Assignment # 2, Basi di Dati 2015/2016

Instructor: Prof. Yannis Velegrakis

Script (from Task 1) Due for Delivery (by email): Sunday Nov 08 at 06.00 morning

Assignment Due for Delivery (on easychair): Fri Nov 20 at 6.00 morning

Note that the UNITN code of conduct strictly forbids any copying (and considers it cheating). If cheating is suspected, the assignments will both get 0 marks and the issue will be reported to the related authorities within the university.

Consider the following Relational Schema

Country(code: Str, name: Str, capital: Str, area: int)
- code (this is the usual country code, e.g. CDN for Canada, F for France, I for Italy)
- name (the country name)
- capital (the capital city, e.g. Rome for Italy)
- area (The mass land the country occupies in square Km)

Population(country: Str, population: int, children: int, adult: int, birth_rate: int, death_rate: int, sex_ratio: int)
- country is FK to the Country table
- population (this is the number of the people living in the country)
- children (The percentage of the population that are between 0 and 14 years old)
- adult (The percentage of the population that are between 15 and 64 years old)
- birth_rate (births in a year per 1,000 people of the population)
- death_rate (deaths per 1,000 individuals per year)
- sex_ratio (sex ratio at birth: male/female * 100)

Economy (country: Str, GDP: int, inflation: int, military: int, poverty: int)
- country is FK to the Country table
- GDP (gross domestic product)
- inflation (annual inflation rate)
- military (military spending as percentage of the GDP)
- poverty rate (percentage of population below the poverty line)

Language (country: Str, language: Str, percentage: int)
- country is FK to the Country table
- language is a spoken language name
- percentage (percentage of population speaking the language)

Task 1: Create a database with the above tables and populate it

Create the above database in PostgreSQL. Make sure you create also all the keys and foreign keys.

Create a script called load.sql that contains the commands to populate the above tables with data. You need to insert data regarding the country, population, economy and languages of 12 countries of your choice (you can insert more if you like). The country code should be one of the international country codes as listed here:

https://en.wikipedia.org/wiki/List_of_international_vehicle_registration_codes

For the rest of the fields you can try to use if you want realistic data taken from the CIA World fact book:

https://www.cia.gov/library/publications/the-world-factbook/

Also try to use some non-traditional countries, meaning, try not to put only US/Canada and EU countries, but some other that are not so popular.

Test that your data has been correctly inserted in the database by executing commands like:

```
select * from Country
```

MAKE SURE that you run your script before sending it to ensure that there are no syntax errors.

Note that the deadline for this Task is different than the deadline for the tasks that follow.

Task 2: Write the following queries in Relational Algebra

A. Find the most popular language in Italy
B. Find the codes of the countries in which English and French are spoken
C. Find all the codes of the trilingual countries (i.e., countries in which at least 3 languages are spoken)
D. Find the code of the second largest country in the world (in terms of land mass area)

Create a pdf file called **RA.pdf** containing the Relational Algebra queries that find what is asked in the points above. The answers should all be typed in a computer (no photos of handwritten solutions are allowed). Latex or MS Word or any other editor is fine. For the relational algebra you may need to use the math mode (in the case of latex). In the case of MS Word, you may use the equation editor or features like subscript, Greek symbols, and other symbols, that make clear the notation of the relational algebra as was mentioned in the class. You can also use any other editor as long as the symbols are clear. Note that using any other name for the file instead of RA.pdf will make this part being ignored.

Task 3: Write the following queries in SQL

1. Find the percentage of elderly population (65 and over) in the country with the highest GDP.
2. Find the dominant language (dominant means: spoken by more than 50% of the population) of the country with the highest male/female ratio.
3. List 5 countries with the highest military spending (not the percentage, but the actual spending), and for each of them list their capital and area.
4. Find the poverty rate in the country/countries with the largest number of languages spoken.
5. Find all countries where English is the dominant language, and the poverty rate is higher than that of the USA.
6. Find the country with the fastest declining population (a decline is a positive value in the difference death_rate - birth_rate).
7. For each language, find the percentage of the world population that speaks it.
8. For each language, find the percentage of the world population that speaks it, but considering only countries whose population is declining.
9. Consider the following hypothesis: the top 10 countries in terms of the percentage of their elderly population are among the richest 20% (in terms of the GDP per capita). Write an SQL query that checks if the hypothesis is true (if it is true, the output must contain those 10 countries).
10. Consider the following hypothesis: 80% of the world population lives in countries that are among the poorest 20%. Write an SQL query that checks if the hypothesis is true (if it is true, the output must be the word 'yes'. If not true, the output should be the word 'no').
11. List all the countries that belong to the top 10 in terms of both of the following criteria: the percentage of people in poverty, and GDP per capita.
12. Assume that all the countries stop military spending, and distribute the money back to their citizens. Find the average, maximum, and minimum increase of GDP per capita due to this action. For the minimum and maximum, also list the country (countries).
13. Order languages by the average percentage of the adult population of countries in which they are spoken by at least 25% of the population (in the decreasing order).
14. Find the richest (highest GDP) country whose name starts with a 'C'.
15. Find the code of the 15th largest country in the world (in terms of land mass area). (If more than one country have the same area, then they are considered each one individually. For example, if Canada is the largest country and has a land mass of 1000 square Km, and India has also 1000 sq meters, and USA has 900, then USA is considered the 3rd largest country).
16. Find the code of the 15th largest country in the world (in terms of land mass area). (If more than one country have the same area, then they are all considered in the same place. For example, if Canada is the largest country and has a land mass of 1000 square Km, and India has also 1000 sq meters, and USA has 900, then USA is considered the 2nd largest country).
17. Bonus Query: Print the letter "A" two times.

Create 17 text files named as X.sql, where X is a number between 1 and 17. Each of these files should contain the SQL query that produces what the respective point above is asking. For instance, the file 14.sql will contain the SQL query that finds the richest country whose name starts with "C". Each file should be able to run in PostgreSQLthrough the command: **psql < X.sql**

Delivery

Note that the load.sql file has a different delivery date than the rest, and that giving different names to the files will simply makes the files to be ignored.

The assignment can be done in groups of 2 persons. Groups of 1 person may be allowed if there is a serious problem and only after a communication with the instructor. You need to send an email to the instructor explaining the reason or the problem that does not allow you to form a group of 2.

To submit the assignment, create in google drive a directory named as your matricola number. (only one member of the group needs to do that….). Inside that directory create a text file called **Partner.txt** in which you put the matricola of the other member of the group

Put also in that directory the **load.sql** file, the **RA.pdf**, and all the **X.sql** files.

Share the directory (**in read only mode**) with the following three persons:

d.papadimitriou@unitn.it, matteo.lissandrini@unitn.it and velgias@unitn.it

The delivery time is considered the last time the file was modified. Modified files after the deadline will be simply ignored.