Data Administration in Information Systems

Exam – July 10, 2023

1. A database table R(A, B, C) has no indexes and is stored by increasing order of the numerical values in column C.
   a) How many seeks and transfers are required to find the value of B that corresponds to the smallest value of C? Justify.
   b) How many seeks and transfers are required to find the value of B that corresponds to the largest value of C? Justify.

2. A database table R(A, B, C) has a clustered B+ tree index on primary key A.
   a) If the B+ tree index has height *h*, how many block accesses are needed to locate a given value of A? Justify.
   b) If the B+ tree index has height *h*, how many block accesses are needed to locate every record with A ≥ x, where x a given value of A? Justify.

3. A database table R(A, B, C) has a non-clustered hash index on C, without overflow buckets.
   a) How many block accesses are needed to find the value of B that corresponds to a given value of C? Justify.
   b) How many block accesses are needed to locate every record with C ≥ v, where v a given value of C? Justify.

4. When sorting a large table, we may need multiple merge passes if the number of blocks b is larger than memory M.
   a) The number of merge passes is given by $\lceil \log_{M-1}(b/M) \rceil$. Why is there a logarithm in this formula? Explain.
   b) If $b < M$, what does this mean, and what are the implications, also in terms of merge passes? Explain.

5. The order of operations in an execution plan might have an effect on performance.
   a) In general, row selections should be performed before table joins. Give two reasons for that.
   b) Another recommendation is to perform column selections (i.e. projections) before table joins. However, this is not always possible to do. Why? Justify.

Answer the following questions in a separate sheet of paper. _____

6. At the end of a working day, a bank launches a SQL script to execute a series of money transfers between accounts.
   a) Describe an advantage and a disadvantage of using a multiple granularity locking scheme in this scenario.
   b) Describe an advantage and a disadvantage of using a multi-version timestamp protocol in this scenario.

7. A database system crashed. Upon reboot, it tries to recover. However, it crashes again during the recovery process.
   a) Give a reason why the recovery process itself could cause the database system to crash again. Explain.
   b) Between the two crashes, the system writes a CLR. How does this change the next recovery attempt? Explain.

8. In this course, we have worked with materialized views.
   a) Explain how materialized views can be used to improve the performance of SQL queries, when aggregates are involved.
   b) Explain why a query optimizer might decide not to use a materialized view, even though the materialized view, among other results, also provides the results desired by a query.

9. The concepts of composite index and of covering index are slightly different.
   a) Is it possible to have a composite index that is not a covering index for the query? Justify.
   b) Is it possible to have a covering index for a query that is not a composite index? Justify.

10. A critical query is one that takes a long time to run, when compared to other queries.
    a) How can you find critical queries with the tools that we used in the labs? Explain.
    b) With the tools that we used in the labs, where can you get ideas about how to improve the performance of a critical query? Explain.