

Name: Troy Zhongyi Zhang  
Netid: [zhongyiz@uchicago.edu](mailto:zhongyiz@uchicago.edu)

## **DATA ENGINEERING PLATFORMS (MSCA 31012)**

---

[sbharadwaj@uchicago.edu](mailto:sbharadwaj@uchicago.edu) | [jchan530@uchicago.edu](mailto: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)
```


Screenshot for successful importing files:

```
[(base) Zhongyis-MBP:Assignment4 zhongyizhang$ pwd  
/Users/zhongyizhang/Desktop/Assignment4  
[(base) Zhongyis-MBP:Assignment4 zhongyizhang$ mongoimport --db sakila --collecti  
on customers --drop --file "/Users/zhongyizhang/Desktop/Assignment4/dataset/cust  
omers.json"  
2019-08-17T10:12:38.078-0500    connected to: localhost  
2019-08-17T10:12:38.079-0500    dropping: sakila.customers  
2019-08-17T10:12:38.194-0500    imported 599 documents  
[(base) Zhongyis-MBP:Assignment4 zhongyizhang$ mongoimport --db sakila --collecti  
on films --drop --file "/Users/zhongyizhang/Desktop/Assignment4/dataset/films.js  
on"  
2019-08-17T10:13:43.237-0500    connected to: localhost  
2019-08-17T10:13:43.237-0500    dropping: sakila.films  
2019-08-17T10:13:43.295-0500    imported 1000 documents  
[(base) Zhongyis-MBP:Assignment4 zhongyizhang$ mongoimport --db sakila --collecti  
on stores --drop --file "/Users/zhongyizhang/Desktop/Assignment4/dataset/stores.  
json"  
2019-08-17T10:14:20.549-0500    connected to: localhost  
2019-08-17T10:14:20.549-0500    dropping: sakila.stores  
2019-08-17T10:14:20.594-0500    imported 2 documents  
(base) Zhongyis-MBP:Assignment4 zhongyizhang$
```

Questions: – {50 Points}

**1.List total number of customers living in California?**


`db.customers.find( { District: "California" } ).count()`

 0.216 s

1 9

Or

`db.customers.count( {District:"California"})`

 0.201 s

1 9

## 2. List all movies that are rated NC-17

```
db.films.find({ Rating: "NC-17" })
```

films		0.021 s	210 Docs		20		p. 1		1 - 20		Table	
	_id	Actors	Category	Description	Length	Rating	Rental Dur					
1	14	Array[4]	Classics	A Emotional Drama of a A Shark And a Database Administrator who must Vanquish a Pioneer in Soviet Georgia	94	NC-17	6					
2	16	Array[5]	Foreign	A Fast-Paced Drama of a Robot And a Composer who must Battle a Astronaut in New Orleans	180	NC-17	6					
3	15	Array[6]	Foreign	A Brilliant Drama of a Cat And a Mad Scientist who must Battle a Feminist in A MySQL Convention	46	NC-17	5					
4	27	Array[9]	Sports	A Amazing Reflection of a Database Administrator And a Astronaut who must Outrace a Database Administrator in A Shark Tank	179	NC-17	7					
5	10	Array[8]	Sports	A Action-Packed Tale of a Man And a Lumberjack who must Reach a Feminist in Ancient China	63	NC-17	6					
6	29	Array[7]	Action	A Fateful Yarn of a Womanizer And a Feminist who must Succumb a Database Administrator in Ancient India	168	NC-17	5					
7	3	Array[5]	Documentary	A Astounding Reflection of a Lumberjack And a Car who must Sink a Lumberjack in A Baloon Factory	50	NC-17	7					
8	31	Array[4]	Family	A Awe-Inspiring Reflection of a Pastry Chef And a Teacher who must Overcome a Sumo Wrestler in A U-Boat	92	NC-17	5					
9	34	Array[12]	Horror	A Touching Epistle of a Madman And a Mad Cow who must Defeat a Student in Nigeria	62	NC-17	6					
10	38	Array[3]	Action	A Beautiful Yarn of a Pioneer And a Monkey who must Pursue a Explorer in The Sahara Desert	68	NC-17	6					
11	42	Array[7]	Sports	A Stunning Reflection of a Robot And a Moose who must Challenge a Woman in California	170	NC-17	5					
12	47	Array[8]	Foreign	A Boring Character Study of a A Shark And a Girl who must Outrace a Feminist in An Abandoned Mine Shaft	153	NC-17	5					
13	51	Array[6]	Music	A Insightful Panorama of a Forensic Psychologist And a Mad Cow who must Build a Mad Scientist in The First Manned Space Station	75	NC-17	5					
14	53	Array[7]	Family	A Epic Drama of a Madman And a Cat who must Face a A Shark in An Abandoned Amusement Park	87	NC-17	5					
15	62	Array[5]	Documentary	A Astounding Panorama of a Lumberjack And a Dog who must Redeem a Woman in An Abandoned Fun House	106	NC-17	5					

## 3. List the count of movies by category

```
db.films.aggregate([
  {
    $group: { _id: "$Category",
              count: {$sum: 1}
            }
  }
])
```

films 0.015 s   16 Docs			
	_id	count	
1	Games	61	
2	Children	60	
3	Travel	57	
4	Comedy	58	
5	New	63	
6	Family	69	
7	Sports	74	
8	Sci-Fi	61	
9	Drama	62	
10	Action	64	
11	Foreign	73	
12	Animation	66	
13	Classics	57	
14	Documentary	68	
15	Music	51	
16	Horror	56	

**4. Find the top 2 movies with movie length greater than 25mins and which has commentaries as a special feature**

// List ordered by \_id, first appeared first results come (or ordered by \_id):

```
db.films.find(
  {$and: [
```

films	0.011 s	2 Docs						20		p. 1	1 - 2	Table	
_id	Actors	Category	Description	Length	Rating	Rental Duration	Replacement Cost	Special					
1	15	Array[6]	Foreign	A Brilliant Drama of a Cat And a Mad Scientist who must Battle a Feminist in A MySQL Convention	46	NC-17	5	10.99	Trailers,				
2	20	Array[6]	Music	A Boring Drama of a Woman And a Squirrel who must Conquer a Student in A Baloon	79	R	4	23.99	Comme				

```
db.films.find(
    {"Length": {$gt: "25"}, "Special Features": {$regex:
"Commentaries"}}
).sort({Length: -1}).limit(2)
```

```
// Use case 1:  
// Show the customer's full name (concatenated the first name and last  
// name into one field) if this customer is in the US and phone number  
// includes 26  
  
db.customers.aggregate([ { $match: { $and: [ { Country: "United States" },  
{ Phone: { $regex: "26" } } ] } }, { $project:  
{ "Full Name": { $concat: [ "$First Name", " ", "$Last Name" ] } } } ] )
```

customers 0.006 s   2 Docs		
	_id	Full Name
1	6	JENNIFER DAVIS
2	305	RICHARD MCCRARY

// Use case 2:

// List 10 movies in the music category ordered by length from the shortest to longer length movies

```
db.films.find( {Category:"Music"} ).sort( {Length: 1} ).limit(10)
```

films 0.013 s   10 Docs								
	_id	Actors	Category	Description	Length	Rating	Rental Duration	Replacement
1	752	Array[8]	Music	A Thoughtful Documentary of a Crocodile And a Robot who must Outrace a Womanizer in The Outback	101	NC-17	6	27.99
2	76	Array[4]	Music	A Fast-Paced Saga of a Frisbee And a Astronaut who must Overcome a Feminist in Ancient India	103	NC-17	4	23.99
3	983	Array[5]	Music	A Unbelievable Documentary of a Teacher And a Monkey who must Defeat a Explorer in A U-Boat	105	PG	7	18.99
4	220	Array[3]	Music	A Thoughtful Story of a Mad Cow And a Womanizer who must Overcome a Mad Scientist in Soviet Georgia	106	NC-17	7	13.99
5	592	Array[7]	Music	A Fast-Paced Story of a Waitress And a Cat who must Fight a Girl in Australia	107	PG	6	28.99
6	54	Array[6]	Music	A Awe-Inspiring Drama of a Car And a Pastry Chef who must Chase a Crocodile in The First Manned Space Station	113	R	5	15.99
7	987	Array[6]	Music	A Action-Packed Reflection of a Composer And a Mad Scientist who must Face a Pioneer in A MySQL Convention	116	PG	3	13.99
8	133	Array[7]	Music	A Fateful Reflection of a Moose And a Husband who must Overcome a Monkey in Nigeria	117	NC-17	7	14.99
9	477	Array[5]	Music	A Stunning Reflection of a Boat And a Pastry Chef who must Succumb a A Shark in A Jet Boat	118	PG	5	15.99
10	673	Array[9]	Music	A Epic Saga of a Hunter And a Technical Writer who must Conquer a Cat in Ancient Japan	118	PG-13	3	19.99

## 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

match (Producer)-[:PRODUCED]->(movies:Movie) where movies.title='When Harry Met Sally' return distinct Producer.name

```
$ match (Producer)-[:PRODUCED]->(movies:M
```



Table



Text



Code

Producer.name
"Nora Ephron"
"Rob Reiner"

or

```
$ match (Producer)-[:PRODUCED]->(movies:Movie) where movies.tit
```



Table



Text



Code

**Producer.name**




"Nora Ephron"

"Rob Reiner"

Started streaming 2 records in less than 1 ms and completed after 2 ms.

## 2. Find directors who have directed more than 2 movies

MATCH (director:Person)-[:DIRECTED]->(m) WITH director,  
COUNT(m) AS CountMovie WHERE CountMovie>2 RETURN  
DISTINCT director.name

\$ MATCH (director:Person)-[:DIRECTED]->(m) WITH director, CO	
 Table	<b>director.name</b>
	"Andy Wachowski"
 Text	"Lana Wachowski"
	"Rob Reiner"
 Code	"Ron Howard"
Started streaming 4 records after 49 ms and completed after 60 ms.	

## 3. Find the actors with 5+ movies, and the movies in which they acted

match(actor:Person)-[:ACTED\_IN]->(movie:Movie) with actor,  
collect(movie.title) as movies, count(movie) as mc where mc >= 5 return  
actor.name, movies



\$ match(actor:Person)-[:ACTED\_IN]->(movie:Movie) with actor, collect(movie.title) as movies, count(movie) as mc where mc...

Table

Text

Code

actor.name

movies

"Hugo Weaving"

["The Matrix", "The Matrix Reloaded", "The Matrix Revolutions", "Cloud Atlas", "V for Vendetta"]

"Keanu Reeves"

["The Matrix", "The Matrix Reloaded", "The Matrix Revolutions", "The Devil's Advocate", "The Replacements", "Johnny Mnemonic", "Something's Gotta Give"]

"Jack Nicholson"

["A Few Good Men", "As Good as It Gets", "Hoffa", "One Flew Over the Cuckoo's Nest", "Something's Gotta Give"]

"Meg Ryan"

["Top Gun", "You've Got Mail", "Sleepless in Seattle", "Joe Versus the Volcano", "When Harry Met Sally"]

"Tom Hanks"

["You've Got Mail", "Sleepless in Seattle", "Joe Versus the Volcano", "That Thing You Do", "Cloud Atlas", "The Da Vinci Code", "The Green Mile", "Apollo 13", "Cast Away", "Charlie Wilson's War", "The Polar Express", "A League of Their Own"]

"Hugo Weaving"

["The Matrix", "The Matrix Reloaded", "The Matrix Revolutions", "Cloud Atlas", "V for Vendetta"]

"Keanu Reeves"

["The Matrix", "The Matrix Reloaded", "The Matrix Revolutions", "The Devil's Advocate", "The Replacements", "Johnny Mnemonic", "Something's Gotta Give"]

"Jack Nicholson"

["A Few Good Men", "As Good as It Gets", "Hoffa", "One Flew Over the Cuckoo's Nest", "Something's Gotta Give"]

Started streaming 70 records after 14 ms and completed after 18 ms.

## 4. Movies and actors exactly 3 "hops" away from the movie Hoffa

MATCH (Hoffa:Movie {title:'Hoffa'})-[\*3]-(hollywood) RETURN DISTINCT hollywood

\$ MATCH (Hoffa:Movie {title:'Hoffa'})-[*3]-(hollywood) RETURN DISTINCT hollywood	
Graph	hollywood
Table	<pre>{   "name": "John C. Reilly",   "born": 1965 }</pre>
Text	
Code	<pre>{   "name": "J.T. Walsh",   "born": 1943 }</pre>
	<pre>{   "name": "Jack Nicholson",   "born": 1937 }</pre>
Started streaming 308 records after 7 ms and completed after 38 ms.	

```
MATCH (pers:Person)-[act:ACTED_IN]-(film:Movie)
WITH pers, act.roles as roles, film
WHERE size(roles) > 1
RETURN pers, roles, film
```

\$ MATCH (pers:Person)-[act:ACTED\_IN]-(film:Movie) WITH pers, act.roles as roles, film WHERE size(roles) > 1 RETURN pers,...

Graph

Table

Text

Code

pers	roles	film
<pre>{   "name": "Meg Ryan",   "born": 1961 }</pre>	[ "DeDe", "Angelica Graynamore", "Patricia Graynamore" ]	<pre>{   "title": "Joe Versus the Volcano",   "tagline": "A story of love, lava and burning desire.",   "released": 1990 }</pre>
<pre>{   "name": "Tom Hanks",   "born": 1956 }</pre>	[ "Zachry", "Dr. Henry Goose", "Isaac Sachs", "Dermot Hoggins" ]	<pre>{   "title": "Cloud Atlas",   "tagline": "Everything is connected",   "released": 2012 }</pre>

Started streaming 84 records after 1 ms and completed after 10 ms.