Database and Data

Installation and setup guide

Table of Contents

[**Introduction** 2](#_Toc461725288)

[**Setup Introduction** 2](#_Toc461725289)

[**MySQL Installation** 3](#_Toc461725290)

[Windows 3](#_Toc461725291)

[Mac OS 9](#_Toc461725292)

[Installing the MySQL Database 9](#_Toc461725293)

[**PyMySQL Installation** 11](#_Toc461725294)

[**Connecting to MySQL from Python** 12](#_Toc461725295)

[**Downloading and Importing Exercise Data** 15](#_Toc461725296)

[To import using the MySQL command line 15](#_Toc461725297)

[To import using MySQL Workbench 16](#_Toc461725298)

# **Introduction**

The goal of the “BIG DATA AND SOCIAL SCIENCE RESEARCH: THEORY AND PRACTICAL APPROACHES” book is to bring computer scientists and social scientists together to provide a practically oriented overview of the analytical and statistical tools associated with big data for social science students. We use real data used for real world policy problems to analyze some of the vast new sets of data on human beings. We show how this data can be collected, integrated, and analyzed in a scientific fashion. In the accompanying exercise notebooks, we walk students through key programming and analysis techniques so they have working examples they can apply to their own work.

# **Setup Introduction**

In most of the book’s exercises, you will work with data retrieved from a "homework" database that you will be downloading and installing in a local database. This document shows you how to install and configure a MySQL database server, the MySQL Workbench program, Python packages for interacting with MySQL in Python programs, and the data for the class. It also includes some brief notes on using PostgreSQL instead. We provide PostgreSQL database SQL files, but we don’t provide detailed documentation.

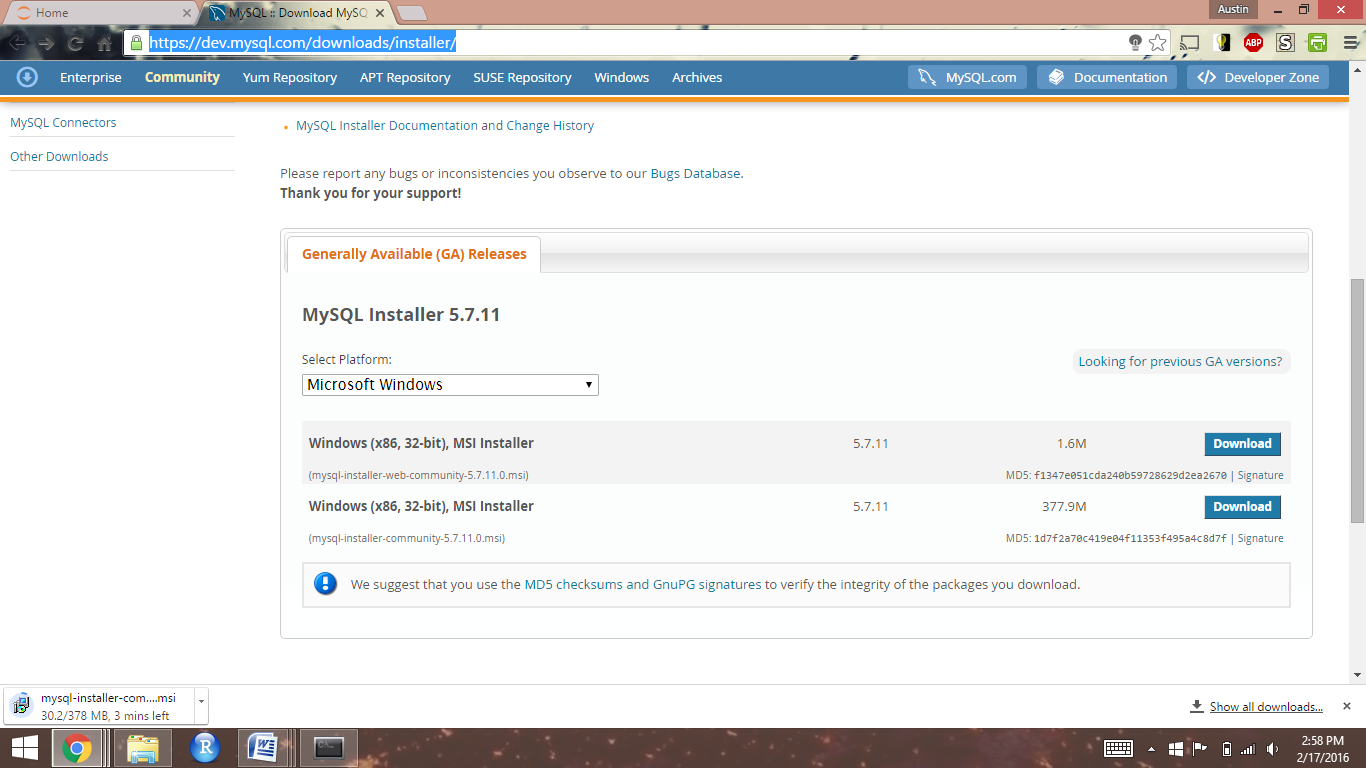
*Note:*

* *Before installing MySQL, you should use the document Anaconda\_Installation\_Guide.docx to install Python, IPython, and Jupyter.*
* *The installation instructions below use Windows screenshots. Installation on a Mac follows the same steps, though some of the installers might work a bit differently.*

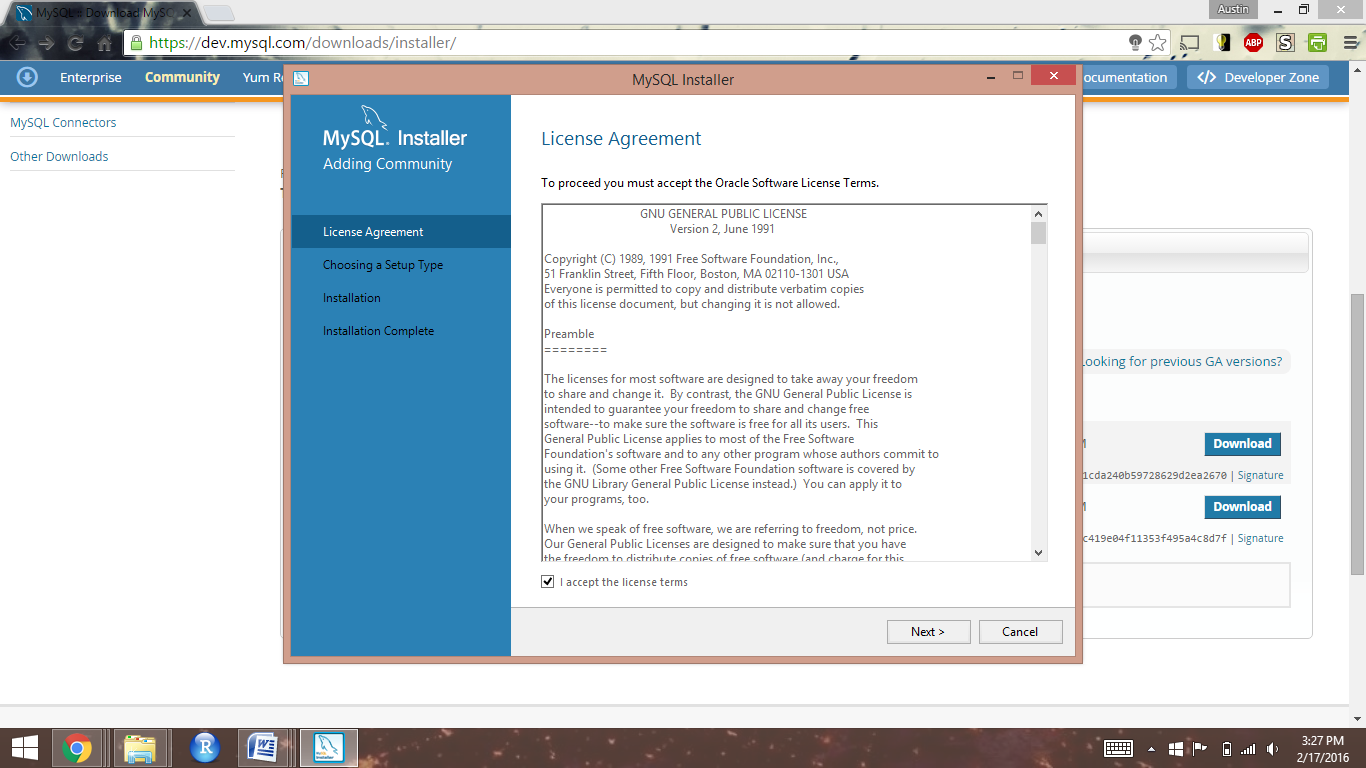
# **MySQL Installation**

Go to the [MySQL download website](https://dev.mysql.com/downloads/installer/) and download the MySQL Installer package (if you don’t want to create an Oracle ID, just click the “No thanks, just start my download” link at the bottom of the page).

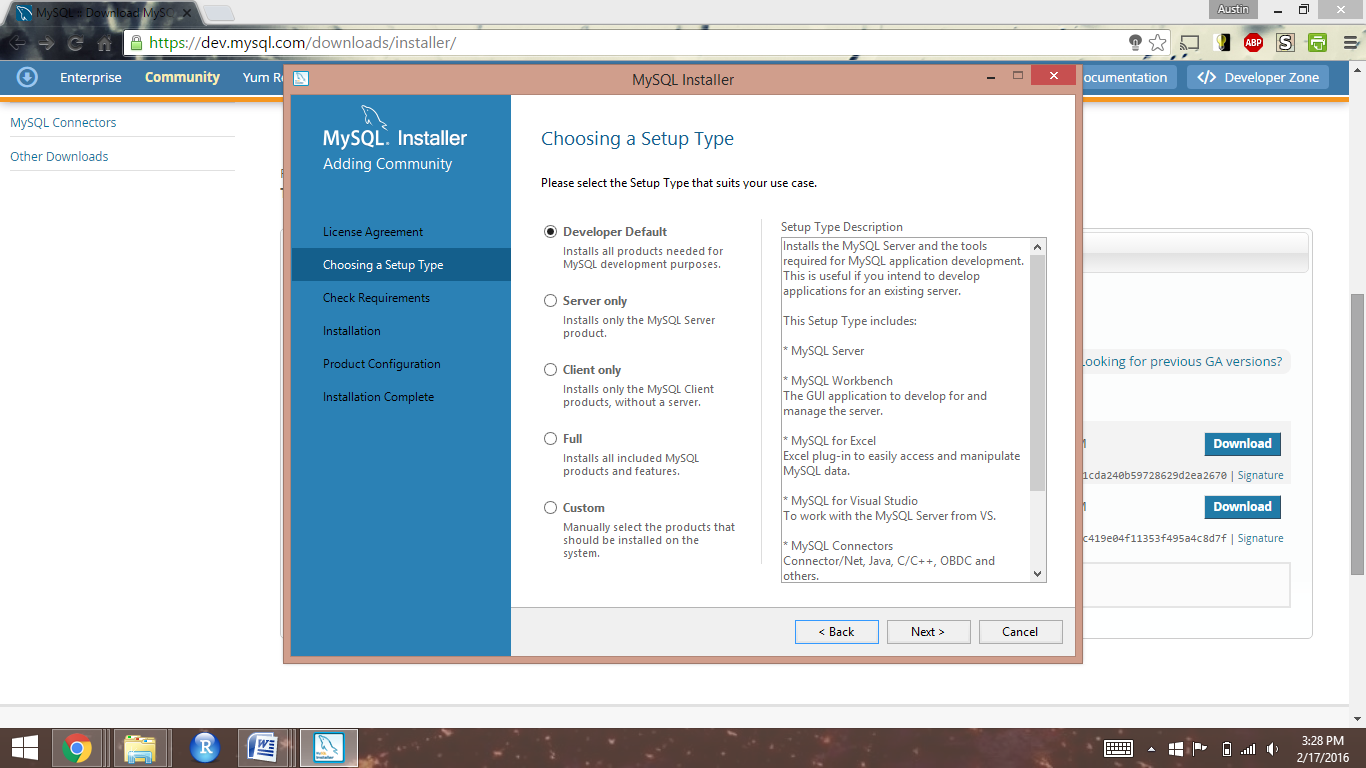
## Windows



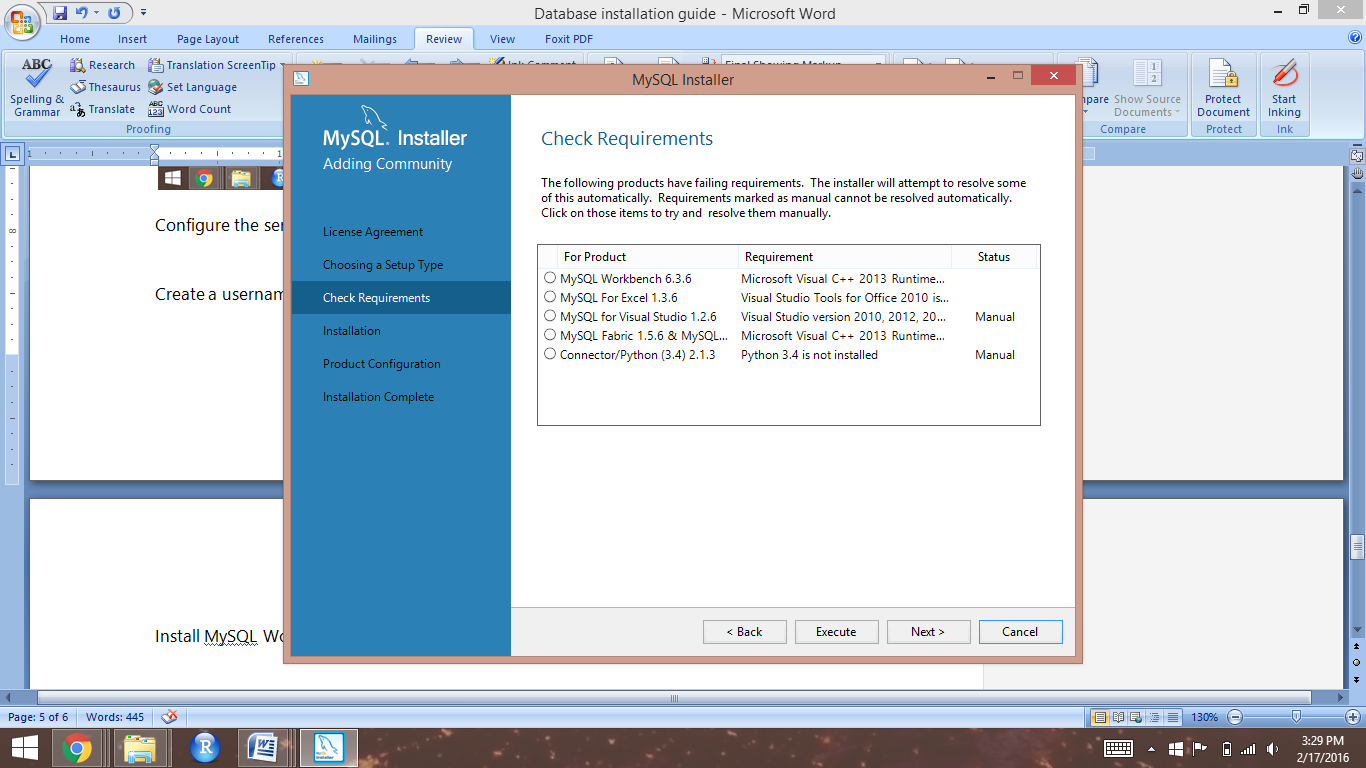
Open the downloaded executable file and move forward with installation



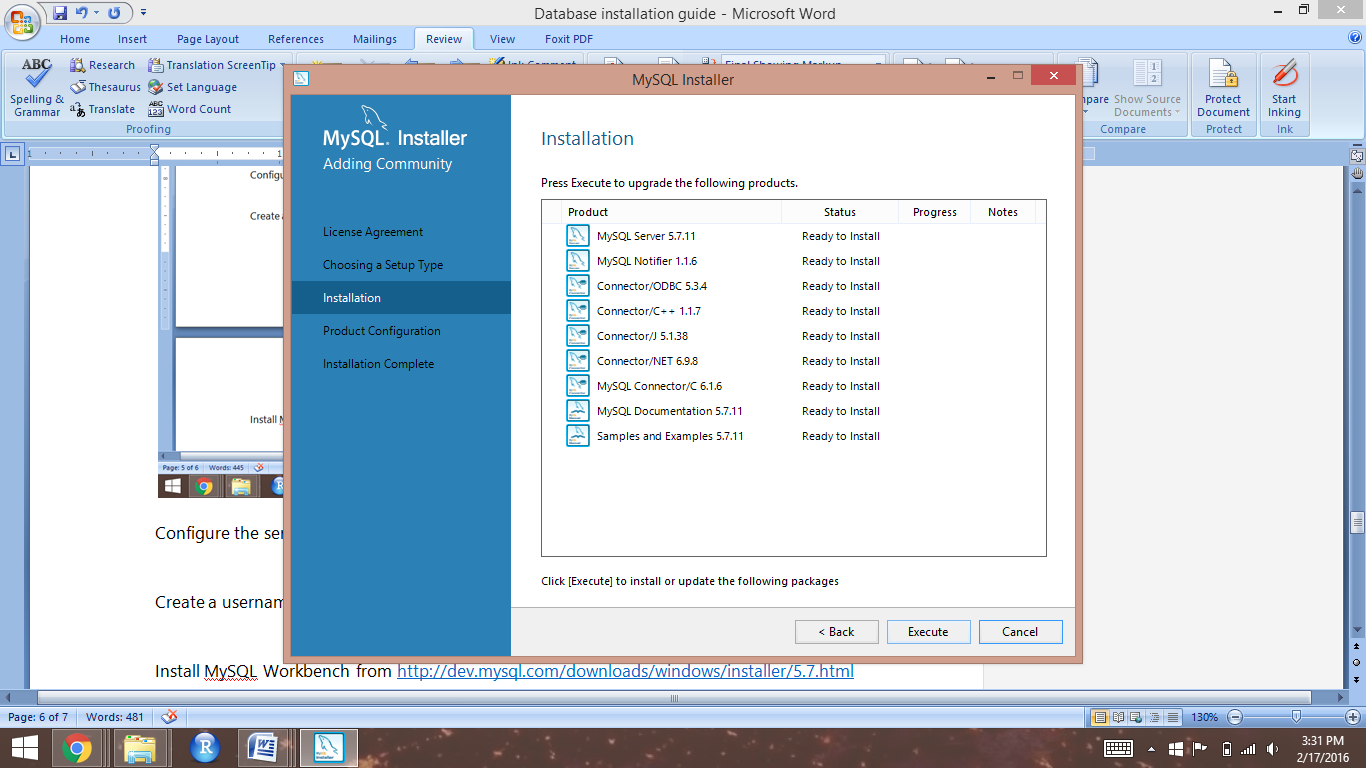
Select "Developer Default" and continue



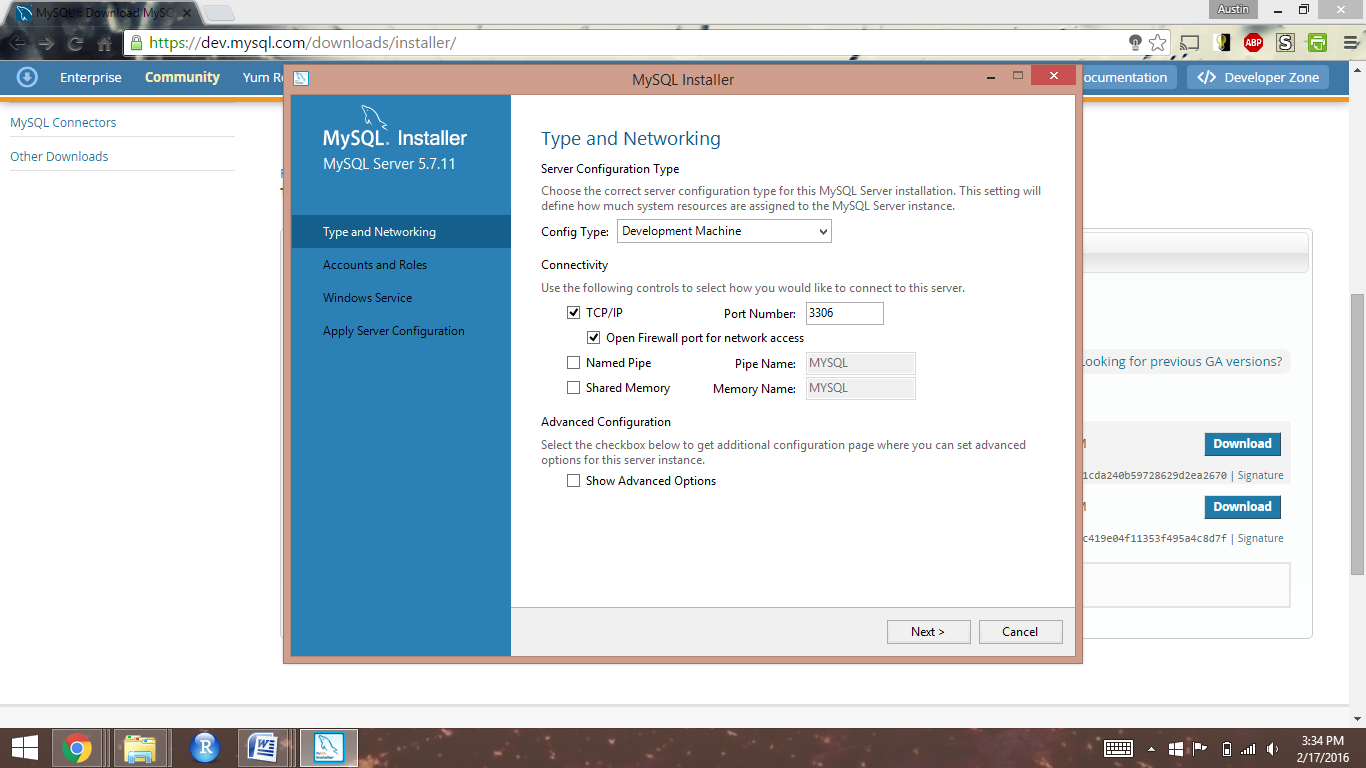
When performing the requirement check, you may encounter that your system lacks requirements for the most up-to-date versions of the MySQL software. This should not be a problem, and you should move forward with the installation.



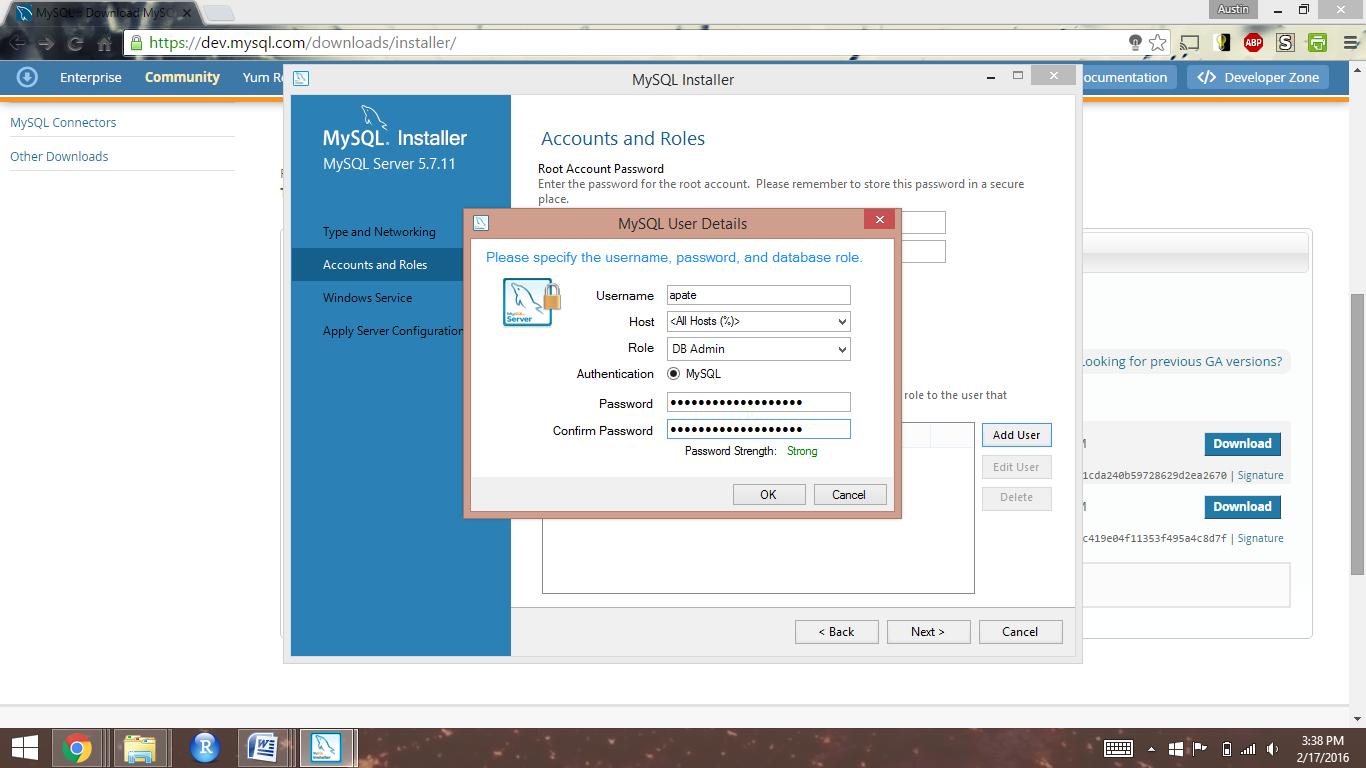
Select "Execute" to install the components of the MySQL developer package



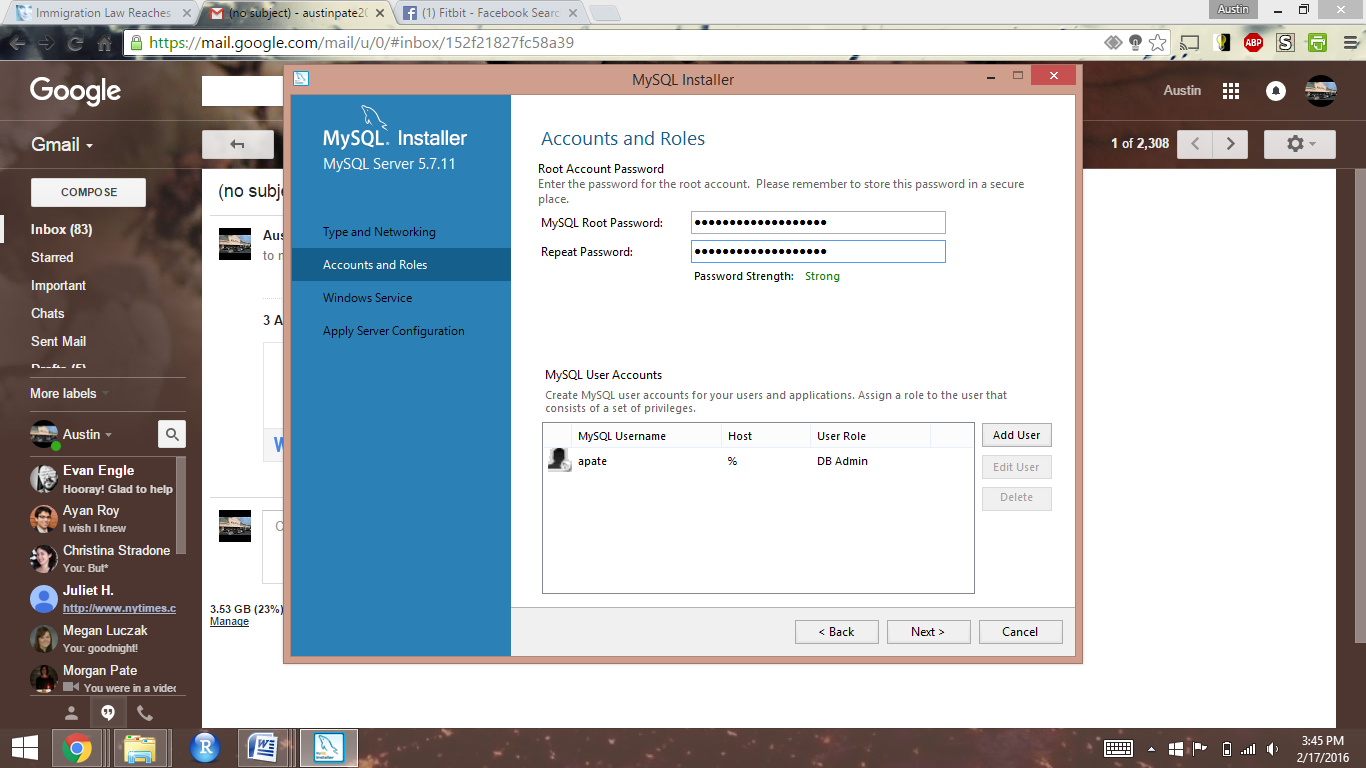
Keep the server configuration default settings and click "Next"



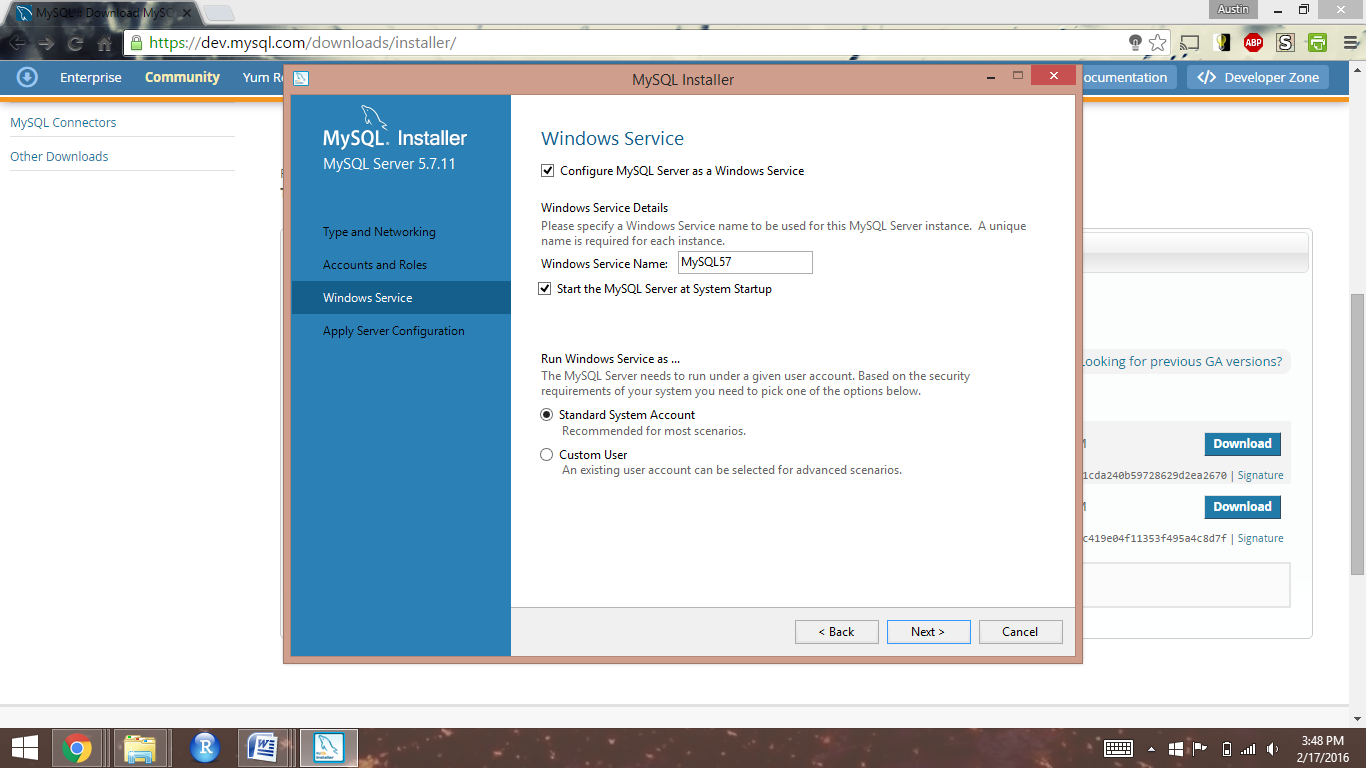
Click "Add User" then create a username and password for your server admin account. Click "OK"



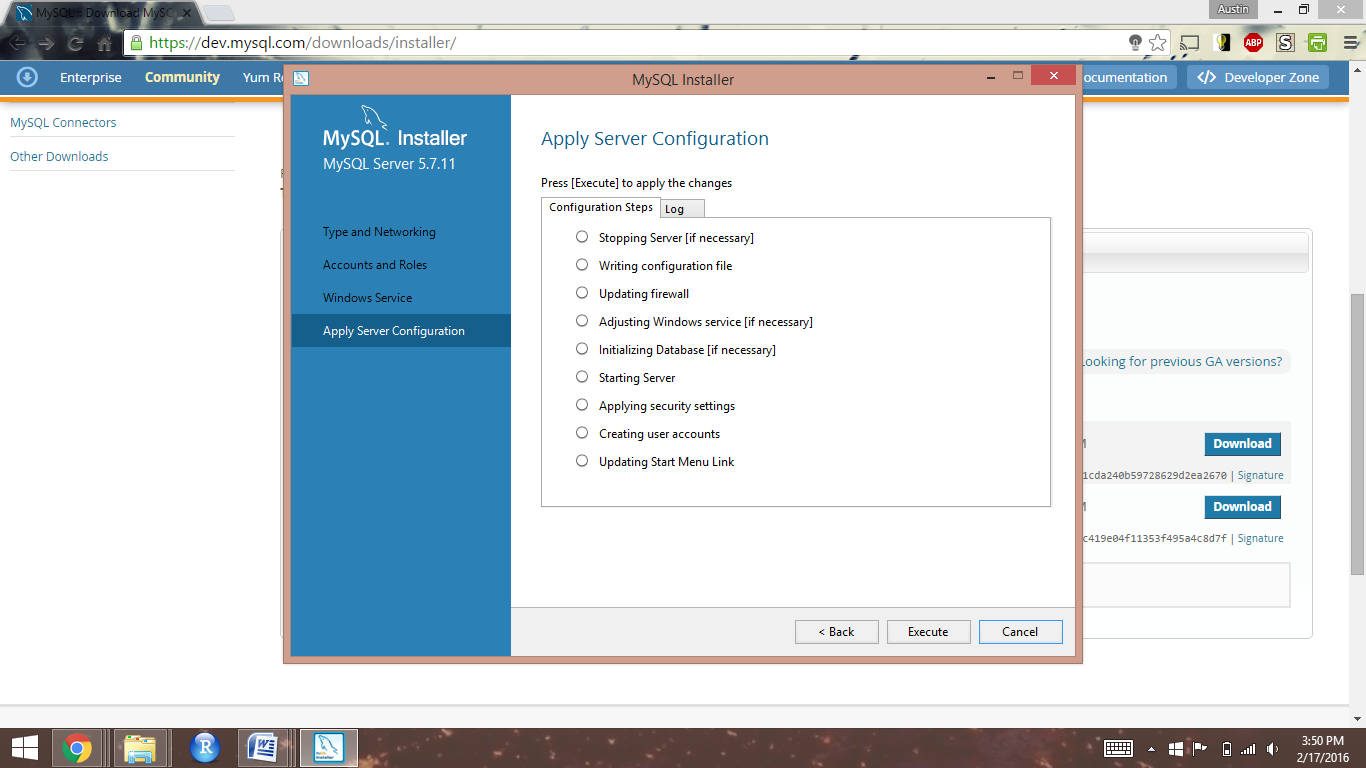
Create a password for your MySQL root account, then click "Next"



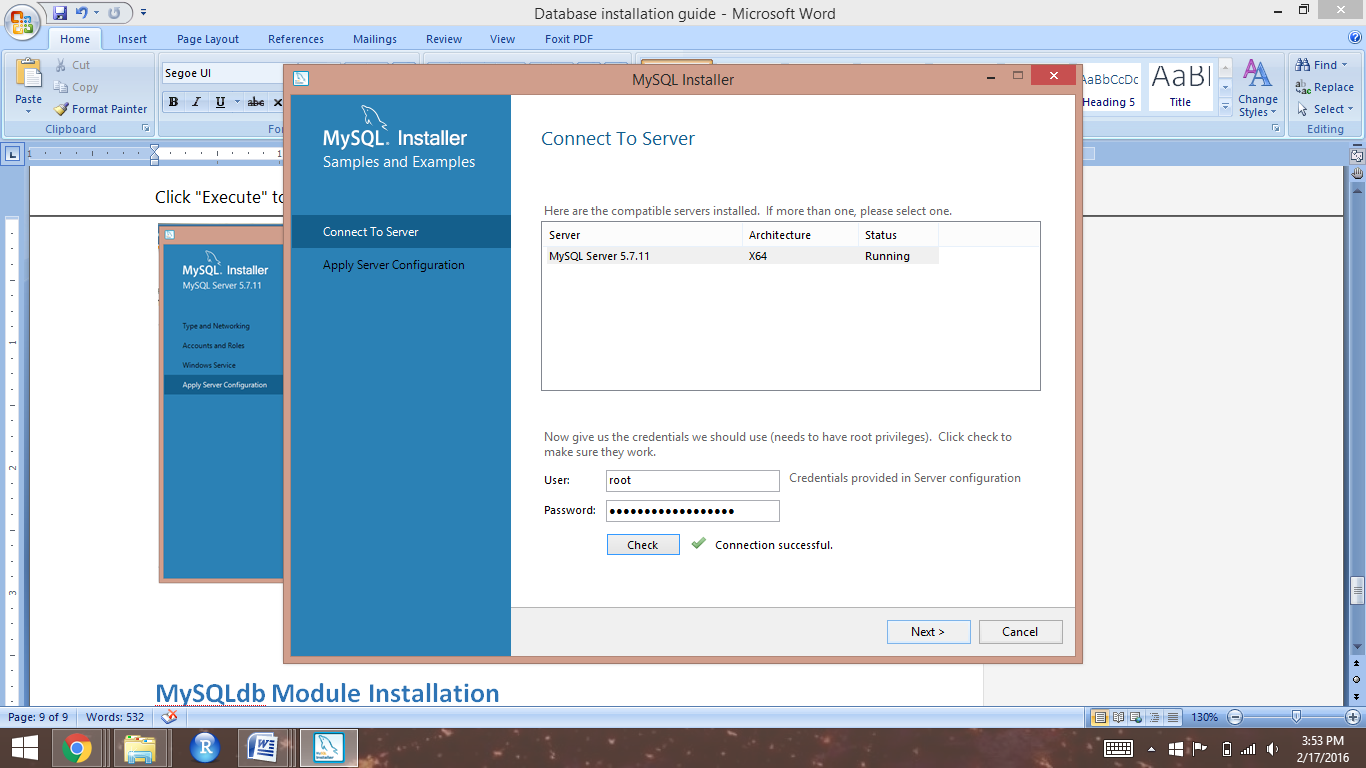
Keep the default settings for Windows Service and the "Standard System Account", then click "Next"



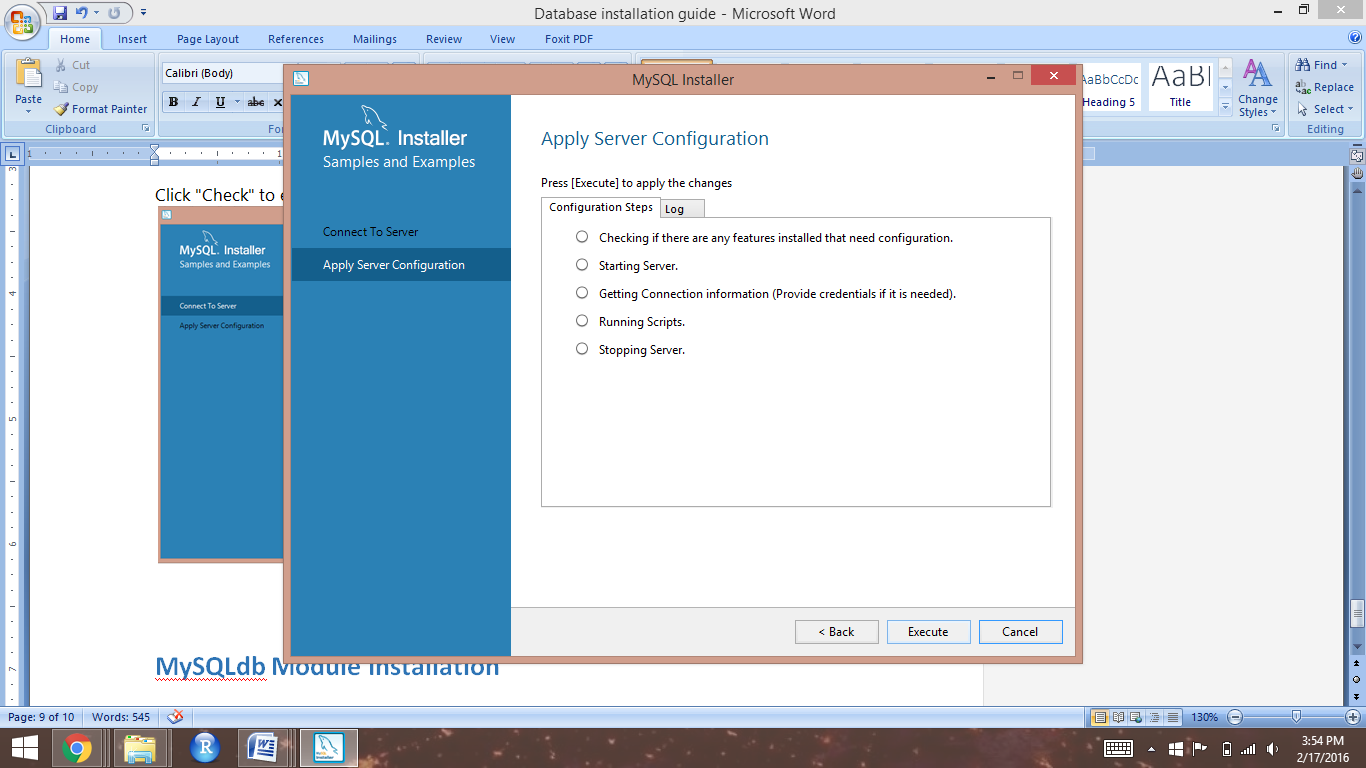
Click "Execute" to apply your server configuration, the "Finish"



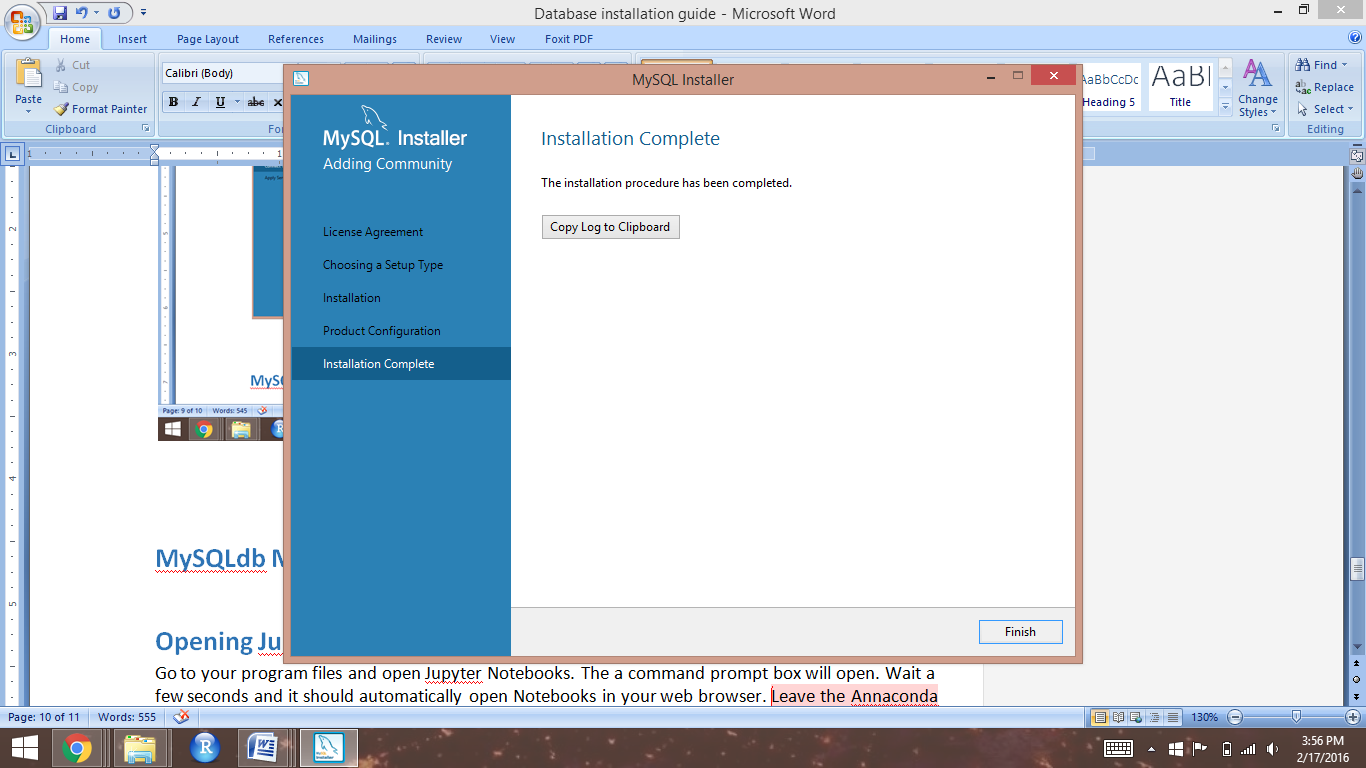
Click "Check" to ensure you can connect to your server, then click "Next"



Click "Execute" to finalize your server configuration, then click "Finish"



Click "Finish" to complete the MySQL installation

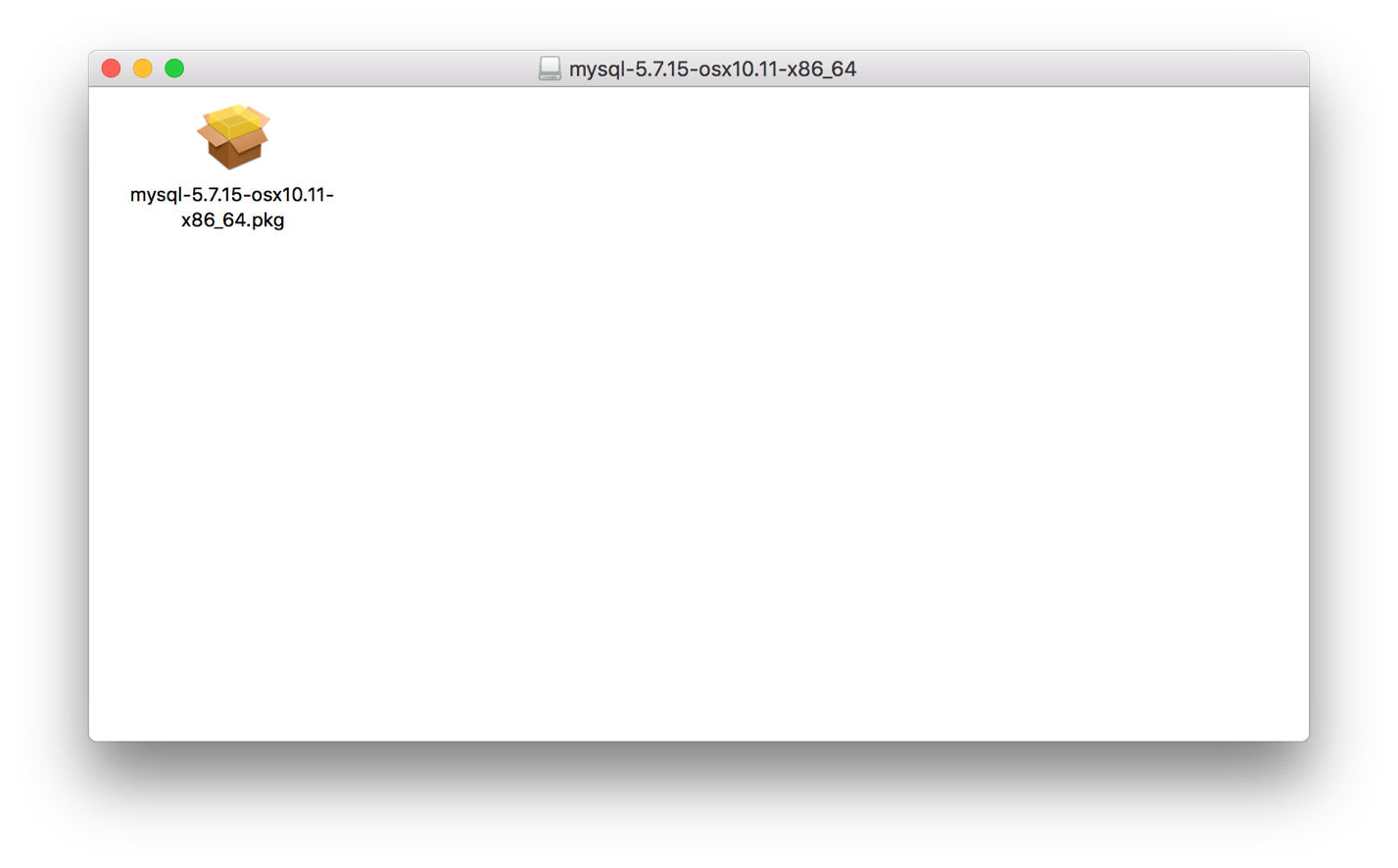


## Mac OS

### Installing the MySQL Database

When installing on a mac, you will have to first download the database server itself, then download and install MySQL Workbench. The [MySQL download website](https://dev.mysql.com/downloads/installer/) will, by default, provide the server, not MySQL Workbench. We document here how to install from the DMG archive file download option.

First, open the DMG file and double-click on the installation package stored inside:



Proceed through the installation. This is a standard Mac installer, so you’ll Continue past a Read Me, accept a license, then you’ll have to enter an admin password so the server can be installed as a service. During installation, a box will pop up containing a temporary root password. Make sure to copy and paste that password somewhere safe, so you don’t lose it.

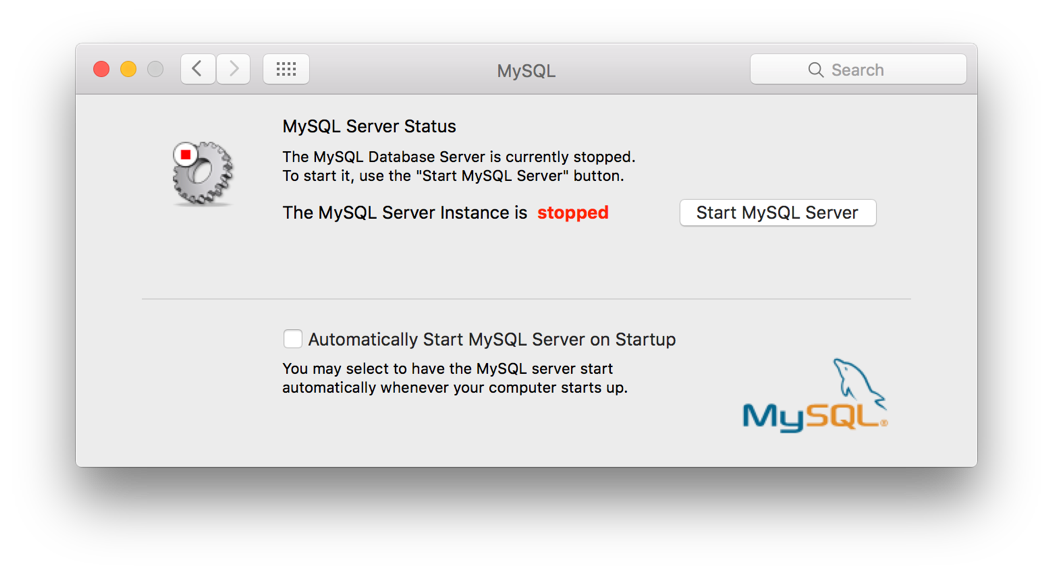
Once you’ve installed the database, a “MySQL” Preferences pane will be installed in your system’s Preferences:



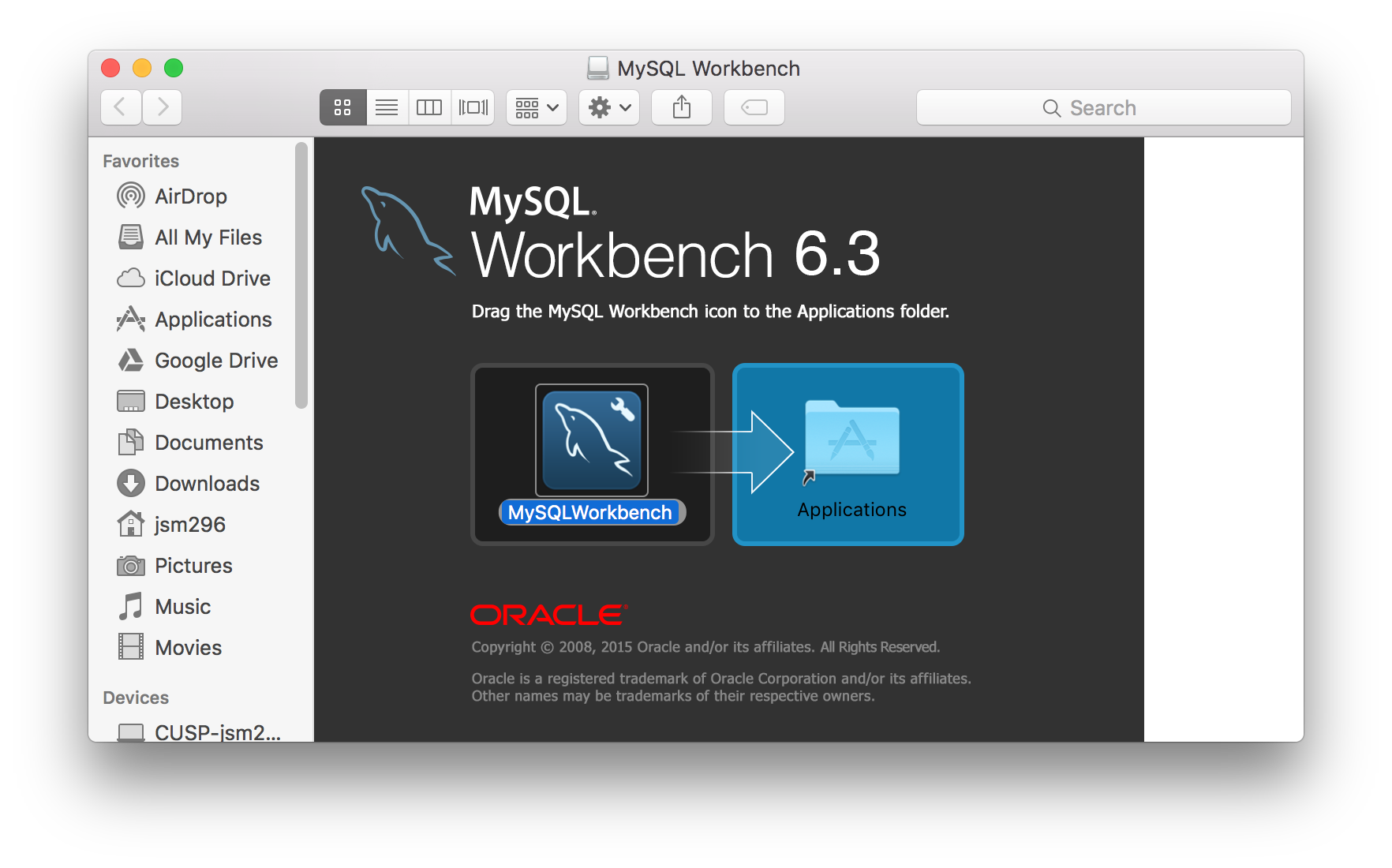
This Preference pane lets you start and stop the MySQL server, and lets you control whether it starts when your computer starts:



By default, MySQL server is set up to start when your computer starts (this is why it prompted you for your password when you installed it). We recommend you disable this feature and only start the database when you need it, using this Preference pane. To keep MySQL from starting when your computer starts, un check the box next to “Automatically Start MySQL Server on Startup”:



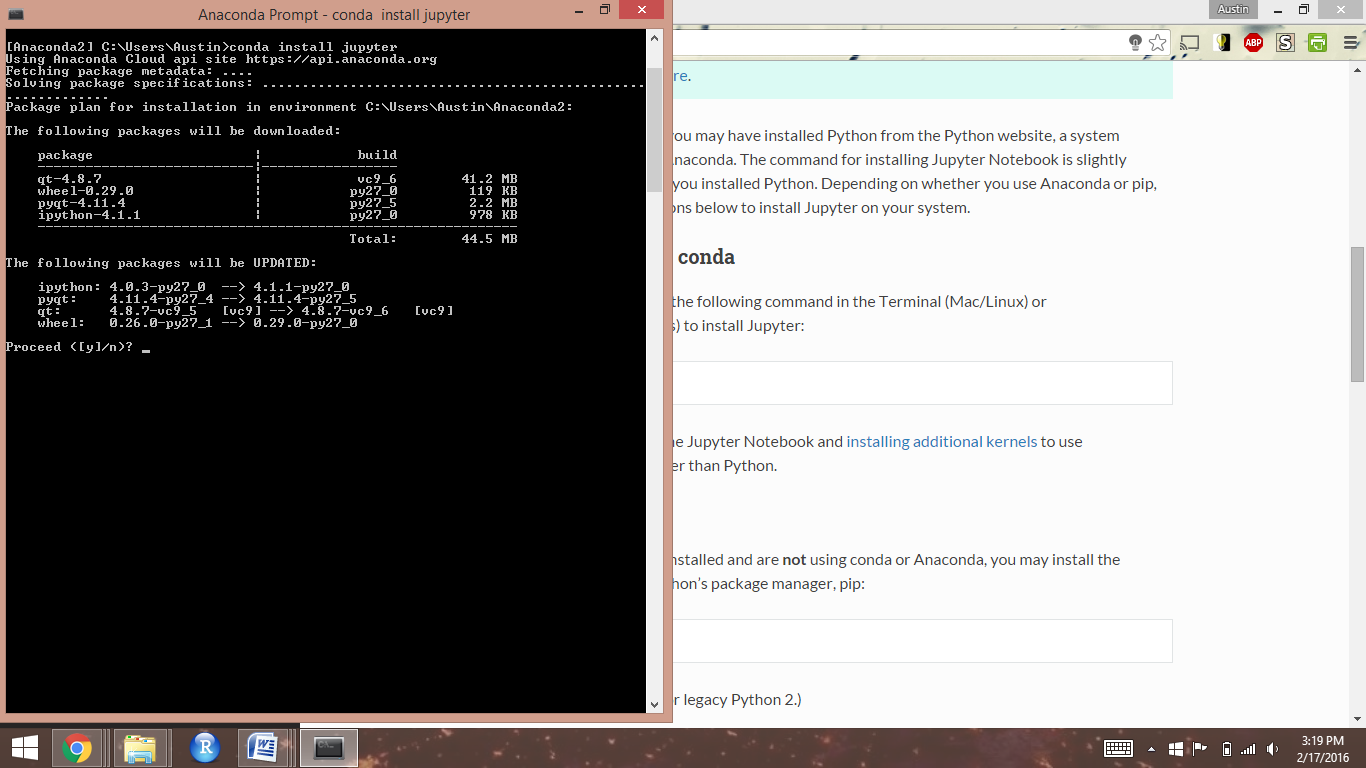
To install MySQL Workbench, download the DMG archive file for it from <http://dev.mysql.com/downloads/workbench/>, open the archive, and copy the MySQL Workbench Application from the archive to your computer’s Applications folder:



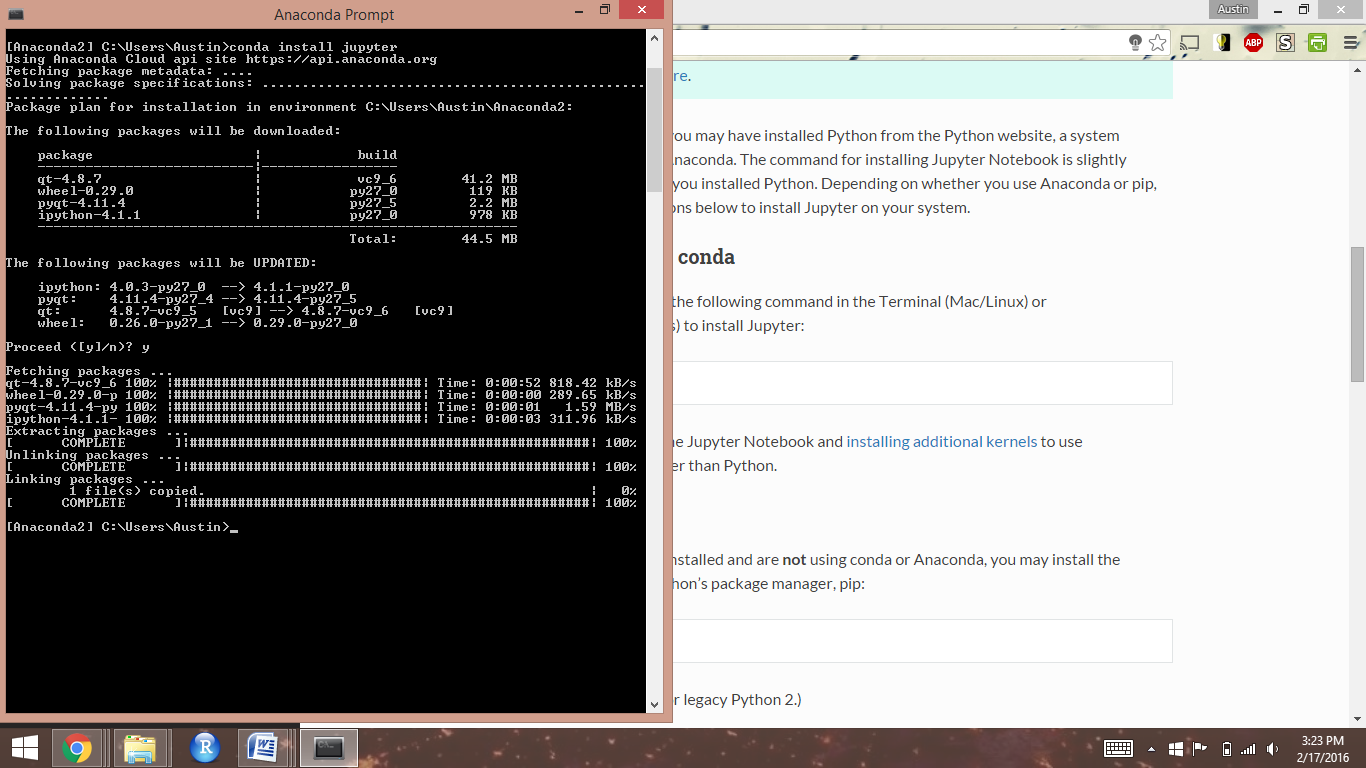
# **PyMySQL Installation**

To install python packages, you’ll open a shell on your computer and use the conda command, provided by Anaconda. First, open a command shell on your computer (the windows command shell is pictured, but the commands are the same whether on Windows or Mac).

To install the PyMySQL package that lets Python interact with MySQL, type “conda install pymysql”, then type “Y” to continue.



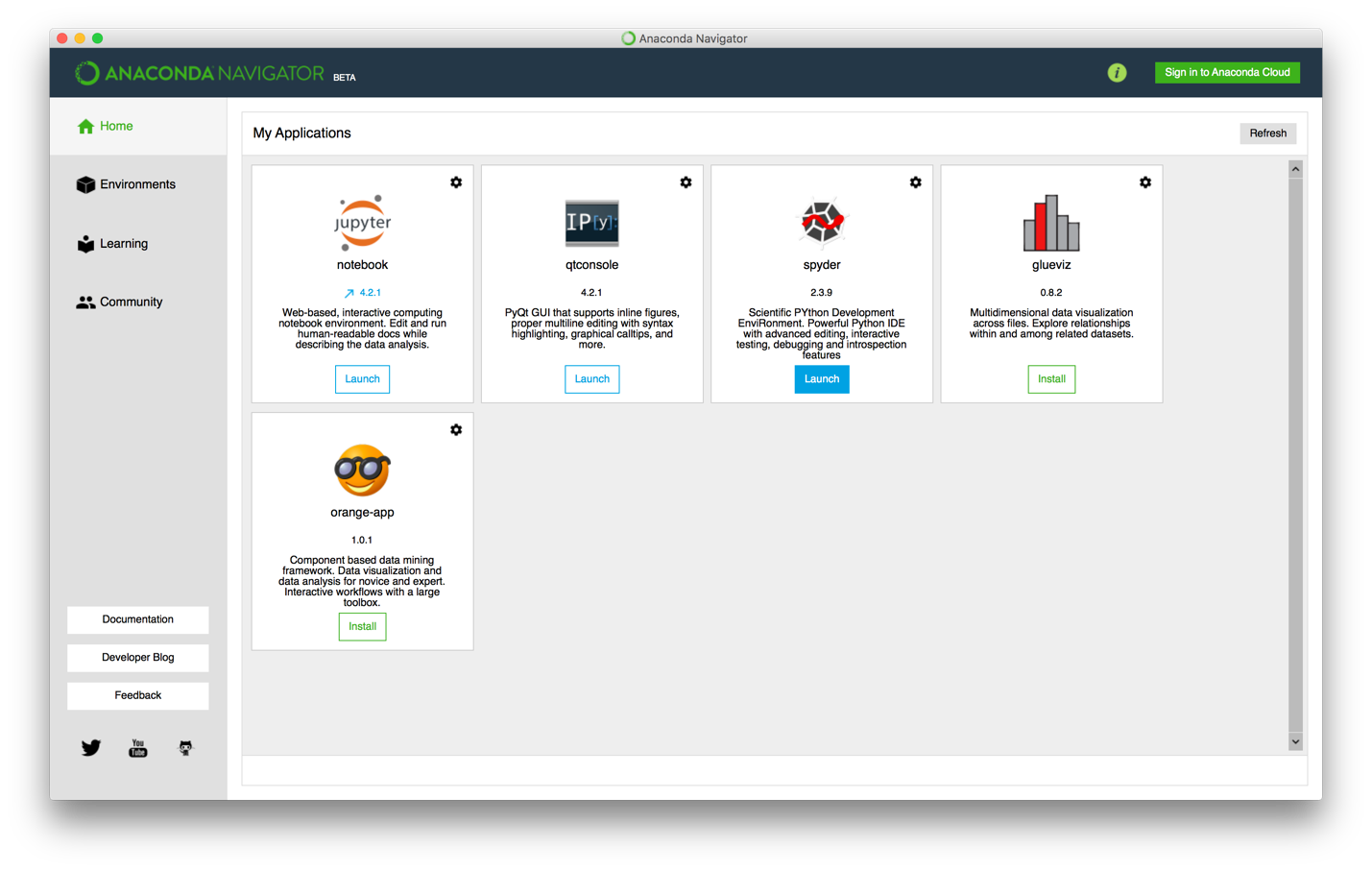
Once the prompt returns to your original directory the installation is complete, and you can close your command window:



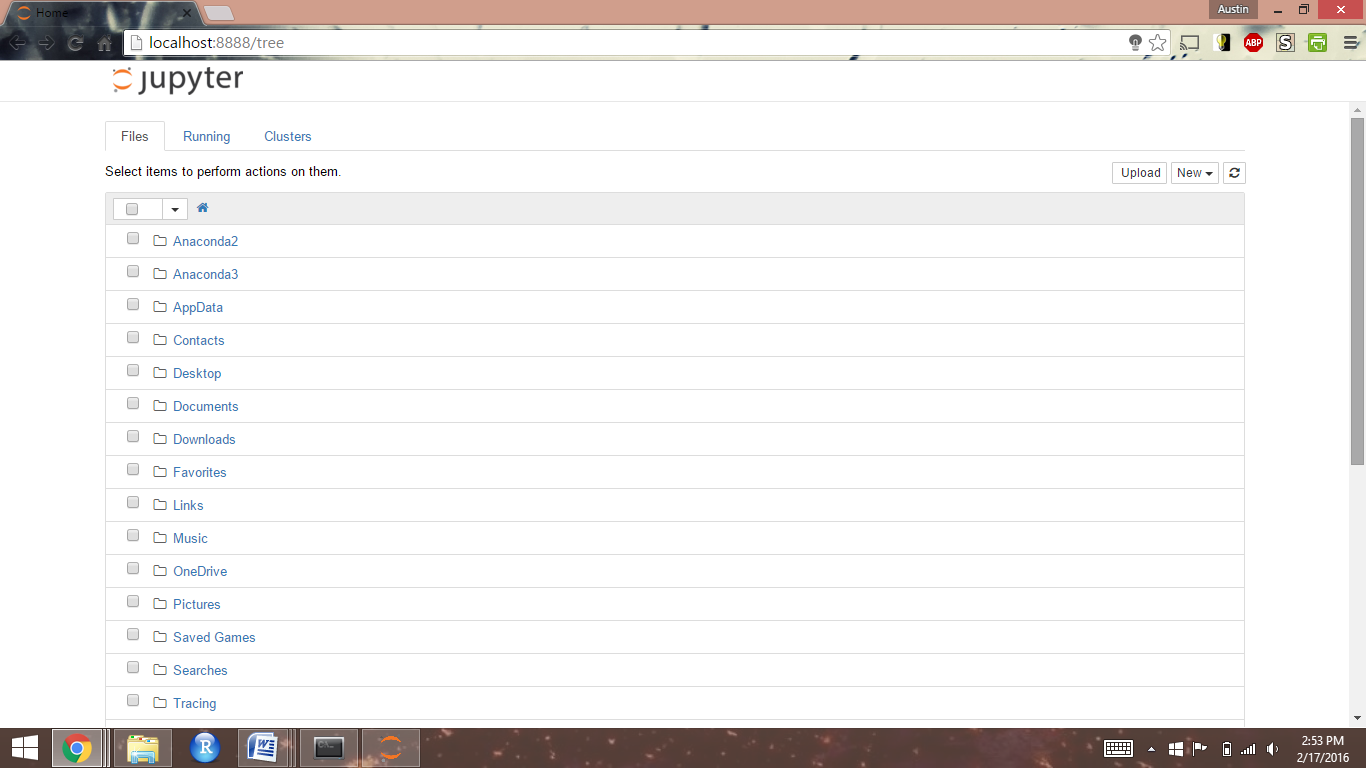
# **Connecting to MySQL from Python**

To test whether PyMySQL is installed correctly and Python can connect with MySQL, we’ll open a Jupyter Notebook. To start, go to your Program Files (Windows) or Applications (Mac OS) and open the Navigator program installed by Anaconda.

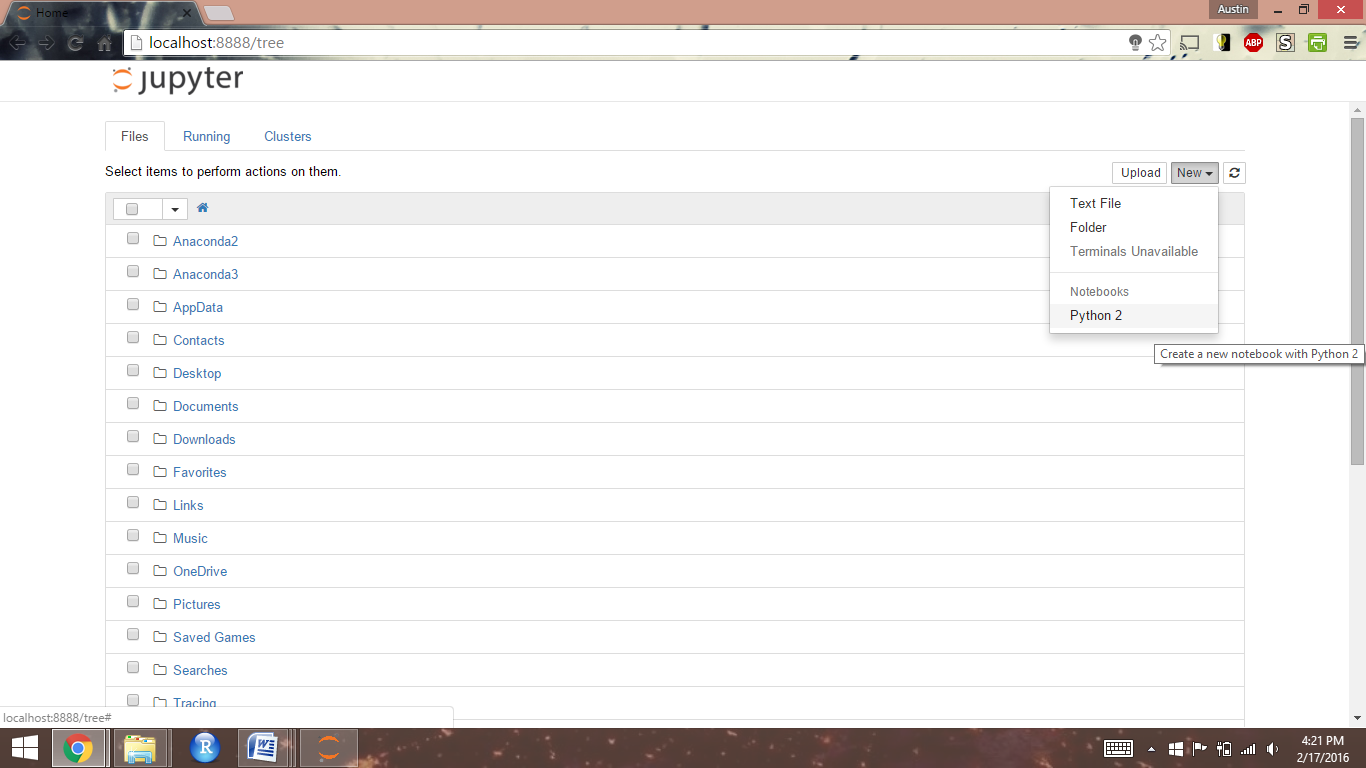
In the Navigator window that opens (same on Windows and Mac), click the “launch” button for “Jupyter notebook”:



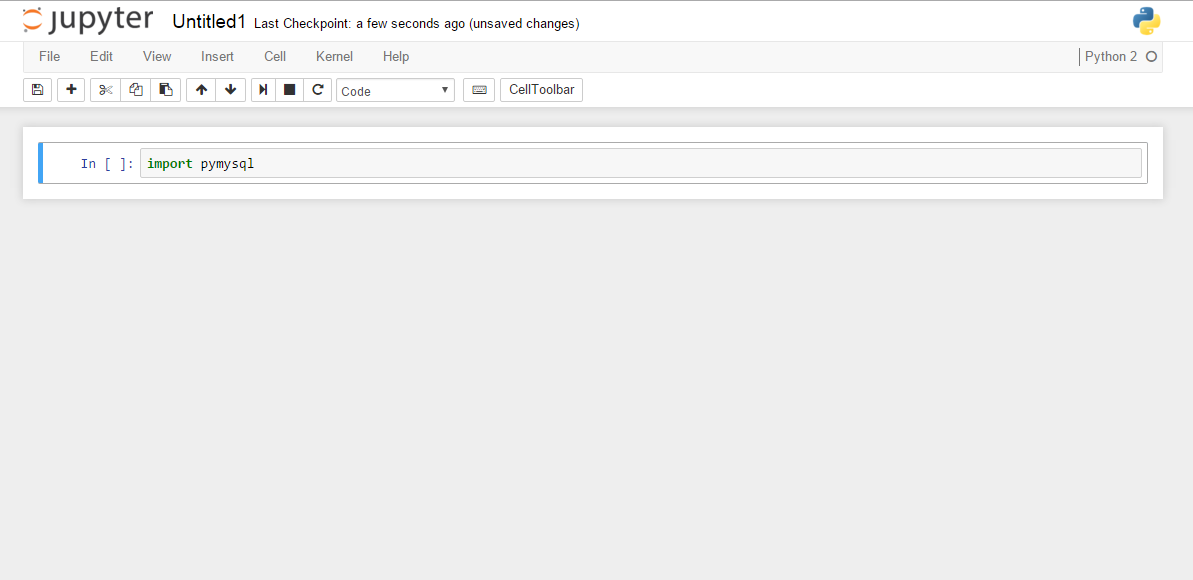
A command prompt box will open. Wait a few seconds and it should automatically open the Jupyter Notebook main page in your web browser. Leave the command prompt box open while you use your Notebooks.



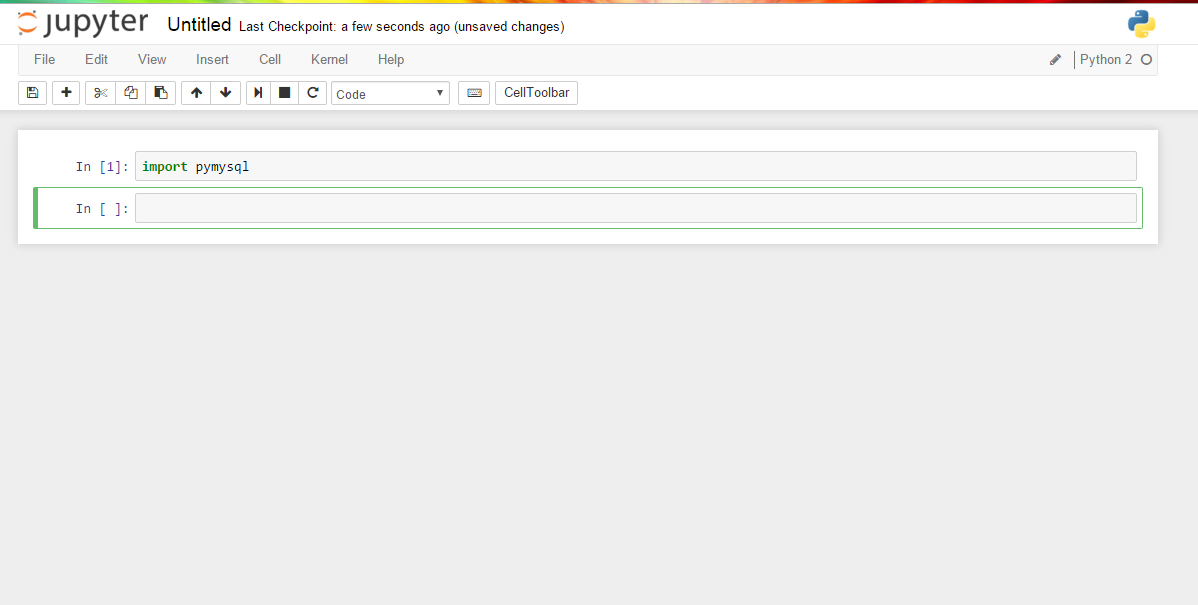
To create a new notebook, click the "New" tab, then click "Python 2" under "Notebooks"



To confirm the successful installation of the PyMySQL module, type "import pymysql" then click the Run icon



If the module was successfully installed, a new empty cell will appear below



# **Downloading and Importing Exercise Data**

Go to <http://serv.cusp.nyu.edu/classes/bigdatabook/bigdataworkbook.zip> and download the MySQL export file containing the databases you will be working with.

Double-click the file to unzip it and extract the enclosed SQL export file.

## To import using the MySQL command line

Open a command terminal in the directory with the SQL export file, then:

On Mac OS type:

/usr/local/mysql/bin/mysql -u root -p

On Windows, type the following, replacing "<version>" with the version information at the end of the MySQL install folder:

C:\> “\Program Files\MySQL<version>\bin\mysql” –u root -p

When prompted, enter the root password you were shown when you installed MySQL.

If this is the first time you’ve connected to the database, you’ll need to change the temporary password set by the MySQL installer. To do this, at the mysq> prompt, run the following command, replacing “<new\_password>” with a new password:

mysql> SET PASSWORD = PASSWORD('<new\_password>');

Then, use the “SOURCE” command to run and load each of the SQL files that was contained in the ZIP file:

mysql> SOURCE homework-grant.sql;

mysql> SOURCE homework-machine\_learning.sql;

mysql> SOURCE homework-nsf\_award.sql;

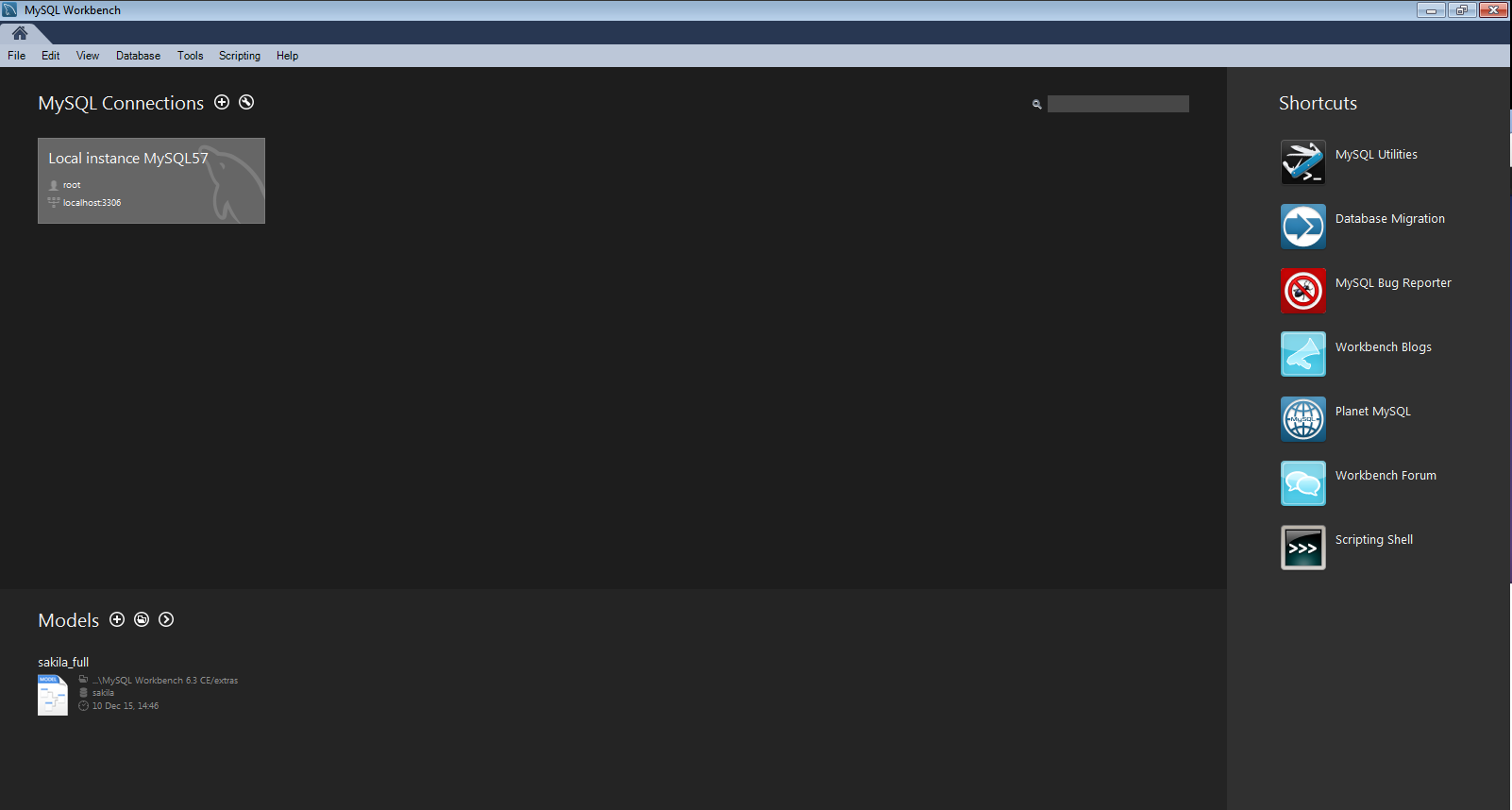
mysql> SOURCE homework-text\_analysis.sql;

mysql> SOURCE homework-uc\_pay\_2011.sql;

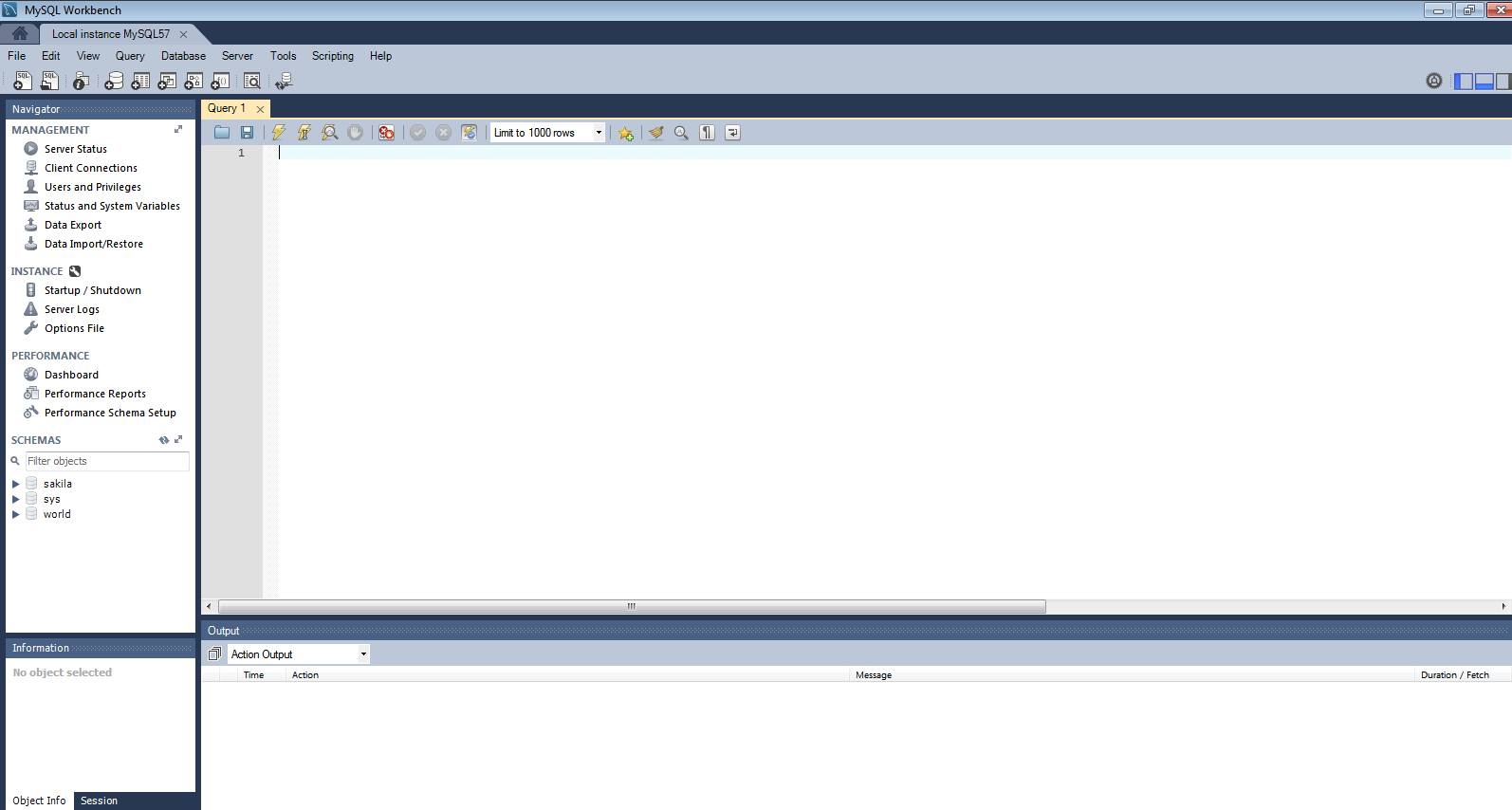
mysql> SOURCE homework-vendor.sql;

## To import using MySQL Workbench

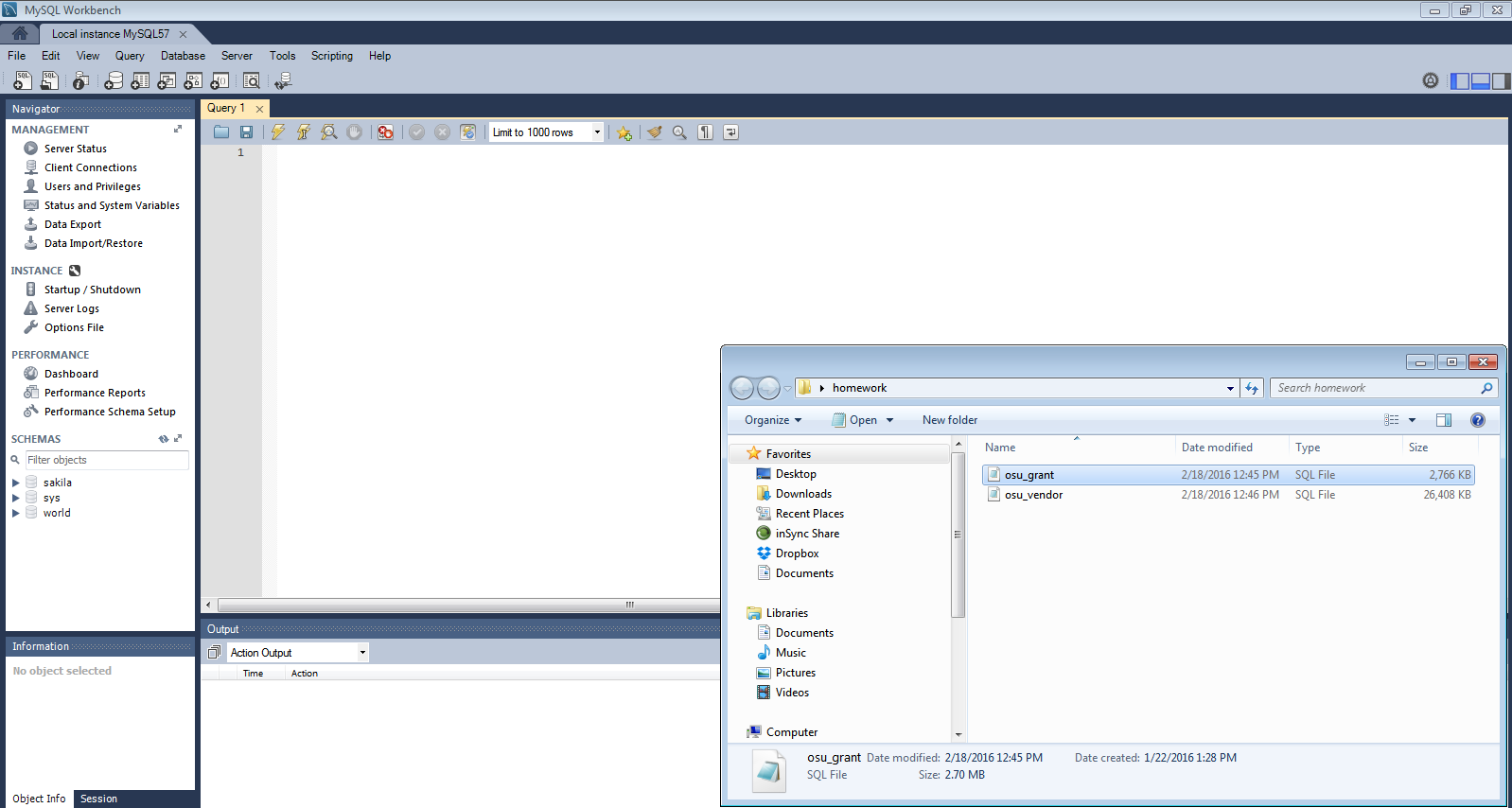
Open MySQL Workbench. There should be an available connection based upon your initial MySQL setup named “Local Instance”. Double click the “Local instance” icon to open your local server. Enter the “root” password provided when you installed MySQL If you are prompted for a password.



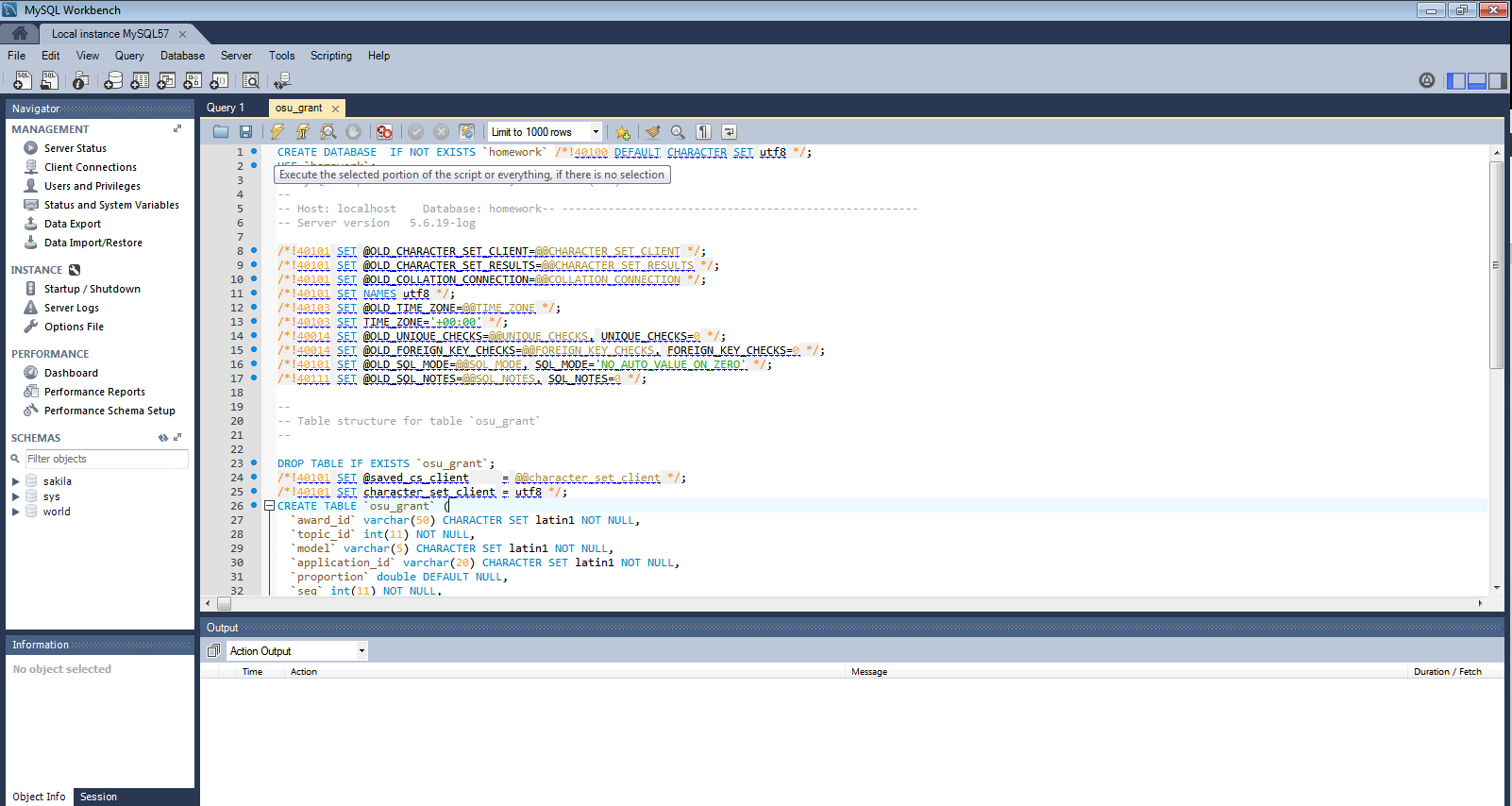
Once you open the connection to your local server, you should be directed to the screen below



Toggle back to the downloaded database files, then drag-and-drop each of the “\*.sql” files into the “Query 1” window



Once the window has populated with each file script, click the first lightning bolt button on at the top to execute the script, creating the homework database and that SQL file’s tables.



Confirm you have successfully created and imported the homework database by clicking the refresh button next to “Schemas” on the left-hand toolbar. You should be able to see the tables:

* machine\_learning
* nsf\_award
* text\_analysis
* uc\_pay\_2011
* ugrant
* vendor

listed under the homework database.

# Using PostgreSQL

We are working on adding support for PostgreSQL database. To start, it won't be as well documented as MySQL, but it isn't that hard to set up. You'll need to install PostgreSQL, then download the following PostgreSQL database export, and then install it in a database named "homework" in your database.

Good free clients are a little harder to come by, but the Java-based cross-platform dbeaver is a good one to try out: http://dbeaver.jkiss.org/

Data Files: http://serv.cusp.nyu.edu/classes/bigdatabook/homework.pgsql.zip