: ☐

Database fields

Partition data over tables: ☐

Partitioning field:

Partition data per month: ☒

Partition data per day: ☐

Use batch update for inserts: ☒

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:

OK Cancel SQL

Simple SQL editor

SQL statements, separated by semicolon ';'

```
CREATE TABLE penta.emp1
(
  EID DOUBLE
  , `Full Name` TINYTEXT
  , `Job Title` TINYTEXT
  , Department TINYTEXT
  , `Business Unit` TINYTEXT
  , Gender TINYTEXT
  , Ethnicity TINYTEXT
  , Age DOUBLE
)
```

Line 1 column 0

Execute Clear cache Close

Click on Execute

Results of the SQL statement...

The SQL statements had the following results

```
SQL executed: CREATE TABLE penta.emp1
(
  EID DOUBLE
  , `Full Name` TINYTEXT
  , `Job Title` TINYTEXT
  , Department TINYTEXT
  , `Business Unit` TINYTEXT
  , Gender TINYTEXT
  , Ethnicity TINYTEXT
  , Age DOUBLE
)
1 SQL statements executed
```

OK Cancel

Click OK

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

The screenshot displays a transformation workflow in SAP Data Services Designer. The workflow consists of the following steps:

- Microsoft Excel input** and **Microsoft Excel input 2** (Input nodes)
- Sort rows** and **Sort rows 2** (Sort nodes)
- Merge join** (Join node)
- Select values** (Filter node)
- Table output** (Output node)

Below the workflow, the **Execution Results** tab is active, showing a log of the transformation execution:

```

2023/10/30 14:51:22 - Spoon - Started the transformation execution.
2023/10/30 14:51:23 - Transformation 5 - Dispatching started for transformation [Transformation 5]
2023/10/30 14:51:23 - Table output.0 - Connected to database [mcon] (commit=1000)
2023/10/30 14:51:23 - Microsoft Excel input 2.0 - Finished processing (I=15, O=0, R=0, W=15, U=0, E=0)
2023/10/30 14:51:23 - Microsoft Excel input.0 - Finished processing (I=15, O=0, R=0, W=15, U=0, E=0)
2023/10/30 14:51:23 - Sort rows 2.0 - Finished processing (I=0, O=0, R=15, W=15, U=0, E=0)
2023/10/30 14:51:24 - Sort rows.0 - Finished processing (I=0, O=0, R=15, W=15, U=0, E=0)
2023/10/30 14:51:24 - Merge join.0 - Finished processing (I=0, O=0, R=30, W=15, U=0, E=0)
2023/10/30 14:51:24 - Select values.0 - Finished processing (I=0, O=0, R=15, W=15, U=0, E=0)
2023/10/30 14:51:24 - Table output.0 - Finished processing (I=0, O=15, R=15, W=15, U=0, E=0)
2023/10/30 14:51:24 - Spoon - The transformation has finished!!
  
```

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

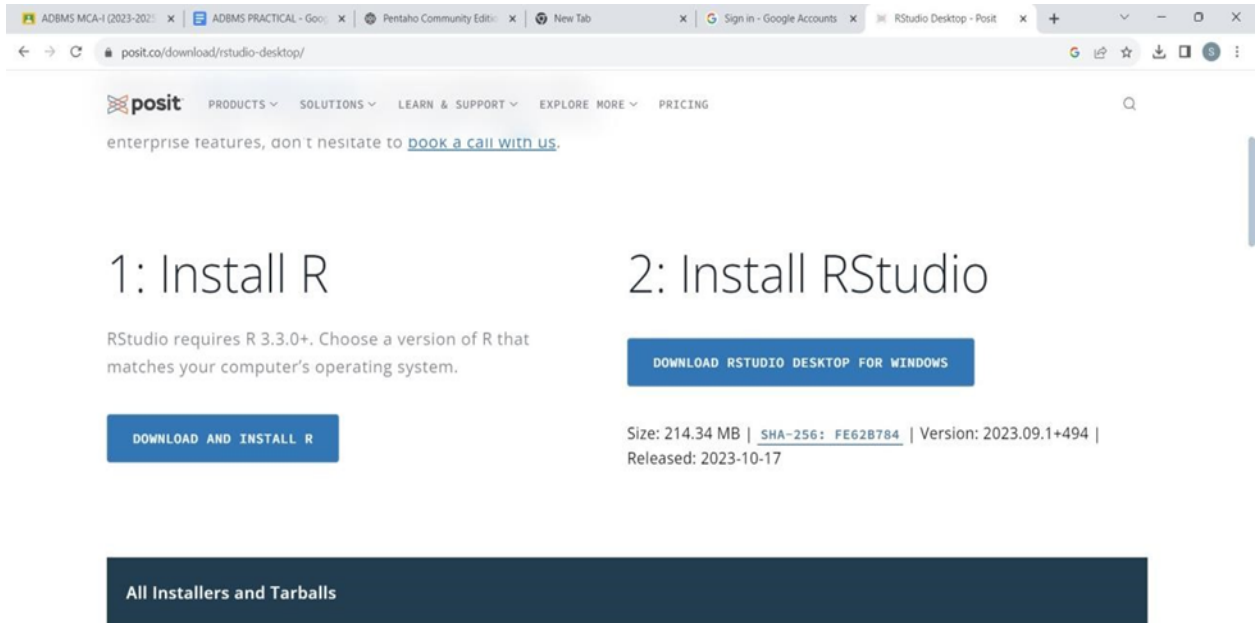
```

15 rows in set (0.00 sec)

mysql>
  
```

Practical no.5

R and RStudio Installation



The screenshot shows the Posit website's download page for RStudio Desktop. The page is divided into two main sections: "1: Install R" and "2: Install RStudio".

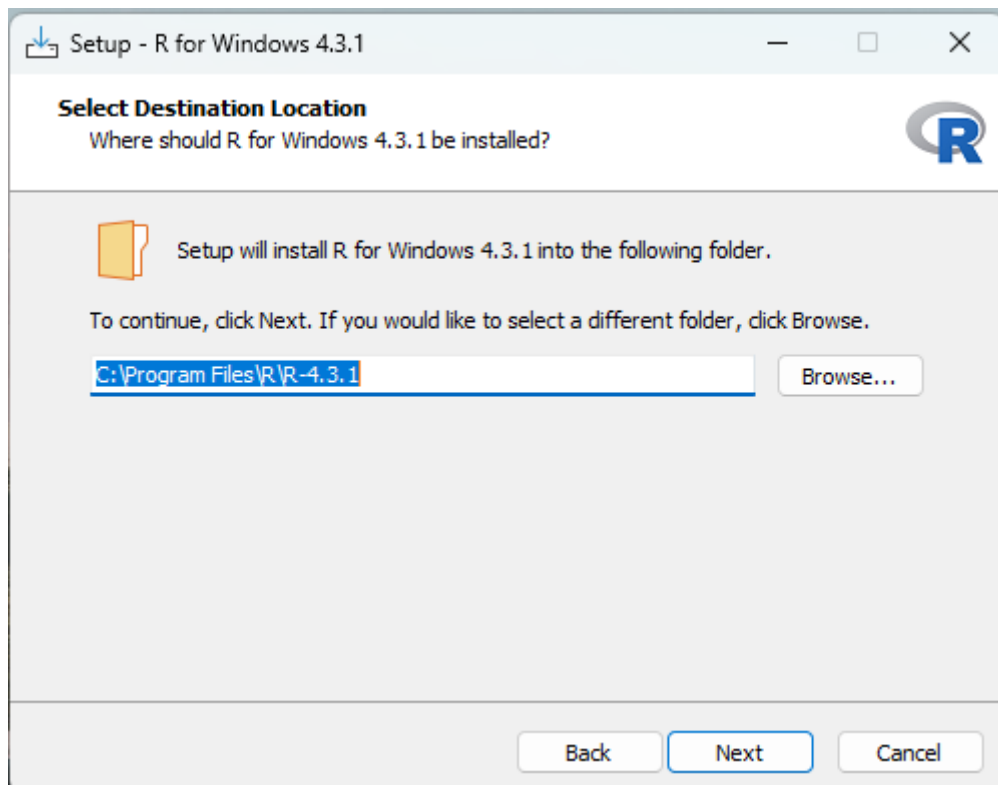
1: Install R
RStudio requires R 3.3.0+. Choose a version of R that matches your computer's operating system.
[DOWNLOAD AND INSTALL R](#)

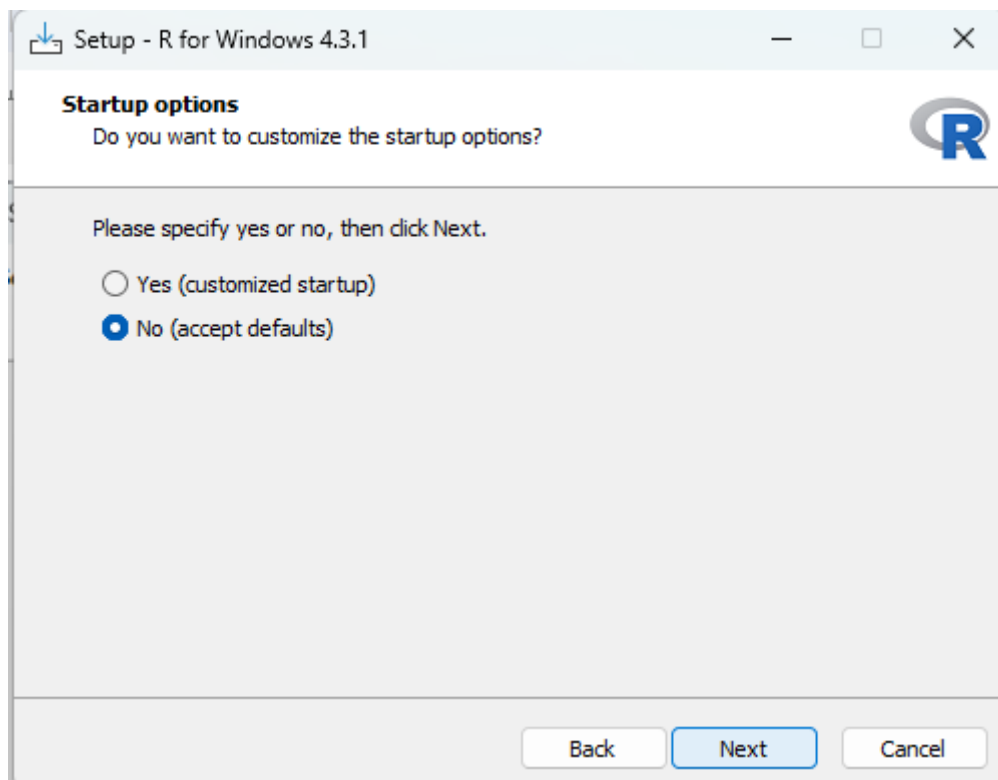
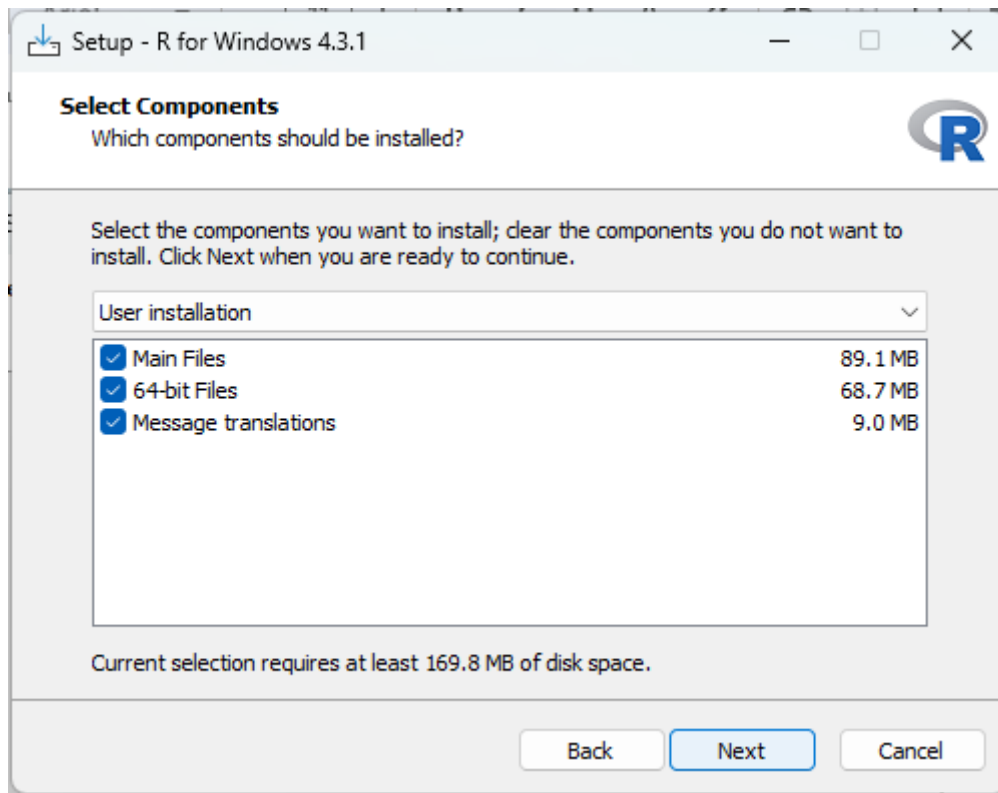
2: Install RStudio
[DOWNLOAD RSTUDIO DESKTOP FOR WINDOWS](#)
Size: 214.34 MB | SHA-256: FE62B784 | Version: 2023.09.1+494 | Released: 2023-10-17

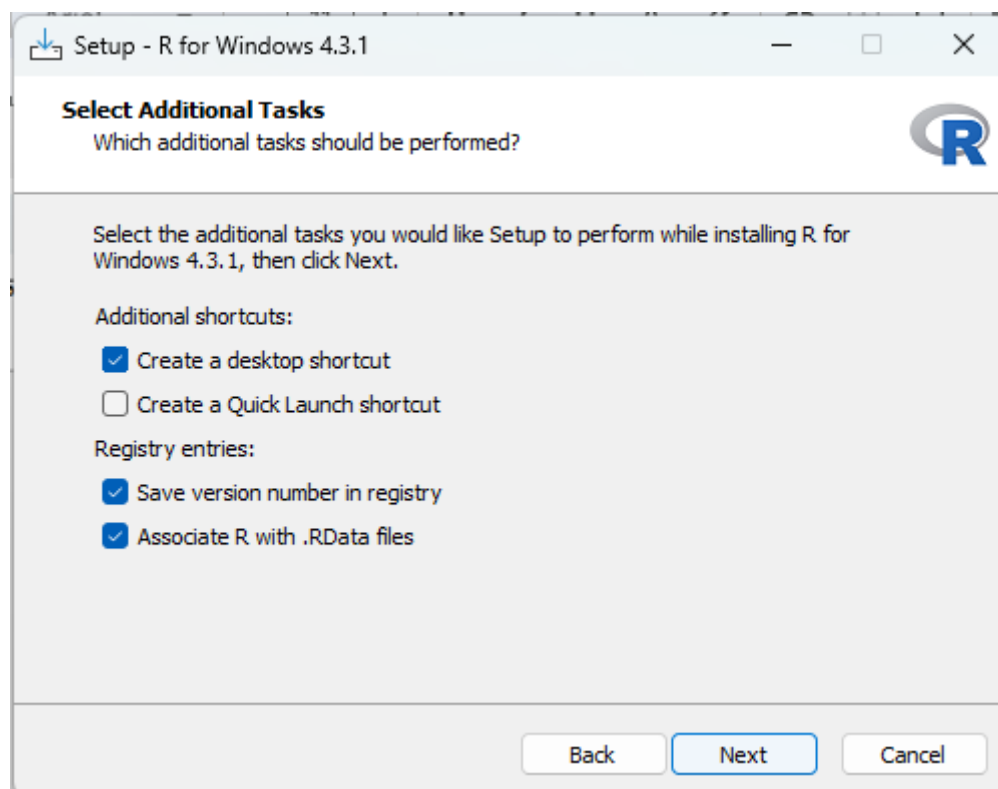
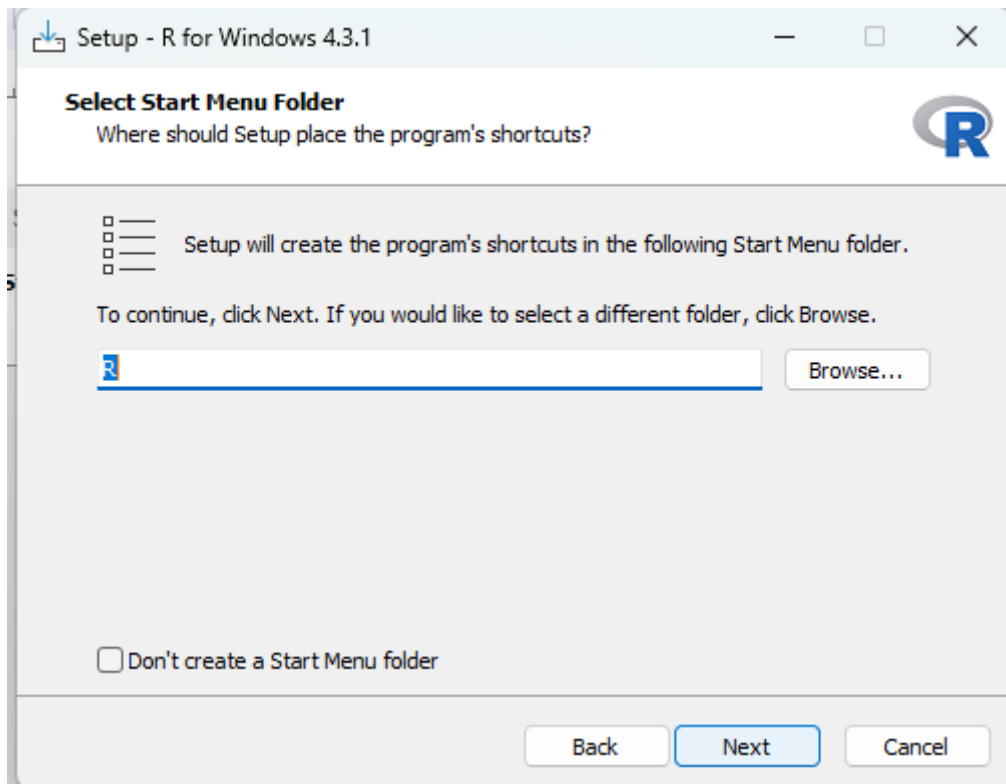
[All Installers and Tarballs](#)

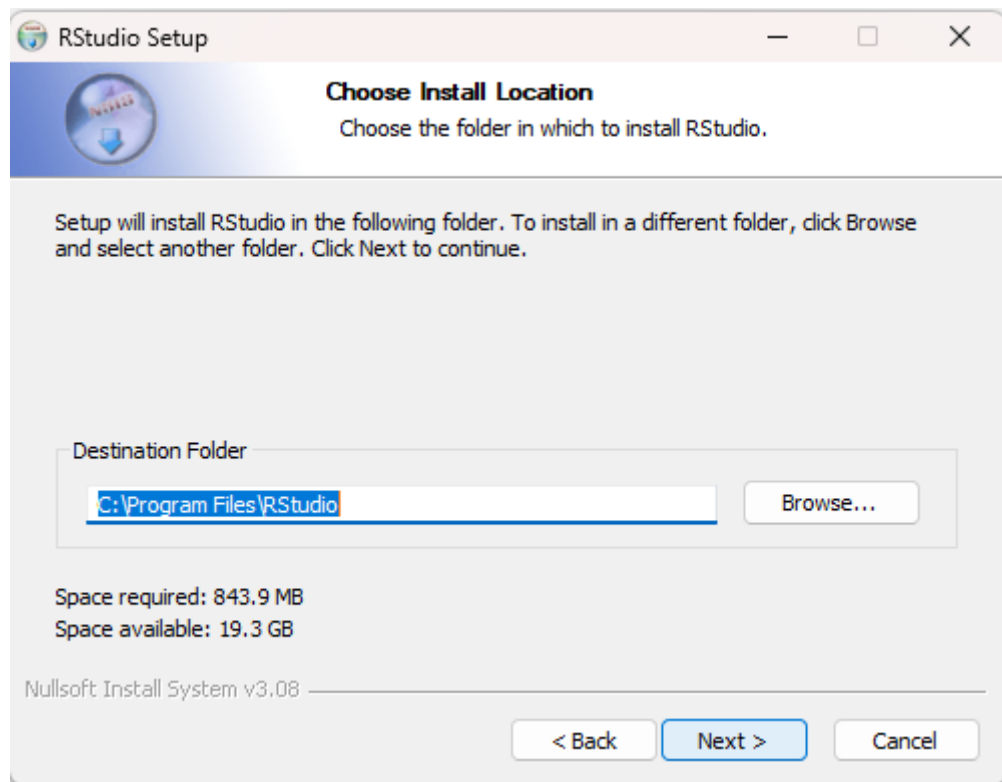
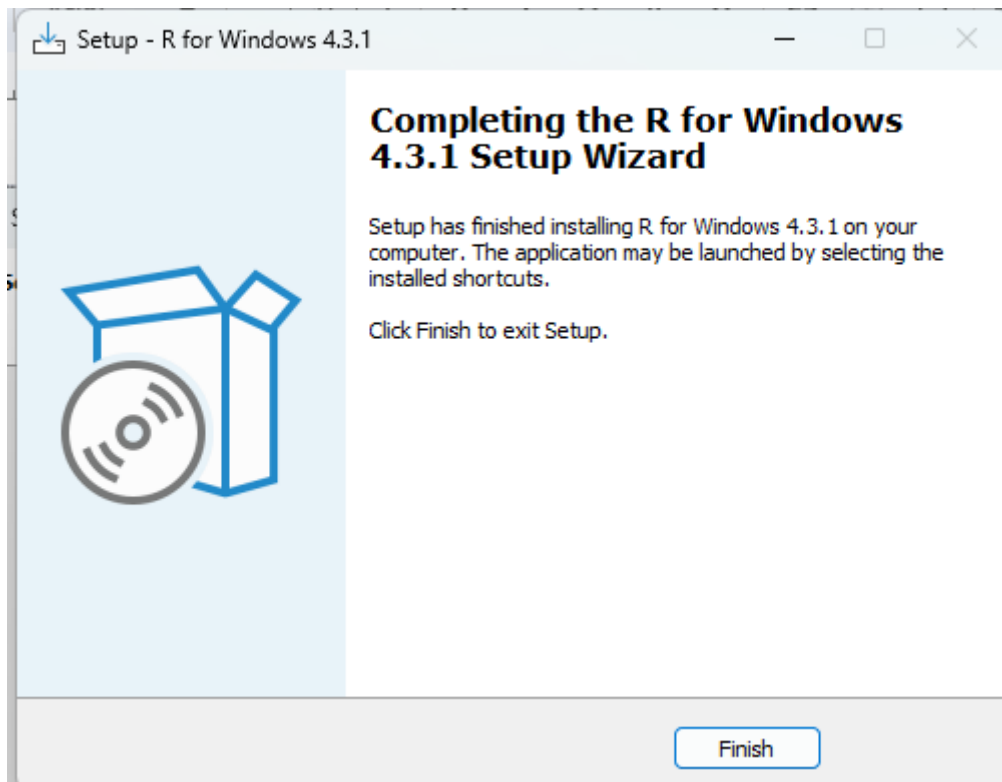


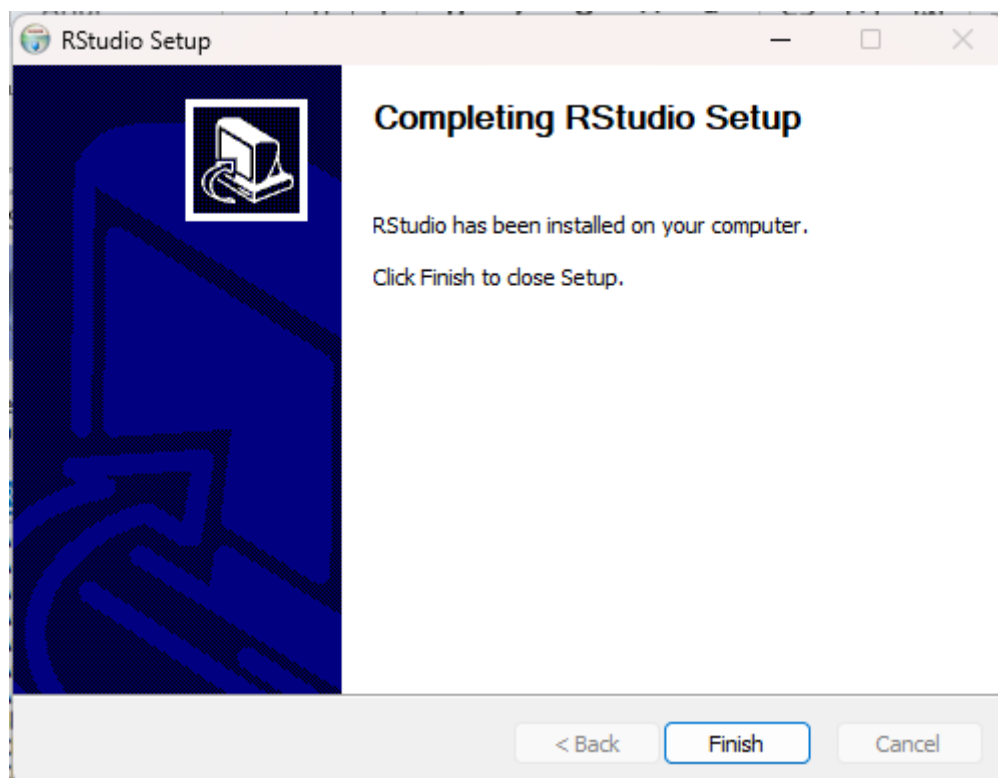
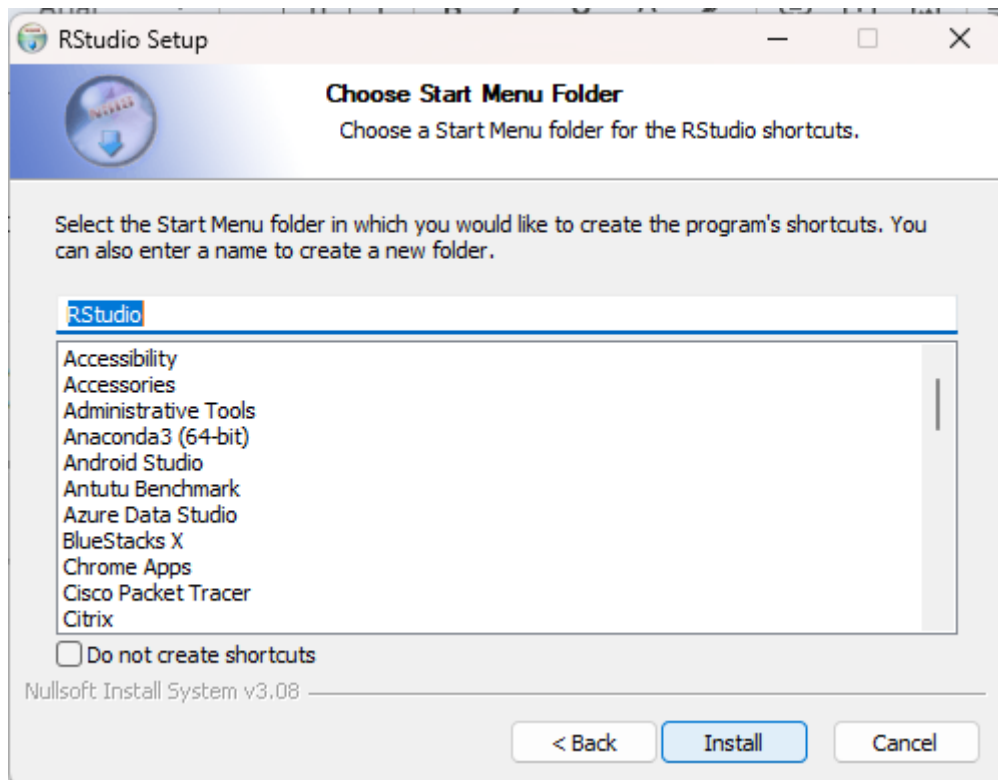
Rstudio installation











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 manoragnjan  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 × 4
  `roll no.` name      surname course
  <dbl> <chr>      <chr>   <chr>
1         1 manoragnjan 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("manoragnjan", "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
3				

```
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    |
+-----+-----+-----+
| manorajan | 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@gmail.com', 'shriya@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