SESSION: 2024/25 DIET 1

Module Title: Database Development

Module Code: M1I325894
Level: 1

Module Leader: Dr Muhammad Ayub Ansari

Coursework Description

Contents

Database Development Scenario

The following entity-relationship diagram represents various relationships between entities for a movies database.

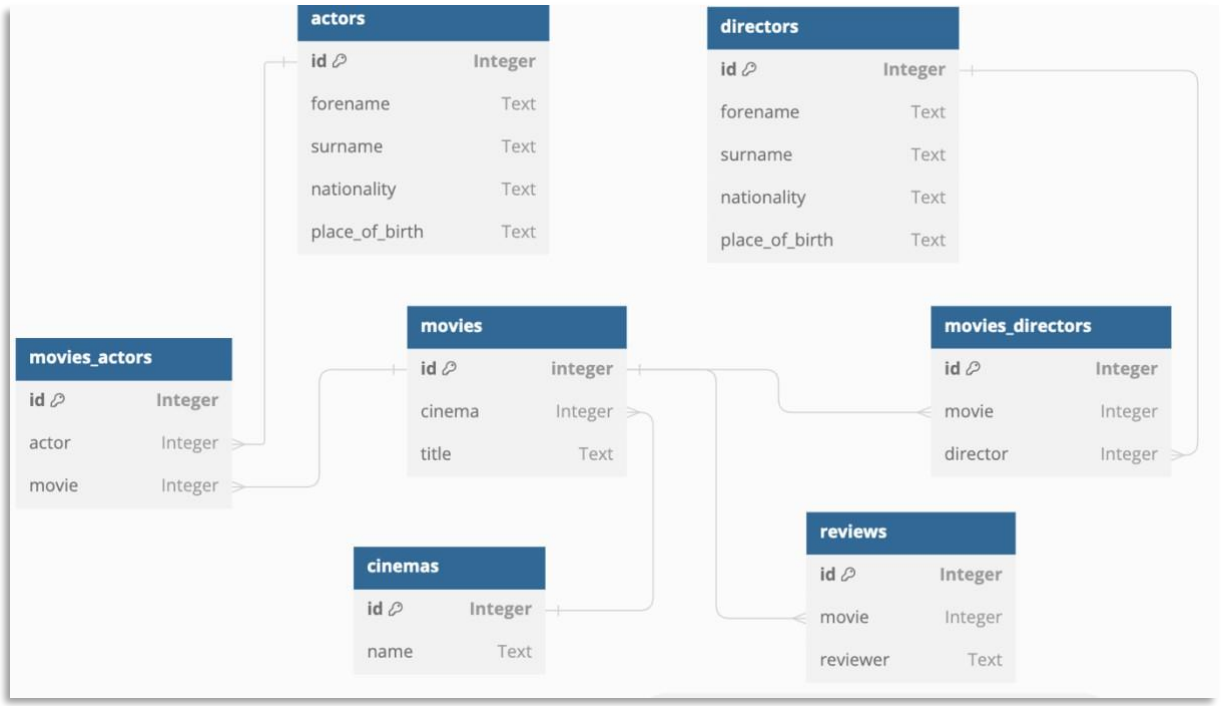

Figure 1: ERD representation.

Work-flow

A folder named <*surname_forename.zip*> is available. Download this and extract it to your machine. Rename the resulting folder appropriately, i.e., if your name is Ronnie Lane, then name it *Lane_Ronnie*. This folder, named after you, will be your *working directory* while you are developing your coursework, similar to the way you should have been working inside a lab folder.

Notice that the folder contains a number of file types (.sql, .csv) and two folders (*schema* and *documents*) Don't let the number of .sql files be a concern; you will realise that each .sql file only requires a relatively small amount of work. After you read through this entire document, you might find it useful to come back to the following 'Preliminary task', section.

Preliminary task

Note: the provided scripts <.sql> should be used to place your SQL code inside. **Do not** rename these files. The file <top_level.sql> has been completed for you; take a look at this in your editor (Notepad++, TextMate, etc), and call it (using sqlite3) to see the output. For example in Powershell type:

```
.\sqlite3.exe movies.db
```

Then, remembering to use the ‘dot’ prefix type:

```
.read top_level.sql
```

This will reveal the output from each script in the call hierarchy:

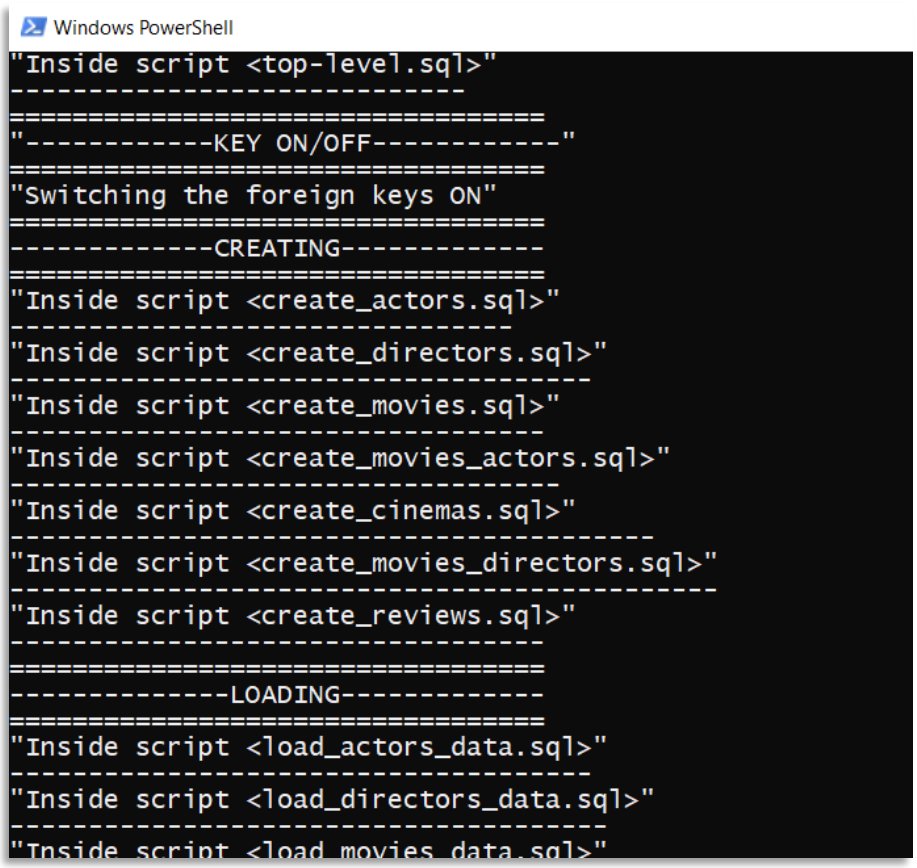

Figure 2: calling top_level.sql

When you have successfully completed the other scripts, calling *top_level.sql* should create the database and load the available data contain in the .csv file.

Once you have done these preliminary tasks, please read the marked coursework task below. Good luck.

Coursework marked tasks

N.B: All SQL must run in the software used on the course, i.e., **sqlite3**.

Task 1 [total marks = 70]

Task 1a: Creating tables and loading data [**35 marks**]

The task is to implement the ERD presented in Figure 1.

For **each table in the ERD** (there are 7), complete the two related SQL scripts – i.e. to *create* a table and then *load data* into it:

- create_*.sql
  - This script will implement the respective schema. The script should contain logic that removes a table, before recreating it, if it exists.
- load_*.sql
  - This file will load the data, which is provided in the respective .csv files.

5 marks for each script = [**35 marks]**.

Use the names in Figure 1to name your attributes. You can then use the data .csv files to load data into your database, once you have implemented the creation of the tables.

In summary, this part of the coursework will require you to complete a total of 14 scripts – i.e. two scripts per table in the diagram (2 x 7 = 14). These scripts, when called by the *top-level.sql* script, will create a database and populate it with the data.

Task 1b: Report [**35 marks**]

Notice also, after all of the calls to the various create and load scripts, a final script <report.sql> is called from <top-level.sql>. You should complete the report.sql, which will contain SQL commands to:

1. List the forename and surname of actors whose forename begins M and whose surname begins with S [**5 marks**].
2. List the movies in reverse alphabetical order of the title and limit the output to 5 movies [**5 marks**].
3. List the number of movies that are being shown at the Cameo [**5 marks**].
4. List the titles of the movies reviewed by 'Rotten Tomatoes' [**5 marks**].
5. List the percentage of the reviews provided by Film Critic [**5 marks**].
6. List the movie titles directed by Scorsese [**5 marks**].
7. List the actors in the Pulp Fiction movie [**5 marks**].

A call to <report.sql> is provided in the <**top_level.sql**> script, so as you implement the above calls the report part of the output will grow.

Task 2 [Total = 30 Marks]

Inside the document named <answers.docx> there are 2 questions. Please write your answer to those questions in the document.

1. *In light of the ERD provided with this coursework, and your experience of implementing a solution, discuss what primary and foreign keys are and why they are useful.* **[300 words]. [15 Marks]**
2. *A graph database could be used more effectively to represent movies information instead of a relational database. Discuss.* **[300 words] [15 Marks]**

Marking scheme summary

Task 1	Task 2	Total
70	30	100%

Submission

Do not:

- Rename the .sql files
- Change any of the data in the .csv files

If you do this could seriously affect your mark

Process

In addition to making sure you have completed all tasks, ensure that you have:

- Followed and completed all the above tasks.
- Signed the cover page <**cover-page.docx**> see *documents* subfolder.
- Zipped the working directory: **Do not** zip it up into any other format other than a <**.zip**>; if I was submitting this coursework the name of my zipped folder would be:
    - <***ansari_muhammad.zip***>

…a link will be provided on GCULearn; submit your zipped folder **before** the deadline if you would like your work marked.