CS 564, Fall 2013: SQL Assignment

Due Date: Nov 19, 2013 by 2:00PM. No late days.
Project Grade Weight: 6% of the total grade

[This is an individual assignment, do not work in groups!]

Introduction

In this assignment you will use SQLite3 to build a simple database application. You can find documentation on SQLite3 at [http://www.sqlite.org/](http://www.sqlite.org/). SQLite3 is not as functional as PostgreSQL or mySQL or the commercial relational DBMSs, but it is much easier to use and program. That is why there are rumored to be many more installations of SQLite than PostgreSQL or any other traditional DBMS. Also, you can easily install SQLite3 on your own machine, but make sure your final code runs on SQLite3 installed on the CS mumble machines.

For this assignment, we will use the US Census Population Estimate dataset, which lives on the web at: [http://www.census.gov/popest/data/index.html](http://www.census.gov/popest/data/index.html). We have downloaded and cleaned this dataset, and you can download the dataset from: [http://www.cs.wisc.edu/~jignesh/cs564/projects/sql/Pop_Estimate.tar.gz](http://www.cs.wisc.edu/~jignesh/cs564/projects/sql/Pop_Estimate.tar.gz)

<table>
<thead>
<tr>
<th>File name</th>
<th>Query</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query1.txt</td>
<td>Find the states whose 2011 housing estimate (<strong>HUEST_2011</strong>) is greater than the national average. &lt;br&gt;Print the (state name) as the output.</td>
</tr>
</tbody>
</table>
Query2.txt | For each division, print the name of the division 1~9 (DIVISION_Name) and the number of states in the division.  
Print the (division name, count) as the output.
---|---
Query3.txt | Find the states for which the ratio of the “4/1/2010 resident total population estimates base”(ESTIMATESBASE2010)/“April 1, 2010 housing unit estimates base”(HUESTBASE2010) is greater the nationwide ratio. The ratios should be compared in floating point number, not in integer.  
Print the (state name) as the output.
---|---
Query4.txt | Display the population (CENSUS2010POP) categorized by race and sex in descending order of population.  
Print the (Race_Desc, Sex_Desc, population) as the output.
---|---
Query5.txt | For each state, find the name of the county that has the greatest difference between the birth rate and the death rate in 2012. Display the name of the state and the county in lexicographic order of “Region_Desc”. For states with the same region, sort them in decreasing order of the 2011 population estimate (POPESTIMATE2011) of the county.  
Print the (state name, county) as the output.
---|---
Query6.txt | For each states, display the population (POPESTIMATE2011) for the age group 20~65 (including 20 and 65).  
Print the (state name, population) as the output.
---|---
Query7.txt | For all age groups, calculate the absolute difference of 2011 population estimate (POPESTIMATE2011) and 2010 population estimate(POPESTIMATE2010). Also, display the indicator that indicates if the population has increased or decreased, or remained the same. The value of indicator should be one of the following: {increased, same, decreased}.  
Print the (age, difference, Indicator) as output.
---|---

If you can’t write any of the queries above, provide an explanation of the features that is lacking in SQLite3 that prevents you from being able to write that query. Write these explanations in a file called Queries-readme.txt.

Note: To check the query output, we will first load the data using your load program, and then run:

    sqlite3 census.db < Query3.txt

Thus, your text file can have multiple queries, but it must only output your final desired output, and clean up any temporary tables that it creates. You should try to minimize the # of distinct SQL query blocks in each query file – try to write a single query for each of the queries above, whenever possible.

Copy your query results to a file called results.txt.

**Submission Instructions**
Create only the following files for this assignment: load.h, load.cpp, Query1.txt, Query2.txt, Query3.txt, Query4.txt, Query5.txt, Query6.txt, Query7.txt results.txt, Queries-readme.txt. Please use comments in all your code (including in your SQL file).
To hand in your work please go to learn@uw and select the dropbox folder Assignment 4 (SQL) to upload your files by the deadline.