CS306 Database Systems

2022-2023 Spring Semester Group Project

Step 3

For Step 3 of your project, you need to analyze your data and gain insights while also applying your SQL knowledge. It is crucial to have successfully created your relational model in the MySQL database and ensured data integrity by fully loading your datasets. You may also incorporate additional data sources if necessary to further support your story at this stage.

If you encounter null value problems when inserting your data into the tables, you can fill any null values in the data files with default values such as 0 for numeric fields and NA for character fields. Additionally, you can use the "load data infile" command to quickly insert your data. Once your datasets are complete and accurate, you can proceed to the following steps.

1. Discover Insights: Analyze your data and find patterns, trends, and correlations. When achieving this, please complete the following steps to demonstrate your SQL skills:

    a. Create Views: Create views based on the criteria you need to analyze. The number of views should be equal to the number of group members you have. For example, in the case of COVID-19, you can create a view of high_death_rate_countries by grouping the countries that have a death rate of X% or more over their population, and a view of low_death_rate_countries by grouping countries with a death rate of Y% or less over their population. You can also create views for highly_vaccinated_countries, lock_down_applied_countries, high_income_countries, low_income_countries, and so on. Or if you would like to group by date, you can create lock_down_period so on.

    b. Joins and Set Operators: Use set operators between two relations/views in your SQL statement. For example, use the “Except” operator to find the difference set of high_income_countries over low_death_rate_countries. Then use Outer Join to achieve the same result. Make sure you can show the same count of tuples in your log file.

    c. "In" and "Exists": Write a nested select statement using the "In" operator, and then replace it with the "Exists" operator and show you have the same result (same count of tuples in your log file).

    d. Aggregate Operators: Use all five aggregate operators (SUM, AVG, COUNT, MIN, MAX) in select statements. The number of SQL statements should be equal to the number of group members you have. You need to include group by and having clauses by joining at least two tables in each statement.

2. Constraints and Triggers: Determine the minimum and maximum values available in a table of your choice and a numeric field. Add a general constraint to this table with the alter table command and name your constraint. It should prevent entering values other than min and max

values. Then, try to enter a value other than the specified values in this field with the insert command. Create "before insert" and "before update" triggers on the same table, and if the value to be entered is outside of the ranges, the values should be fixed to the max and min values in the trigger script. Your SQL action log file should show your insertion log and returned error messages.

3. Stored Procedure: Create a stored procedure with one parameter (e.g. iso_code) and make sure your SQL command work differently for different input values of the parameter. For example, if the parameter iso_code is 'US', the SQL command could return data related to COVID-19 cases and deaths in the United States, while if the parameter is 'GB', the SQL command could return data related to COVID-19 cases and deaths in the United Kingdom. Call your stored procedure with two different parameter value and show the output in your log file.

Import your SQL files and SQL action log files into your repository. The SQL action log file should show all your statements' return logs. Please remove all unnecessary lines and comment the result for your cases. We will review the work you commit before the due date in your GitHub repositories. Please ensure that you upload your work before the deadline and that it is not accidentally deleted.

In your PDF report, list your SQL sentences for each category, explain your purpose for writing each command in English, and describe the result of each SQL statement. Add your create view statements, stored procedure statement, constraint and trigger statements, and compare creating general constraint and creating trigger methods with pros and cons in terms of enforcing data integrity. Ensure that you include a link to your repository in your PDF report and state your SQL file name and SQL action log file name for this step.

If you have any questions or need further clarification, please do not hesitate to reach out to me or TA Semih Yılmaz at yilmazsemih@sabanciuniv.edu.

Please note that only one group member is required to make the submission. Please clearly state your group name, the names, and IDs of all participating group members in your PDF report. However, if any of your group members do not participate in the workshops or work harmoniously, you may not include their names in the submission for that project step, based on a consensus decision.

Late submissions will not be accepted, so please be careful with your work and the deadline.

At the end of the project, you will need to report and present your data story, which is the process of using data to tell a compelling narrative. This involves weaving together data points, visuals, and context into a coherent story that provides insight into your topic. To accomplish this, you will need to discover insights and use data visualization techniques.

Data visualization is the practice of displaying data and information in a visual format, such as charts, graphs, maps, and infographics. The main objective of data visualization is to help people understand complex data and identify patterns or trends that might be difficult to discern from raw data.

Good luck with your project!