DATA ENGINEERING PLATFORMS (MSCA 31012)

sbharadwaj@uchicago.edu | jchan530@uchicago.edu

Submissions

- Single file (txt/word/Pdf) that contains the queries for both MongoDB and Neo4j along with any assumptions made

Part A (MongoDB): Manipulating, Sorting and Grouping & Summarizing data

Data (Sakila dataset):

- customers.json
- films.json
- stores.json

Note: Import dataset (sample scripts to load data)

mongoimport --db sakila --collection customers --drop --file "C:\Users\SBharadwaj\Desktop\Shree\DEPA\03-Assignments\4\solution\customers.json"
(imports 599 documents)

mongoimport --db sakila --collection films --drop --file "C:\Users\SBharadwaj\Desktop\Shree\DEPA\03-Assignments\4\solution\films.json"
(imports 1000 documents)

mongoimport --db sakila --collection stores --drop --file "C:\Users\SBharadwaj\Desktop\Shree\DEPA\03-Assignments\4\solution\stores.json"
(imports 2 documents)

Questions: – {50 Points}

1. List total number of customers living in california ?
2. List all movies that are rated NC-17
3. List the count of movies by category
4. Find the top 2 movies with movie length greater than 25mins and which has commentaries as a special feature

5. Provide 2 additional queries and indicate the specific business use cases they address

Part B (Neo4J): Linking, Manipulating & viewing relationships within data

Data (Movies dataset):

- Moviesdb.txt

Questions: – {50 Points}

1. Find all Producers that produced the movie When Harry Met Sally
2. Find directors who have directed more than 2 movies
3. Find the actors with 5+ movies, and the movies in which they acted
4. Movies and actors exactly 3 "hops" away from the movie Hoffa
5. Provide 2 additional queries and indicate the specific business use cases they address