CS 1656 – Introduction to Data Science (Fall 2016)

Department of Computer Science – University of Pittsburgh
Instructor: Prof. Alexandros Labrinidis – Teaching Assistant: Zuha Agha

Graph Databases: Neo4j & Cypher

Installation

Step 1: Install neo4j community edition 3.0.7 for your operating system from the link provided below.

https://neo4j.com/download/community-edition/

Step 2: Once neo4j is successfully installed, it will open up the neo4j browser at http://localhost:7474/browser/. You will be prompted to set up a username and password the first time the browser is opened. The default username is neo4j and default password is neo4j as well.

Step 3: Install official neo4j python driver which allows connecting to a neo4j graph database and run cypher queries from a python program. Type in the command below to install the driver.

```
pip install neo4j-driver
```

There are other python community drivers available as well which can be found at the link below,

https://neo4j.com/developer/python/

CS 1656 – Introduction to Data Science (Fall 2016)

Department of Computer Science – University of Pittsburgh
Instructor: Prof. Alexandros Labrinidis – Teaching Assistant: Zuha Agha

Dataset Download & Setup

For this assignment, you will be using a graph database of movies found at the link below,

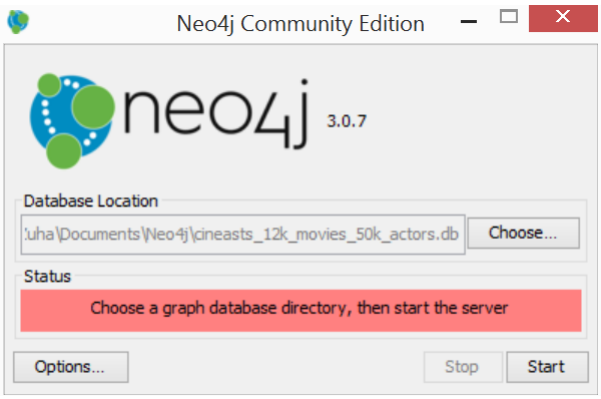
https://neo4j.com/developer/movie-database/

Click on Download for Neo4j 2.1 link under download and extract the zip folder where ever you want. You may copy the directory in your **/path/to/neo4j/data** directory and override the **deafault.graphdb** directory if you want. The directory **default.graphdb** is the default data directory when you start the neo4j server. Then follow the steps below,

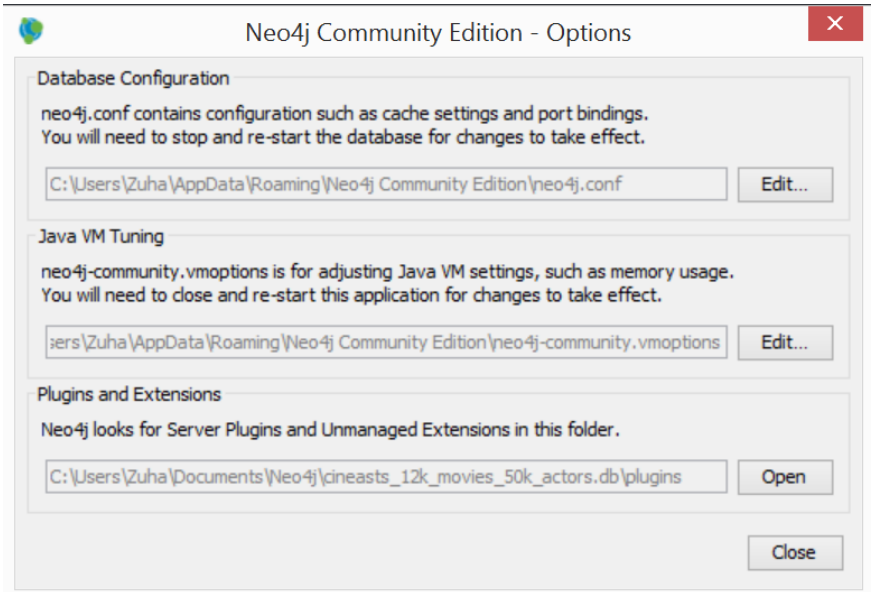
Step 1: Go to you neo4j/bin directory and run neo4j-ce.exe. The following dialogue will appear,

If the server is already running, click on 'Stop' button to stop the server.

Step 2: Click on Choose and select the movie dataset directory you downloaded.



Step 3: Click on option and the following dialogue will appear.



Step 4: Click on 'Edit' in Database configuration and make the changes in the configuration file as specified in the next steps.

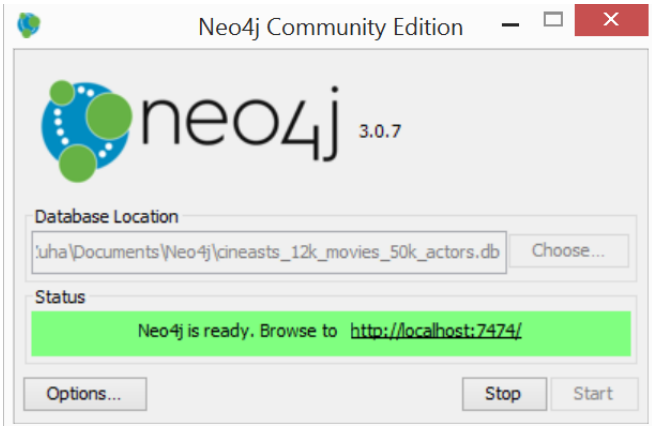
Step 5: Change dbms.secuirty.auth_enabled = true to false as shown below. This would allow you to the connect to the database without authenticating your username and password every time.

```
# Require (or disable the requirement of) auth to access Neo4j
dbms.security.auth_enabled=false
```

Step 6: Uncomment dbms.allow_format_migration = true (by removing the # sign) as shown below.

```
# Enable this to be able to upgrade a store from an older version.
dbms.allow_format_migration=true
```

Step 7: Click start to connect to the database and you will see the following status change in the dialogue when neo4j is ready. You are now ready to write cypher queries in python driver for neo4j. You may run your queries in the neo4j browser as well to observe the output but the final program MUST be submitted in python.



Querying Neo4j Graph Database in Python

The sample python program to execute cypher queries on the movie database using neo4j-driver is provided in **cypher_sample1.py.**