Toronto Metropolitan University

CIND-110
Data Organization for Data Analysts

Assignment I
Design and Maintain Relational Databases

Assignment Context:

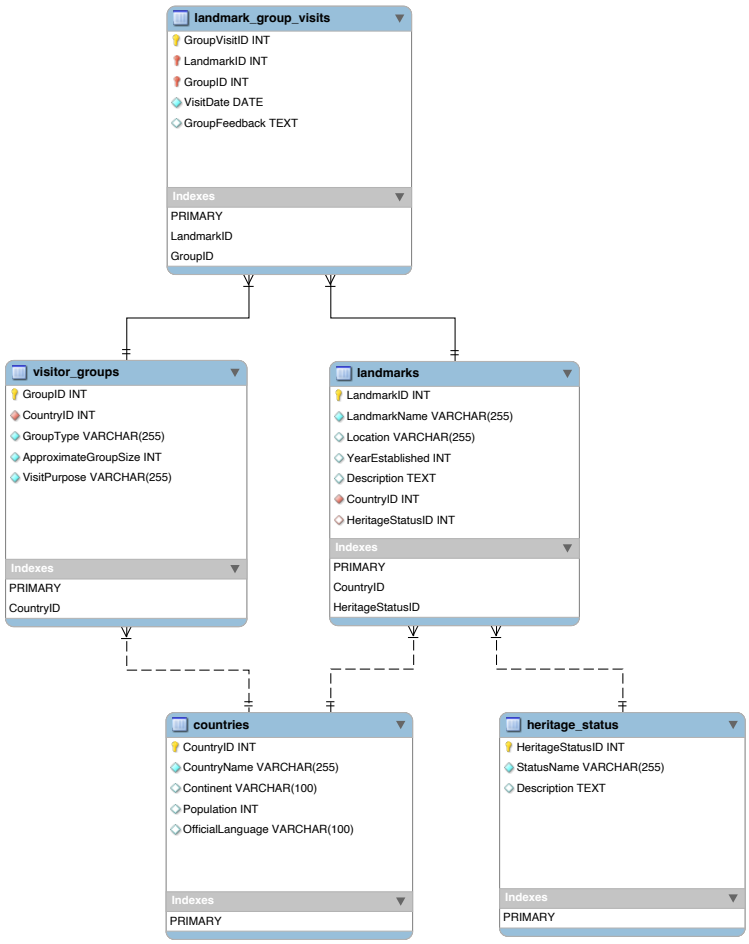
- A department is planning to migrate its local servers to a computational cloud platform, and the administration has requested each unit's DBA to ensure the secure migration of their respective databases to the new platform.
- A data analyst has been tasked with extracting and analyzing some data from an existing database with an outdated logical model stored on a particular server.
- To extract required data, a data analyst needs to check the most recent logical model of the existing database.
- A department unit aims to maintain its local database with more information and ensure its alignment with the information on the main central database.

Starts: Wednesday September 11, 2024, 11:59 PM
Due: Wednesday October 16, 2024, 11:59 PM
This assignment counts for 15% of the total grade

General Instructions:

- This assignment aims to give you hands-on experience in forward and reverse engineering processes. You will also learn how to create and execute basic and advanced queries with the `LandmarksDB` dataset. You can access the dataset as an `SQL` file from the `Assignment_1` tab on the course shell.

- The Entity-Relationship (ER) diagram below illustrates the design of the `LandmarksDB` database schema, including the names of the entities, the attribute data types, and the types of relationships.



- Download the `LandmarksDB.sql` file and execute it using the `MySQL Workbench` tool to load the database to your virtual desktop environment or local machine.

- Ensure that the primary and foreign keys are identified, and entity and referential integrity constraints are retained for the `LandmarksDB` database, all based on the above ER diagram. If necessary, please add any missing keys.

- After identifying the cardinality ratios and their directions for each relationship (whether they are one-to-many, one-to-one, or many-to-many), create them and synchronize the new ER model to the `LandmarksDB` database source.

Part I: [Total Points: 30]

- **Question 1 [15 Pts.]**

  Use the `MySQL Workbench` tool to perform a **reverse engineering** process on the entire database, including the five tables and their cardinality ratios. Save the resulting logical data model as an `MWB` file and store it on your TMU Google Drive.

  1. To **reverse engineer** the database, please follow the instructions provided in the official MySQL Workbench documentation at the following link:
     https://dev.mysql.com/doc/workbench/en/wb-reverse-engineer-live.html

  2. **Submission**: Submit the resulting logical data model in `MWB` format along with a screenshot in either `JPG` or `PNG` format. The logical model should represent all the database entities, attributes, keys, entity constraints, and referential integrity constraints.

- **Question 2 [15 Pts.]**

  Use the `MySQL Workbench` tool to **export** the entire database and save the output script as an `SQL` file onto your TMU Google Drive.

  1. To **export** data using MySQL Workbench, please follow the instructions provided in the official MySQL Workbench documentation at the following link:
     https://dev.mysql.com/doc/workbench/en/wb-admin-export-import-management.html

  2. **Submission**: Submit the output script in `SQL` format. The script should include the structure of the entities in addition to the data stored in these entities.

Part II: [Total Points: 70]

- For each of the following 7 questions: Write your answer as a `MySQL` query statement, execute it, and take a screenshot of the results. If the query result has more than 10 records, limit the output to the top 10 before taking the screenshot. Then, paste the answer and the respective screenshot into a Word or Google document file with the question number.

- **Submission**: Submit your answer sheet for this part either in `PDF` or `DOCX` file format.

1. [**8 pts.**] List all landmarks located in France and the USA, along with their location details and the associated country name.

2. [**8 pts.**] List the total number of landmarks in each country.

3. [**10 pts.**] Identify all landmarks established after 1900 that have a heritage status.

4. [**10 pts.**] Generate a list of countries, their continents, the total number of landmarks, and the total number of visitors (group size) to their landmarks. Sort the results by total visitors in descending order, excluding countries with no recorded visitors.

5. [**10 pts.**] Identify unique pairs of landmarks located in the same country that share the same location (e.g., the same city). Ensure that no redundant pairs are included (e.g., if landmark A and B are paired, only list A-B and not B-A).

6. [**12 pts.**] Identify the Most Popular Season for Landmark Visits for Each Continent.

   Hint: You can determine the season using the `CASE` statement and by grouping visits by continent and season. You may also use aggregation functions like `SUM()` to calculate the total number of visitors and `MAX()` to identify the season with the highest number of visitors for each continent.

7. [**12 pts.**]Identify countries with landmarks that include the terms "citadel," "fort," or "castle" in their description and show the percentage of visits to these landmarks compared to the total number of visits to all landmarks in the country. Include countries without visit data, displaying their visit counts and percentages as 0.

   Hint: Use `LIKE` to filter landmarks that contain the specified keywords in their description. You may also use a nested query to calculate the total number of visits for each country, and apply `IFNULL()` to handle cases where no visit data is available.

Here is a link for more details on **how to submit your assignment** to the course shell:
https://www.torontomu.ca/courses/students/tutorials/assignments/

This is the end of the assignment