rec10

November 4, 2020

1 CS 1656 – Introduction to Data Science

1.1 Instructor: Alexandros Labrinidis

1.1.1 Teaching Assistant: Evangelos Karageorgos

1.1.2 Additional credits:Xiaoting Li, Tahereh Arabghalizi, Zuha Agha, Anatoli Shein, Phuong Pham

1.2 ## Recitation : Cypher+Neo4j+Python

In this recitation, we will query a Neo4j graph database using with Cypher language and Python. Neo4j is a highly scalable, native graph database purpose-built to leverage not only data but also its relationships. Cypher is a declarative graph query language that allows for expressive and efficient querying and updating of the graph store.

In [10]:
```
# Use the following to get the neo4j database password from the user
import getpass
print ("Give me your neo4j password:")
neopass = getpass.getpass()
#print ("You typed:", neo4jpass)
```

```
Give me your neo4j password:
ůůůůůůůů
```

In [12]:
```
from neo4j import GraphDatabase, basic_auth
import time
import datetime

# More information on neo4j python API at:
# http://neo4j.com/docs/api/python-driver/current/

#Connect to the database
uri = "bolt://localhost:7687"
driver = GraphDatabase.driver(uri, auth=("neo4j", neopass))

#Start new session
session = driver.session()
```

```
        #Start new transaction
        transaction = session.begin_transaction()
```

Here are two helpful functions to convert cypher datetime objects to timestamps and reverse

```
In [13]: def fromTimestamp(timestamp):
             if isinstance(timestamp, str):
                 timestamp = int(timestamp)
             return (datetime.datetime(1970, 1, 1) + datetime.timedelta(milliseconds=timestamp))

         def toTimestamp(dateval):
             return time.mktime(datetime.datetime.strptime(dateval, "%m/%d/%Y").timetuple())*100
```

1.2.1 Queries

Q1) Find the actor named "Tom Hanks". OUTPUT: name, birthday, birthplace.

```
In [14]: result = transaction.run("""
             MATCH (tom:Actor {name: 'Tom Hanks'})
             RETURN tom.name, tom.birthday, tom.birthplace
         """)
         [(r[0], fromTimestamp(r[1]), r[2]) for r in result]

Out[14]: [('Tom Hanks', '07/08/1956 23:00:00', 'Concord, California, USA')]
```

1.2.2 Tasks

Q2) Find the movie with title "Avatar". OUTPUT: studio, release date.

Q3) Find movies released in the 1990s. Hint: In cypher, you can use the function toFloat() to convert a datetime value to a scalar value (for example, ... WHERE toFloat(val) > 1400288100); This makes it easier to compare it to other values generated by the provided python function toTimestamp(). OUTPUT: the movie title.

Q4) List all Tom Hanks movies. OUTPUT: movie title, studio, release date.

Q5) Who directed "Avatar". OUTPUT: name.

Q6) Find all Tom Hanks' co-actors. OUTPUT: co-actor's name.

Q7) How people are related to "Avatar". OUTPUT: name, the type of relationship

Q8) Find actors that have not worked with Tom Hanks but have worked with Tom Hanks' co-actors. Make sure your results are unique by using the DISTINCT clause. OUTPUT: name.

Q9) Find someone to introduce Tom Hanks to Tom Cruise. Someone who worked both with Tom Hanks and Tom Cruise in different movies. OUTOUT: name.

Let's close the session,the transaction and the driver.

```
In [15]: transaction.close()
         session.close()
         driver.close()
```