

# Data Engineering Platforms

## MSCA 31012

### Installation Documentation for Mac

#### Contents:

1. [OpenRefine](#)
2. [FileZilla](#)
3. [Anaconda](#)
4. [MySQL](#)
5. [MongoDB](#)
6. [Neo4j](#)
7. [Tableau](#)
8. [GCP Setup](#)

**Warning:** Please read the documentation thoroughly. Some installations have special instructions for this course, including stipulations to wait until a certain week to install for demo versions.

**NOTE:** Always install the latest versions. This document may show a slightly older version number in the screenshots, but still use the latest version.

# OpenRefine Installation & Setup Instructions

**Purpose:** OpenRefine is a data cleansing and transformation tool.

Source: <https://docs.openrefine.org/manual/installing>

OpenRefine is a desktop application in that you download it, install it, and run it on your own computer. However, unlike most other desktop applications, it runs as a small web server on your own computer and you point your web browser at that web server in order to use Refine. So, think of Refine as a personal and private web application.

## Requirements

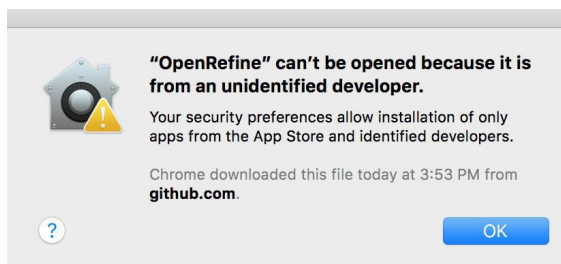
Java JRE is a prerequisite for OpenRefine

To check if you already have a JRE installed, open **Terminal** and type **java -version**. If not found please follow JRE installation instructions below.

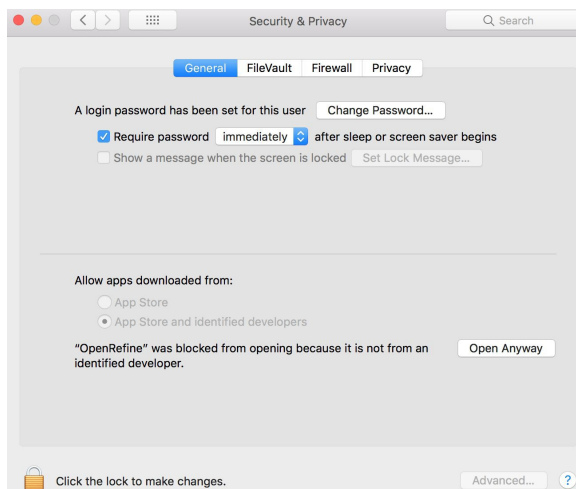
[https://docs.oracle.com/javase/8/docs/technotes/guides/install/install\\_overview.html#CJAGAACB](https://docs.oracle.com/javase/8/docs/technotes/guides/install/install_overview.html#CJAGAACB)

## Install Summary

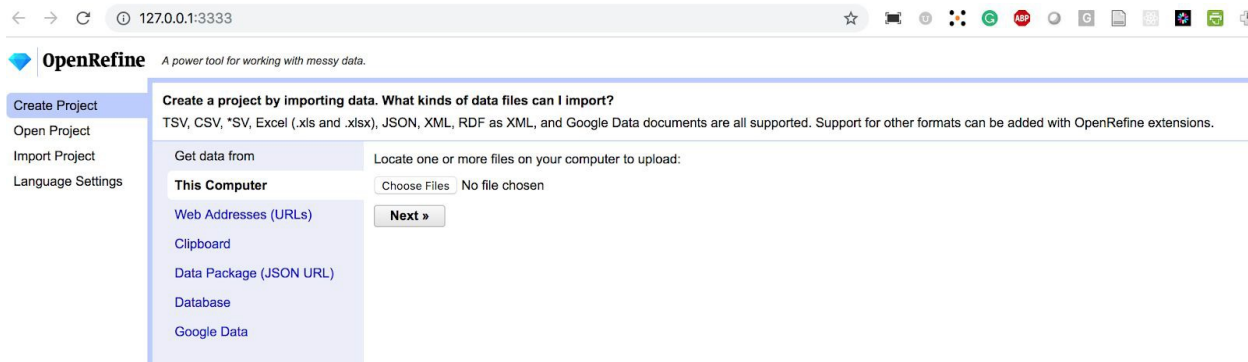
1. OpenRefine requires you to have a working Java JRE, otherwise you will not be able to start OpenRefine.
2. [Download OpenRefine here](#).
3. Install it as detailed below for your operating system ([Windows](#), [Mac OSX](#), [Linux](#)).
4. Go to your Applications, and click on OpenRefine.
  - a. The following error may occur:



- b. If you receive the above error, go to your System Preferences and click on Security & Privacy.
- c. Click on the General tab and click on Open Anyway:



- d. A pop will open asking if you're sure to open. Click Open.
5. You can now go back to Applications and open OpenRefine. A new tab on your web browser should open.



6. As long as OpenRefine is running, you can point your browser at <http://127.0.0.1:3333/> to use it, and you can even use it in several browser tabs and windows.
7. If you're running a proxy or get a BindException, you can change the IP configuration with -i and -p, see **Running & Configuration below**, or use refine -help for options.

---

By default (and for security reasons) Refine only listens to TCP requests coming from localhost (127.0.0.1 on port 3333). If you want to respond to TCP requests coming to any IP address the machine has, run refine like this from the command line:

```
./refine -i 0.0.0.0
```

On Mac OS X, you can add a specific entry to the Info.plist file located within the app bundle (/Applications/OpenRefine.app/Contents/Info.plist):

```
<string>-Drefine.host=0.0.0.0</string>
```

# **FileZilla Client Installation**

**Purpose:** FileZilla will be used to connect to the RCC Midway server via SFTP

## **Installation & Setup**

### **1. Download FileZilla Client**

- a. Navigate to <https://filezilla-project.org>
- b. Click the grey “Download FileZilla Client” button (**not** FileZilla Server)
- c. Click the green “Download FileZilla Client” button on the next page to download the client for your given operating system
- d. If a pop-up is presented, click the green “Download” button for FileZilla (**not** FileZilla Pro)

### **2. Install FileZilla Client**

- a. Once the program has downloaded, run the executable
- b. Use the default installation parameters EXCEPT for the following:
  - By default, FileZilla will check the boxes to opt you in to “offers” like McAfee or Yahoo. Uncheck these boxes during the installation process to avoid installing the unnecessary software.

### **3. Setup FileZilla Client**

Screenshots can be found here: <https://2fa.rcc.uchicago.edu/getstarted.html> > FileZilla tab

- a. Open FileZilla
- b. Add a new Site
  - Navigate to File > Site Manager
  - Click “New Site” to begin configuring a new connection
  - Click “Rename” to give your connection a meaningful name such as Midway
  - Input the following parameters:
    - Protocol: SFTP
    - Host: midway2.rcc.uchicago.edu
    - Port: 22
    - Logon Type: Interactive (you will need to use 2FA)
    - User: Your CNET ID
  - Select Transfer Settings tab and check the “Limit number...” and type 1
  - Click OK to save the settings
- c. Navigate to Edit > Settings
  - From the Settings menu, select Transfers from the left window pane
  - Input 1 in the “Maximum simultaneous transfers” box
- d. Connect to your new saved connection
  - File > Site Manager
  - Select your saved connection on the left (Midway)
  - Click Connect
- e. If prompted with a dialogue about an “Unknown host key”, check the box to “Always trust this host” and click OK.
- f. Type your password in the first prompt
- g. Select your 2FA method in the second prompt and complete the 2FA process
- h. Confirm you see your file tree on the right side to indicate you successfully connected

# Anaconda

**Purpose:** Anaconda is a Python distribution and package management tool. Python and Jupyter notebooks will be used for parts of this course.

## Download & Install Anaconda

1. Download **Anaconda Distribution** for Python 3.9:

<https://www.anaconda.com/products/distribution#Downloads>



### NOTE:

- a. For those using M1/M1 pro/M1 max/M2 chips, please download the M1 version.
  - b. If you are not familiar with command line, we recommend to use graphical installer.
2. Run the Installer
  3. During the Install, you may be asked if you want to download Microsoft VSCode. This is optional. VSCode is a text editor.
  4. Once it's finished installing, go to your terminal shell. 'cd' back to your local directory. Type 'jupyter notebook'. This will open a new tab in your default browser.
    - a. If you ever want to open a jupyter notebook in a different directory than your local, cd to that directory first and then type jupyter notebook.

# MySQL Installation

## Installation Instructions

1. Before starting the MySQL installation make sure to uninstall any pre-existing installations of MySQL database and install the latest available version.

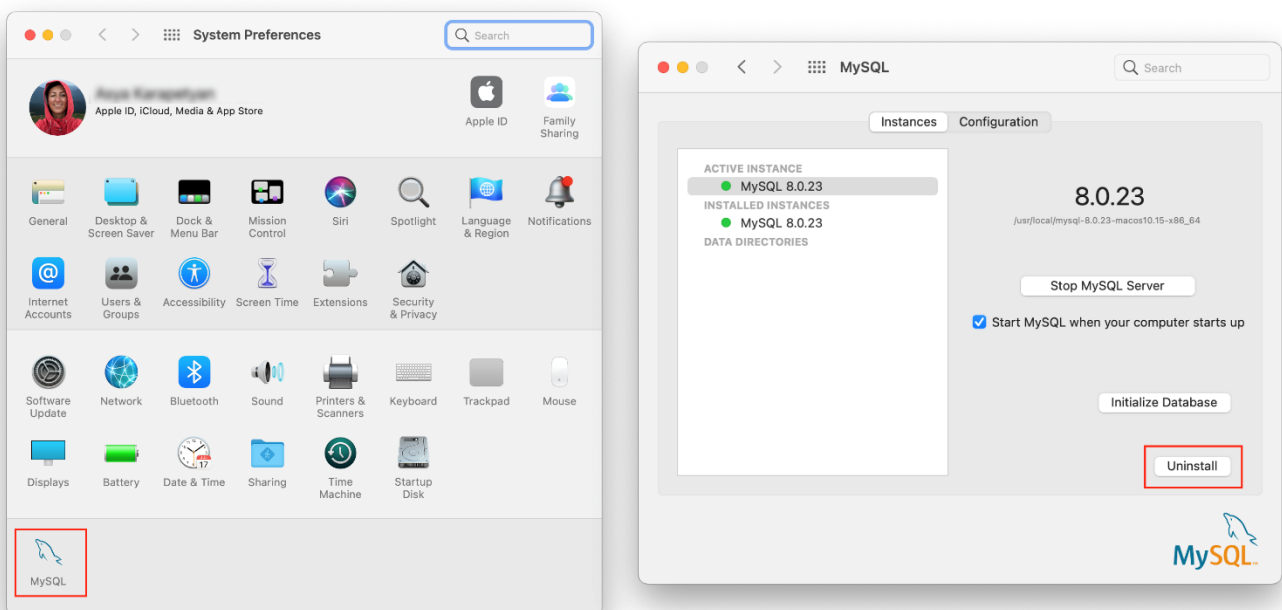
- a. Start Terminal and run the following commands

- `find / -name mysql`

```
local — -bash — 132x50
GLHR-MB-c02s30efg8wm:local sbharadwaj-local$ find / -name mysql
find: /usr/sbin/authserver: Permission denied
/usr/local/mysql-5.7.14-osx10.11-x86_64/bin/mysql
/usr/local/mysql-5.7.14-osx10.11-x86_64/include/mysql
find: /usr/local/mysql-5.7.14-osx10.11-x86_64/data: Permission denied
/usr/local/mysql
find: /.Spotlight-V100: Permission denied
find: /Library/Application Support/Apple/ParentalControls/Users: Permission denied
find: /Library/Application Support/Apple/AssetCache/Data: Permission denied
find: /Library/Application Support/ApplePushService: Permission denied
find: /Library/Application Support/JAMF/Usage: Permission denied
find: /Library/Application Support/JAMF/run: Permission denied
find: /Library/Application Support/JAMF/Downloads: Permission denied
find: /Library/Application Support/JAMF/tmp: Permission denied
find: /Library/Application Support/Symantec/Settings: Permission denied
find: /Library/Application Support/com.apple.TCC: Permission denied
find: /Library/SystemMigration/History/Migration-258EE2A2-F20E-4B2B-B467-4E4DF95E6035/QuarantineRo
emission denied
find: /Library/Logs/DiagnosticReports: Permission denied
find: /Library/Caches/MozyPro/files_tmp: Permission denied
find: /Library/Caches/com.google.SoftwareUpdate.0: Permission denied
find: /Library/Caches/com.apple.iconservices.store: Permission denied
find: /.Trashes: Permission denied
find: /System/Library/DirectoryServices/DefaultLocalDB/Default: Permission denied
find: /System/Library/User Template: Permission denied
^C
GLHR-MB-c02s30efg8wm:local sbharadwaj-local$
You have new mail in /var/mail/sbharadwaj-local
GLHR-MB-c02s30efg8wm:local sbharadwaj-local$
```

- b. Terminate MySQL service.

- Open System Preference
- Click MySQL Icon, and enter the MySQL tab.
- Click the Uninstall button



- c. Clean Preference files. Open Terminal, and type:

- `sudo rm /usr/local/mysql`
- `sudo rm -rf /usr/local/var/mysql`
- `sudo rm -rf /usr/local/mysql*`
- `sudo rm ~/Library/LaunchAgents/homebrew.mxcl.mysql.plist`

- `sudo rm -rf /Library/StartupItems/MySQLCOM`
  - `sudo rm -rf /Library/PreferencePanes/My*`
  - `rm -rf ~/Library/PreferencePanes/My*`
  - `sudo rm -rf /Library/Receipts/mysql*`
  - `sudo rm -rf /Library/Receipts/MySQL*`
  - `sudo rm -rf /private/var/db/receipts/*mysql*`
- d. If file `/etc/hostconfig` exists, edit it to remove the line: `MYSQLCOM=-YES-`.
  - e. Move MySQL workbench to trash
  - f. Reboot.

\*\*Make a note of the below useful URLs that give additional insights into the installation process. If you run into any issues with the latest version of MySQL please revert back to the 5.6 version.

<https://dev.mysql.com/doc/refman/5.6/en/osx-installation-pkg.html>

<https://dev.mysql.com/doc/refman/5.6/en/osx-installation-launchd.html>

<https://dev.mysql.com/downloads/workbench/>

2. Download MySQL Installer Package from: <https://dev.mysql.com/downloads/mysql/>  
Choose the **latest** version of the DMG Archive for Mac download

#### Generally Available (GA) Releases

### MySQL Community Server 8.0.12

Select Operating System:

macOS

[Looking for previous GA versions?](#)



Packages for High Sierra (10.13) are compatible with Sierra (10.12)

macOS 10.13 (x86, 64-bit), DMG Archive

8.0.12

177.2M

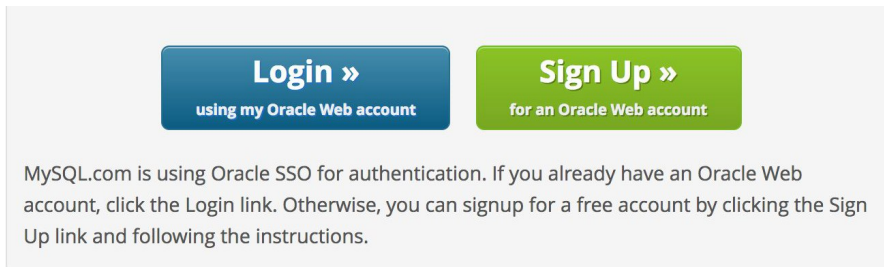
[Download](#)

(mysql-8.0.12-macos10.13-x86\_64.dmg)

MD5: ee79241a8395226c9f42f00a3a476e61 | [Signature](#)

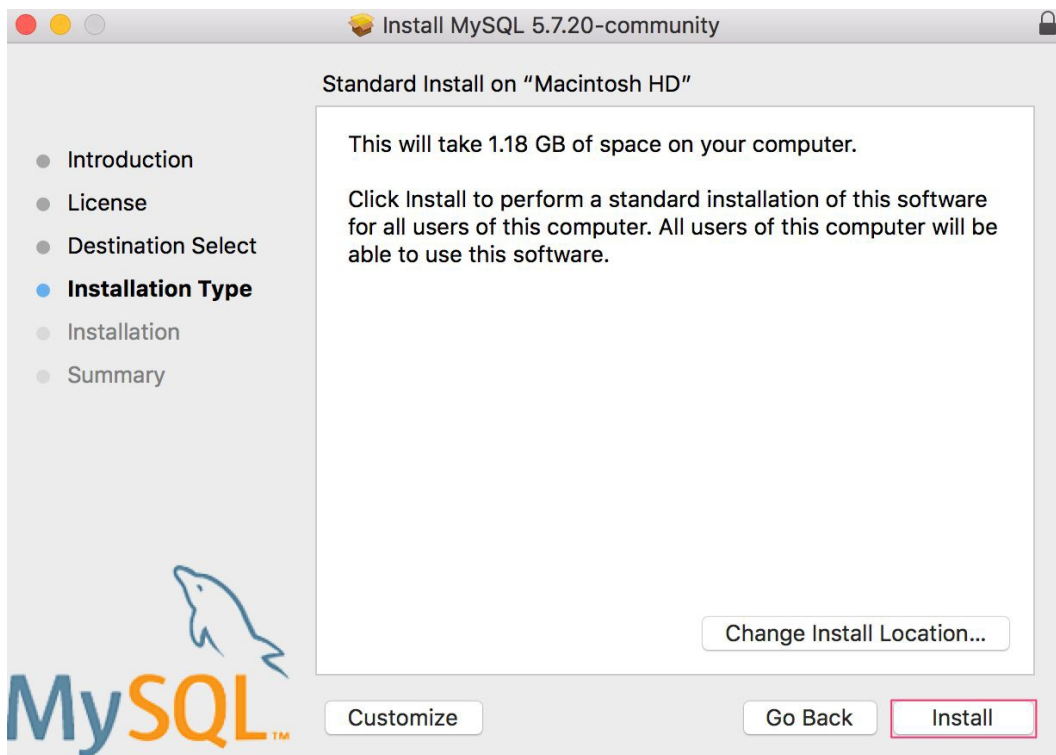
**Note:** For those using M1/M1 pro/M1 max chips, download the ARM version, rather than x86.

3. The next page will have 2 buttons to login and signup. You can scroll to the bottom and bypass this by choosing No thanks, just start my download.



**No thanks, just start my download.**

4. Save the file to a folder on your machine. This will prevent you from having to download it again if something goes wrong with the install.
5. Double click on the pkg to open up the installer and follow the steps below.
  - a. Click Continue on the 'The package will run a program to determine if the software can be installed.
  - b. Click Continue on the Installation welcome prompt.
  - c. Click continue on the License Agreement
  - d. Agree to the license agreement
  - e. Click Install. The installation might need admin privileges so please enter admin credentials to install software.

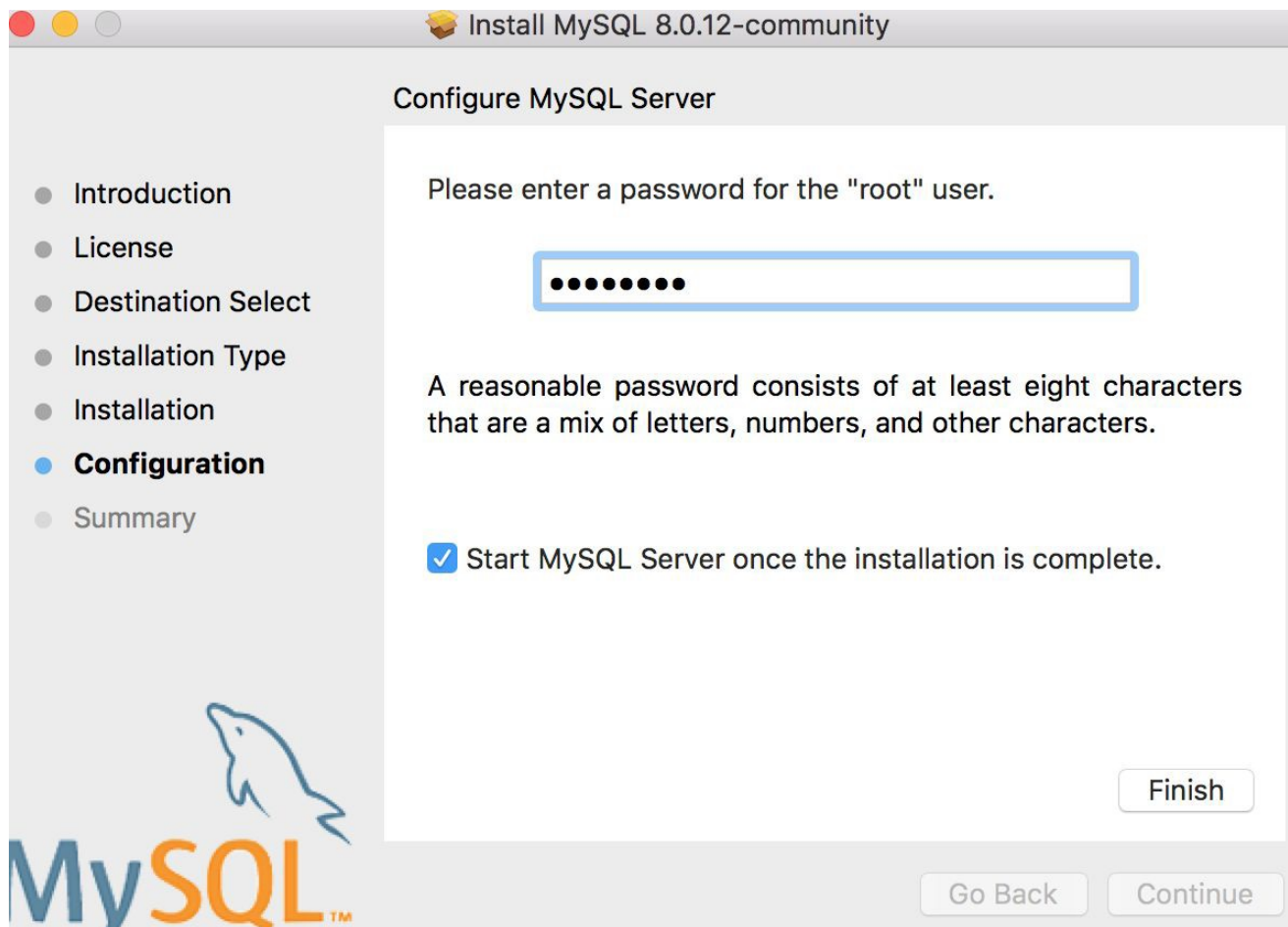




6. Once the Install is complete, the Configuration is next.
  - a. On the Configuration screen, select Use Legacy Password Encryption



- After clicking Next enter a password. PLEASE use '**rootroot**' as your password. Using this common password will help with in-class scripts. Then click Finish.



- The installer should now be complete. Click Close.

- Verify the installation

```
GLHR-MB-c02s30efg8wm:local sbharadwaj-local$ ls -ltr
total 0
drwxr-xr-x  3 root          wheel   96 Nov 26  2017 jamf
drwxr-xr-x  3 root          wheel   96 Nov 26  2017 remotedesktop
drwxr-xr-x  5 root          wheel  160 Nov 26  2017 man
drwxr-xr-x  5 root          wheel  160 Nov 26  2017 share
drwxr-xr-x 30 root          admin  960 Nov 26  2017 lib
drwxr-xr-x 29 root          wheel  928 Nov 26  2017 include
drwxr-xr-x 40 root          wheel 1280 Jan 31  2018 bin
drwxr-xr-x  8 sbharadwaj-local staff 256 Sep 18 21:22 mongod
lrwxr-xr-x  1 root          wheel   30 Sep 20 20:48 mysql -> mysql-8.0.12-macos10.13-x86_64
drwxr-xr-x 13 root          wheel  416 Sep 20 20:48 mysql-8.0.12-macos10.13-x86_64
You have new mail in /var/mail/sbharadwaj-local
GLHR-MB-c02s30efg8wm:local sbharadwaj-local$
```

10. Go to your System Preferences and click on MySQL.
11. **You may get an error when trying to open MySQL indicating MySQL preference pane is not working.**

**If this occurs:**

- a. Go to Terminal to /usr/local/mysql/support-files. Find the mysql.server file:

```
/usr/local/mysql/support-files
[GLHR-MB-c02s30efg8wm:support-files sbharadwaj-local$ ls -ltr
total 48
-rw-r--r--  1 root  wheel   773 Jun 28 11:18 magic
-rwxr-xr-x  1 root  wheel  1061 Jun 28 12:53 mysqld_multi.server
-rwxr-xr-x  1 root  wheel  2048 Jun 28 12:53 mysql-log-rotate
-rwxr-xr-x  1 root  wheel 10622 Sep 20 21:37 mysql.server
[GLHR-MB-c02s30efg8wm:support-files sbharadwaj-local$
```

- b. Use VI to update the below fields:

```
edit /usr/local/mysql/support-files/mysql.server
```

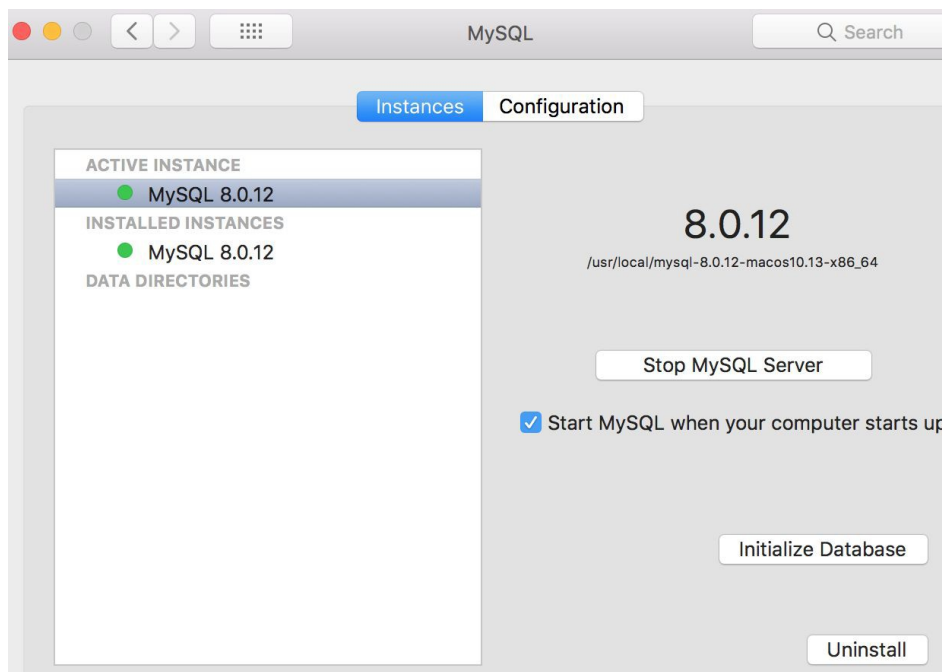
```
change lines ~ 46 & 47
```

```
basedir=
datadir=
```

to

```
basedir=/usr/local/mysql
datadir=/usr/local/mysql/data
```

- c. If you need to take this step, Restart your laptop.
12. After you double click on the MySQL icon under system preferences, the below screen will pop up. When the Active and Installed Instance is Green, the it is running correctly and successful, else please uninstall and reinstall MySQL Server



## MySQL Workbench

1. Download the latest MySQL Workbench Package from:  
<https://dev.mysql.com/downloads/workbench/>

2. Click “No thanks, just start my download” at the bottom

Note: The rest of the screen shots might show the older versions of installation. However, proceed with the latest version

3. Once downloaded, drag to install and authenticate.

4. Open MySQLWorkbench

5. Click on the Local instance card. You'll be prompted for a password. Please use the password from step 10 (should be **root**) to connect MySQL workbench to the MySQL Server. Click on the 'Save password to keychain' so you won't have to type in your password each time.

# Welcome to MySQL Workbench

MySQL Workbench is the official graphical user interface (GUI) tool for MySQL. It allows you to create and browse your database schemas, work with database objects and insert data as well as design and run SQL queries to work with stored data. You can also migrate schemas and data from other database vendors to your MySQL database.

[Browse Documentation >](#)

[Read the Blog >](#)

[Discuss](#)

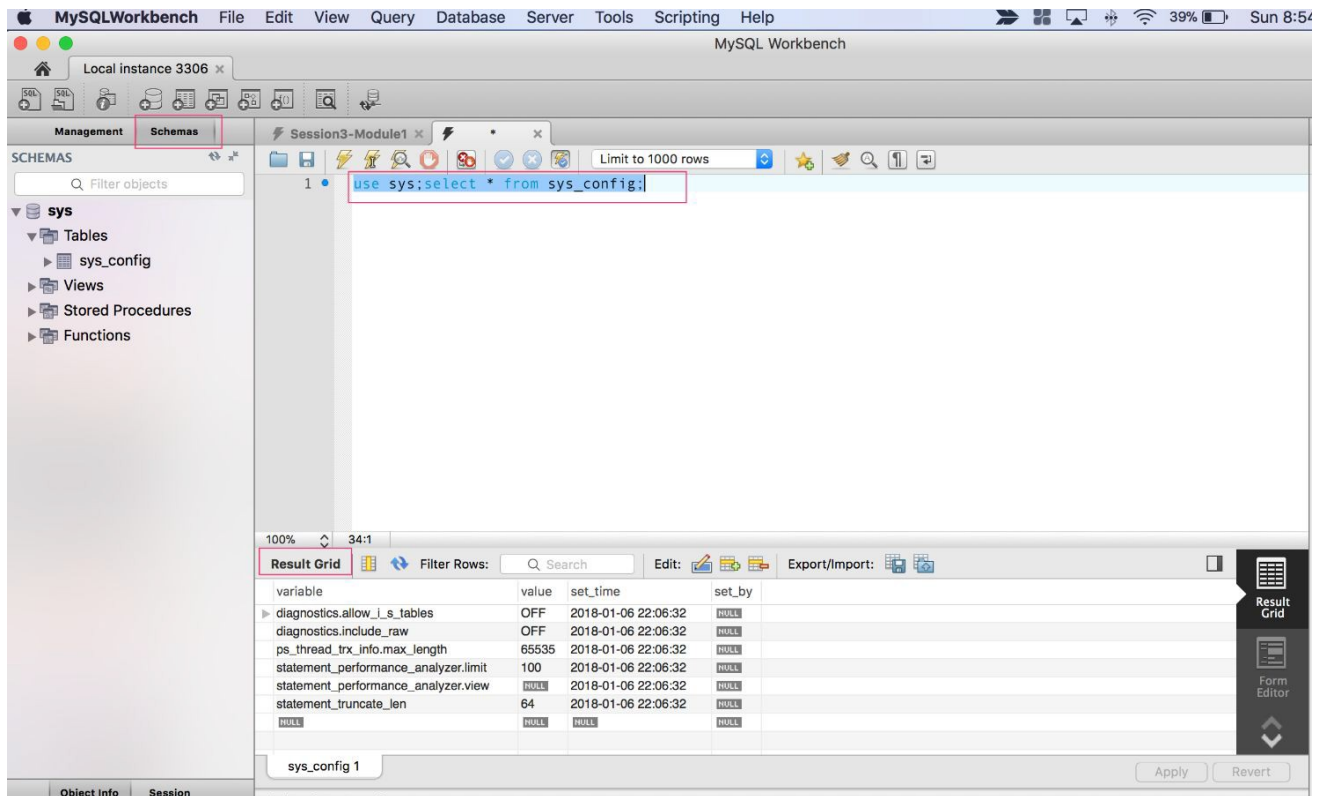
MySQL Connections  

Local instance 3306 (auto sa...)

 root

 localhost:3306

6. Type the following query and then run it (click lightning bolt or select and use CMD+ENTER to run)



7. In case of password related errors, please follow the steps below to reset the root password

- Stop MySQL server
- `sudo /usr/local/mysql/support-files/mysql.server start --skip-grant-tables`
- `/usr/local/mysql/bin/mysql`
- `mysql> FLUSH PRIVILEGES;`
- `mysql> ALTER USER 'root'@'localhost' IDENTIFIED BY 'root';`
- Ctrl + z
- `/usr/local/mysql/bin/mysql -u root -p`
- enter the new password i.e root

# MongoDB Installation & Setup

**Purpose:** MongoDB is the NoSQL document store database used in this course.

## Install MongoDB using Homebrew

1. Follow the instructions for installing via Homebrew found here:  
<https://docs.mongodb.com/manual/tutorial/install-mongodb-on-os-x/>

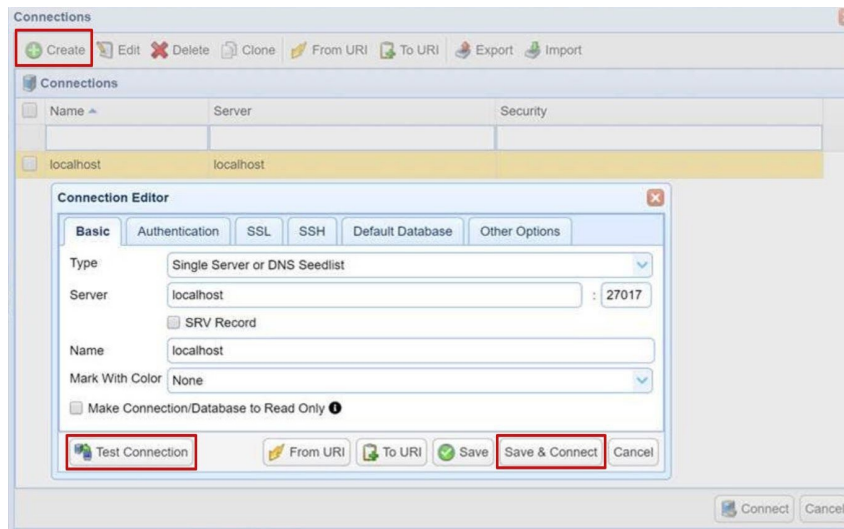
NOTE:

- a. Make sure your homebrew is up to date.
- b. Make sure while installing you select an option "Run service as Network Service user"

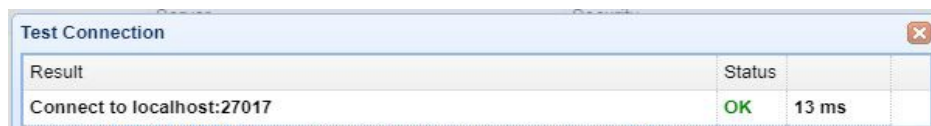
## Install NoSQLBooster Client & Connect

1. Download and install the NoSQLBooster client here: <https://nosqlbooster.com/>
2. Make sure MongoDB is running as a MacOS service (if you haven't completed this step already from above), using the homebrew command found here:  
<https://docs.mongodb.com/manual/tutorial/install-mongodb-on-os-x/#run-mongodb>
3. Open up the NoSQLBooster application.

4. Connect to your MongoDB Server
  - a. Click "Create"



- b. The server parameters should auto-fill
- c. Click Test Connection. Verify the Status is OK, then click "Close"



NOTE: You may see some warnings. Those are OK, as long as the Status says OK in green.

- d. Click "Save & Connect"
- e. You are now connected to MongoDB in NoSQLBooster



# Neo4j Installation

**Purpose:** Neo4j is the graph database used for this course.

**WARNING:** This is a trial installation for 30 days. WAIT to install until just prior to the week we use Neo4j in class (week 8).

## Download Neo4j

1. Download the free book on graph databases:  
<https://neo4j.com/graph-databases-book/?ref=home>
2. Download Community server version of Neo4j by clicking the download button at:  
<https://neo4j.com/download-center/#community>

The screenshot shows the Neo4j download center page for the Community Server edition. The page has a navigation bar with links to PRODUCTS, USE CASES, CUSTOMERS, PARTNERS, EVENTS & RESOURCES, and DEVELOPERS. The main content area is divided into three tabs: Enterprise Server, Community Server (which is selected and highlighted with a red box), and Neo4j Desktop. Under the Community Server tab, the page displays information for Neo4j Community Edition 4.0.6, released on 16 June 2020. Below this, there is a table with two columns: OS and Download. The table lists download links for Linux/Mac (Neo4j 4.0.6 (tar) SHA-256) and Windows (Neo4j 4.0.6 (zip) SHA-256). Further down, there is a section for Neo4j Repositories, which includes links for Debian/Ubuntu (Neo4j on Debian and Ubuntu Cypher Shell), Linux Yum (Neo4j Stable Yum Repo), and Docker (Neo4j Docker Image). At the bottom of the screenshot, there is a section for Neo4j Community Edition 3.5.19, released on 15 June 2020. This section also has a table with OS and Download columns, listing download links for Linux/Mac (Neo4j 3.5.19 (tar) SHA-256) and Windows (Neo4j 3.5.19 (zip) SHA-256). The entire section for Neo4j Community Edition 3.5.19 is highlighted with a red box.

OS	Download
Linux/Mac	<a href="#">Neo4j 4.0.6 (tar)</a> SHA-256
Windows	<a href="#">Neo4j 4.0.6 (zip)</a> SHA-256

Neo4j Repositories

Debian/Ubuntu	<a href="#">Neo4j on Debian and Ubuntu</a> Cypher Shell
Linux Yum	<a href="#">Neo4j Stable Yum Repo</a>
Docker	<a href="#">Neo4j Docker Image</a>

Neo4j Community Edition 3.5.19  
15 June 2020 [Release Notes](#) | [Read More](#)

OS	Download
Linux/Mac	<a href="#">Neo4j 3.5.19 (tar)</a> SHA-256
Windows	<a href="#">Neo4j 3.5.19 (zip)</a> SHA-256

3. Register with your name and email (if prompted)
4. Run the installer.
5. Launch Neo4j

## **Setup Neo4j**

1. Launch Neo4j Desktop
2. Accept the license terms
3. Confirm the location to store application data if prompted (default is fine)
4. Click to log in using a social account, then select Google.  
Log in using your uchicago.edu credentials.
5. Let the setup finish.



# **Tableau Installation**

**Purpose:** Tableau is a data visualization tool.

## **License & Install:**

1. Sign up for a 1-year educational license at <https://www.tableau.com/academic/students>
2. Click “Get Tableau for Free” and enter your information with uchicago.edu email
3. Verify your student status as prompted
4. Download and install Tableau as prompted

**Note:** After installing the tableau, if you are unable to open it, most likely you need to update your OS to the latest version.

# **Google Cloud Platform (GCP) Cloud SQL Setup**

**Purpose:** GCP will be used for your final project to collaborate with your team and get exposure to cloud computing.

## **Obtain GCP Credits (Two Options - try to get both)**

1. Sign up for \$300 of free GCP credits at <https://cloud.google.com/free>  
**NOTE:** This promotion is only for google accounts that have not used GCP before. If you have previously used GCP and obtained these credits, create another google account. It is recommended to use your university Google account.
2. Obtain the \$50 Education credits through the steps mentioned in "GCP-CreditsS.docx" in Canvas.  
**NOTE:** When accepting the \$50 credits, make sure to use the same Google account as you used for the \$300 credit above.

## **Setup GCP Project**

1. After signing up for credits, your console should open.  
If not, navigate to <https://console.cloud.google.com>
2. If you are greeted with a welcome screen, click on "Google Cloud Platform" in the blue bar at the top left to go to your main dashboard
3. Click on "My First Project" in the top blue bar and then click "New Project"
4. Give your project a name and click "Create"