

# CSE 5720-60

## Summer 2024

### Project 1- MySQL

Acknowledgment: The project was originally designed by Dr. Carey with slight modifications tailored for MySQL.

In this project, we are going to create a database and tables using MySQL and import the data into tables. Then you are required to form SQL queries for problem statements shown in step 5. Due by 11:59PM on **June, 18 (Tuesday)**. Turn in the 2 deliverables to folder “Project 1” on Canvas.

#### Deliverables

1. An SQL script that contains the queries listed in the same order as shown in step 5.
2. .csv files that contain the results obtained.

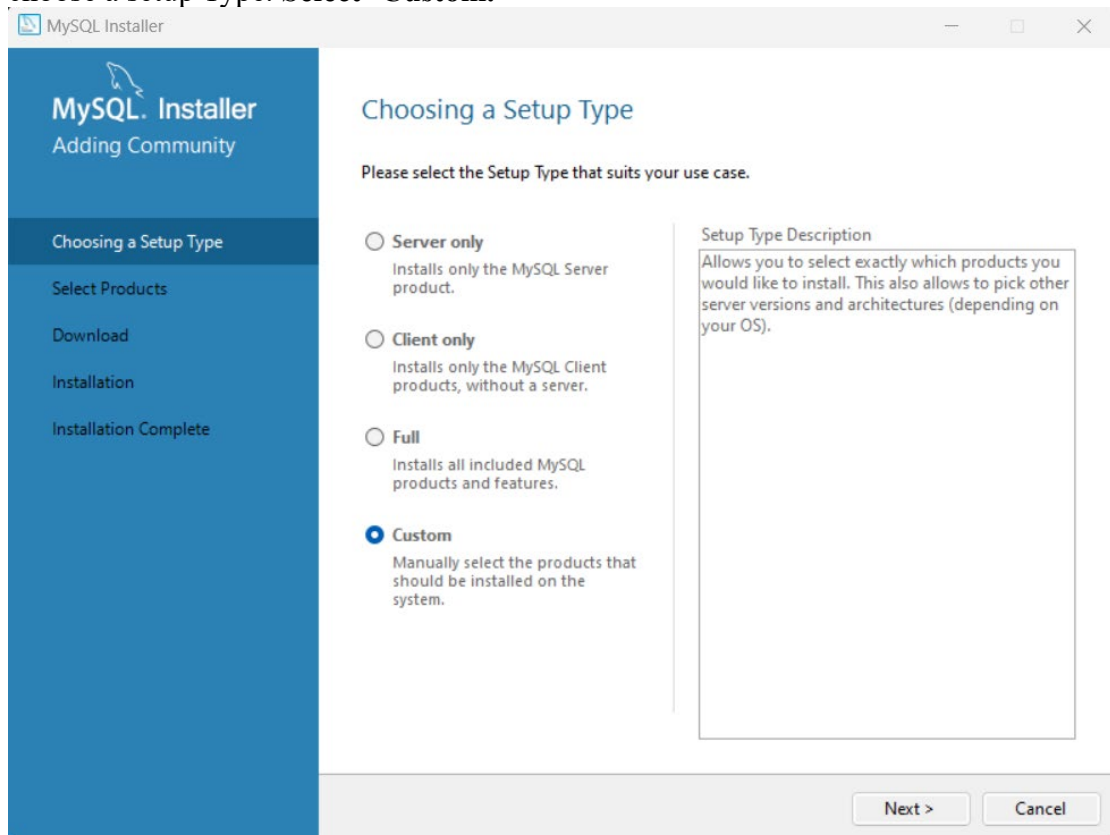
#### STEP 1 - Install MySQL

Go to MySQL download page (<http://dev.mysql.com/downloads/>) to download MySQL Installer for Windows. Select the correct installer based on your system.

For Mac OS X, download the DMG files, extract it, and install by the order: mysql-8.0.xx-osx10.6-x86\_64.pkg, MySQLStartupItem.pkg, MySQL.prePane

The following description is based on MySQL server 8.0 for Windows.

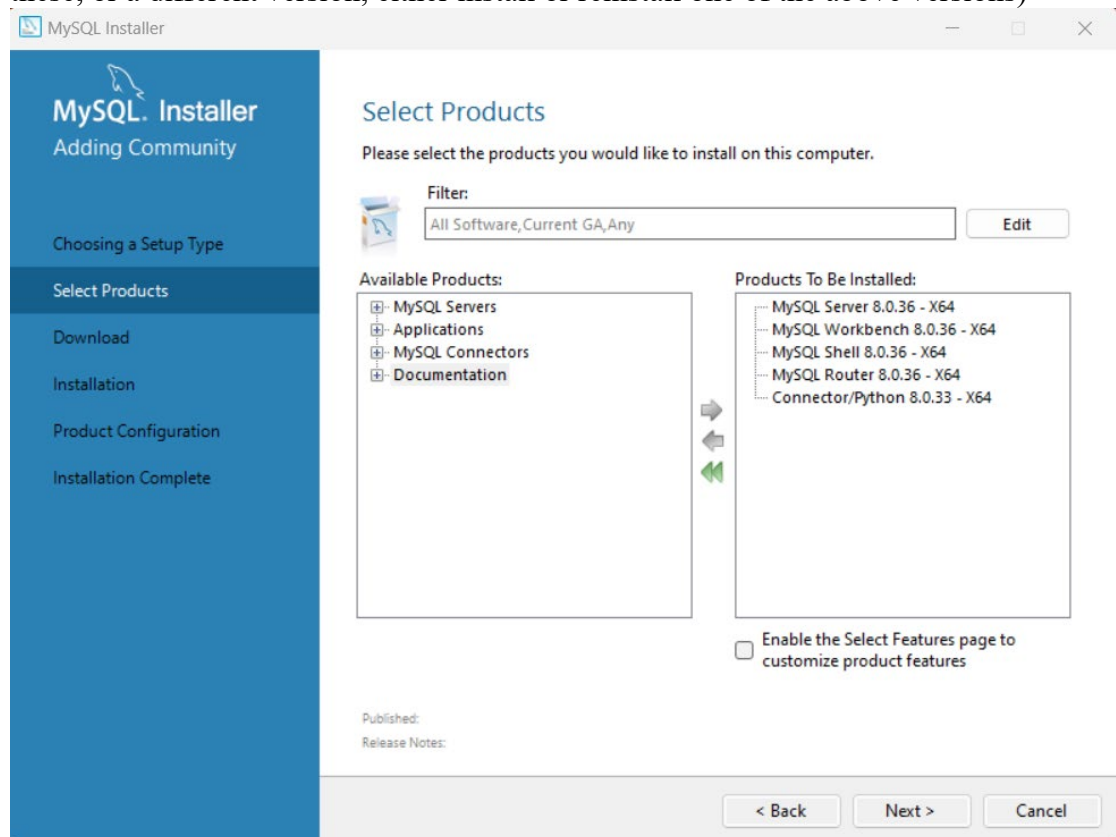
1. Execute the downloaded file and complete the installation, then it will ask you to choose a setup Type. **Select ‘Custom.’**



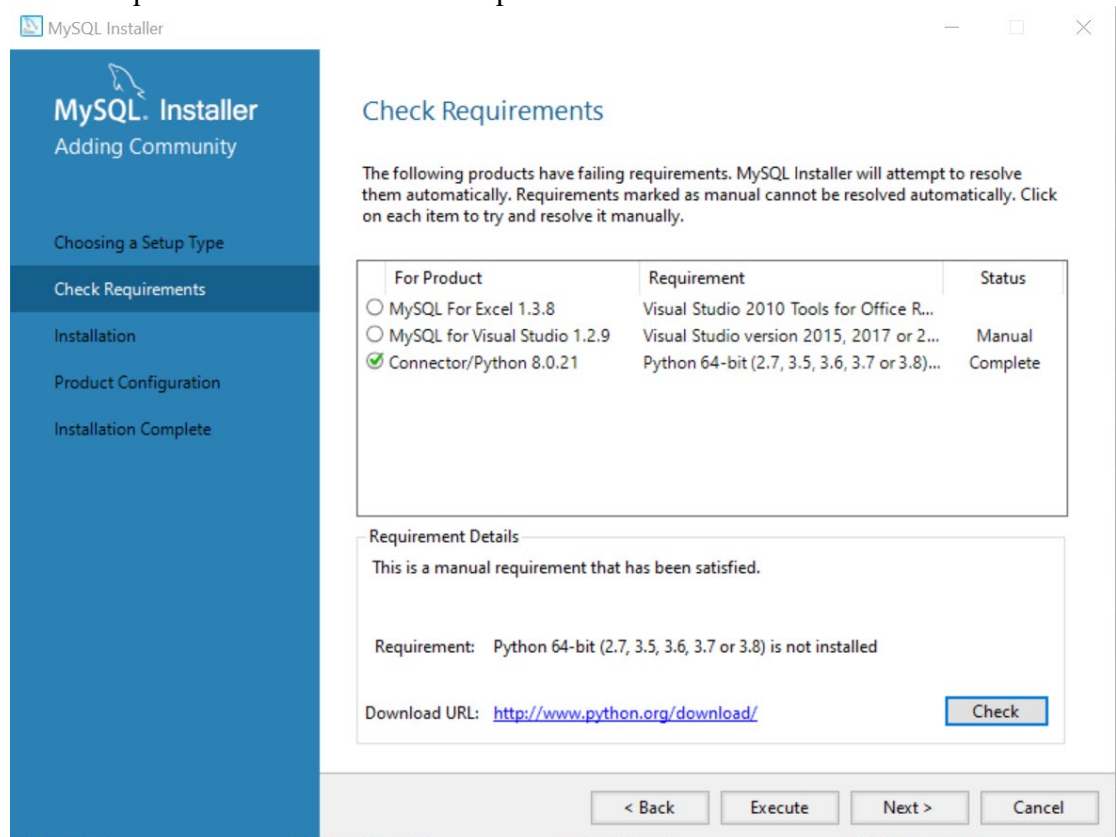
2. Add the following product to the install list: MySQL Server 8.0.x, MySQL Workbench 8.0.x, MySQL Shell 8.0.x, MySQL Router 8.0.x, Connector/Python

8.0.x, where x indicates the latest available version.

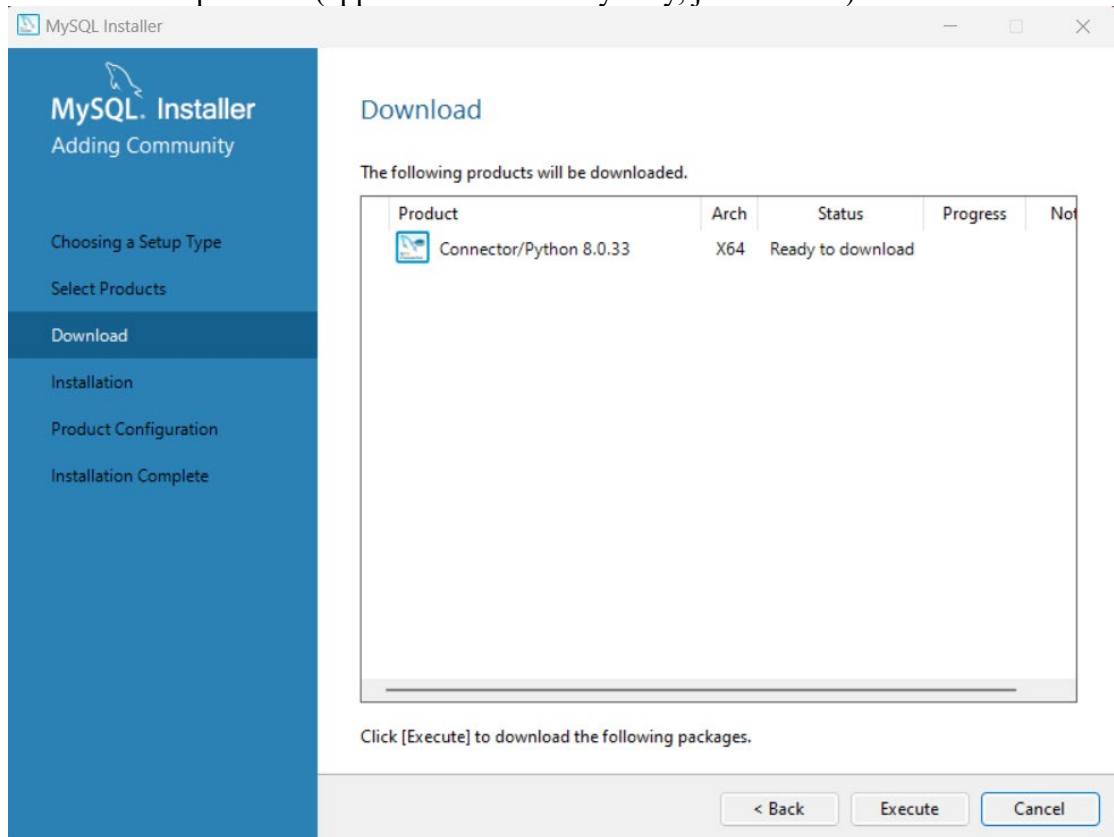
(NOTE: The Python Connector for MySQL 8.0 requires you to have Python version 3.11, 3.10, 3.9, 3.8, or 3.7 installed on your device. If you have none of these, or a different version, either install or reinstall one of the above versions)



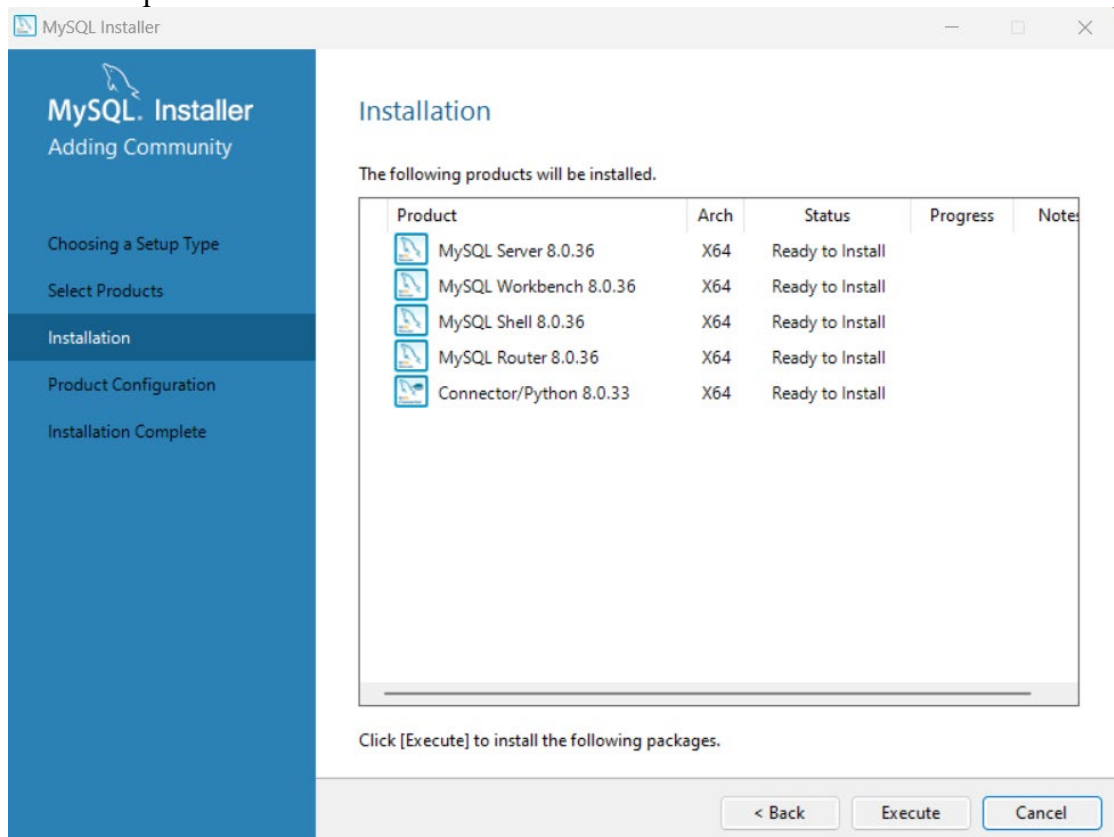
3. Check requirements and install the required software.



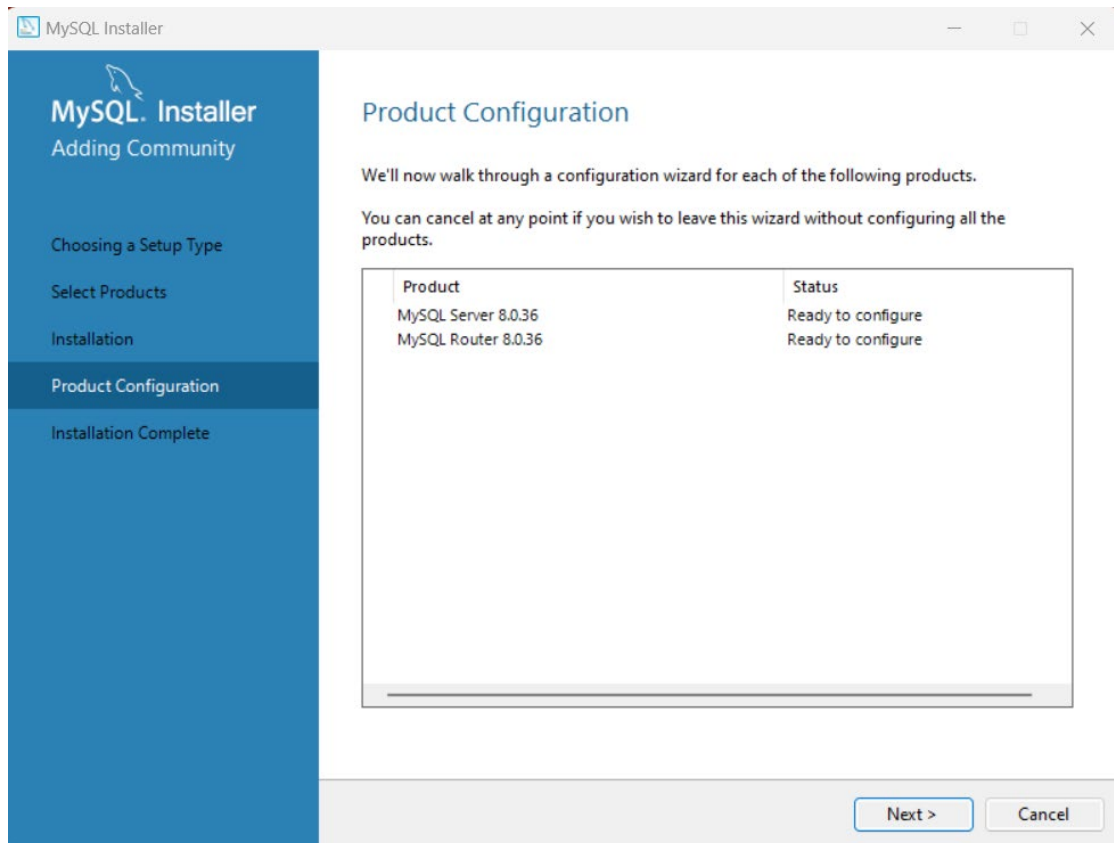
- Download the products (appearance of list may vary, just execute).



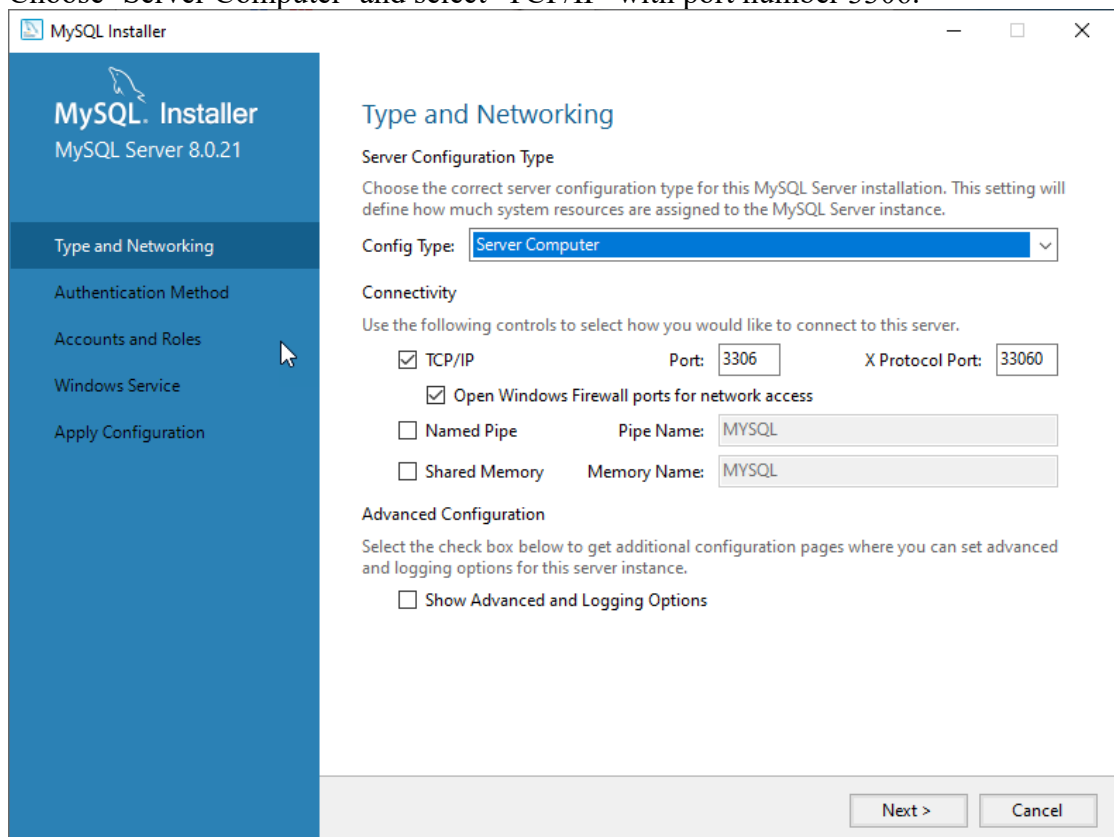
- Install the products.



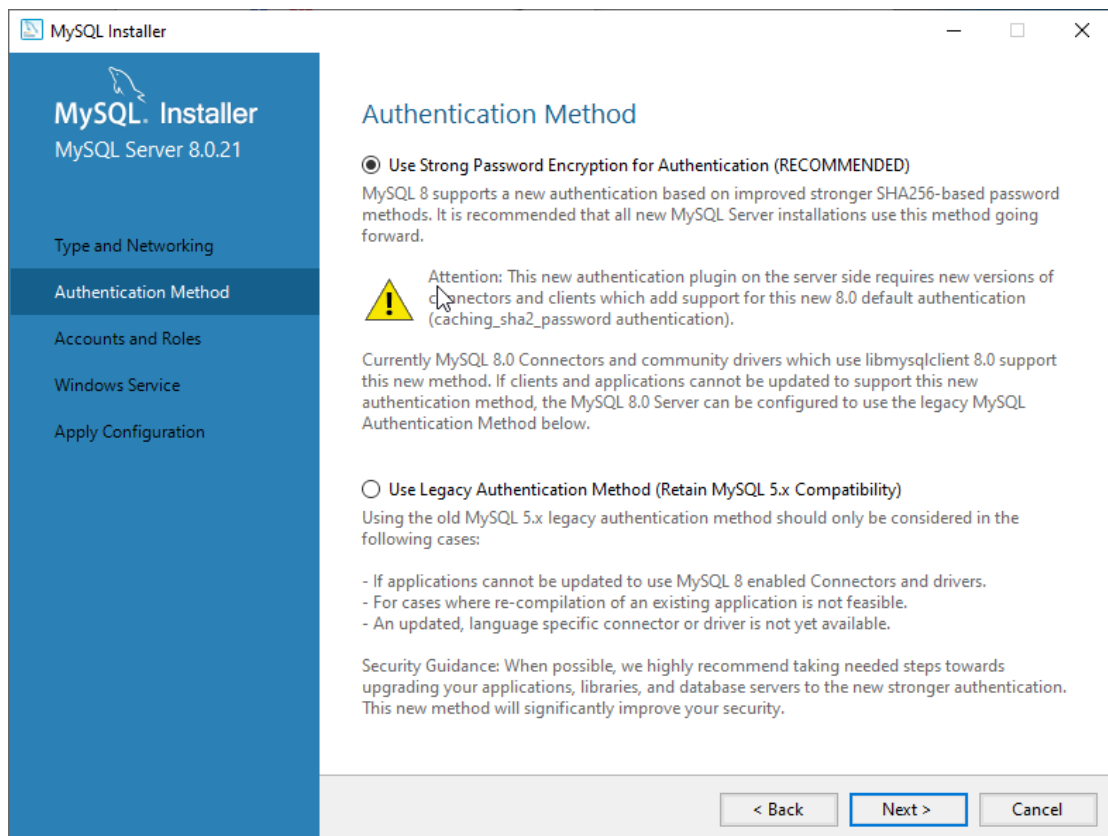
- Select next for the product configuration.



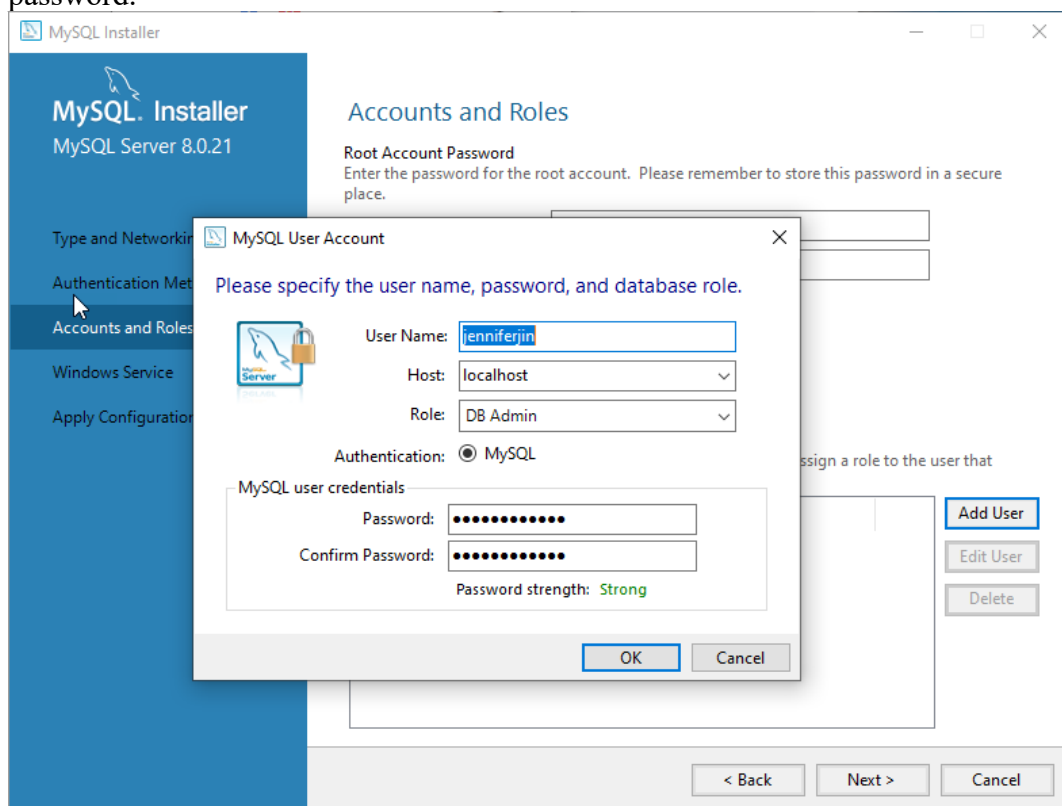
7. Choose 'Server Computer' and select 'TCP/IP' with port number 3306.



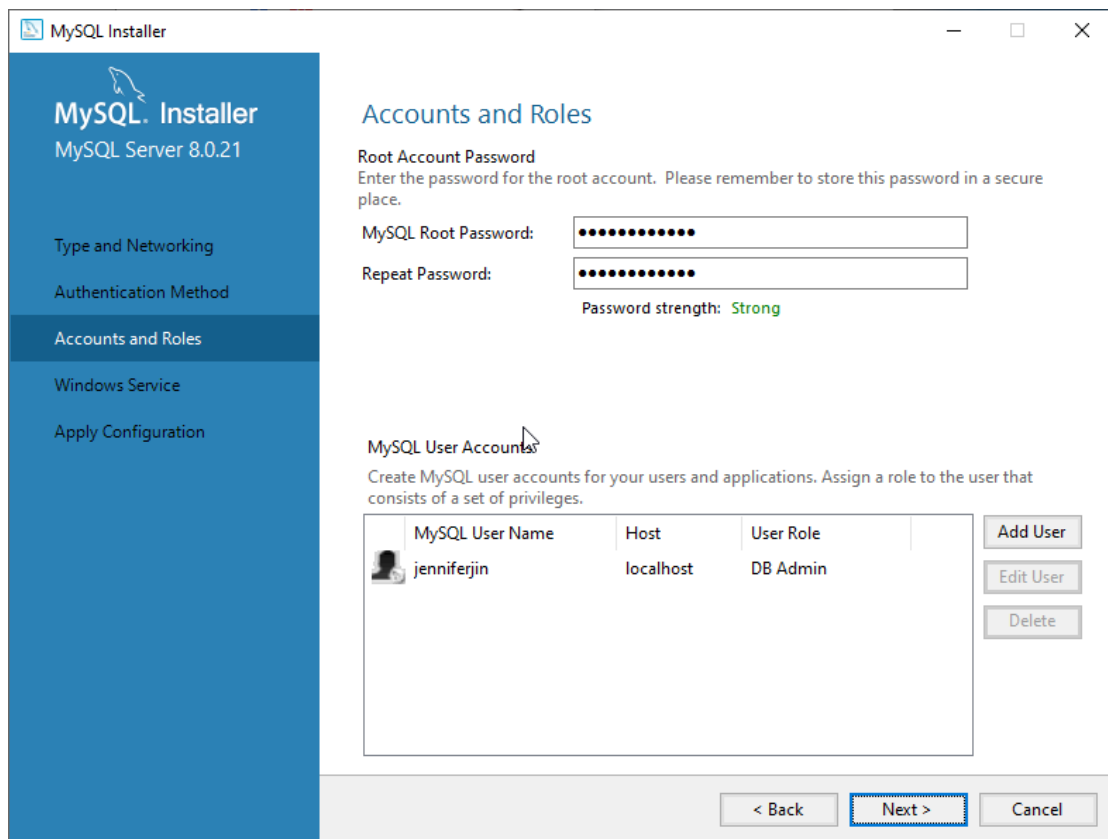
8. Select 'Use strong password encryption for authentication.'



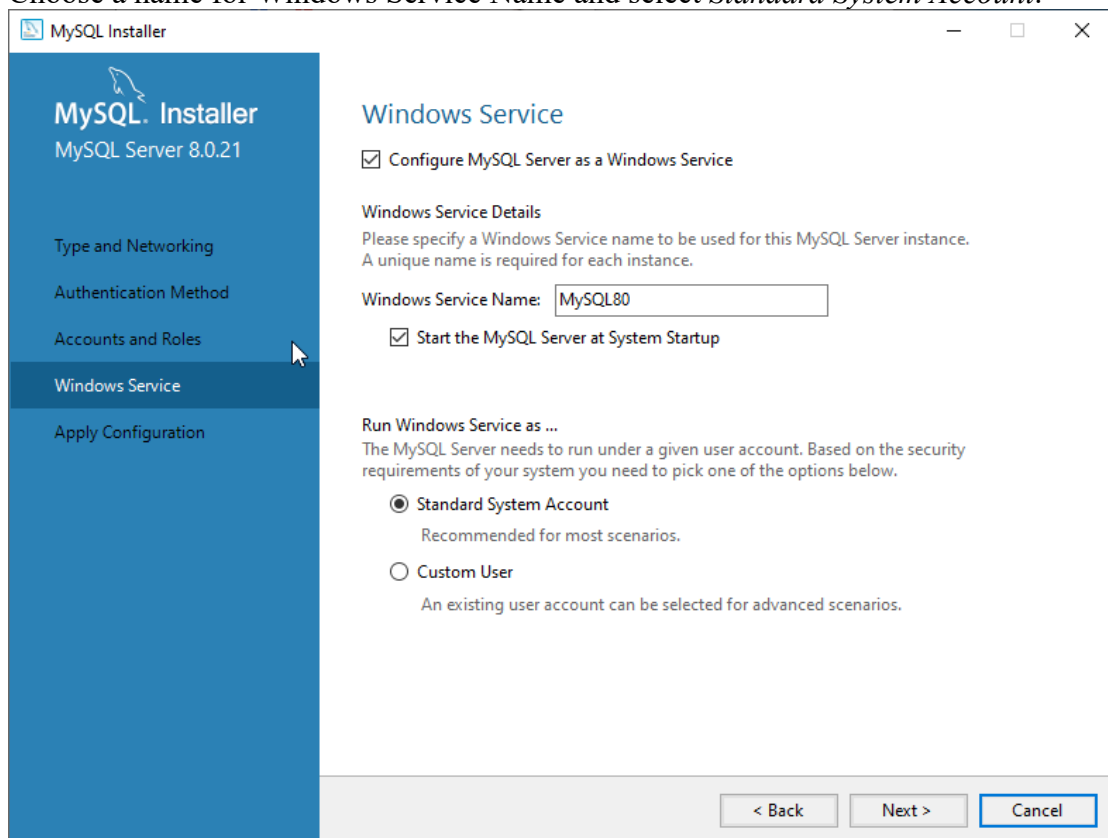
9. Select a user name, choose *localhost* for host, *DB Admin* for Role and type your password.



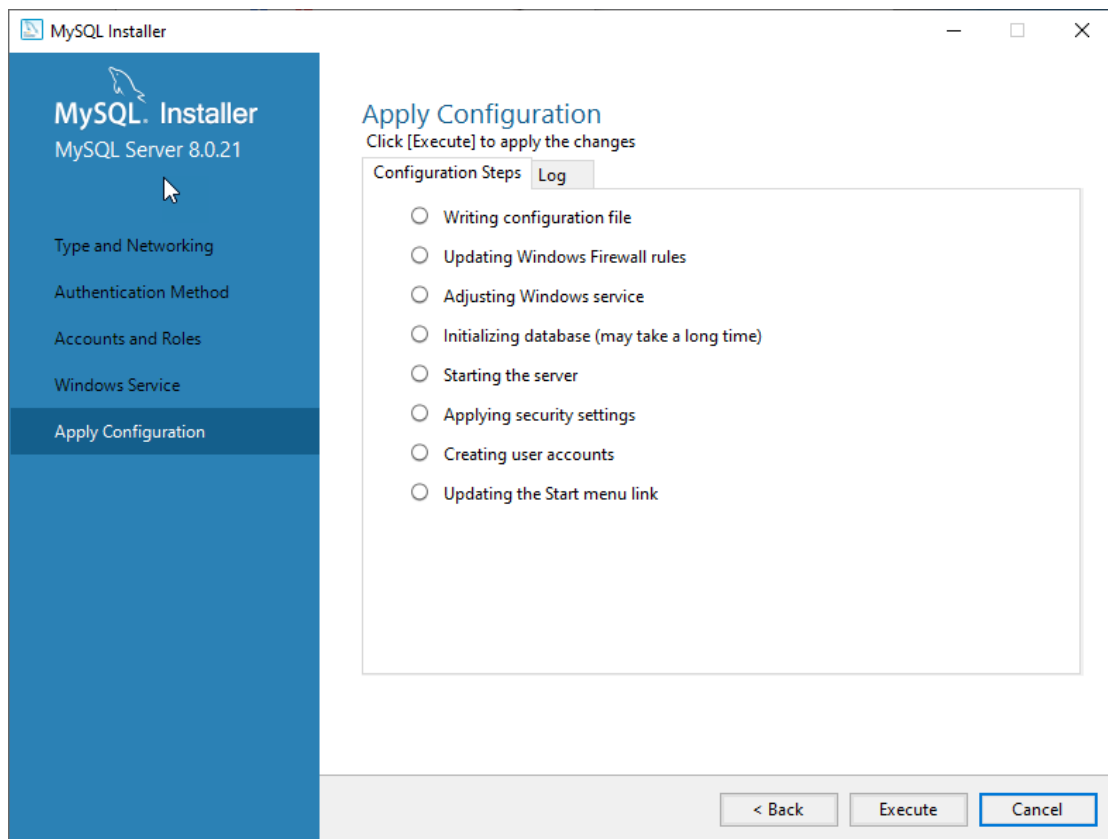
10. Type your password and click next.



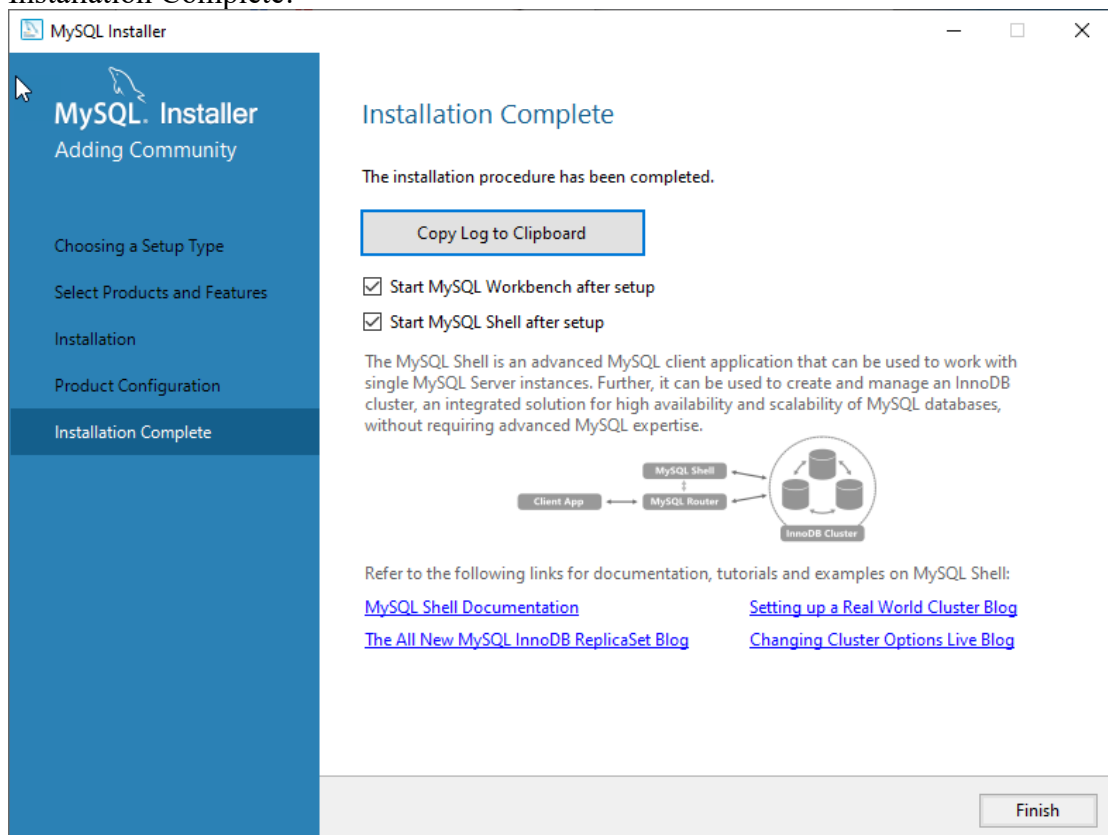
11. Choose a name for Windows Service Name and select *Standard System Account*.



12. Apply configuration.

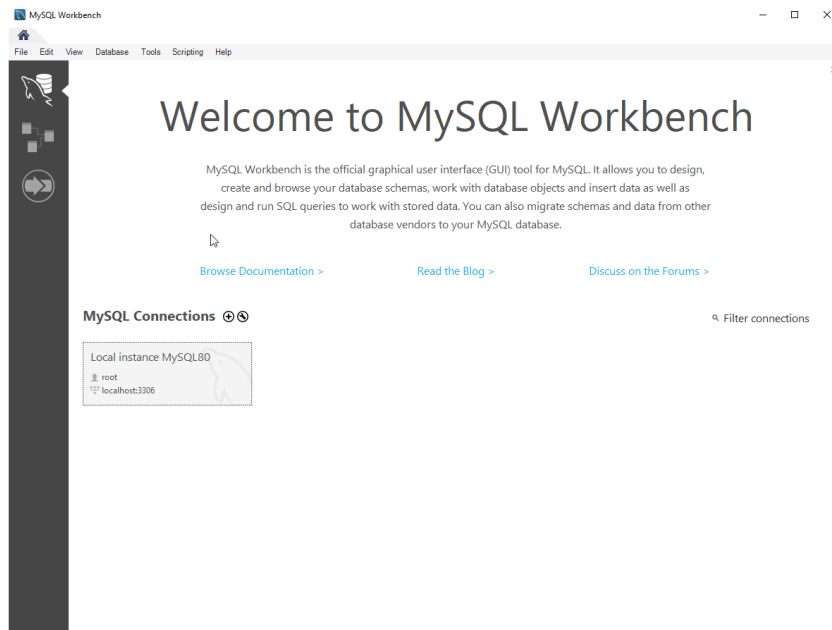


### 13. Installation Complete!



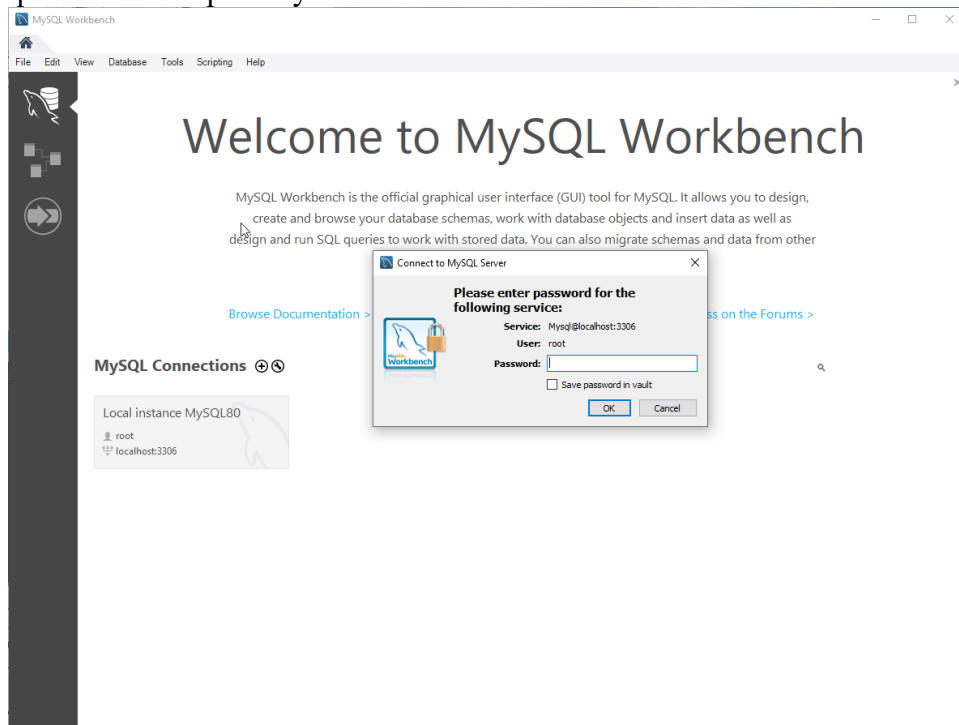
### STEP 2 - A GUI Tool for MySQL – MySQL Workbench

Download the MySQL Workbench from MySQL download page. After the installation, run it and you should see the following:



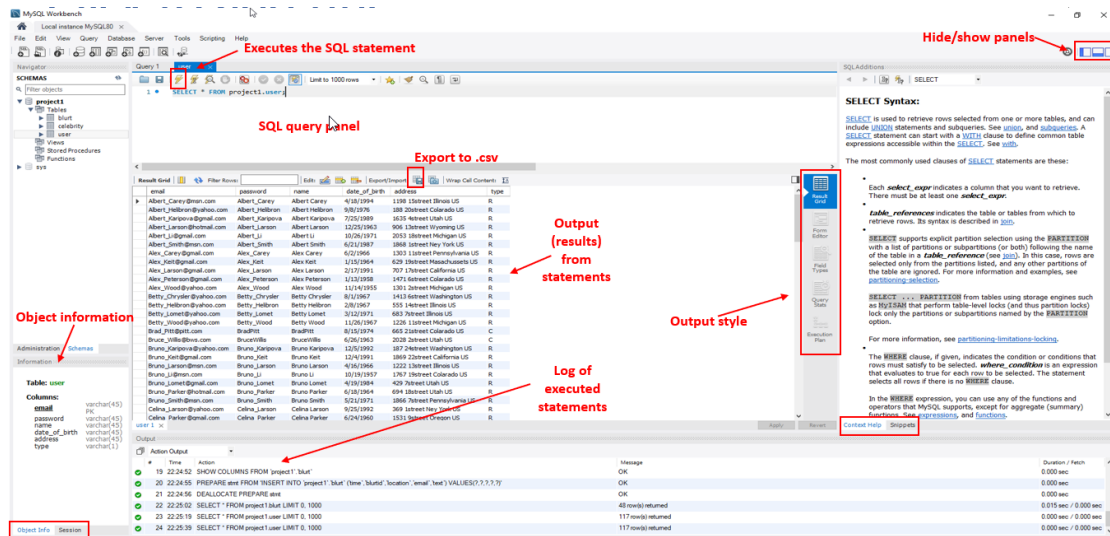
Double click on the local instance. Note that the port should be the same as when you installed MySQL (3306 in the case). If not, edit the connection by:  
**Windows and Linux:** hover over the right side of a connection title and click the title.  
**OS X:** hover over a connection title and click the little (i) in appears in the bottom right corner

Type in your password and connect to the database. Now you can perform queries or manipulate your database:



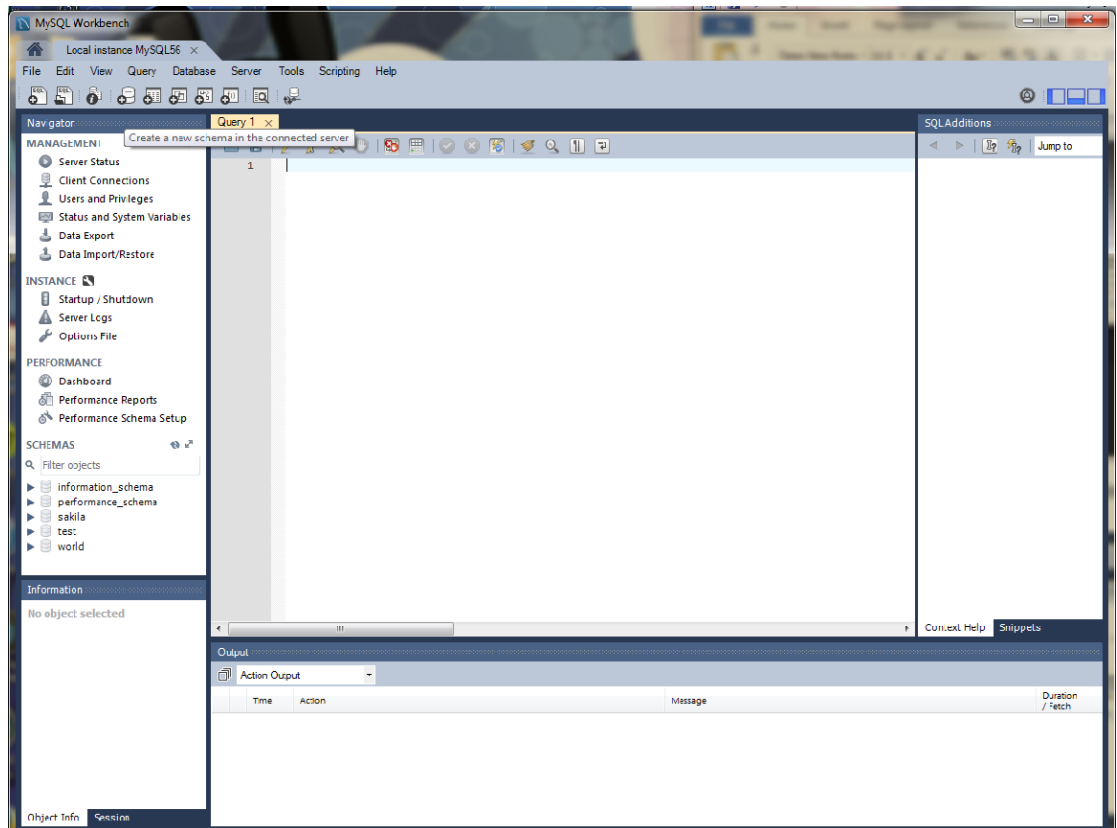
This is the MySQL workbench overview.



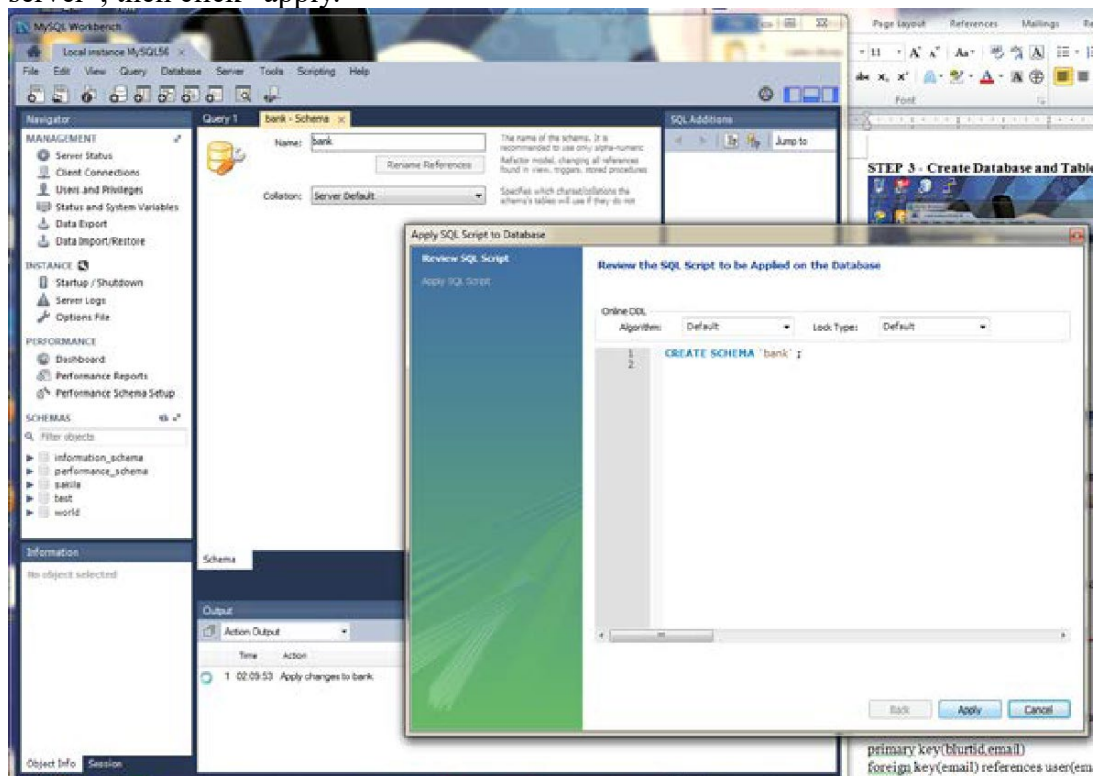


## STEP 3 - Create Database and Tables

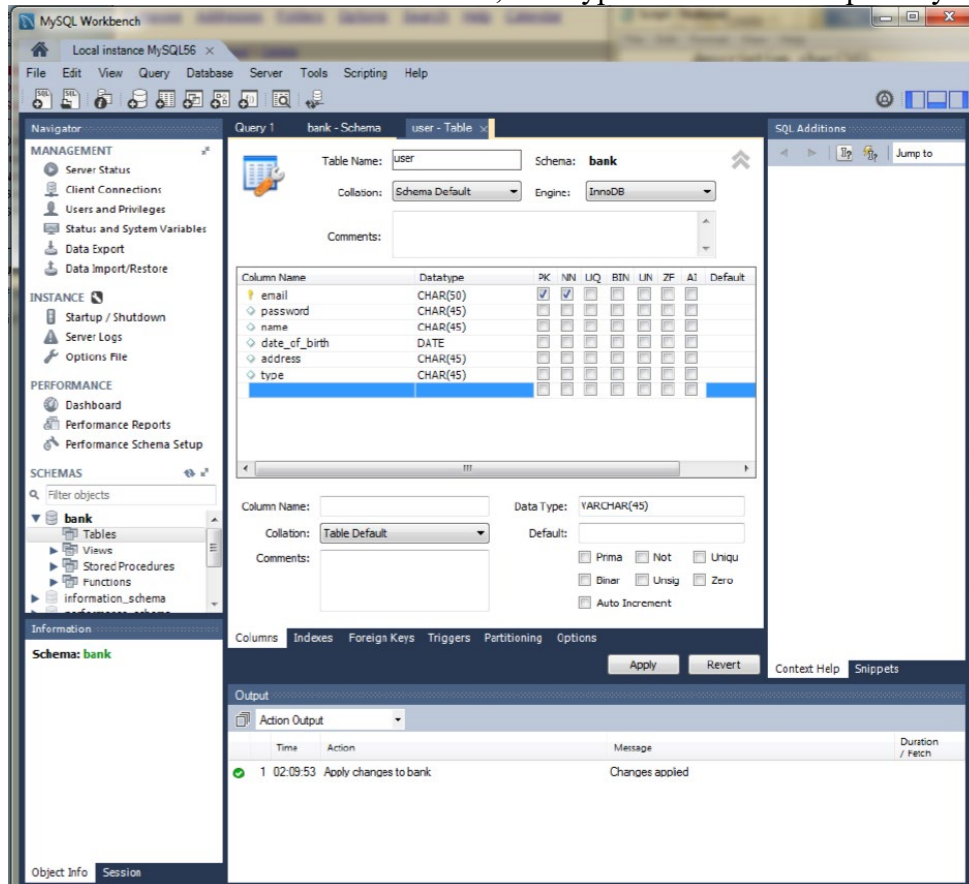
1. Click on "Create a new schema in the connected server" icon on the upper left corner.



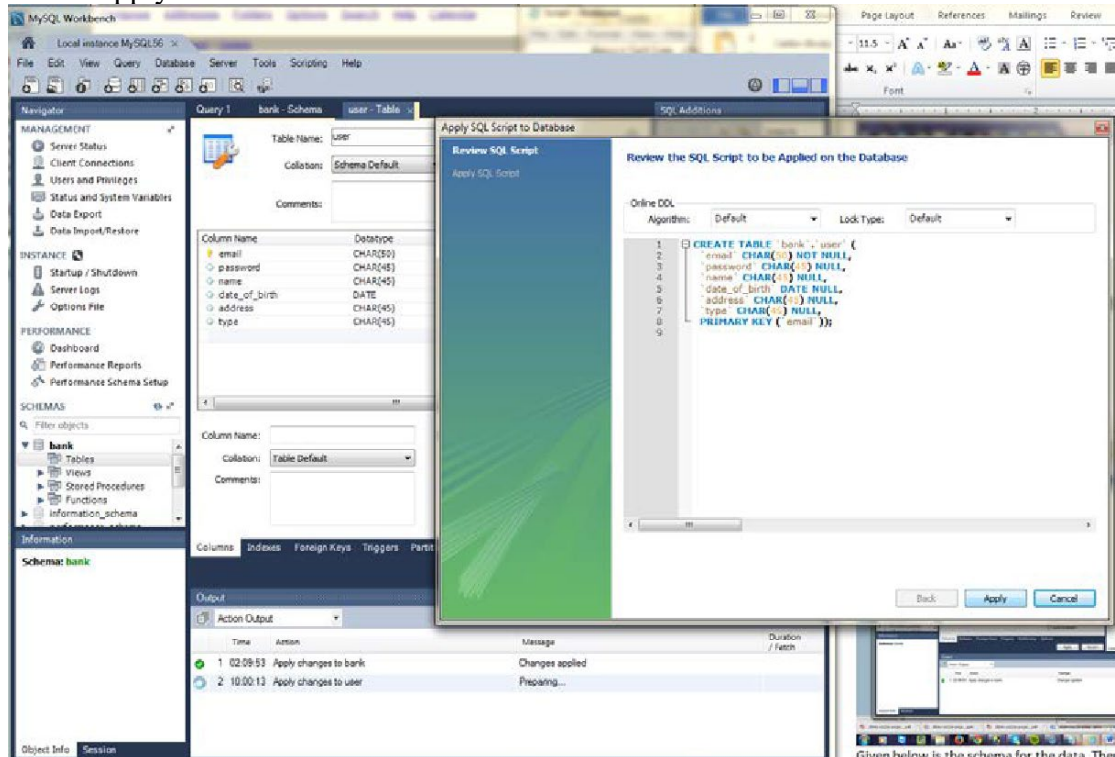
- Under the schema click on “create a new table in the active schema in connected server”, then click “apply.”



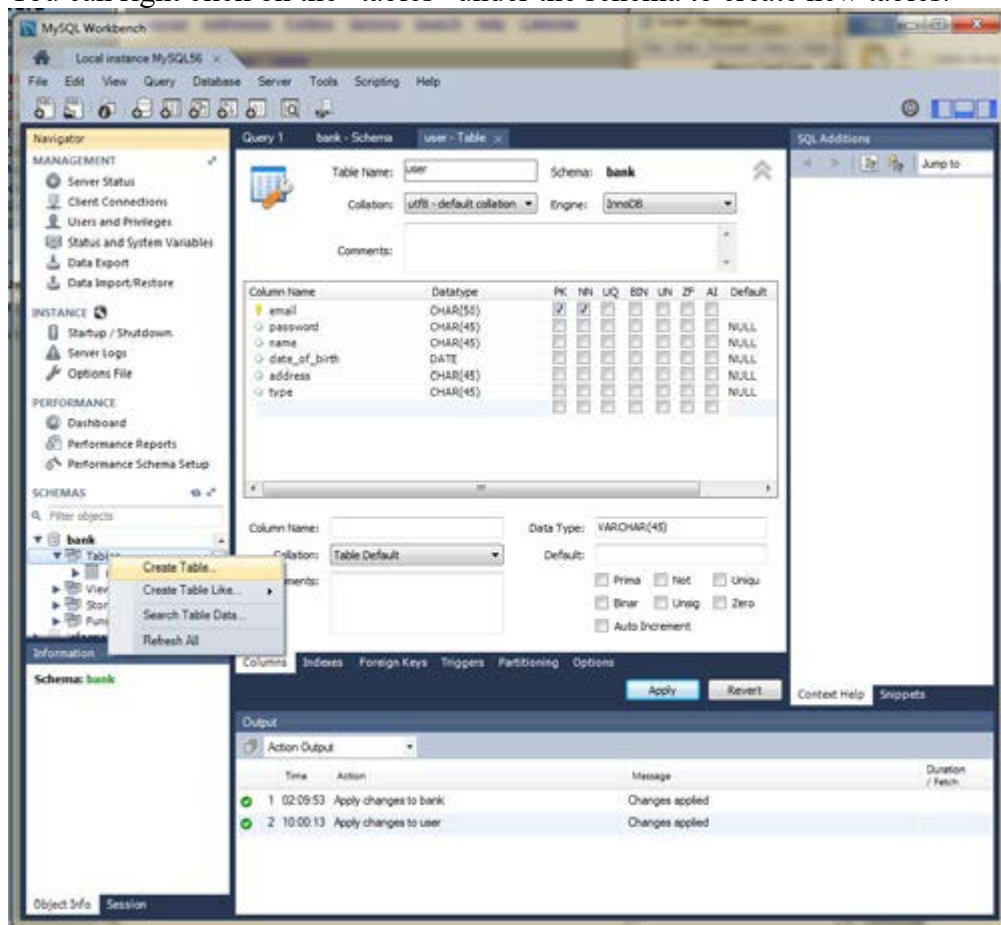
- Fill in the columns with column name, data type and indicate the primary key(s).



4. Click “apply.”



5. You can right click on the “tables” under the schema to create new tables.



Given below is the schema for the data. There are a total of 12 tables, thus 12 CSV files, each corresponding to a relational table.

user (email, password, name, date\_of\_birth, address, type)  
primary key(email)

celebrity (email, website, kind)  
primary key(email)

blurt (blurtid, email, text, location, time)  
primary key(blurtid,email)  
foreign key(email) references user(email)

hobby (email, hobby)  
primary key(email,hobby)  
foreign key(email) references user(email))

follow (follower, followee)  
primary key(follower, followee)  
foreign key(follower) references user(email)  
foreign key(followee) references user(email))

vendor (id, name)  
primary key(id)

vendor\_ambassador (vendorid, email)  
primary key(vendorid)  
foreign key(email) references user(email)  
foreign key(vendorid) references vendor(id))

topic (id, description)  
primary key(id)

vendor\_topics (vendorid, topicid)  
primary key(vendorid, topicid)  
foreign key(vendorid) references vendor(id)  
foreign key(topicid) references topic(id))

blurt\_analysis (email, blurtid, topicid, confidence, sentiment)  
primary key(email, blurtid, topicid)  
foreign key(email, blurtid) references blurt(email, blurtid)  
foreign key(topicid) references topic(id)  
constraint confidence >= 0 and confidence <=10  
constraint sentiment >= -5 and sentiment <=5

advertisement (id, content, vendorid)  
primary key(id)  
foreign key(vendorid) references vendor(id))

user\_ad (email, adid)  
primary key(email, adid)  
foreign key(email) references user(email)

foreign key(adid) references advertisement(id))

The design model:

Users can post their thoughts in form of short messages that we call “blurts”. When signing up, users need to provide their email and a password of their choice. In addition, they need to enter some basic information – name, date of birth, address, email ID and hobbies. Once signed up, they can (besides blurting) “follow” other users. To “follow” a user means subscribing to his/her “blurts”. Users are categorized into “regular” users and “celebrities”. A celebrity has an associated website url and an attribute called “kind” indicating whether he is a politician, actor, singer, etc. Each blurt by a user (regular or celebrity) is assigned an id. Blurt ids are serial and unique to the “blurts” by a given user; the first blurt by a given user would have blurt id 1 and ids are incremented for each successive blurt by the user. Note that blurt ids are unique only to a user, so blurts by two different users may have the same blurt id. Besides an id, each blurt also has its text, timestamp, and user location as additional attributes. The system should have a pre-defined notion of “topics” that are simply subjects that people may blurt about. Examples of topics might include music, pollution, disease, disaster, sports, weather, etc. A topic has a unique id and description (the name of the topic). Each blurt by a user is analyzed to associate with it zero or more topics. Related blurt-topic pairs are stored in blurt\_analysis table. To account for the possible ambiguity arising from the choice of words or language used by a user, an association with a topic has a corresponding confidence level (an integer ranging from 1 – 10 indicating the strength of the association). For example consider the following blurt: “I absolutely hate the rainy weather, can’t go out, listening to the Beatles, just love them” is analyzed to be associated with two topics, weather and music (Beatles). For each topic, the associated sentiment is evaluated and quantified as an integral value ranging between -5 and 5, with higher values indicating a more positive sentiment. Considering the example blurt used above, the topic weather would have an associated sentiment of -5 (hate) while for music the corresponding value is 4 (love). **Note: You don’t need to implement the value constraint as MySQL doesn’t support it.** A vendor has interest in one or more topics and is interested in tracking all users who are blurting about a topic of interest. A vendor may also have a celebrity as its brand ambassador. Vendors create advertisements that have an associated unique id and a textual content. These advertisements are stored in the system and are available to be shown to the regular set of users (that is, not to the celebrities, just to the other “regular” users). Careful matching is done based upon a historical analysis of all blurts by a user. Based upon the analysis, a user may be shown zero or more advertisements.

#### STEP 4 – Import CSV files

Script Template:

```
LOAD DATA LOCAL INFILE "[CSV file name]" INTO TABLE [table name]
COLUMNS TERMINATED BY ',' LINES TERMINATED BY '\n'
```

For each CSV file, replace [CSV file name] and [table name] with actual CSV file name and corresponding table name, e.g.:

```
LOAD DATA LOCAL INFILE "d:\\csvdata\\advertisement.csv" INTO TABLE
advertisement COLUMNS TERMINATED BY ',' LINES TERMINATED BY '\n'
```

Execute 12 scripts using the GUI client

### STEP 5 - Form SQL Queries

For the following statements, you are required to form SQL queries and execute them using the GUI client. Then export the result using the name "Query x.csv", x being the label of each query. Put all the SQL you formed into a file named "Script.txt" in the same order. Then archive the file as "5720project1-xxxxxxxxx.zip", xxxxxxxxxx being your student id, and turn it in on Canvas. The filenames of your result has to follow the instructions exactly or you may get a deduction in your credit.

- 1) Print every hobby, along with the total number of users that included that hobby in their profile.
- 2) Print the name and address of all celebrities who are not movie stars.
- 3) Print a list of every vendor and the number of different advertisements they're running, in descending order.
- 4) Print email, password, name, date\_of\_birth, kind of all celebrities who are Movie Stars and are following someone.
- 5) Print the name and email of every celebrity along with their total number of followers if the amount is over 55. The contents should be in descending order based on the number of followers.
- 6) Find all pair of users (A,B) such that A and B have not blurted on a common topic, but user A is following user B. The query should print the names of A and B in that order.
- 7) Print every location, along with the total amount of blurts, where users made more than 40 collective blurts. The query should print the name of the location and total amount of user blurts in descending order of the amount of blurts.
- 8) For each vendor, print the email of its brand ambassador along with the number of users the brand ambassador is following. Your query should print the vendor name, ambassador email, and number of users the ambassador is following. The output should be in descending order according to the numbers of users being followed.
- 9) Print the following: vendor id, vendor name, vendor ambassador's email, and the total number of users that have been shown the ad(s) from the vendor. Do this for every vendor.