Database Systems, CSCI 4380-01
Homework # 8
Due Friday December 11, 2020 at 11:59:59 PM

Homework Statement. This homework is worth 7% of your total grade. If you choose to skip it, Final Exam will be worth 7% more.

Question 1. In this question, you are asked the reverse of previous homework. Improve the expected run time of **two queries** from either Homework #4 or Homework #5 with respect to the `mediumstreaming` database.

To facilitate this, I have created a new database for each student with name `db2_<username>` and created a copy of mediumstreaming database in this personal space. Please do not drop any tables as in Homework #6.

You can improve the run time in one of two ways:

- Rewrite the query solution such that it is different than your previous solution and my own solution, and has lower query cost. To do this, in your solution, list your initial query (from your submission), my solution and your new query. For each query, list the first few lines of the query plans.

- An easier method is to create one or two indices and use my solution or your own solution. To document the improvement, list (a) the query, (b) index creation commands, (c) a few lines of the query plan before you create the index and (d) full query plan after you create the index. It is important that the full query plan in (d) shows that your index is being used.

In either case, the query plans after your changes should be lower than before for the second cost value (the time to get all the answers). However, it is not required that the reduction in cost is large. Even smaller improvements will be accepted.

When creating indices, consider a few simple rules of thumb that you may have learn from the previous homework:

- Relations with lots of tuples and span many disk pages are likely to benefit more from indices.

- Conditions that are more selective are likely to be more useful. In this include any join conditions as well.

- In case very selective conditions don't exist, you can still target index only scans. In this case, ordering of attributes is important.

When I tested, I found small improvements with indices in almost all queries, but a large improvement in only one query. Homework #4 queries lend themselves to better improvements.

Please document this for two queries!

Question 2. You are given the following schedules. For each schedule, (a) check if it is serializable by drawing the conflict graph, and (b) discuss if it is possible to obtain this schedule using Two Phase Locking.

`S1: r1(x) r2(z) r1(y) w2(w) w2(z) r3(z) w3(x) r1(w) w1(y) w3(z)`

S2: `r1(x) r1(y) r2(z) r3(z) r3(x) w3(x) w2(w) w2(z) r1(w) w1(y)`

Question 3. Suppose you are using REDO/UNDO recovery, and the following are the contents of the log and the disk after crash.

(a) Which log entries should be redone, which should be undone and in which order?

(b) Based on this information, can you conclude if FORCE or NO FORCE is used? Discuss.

(c) Based on this information, can you conclude if STEAL or NO STEAL is used? Discuss.

LOG:

LSN	Entry
100	T1 update P2
101	T2 update P1
102	T2 commit
103	T3 update P1
104	T3 update P3
105	T1 update P4
106	T3 commit

Data pages:

Data Page	LSN of Last recorded log entry
P1	101
P2	100
P3	104
P4	-

SUBMISSION INSTRUCTIONS.

Submit a single text or PDF file that documents your answers to the homework.