

Homework #5

due Thursday, November 18 , 2010 at 2 pm

Database Systems, CSCI-4380-01

Each student must work on this homework alone.

1 Homework Description

In this homework, you will be querying the database from Homeworks #2, #3 and #4, i.e. Kip's political campaign. The schema is the same one from previous homeworks. Recall that the database allows users to specify their interest in issues, create and participate in polls, events and pose questions to the candidate. Write the following queries using SQL. The bonus queries are worth 10% more than the regular queries. Answer only one of a query or its bonus version (Q_i or $Q_{i.1}$).

Write the following programs using pl/pgsql:

Question 1. Write a function that takes as input a user id, computes and returns the status of the user based on the following formula:

- If the user's current status is shouting, grumbling or possibleAbuse, then the function returns the current status. Otherwise, the remainder of the logic is executed.
- Users get 1 point for each event they have created and for each group of 20 people who RSVP'ed for their events.
- If this number is less than 5, then the function should return the status "offline".
- If the number above is greater than 5, then compute the number of friends they have (A) and the total number of RSVP's their friends' events got (B). Let $X = B/(A*20)$. If X is less than 10, then return the status "online". If X is greater than 10, then return the status "campaigning".

Question 2. Use the function from Question 1 to update the status of all users who have had an activity in the last 6 months (i.e. created an event, friended someone, created a poll, created a question, voted on a poll or question).

Question 3. Suppose you are given a relation with 500 million tuples. Suppose you want to create an index on attributes (X,Y) where each index node may contain at most 500 entries (key value, pointer pairs).

- How many nodes total are in this B-tree if all the nodes (except for the root) contain about 250 entries?
- What is the size of the smallest B-tree (in terms of the total number of nodes) that this relation can fit in?
- What is the size of the largest B-tree (in terms of the total number of nodes) that this relation can occupy?

Question 4. Suppose you are given a B-tree index for attributes (A,B,C) of R. Each node of the B-tree contains 1,000 entries (except for the root) and the B-tree has height 3 with a total of 5,000 nodes at the leaf level. What is the cost of answering the following queries by using this index? For each query, list (A) total number of index nodes that need to be scanned, (B) total number of tuples of R that need to be read from R.

- $A = 5$ and $B = 5$ and $C = 5$
- $A = 5$ and $5 \leq B \leq 10$ and $C = 5$
- $A = 5$ and $B = 5$ and $5 \geq C \leq 10$
- $5 \leq A \leq 10$ and $B = 5$ and $C = 5$
- $A = 5$ and $5 \leq D \leq 10$
- $5 \leq A \leq 10$ and $5 \leq D \leq 10$

The following tells you how many tuples are expected for each condition ($N(C)$ is the number of tuples expected for condition C):

$$\begin{aligned} N(A = 5) &= 10,000 \\ N(B = 5) &= 1,000 \\ N(C = 5) &= 5,000 \\ N(5 \leq A \leq 10) &= 20,000 \\ N(5 \leq B \leq 10) &= 5,000 \\ N(5 \leq C \leq 10) &= 10,000 \\ N(5 \leq D \leq 10) &= 50 \end{aligned}$$

For any condition $C1$ and $C2$, use $N(C1) * N(C2) / 10^4$.

2 Deliverables

Turn in a single text file (.sql) containing the two functions as in the previous homeworks. In addition, return a second text file containing all the answers including the text for your answers to Questions 1 and 2.