Sistemas de Informação e Bases de Dados 2021/2022

Project Assignment - Part 3

The third assignment consists of the development of SQL queries, integrity constraints and the creation of a web application prototype and OLAP queries. Use the supplied **`schema-part3.sql`** file to create the database for this part.

1. Database Loading

Once created, the database should be consistently filled with the records required to ensure that all SQL queries, presented below, have a **non-empty result**. Record creation and the database loading can be carried out through whatever method you find more adequate (manually, Excel spreadsheet, SQL script, Python, or other).

2. Integrity Constraints

Write the code to implement the following integrity constraints - which are presented assuming your knowledge of previous phases - with the SQL procedural extensions (Stored Procedures and Triggers) in the table schema provided:

(IC-1) Two reservations for the same boat can not have their corresponding date intervals intersecting.

(IC-2) Any location must be specialized in one of three - disjoint - entities: marina, wharf, or port.

(IC-3) A country where a boat is registered must correspond - at least - to one location.

The integrity constraints **definable** without resorting to procedural extensions (Stored Procedures and Triggers), should be implemented using other mechanisms if appropriate. However, mechanisms such as ON DELETE CASCADE and ON UPDATE CASCADE are not allowed.

3. SQL

Present the most succinct SQL[1] query for each of the following questions. If appropriate, you can use the view created previously.

[1] You cannot use SQL instructions that are not part of the standard (such as the `LIMIT instruction`).

1. Who is the owner with the most boats per country?
2. List all the owners that have at least two boats in distinct countries.
3. Who are the sailors that have sailed to every location in 'Portugal'?
4. List the sailors with the most trips along with their reservations
5. List the sailors with the longest duration of trips (sum of trip durations) for the same single reservation; display also the sum of the trip durations.

4. View

Create a view that summarizes the most important information regarding boat trips, combining information from different tables in the relational model. The view should have the following schema:

```
trip_info(
   country_iso_origin, country_name_origin,
   country_iso_dest, country_name_dest,
   loc_name_origin,
   loc_name_dest,
   cni_boat,
   country_iso_boat, country_name_boat,
   trip_start_date )
country_iso_origin : Foreign Key(Country)
country_iso_dest : Foreign Key(Country)
cni_boat, country_iso_boat: Foreign Key(Boat)
```

5. Application Development

Create a web application prototype to demonstrate to the possible prospect clients. You application should be developed using Python CGI scripts and HTML pages that allows users to:

a) Register and remove owners
b) Register and remove boats including boats with a VHF station
c) Insert, list and remove boat sailors
d) Create and remove reservations

The solution should prize security, preventing attacks by SQL INJECTION. Additionally, the **atomicity of related operations** in the database should be ensured. CSS can be used but is not ma

6. Data Analytics Queries

Using the view from Question 4, write two SQL queries that allow you to analyze the total number of trips according to different groups depending on:

1. The start date (i.e., per year, per month independently of year, and per exact date);
2. The location of origin (i.e., per location within countries, per country, and in total).

The submitted solution must use ROLLUP, CUBE, GROUPING SETS instructions, or the UNION of GROUP BY clauses.

7. Indexes

Present the SQL index(es) creating instruction(s) to improve the querying times for each of the cases listed below, explaining what are the operations that would be optimized and how.

Indicate, with proper justification, what type of index(es), over which attribute(s) and over which table(s), it would make sense to create, in order to speed up each query execution. Assume that the size of the tables exceeds the available memory by several orders of magnitude.

Assume that there are no indexes over the tables, aside from those implicit when declaring primary and foreign keys.

7.1 - List the names of all boats with a registration year above some given year, and registered in a country with a given name:

```
SELECT boat.name
FROM boat INNER JOIN
     country ON boat.iso_code=country.iso_code
WHERE year >= <some year>
AND country.name = <some country>;
```

7.2 - Count the number of trips with a start date within two points in time, to a location whose name features a given textual pattern:

```
SELECT count(*)
FROM trip INNER JOIN location ON
      trip.end_latitude=location.latitude
      AND trip.end_longitude=location.longitude
WHERE start_date BETWEEN <some date> AND <some date>
AND location.name LIKE '%SOME_PATTERN%';
```

Report

The project will be graded based on the report submitted and the discussion. The report should contain all answers to the items requested above. In the table below the points awarded to each portion of the work are listed.

Item	**Points**
SQL	5
Integrity Constraints	2
Application	6
Indexes	3
Data analytics	4

The report should begin with a cover page with the title "**Database Project, Part 3**", the **name and number of the students, the contribution from each member in relative percentage, along with** the **effort (in hours)** that each member put into the project, the **group number**, the group **shift** and the lab teacher's name. Besides the cover page, the report should have, at most, **8 pages**.

Delivery

The submission in Fénix should be a **zip** file with the following structure:

reportGG.pdf (where GG is the group number)	The report in **pdf** where **GG** is the group number, containing an **explanation of the architecture of the web application with a link to a working version**, the **relations between the various files** and the **indexes**. You should not include the instructions used to populate the database. The report does not need to include the OLAP component. ⚠ Groups must make sure to have the application working online until after the discussion.
queries.sql	File with the SQL queries.
populate.sql	File to populate the database with the test data.
ICs.sql	File to create the integrity constraints (triggers & stored procedures).
view.sql	File with the instructions to create the view

analytics.sql	File with the data analytics queries
web/	Folder with the Python and HTML files.

The project must be delivered in ZIP formatted with the name `delivery-03-GG.zip`[2] (where **GG** is the group number), through Fénix until 23h59 of the delivery date.

Note: Penalties apply to groups that do not follow the delivery instructions. Evaluation elements that are not found according to the instructions prescribed above **will not be taken into account for grading**.

[2] ⚠ The file format must be exclusively ZIP or GZ. Other archive formats (such as RAR) will not be accepted.