Ceng352 - Database Management Systems
Project 1 : SQL Queries

Spring 2019

1 Introduction

In this project you are asked to write several SQL queries on a relational movie database. The data in this database is from the IMDb dataset, which is officially released public subset of IMDb data that is available for access to customers for personal and non-commercial use. The database exists already on our MySQL server. The server information is in the following section of this document. Your user credentials to access the server will be sent to your METU mails.

The database consists of six tables:

```
Actors (nconst, primaryName, birthYear, deathYear, gender)
Directors (nconst, primaryName)
Movies (tconst, originalTitle, startYear, averageRating, numVotes)
Genres (tconst, genre)
Directed (tconst, director)
Casts (tconst, nconst, characters)

Genres.tconst refers to Movies.tconst,
Directed.tconst refers to Movies.tconst,
Casts.tconst refers to Movies.tconst,
Directed.director refers to Directors.nconst,
Casts.nconst refers to Actors.nconst
```

For each of the following 15 questions, you are asked to write the correct query to find and list asked information from the database.

1. List the actors who played in the movie 'The Shawshank Redemption'. Your query should return distinct names of the actors. [4 rows]
2. List the female actors who played in an action movie in a leap year. A year is a leap year if it is divisible by 4 but not by 100. If a year is divisible by 4 and by 100, it is not a leap year unless it is also divisible by 400. You should check that the genre is 'Action' and the year is a leap year. Your query should return the names of the female actors, the titles of the movies, and the years of the movies (no duplicate rows). [11283 rows]
3. List all dead actors who played in a "Star Wars" movie. For this question, a movie is accepted as a "Star Wars" movie if its title includes "Star Wars" as a substring. Your query should return the names of the actors, the titles of the movies, and the ages they died. [6 rows]

4. List the most popular genres and the number of movies in each such genre for each year. There should be more than 10000 movies in a genre in the whole table to be counted as popular. Your query should return the years, names of genres, and the number of movies in each such genre. Results should be in descending order by year. NULL values should not be included. **Results should not include year-genre pairs for which number of movies is zero.** [1727 rows]

5. For each year, list the number of action movies and the percentage of action movies to the overall number of movies that year. Your query should return the years, the total number of action movies in that year and the percentage of action movies in that year for years where the percentage is greater than 10%. The result should have appropriate column headings. [29 rows]

6. For each movie in year 2015, list the number of female and male actors who played in it. The result of your query should be a three column table which includes the names of the movies, the number of male actors and the number of female actors in movies. [16356 rows]

7. For each movie after year 2015, list the youngest actor(s) who played in it. Your query should return the title of the movie, the name of the youngest actor, and the age of the youngest actor at the year of the movie. [23229 rows]

8. List the actors who played only in movies with an average rating of at least 8. Your query should return the names of the actors. [15412 rows]

9. List the directors who played only in movies that they directed. Your query should return the names of the directors. [4637 rows]

10. List the actors who played in movies that are among the top 10 most voted movies. Your query should return the names of the actors, the titles of the movies, and the character the actor played. [40 rows]

11. A decade is a sequence of 10 consecutive years. For example 1954, 1955, ..., 1963 is a decade, and so is 1956, 1957, ..., 1965. For each decade **from 1900-1909 to 2010-2019**, list the movie with the highest number of votes. Your query should return the titles of the movies, and the first and last year of the decades concatenated with a dash (e.g. 1965-1974). **[111 rows]**

12. List the **top 101** directors who worked with the most number of female actors. Your query should return the names of the directors, and the number of female actors. Results should be in descending order by number of female actors. **[101 rows, number of movies in first row is 886, number of movies in last row is 192]**

13. List the directors of the movies where only male actors played. Your query should return the distinct names of the directors. **[36994 rows]**

14. List the directors who worked with some actors who played only in movies that they directed. Your query should return the distinct names of the directors. **[79694 rows]**

15. List the directors who worked with only the actors who played only in movies that they directed. Your query should return the distinct names of the directors. **[13719 rows]**

2 Accessing the Database

You should use following settings to access the pre-described relational movie database on our MySQL server. The database will also be exported as ".csv" files and uploaded to ODTÜClass. You can create and host your own local databases using those files to work on.

```
ip: 144.122.71.144
port: 8084
username: will be sent to your METU mail
password: will be sent to your METU mail
```

3 Submission

You are asked to submit a ".txt" file with all SQL queries in it. You should name your file as

```
your_student_id.txt
```

using your 7-digits METU student ID. The format for the file is as follows:

```
Question 1
SELECT ...
Question 2
SELECT ...
...
Question [Number of Question]
[SQL Query for That Question]
...
Question 15
SELECT ...
```

You don't have to write your queries on single lines. You are allowed (and encouraged) to write them in multiple lines for better readability. You should not write anything for the unsolved questions. You should have lines for only the solved questions. You should submit the ".txt" file to ODTÜClass before the deadline.