CS 1656 – Introduction to Data Science

Department of Computer Science – University of Pittsburgh
Instructor: Prof. Alexandros Labrinidis
Teaching Assistant: Xiaoting Li – Additional Credits: Tahereh Arabghalizi, Zuha Agha, Anatoli Shein

Setup for Graph Databases: Neo4j & Cypher

Step 1: Install neo4j desktop for your operating system from the link provided below.

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

After download the file, follow the "installation and launch guide" in the "Thanks for downloading" page and create a new graph.

Step 2: 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/

Step 3: Start Neo4j Desktop to create a new project. Click "Add Graph". You can either create a local graph or connect to a remote graph. Here we show you how to create a local graph. Click "Create a Local Graph". Then give the graph name and set the password. Click "Create" to create a new graph.

Step 4: For this assignment, you will be using a graph database of movies found at the link below:

http://data.cs1656.org/cineasts_12k_movies_50k_actors.zip

Extract the zip file where ever you want. The extracted directory **cineasts_12k_movies_50k_actors.db** can really go anywhere, but I put it next to the existing graph.db folder:

/path/to/.Neo4jDesktop/…/data/databases/ cineasts_12k_movies_50k_actors.db

An easy way to find the path "**/path/to/.Neo4jDesktop/…/data/databases/"** is to start the graph after **step3**. Once the server is started, click on Manage -> Open Folder to get this path.

Step 5: Edit the following configuration file:

/path/to/.Neo4jDesktop/…/conf/neo4j.conf

You can find this file by clicking on Manage -> Settings.

First, we will switch the default database **graph.db** to our downloaded one **cineasts_12k_movies_50k_actors.db**. To do that, find the commented out the line:

#dbms.active_database=graph.db

And add the following line below it:

dbms.active_database=cineasts_12k_movies_50k_actors.db

Also, make sure that the following lines are **NOT** commented:

dbms.directories.import=import

dbms.security.auth_enabled=true

dbms.allow_upgrade=true

Step 6: Once the server is started, click on Manage -> Open Browser or use the URL given for HTTP port:

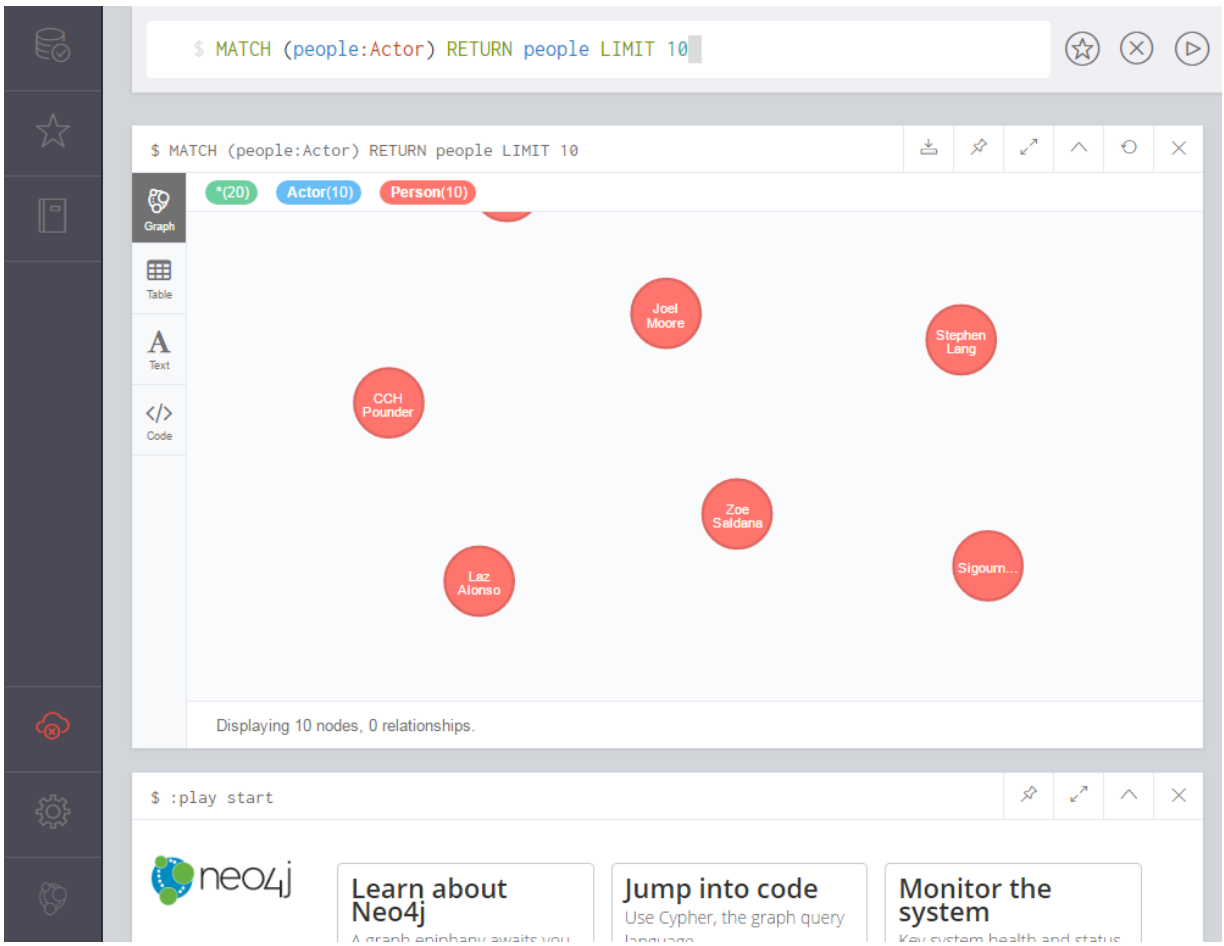
Now you can issue queries to the movies database from the input field at the top of this page[1].

For the example try:

MATCH (people:Actor) RETURN people LIMIT 10

You should see the following screen if it was a success:

[1] If the browser shows an empty graph database, restart the Neo4j Desktop and the graph.


$ MATCH (people:Actor) RETURN people LIMIT 10
$ MATCH (people:Actor) RETURN people LIMIT 10
Graph
Table
Text
Code
*(20)
Actor(10)
Person(10)
Joel Moore
Stephen Lang
CCH Pounder
Zoe Saldana
Laz Alonso
Sigourn...
Displaying 10 nodes, 0 relationships.
$:play start
neo4j
Learn about Neo4j
Jump into code
Use Cypher, the graph query
Monitor the system