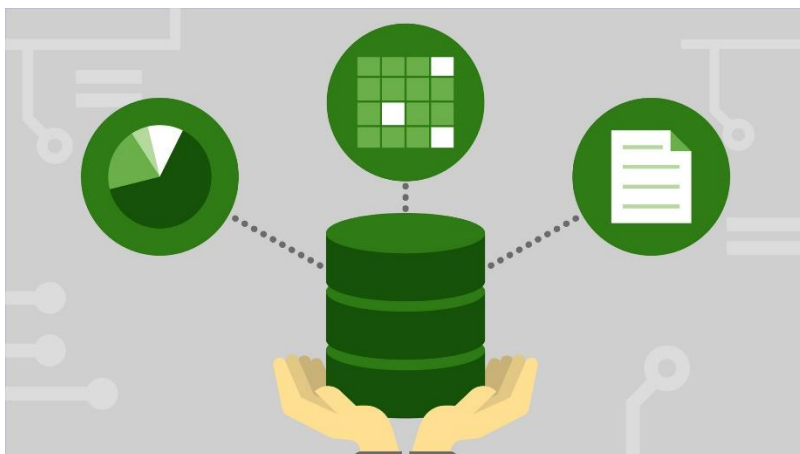


به نام خدا



دانشگاه تهران
پردیس دانشکده‌های فنی
دانشکده برق و کامپیوتر



آزمایشگاه پایگاه داده

دستورکار شماره ۳

شماره دانشجویی

۸۱۰۱۹۵۵۲۶

آبان ۹۹

پارسا صدری سینکی

گزارش فعالیت‌های انجام شده

بخش اول

در بخش اول ابتدا دیتاست MovieLens را دانلود می‌کنیم. jdk نسخه مورد نیاز روی سیستم نصب شد. سپس Neo4j را نسخه community آن را دانلود و نصب کردم. سپس با استفاده از کد Ingestion موجود داده‌های دیتاست را وارد دیتابیس می‌کنیم.

کد نیاز به تعدادی تغییر مثل حذف محدودیت ۱۰۰۰ داده و تغییر نوع ورودی کتابخانه py2neo و تغییر نحوه خواندن فایل‌های دیتاست به utf-8.

بعد از انجام این تغییرات داده‌ها به راحتی وارد دیتابیس شد و آماده اجرا query ها شد.

بخش دوم

1- این query انواع ژانرهای موجود را به صورت یک لیست بر می‌گرداند

```
MATCH (g:Genre)
WITH collect(g.name) as genres
RETURN genres
```

"genres"
["Action", "Adventure", "Animation", "Children's", "Comedy", "Crime", "Documentary", "Drama", "Fantasy", "Film-Noir", "Horror", "Musical", "Mystery", "Romance", "Sci-Fi", "Thriller", "War", "Western"]

2- تعداد فیلم‌های در دیتابیس

```
MATCH (:Movie)
WITH count(*) as count
RETURN count
```

"count"
9742

3- تعداد رای های Silence of the Lambs برای هر نمره

```
MATCH (:User)-[r:RATED]->(m:Movie)
WHERE m.title="Silence of the Lambs, The"
WITH r.rating as rating, count(*) as count
RETURN rating, count
ORDER BY rating
```

"rating"	"count"
0.5	3
1.0	1
2.0	5
2.5	7
3.0	24
3.5	14
4.0	97
4.5	36
5.0	92

4- تعداد رای برای هر فیلم از بیشترین به کمترین تعداد

```
MATCH (:User)-[r:RATED]->(m:Movie)
WITH m.id as id, m.title as title, count(*) as count
RETURN title, count
ORDER BY count DESC
LIMIT 10
```

"title"	"count"
"Forrest Gump"	329
"Shawshank Redemption, The"	317
"Pulp Fiction"	307
"Silence of the Lambs, The"	279
"Matrix, The"	278
"Star Wars: Episode IV - A New Hope"	251
"Jurassic Park"	238
"Braveheart"	237
"Terminator 2: Judgment Day"	224
"Schindler's List"	220

5-

قسمت اول) ژانری که بیشترین تعداد فیلم را دارد

```
MATCH (g:Genre)-[:IS_GENRE_OF]->(m:Movie)
WITH g.name as genre, count(*) as count
ORDER BY count DESC
LIMIT 1
RETURN genre
```

"genre"
"Drama"

قسمت دوم) ژانر با بیشترین میانگین امتیاز

```
MATCH (g:Genre)-[:IS_GENRE_OF]->(:Movie)<-[:RATED]-(:User)
WITH g.name as genre, avg(r.rating) as avgrating
ORDER BY avgrating DESC
LIMIT 1
RETURN genre
```

"genre"
"Film-Noir"

قسمت سوم) ژانر با کمترین میانگین امتیاز

```
MATCH (g:Genre)-[:IS_GENRE_OF]->(:Movie)<-[:RATED]-(:User)
WITH g.name as genre, avg(r.rating) as avgrating
ORDER BY avgrating
LIMIT 1
RETURN genre
```

"genre"
"Horror"

6- فیلم های سال ۲۰۰۰

```
MATCH (m:Movie)
WHERE m.year="2000"
WITH m.title as title
RETURN title
LIMIT 10
```

"title"
"Yards, The"
"Next Friday"
"Supernova"
"Down to You"
"Scream 3"
"Boondock Saints, The"
"Gun Shy"
"Beach, The"
"Snow Day"
"Tigger Movie, The"

7- تعداد تکرار هر تگ از بیشترین به کمترین

```
MATCH (u:User)-[t:TAGGED]->(m:Movie)
RETURN t.tag AS tag, count(t.tag) as NumberOfTags
ORDER BY NumberOfTags desc
LIMIT 20
```

"tag"	"NumberOfTags"
"In Netflix queue"	131
"atmospheric"	36
"thought-provoking"	24
"superhero"	24
"surreal"	23
"funny"	23
"Disney"	23
"religion"	22
"dark comedy"	21
"psychology"	21
"quirky"	21
"sci-fi"	21
"suspense"	20
"crime"	19
"visually appealing"	19
"twist ending"	19
"politics"	18
"music"	16
"time travel"	16
"mental illness"	16

8- میانگین نمره هر تگ از بیشترین به کمترین با حداقل ۵ بار تکرار تگ

```
MATCH (u:User)-[t:TAGGED]->(m:Movie)<-[r:RATED]-(u:User)
WITH t.tag as tag, count(t.tag) as NumberOfTag, avg(r.rating) as rating
WHERE NumberOfTag > 5
RETURN tag, NumberOfTag, rating
order by rating desc
limit 20
```

"tag"	"NumberOfTag"	"rating"
"classic"	8	4.9375
"great soundtrack"	6	4.833333333333333
"future"	6	4.75
"satire"	12	4.708333333333334
"mindfuck"	14	4.642857142857143
"cult film"	7	4.642857142857142
"dark humor"	7	4.642857142857142
"stylized"	9	4.611111111111111
"black comedy"	13	4.576923076923077
"clever"	7	4.571428571428571
"social commentary"	7	4.571428571428571
"twist ending"	19	4.552631578947368
"imdb top 250"	11	4.5
"New York"	6	4.5
"mystery"	6	4.5
"Quentin Tarantino"	7	4.5
"dark"	15	4.466666666666667
"dark comedy"	20	4.4499999999999998
"violence"	8	4.4375
"gritty"	7	4.428571428571428

9- فیلم های با بیشترین میانگین امتیاز با حداقل ۱۰۰ رای

```
MATCH (u:User)-[r:RATED]->(m:Movie)
WITH m.id as id,m.title as title, count(*) as count, avg(r.rating) as avgrating
WHERE count>=100
RETURN title
ORDER BY avgrating DESC
LIMIT 20
```

"title"
"Shawshank Redemption, The"
"Godfather, The"
"Fight Club"
"Godfather: Part II, The"
"Departed, The"
"Goodfellas"
"Casablanca"
"Dark Knight, The"
"Usual Suspects, The"
"Princess Bride, The"
"Star Wars: Episode IV - A New Hope"
"Schindler's List"
"Apocalypse Now"
"American History X"
"Star Wars: Episode V - The Empire Strikes Back"
"Raiders of the Lost Ark"
"One Flew Over the Cuckoo's Nest"
"Reservoir Dogs"
"Pulp Fiction"
"Matrix, The"