

# CS 1656 – Introduction to Data Science

Department of Computer Science – University of Pittsburgh

Instructor: Prof. Alexandros Labrinidis

Teaching Assistant: Evangelos Karageorgos– Additional Credits: Xiaoting Li, 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/>

**Step 2:** Install the 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. You can name the project “cs1656”. Click “Add Database”. You can either create a local database or connect to a remote database. Here we show you how to create a local database. Click “Create a Local Database”. Then give the database a name (“Movie Graph”) and set the password (convenient password for the assignments would be “neo4jpass”). Next, select the database version as “3.5.17” from the drop-down menu. Click “Create” to create the database. DO NOT run it yet, though.

**Step 4:** Next to the “Movie Graph” database title, click on the three dots, and then click on “Manage”. Then, click on the “Settings” tab to view and edit the configuration file. In the configuration, uncomment the line “#dbms.active\_database=graph.db”. Also, uncomment the line “#dbms.allow\_upgrade=true”. Click on “Apply” to save the configuration. Next, click on the “Open Folder” button. The folder that will open will be the project folder for this database.

**Step 5:** 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](http://data.cs1656.org/cineasts_12k_movies_50k_actors.zip)

In the project folder, navigate to data/databases. Extract the “cineasts\_12k\_movies\_50k\_actors.db” folder from the provided database ZIP file into the folder

and rename it to "graph.db". You should have a folder data/databases/graph.db that contains many folders and files that correspond to the database data.

**Step 6:** Go back to the neo4j desktop application, select the cs1656 project, and click on the "Start" button to start the database for the first time.

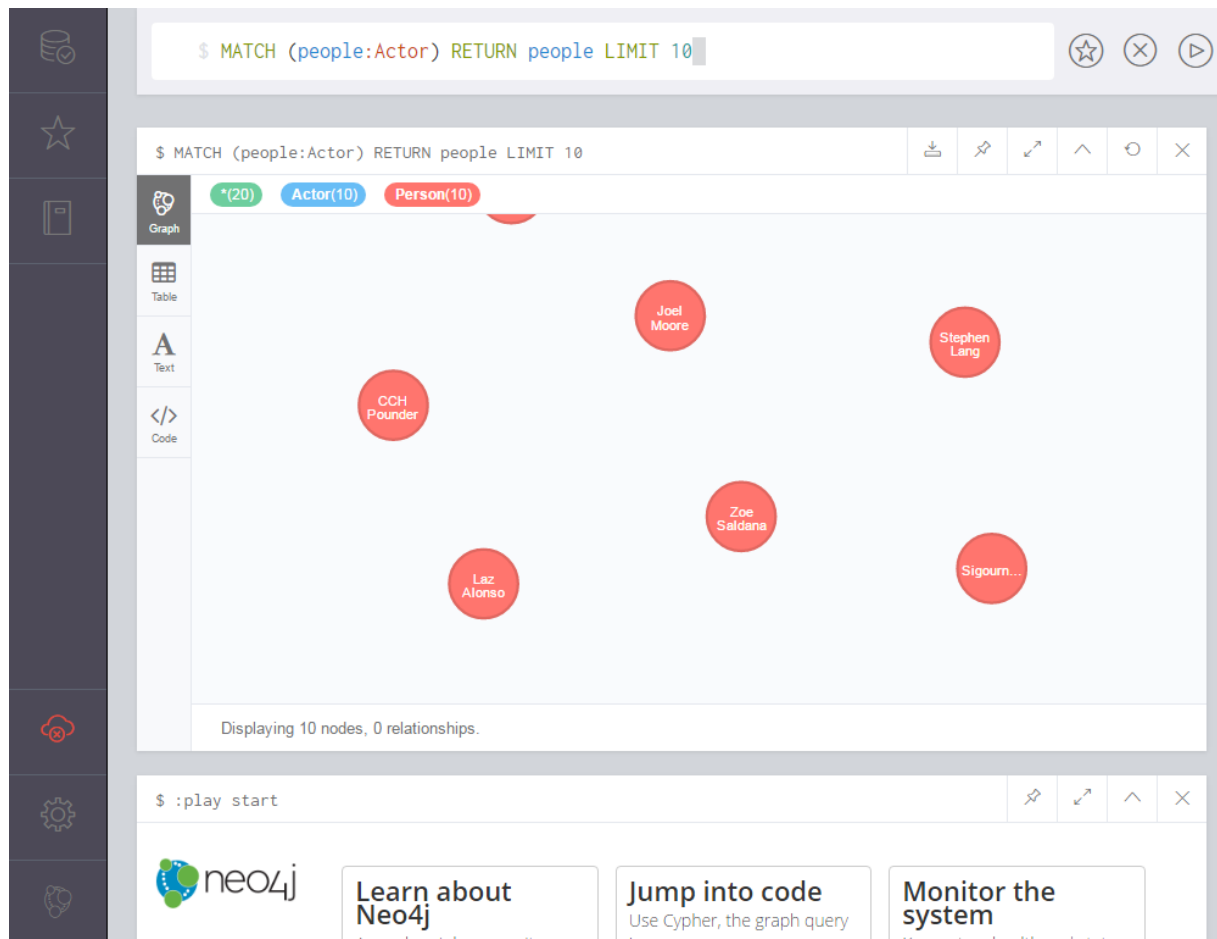
After a while, the database will be activated, and you can click on the "Open" button to launch a separate window to interact with the graph. You can also interact with the graph in a browser under <http://localhost/7474>.

Now you can issue queries to the movies database from the input field at the top of this page<sup>1</sup>.

For the example try:

**MATCH (people:Actor) RETURN people LIMIT 10**

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



<sup>1</sup> If the browser shows an empty graph database, restart the Neo4j Desktop and the graph.