SOFTENG 351 S1 C : Assignment 4

Due Date: Wednesday 10 June 2020 at 4:59pm
50 marks in total = 10% of the finalgrade

In this assignment, we use $V(rel, AttName)$ to denote the number of distinct values of attribute AttName in relation rel.

1. Sort relation Student with StudentIDs

$$7, 5, 11, 23, 15, 41, 35, 10, 9.$$

- Assume that the memory can hold up to $m = 3$ blocks of data, and
- Each block can hold at most 1 record of the relation.

Show the intial runs and passes of the sorting. [4 marks]

2. Below are the vital statistics for four relations, w, x, y, and z.

- w(A,B): $n_w = 1000$, $V(w, A) = 20$, $V(w, B) = 60$
- x(B,C): $n_x = 2000$, $V(x, B) = 50$, $V(x, C) = 100$
- y(C,F,D): $n_y = 3000$, $V(y, C) = 50$, $V(y, D) = 50$
- z(D,E): $n_z = 4000$, $V(z, D) = 40$, $V(w, E) = 100$

Estimate the sizes of the results of the following expressions:

(a) $w \bowtie x \bowtie y \bowtie z$
(b) $\sigma_{A=10}(w)$
(c) $\sigma_{C=20}(y)$
(d) $\sigma_{C=20}(y) \bowtie z$
(e) $w \bowtie y$
(f) $\sigma_{D>10}(z)$
(g) $\sigma_{A=1 \text{ and } b=2}(w)$.

[7 marks]

3. Consider the same settings of relations in Problem 2. To compute expression $w \bowtie x \bowtie y \bowtie z$,

(a) List all the equivalent (relational algebra) expressions of $w \bowtie x \bowtie y \bowtie z$ that satisfy the following constraint:
- The expression does not join two relations that form a Cartesian product e.g., $((w \bowtie x) \times z) \bowtie y$ will not be considered since it has a Cartesian product (disjoint attributes of the two relations to be joined).

[2 marks]

(b) Estimate the sizes for expression $(w \bowtie x) \bowtie (y \bowtie z)$ including its the intermediate results ($w \bowtie x$, $y \bowtie z$), and final results ($(w \bowtie x) \bowtie (y \bowtie z)$) given that each block holds 20 tuples from any relations. [3 marks]

(c) Estimate the I/O cost of $(w \bowtie x) \bowtie (y \bowtie z)$ under the following settings:
- Each relation (w, x, y, z) occupies 100 blocks,
- The memory has $m = 101$ pages,
- $w \bowtie x$ uses one-pass join,
- $y \bowtie z$ uses one-pass join, and
- $(w \bowtie x) \bowtie (y \bowtie z)$ uses sort (multi-way merge sort) merge join.

[4 marks]

4. Consider a relation $r(A, B, C, D)$ that has a clustering index on A and non-clustering indexes on each of the other attributes. The parameters are $b_r = 1000$, $n_r = 5000$, $V(r, A) = 20, V(r, B) = 1000$, $V(r, C) = 5000$, and $V(r, D) = 500$. Give the best query plan, specifically, index-scan or table-scan followed by a filter step, and the corresponding I/O cost (under the default uniform distribution assumption), for the following selections (please only count the I/Os for retrieving the underlying tuples):

(a) $\sigma_{A=1 \text{ AND } B=2 \text{ AND } D=3} r$,

(b) $\sigma_{A=1 \text{ AND } B=2 \text{ AND } C\geq 3} r$,

(c) $\sigma_{A=1 \text{ AND } B\leq 2 \text{ AND } C\geq 3} r$.

[6 marks]

5. Conflicts in scheduling. Draw the precedence graph of the schedule $r_2(X), r_1(Y), w_2(X), r_2(Y), r_3(X), w_1(Y), w_2(Y)$ of three transactions T_1, T_2 and T_3 and then show if the schedule is conflict-serializable. [3 marks]

6. For each of the transactions described below, suppose that we insert one lock and one unlock action for each database element that is accessed.

   T_1: $r_1(A), w_1(B)$

   T_2: $r_2(A), w_2(A), w_2(B)$.

   Tell how many orders of the lock, unlock, read, and write actions are

   (a) Consistent and two-phase locked.

   (b) Consistent, but not two-phase locked.

   (c) Inconsistent, but two-phase locked.

   (d) Neither consistent nor two-phase locked.

   [8 marks]

7. During execution, a transaction passes through several states until it finally terminates. List all possible sequences of states through which a transaction may pass. Explain why each transition state may occur. [3 marks]

8. Explain the concept of deadlocks and 2 possible solutions to deadlocks. [3 marks]

9. The next figure shows the log corresponding to a particular schedule at the point of a system crash for four transactions T_1, T_2, T_3, and T_4. Suppose that we use immediate update protocol with checkpointing. Describe the recovery process from the system crash. Specify which transactions are rolled back, which operations in the log are redone and which are undone and whether any cascading rollback takes places. [7 marks]