MongoDB Assignment

The objective of this assignment is to evaluate your knowledge about MongoDB and its shell. In the following, you will load a data set and write the queries allowing each one to respond to a specific information need.

Preparation

1. Download the following files from boostcamp.

• movielens_movies.json
• movielens_users.json

2. Create a database and name it as Assignment 2.
3. Create two empty collections. One for movies and the other for users.
4. Import the documents(movielens_movies.json and movielens_users.json) by using the following steps in MongoCompass:

   Mongo compass→ Add Data→ Import File→choose the .js file for each collection separately.

Now, the database is ready to be used.

Database schema

The movies collection contains information about movies, that is, their id, title and genre.

```
> db.movies.findOne()
{
      "_id" : 1,
      "title" : "Toy Story (1995)",
      "genres" : "Animation|Children's|Comedy"
}
```

The users collection contains information about users and their ratings of films. Among the information about the users are their id, name, age, occupation, and gender. The ratings given by each user are represented in a document table, each document containing the id of a movie (referring to the id of the movies collection), the rating assigned and the date on which the user left the rating.

```
> db.users.findOne({},{movies : {$slice : 2}});
{
        "_id" : 6038,
        "name" : "Yaeko Hassan",
        "gender" : "F",
        "age" : 95,
        "occupation" : "academic/educator",
        "movies" : [
                {
                        "movieid" : 1419,
                        "rating" : 4,
                        "timestamp" : 956714815
                },
                {
                        "movieid" : 920,
                        "rating" : 3,
                        "timestamp" : 956706827
                }
        ]
}
```

Submission

For submitting the assignment, you must create a report containing each demanded query (write below each question with the possibility of being copied) and the screen shot contained the query and its result. Please note that, for some of permanent modifications, you should also demonstrate the modified parts either on shell or compass.

Example of report

Query 1: Insert a new document in movies collection.

db.movies.insertOne({Example:1})

```
> db.movies.insertOne({Example:1})
{
        "acknowledged" : true,
        "insertedId" : ObjectId("6563a9034a055de2e353d590")
}
>
```

FILTER {Example:1}   ▸ OPTIONS

⬆ ADD DATA ▾   ⬆   VIEW ☰ {} ⊞   Displaying documents 1 - 1

        _id: ObjectId("6563a9034a055de2e353d590")
        Example: 1

Queries

Query 1. How many users are artists (artist) or scientists (scientist)?

Query 2. Who are the ten oldest female authors (writers)?

Query 3. The genres of the movie "Cinderella" should be Animation, Children's and Musical. Modify the corresponding document in a single query so that it contains these three genres without duplicate.

Query 4. For each user, create a num_ratings field which indicates the number of films which he noted.

Query 5. User Barry Erin saw the movie Nixon, which has id 14. It gives it a score of 4. Update the database to take this score into account. Remember that the num_rattings field should represent the number of movies rated by a user.

Query 6. In fact, user Marquis Billie has not seen the movie "Santa with Muscles", which has the id 1311. Delete the rating accidentally entered in the database.

Query 7. How many films were released in the eighties? (The year of release is indicated in parentheses at the end of the title of each film)

Query 8. How many horror movies are there?

Query 9. How many films are both "Musical" and "Romance" type?

Query 10. How many users have rated the movie id 1196 (Star Wars: Episode V - The Empire Strikes Back (1980))?

Query 11. How many users have rated all the films of the first "Star Wars trilogy" (id 260,

1196, 1210)?

Query 12. How many users have rated exactly 48 films?.

Query 13. How many users have rated more than 90 films?

Query 14. What are the last three films rated by Jayson Brad?

Query 15. Show how many films were produced in each year of the 1990s; order the results from the most successful year to the least successful.

Query 16. What is the average rating of the film "Pulp Fiction", which has the id 296?

Query 17. In a single query, return for each user their id, name, maximum, minimum and average scores they have given, and order the result by increasing average score.

Query 18. What is the highest rated genre (the one with the highest average of all scores)?

Query 19. Determine, for each year of production, the film which received the best average score among the films which received at least 1000 scores.

Query 20. What is the most popular genre in terms of number of ratings?