

**Data Administration in Information Systems**Exam – April 24, 2023

1. Consider reading a database table that is stored in a hard disk drive (HDD).
 - a) If we increase the rotation speed of the HDD, how does this affect the seek time and the transfer time? Justify.
 - b) If we run a disk defragmentation tool to defragment the disk, can this help to read the table faster? Explain.
2. Consider a table $R(\underline{A}, B, C)$ with a B+ tree clustered index on primary key A, and a B+ tree non-clustered index on C.
 - a) When looking for records with a specific value of A, how many disk block accesses will be required? Justify.
 - b) When looking for records with a specific value of C, how many disk block accesses will be required? Justify.
3. We want to join two tables $R1(\underline{A}, B)$ and $R2(\underline{B}, C)$ by sorting R1 on B and then merge-joining it with R2.
 - a) If table R1 is twice the size of the available memory, how will sorting be done? Explain the sorting steps.
 - b) Instead of sorting R1 and merge-joining it with R2, would it be better to hash-join both tables? Discuss.
4. An execution plan can change significantly if we make minor changes to a query.
 - a) If we change the selection conditions on a table, how can this affect the execution plan? Explain.
 - b) If we create a materialized view involving some of the tables in the query, how can this affect the execution plan? Explain.
5. When running multiple transactions concurrently, conflicts may appear between read and write operations.
 - a) How can these conflicts result in a schedule that is non-serializable? Explain with an example.
 - b) If we allow a non-serializable schedule to proceed anyway, what kind of problems can occur? Explain.

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

6. Many database systems use a locking system and locking protocols to isolate transactions.
 - a) What are the benefits of using a tree- or graph-based protocol to lock database objects? Describe.
 - b) In the multiple granularity locking scheme, there are several lock variants (S, X, IS, IX, etc.). What is the purpose of having all these variants? Why isn't S and X enough? Justify.
7. In a database recovery algorithm, there are log records and compensation log records (CLRs).
 - a) Why do log records contain the old value for an object, but CLRs do not? Justify.
 - b) Does the dirty page table include dirty pages for all transactions or only for the transactions that did not commit? Why? Justify.
8. In database tuning, we can optimize performance by schema tuning or query tuning.
 - a) If a table is vertically partitioned, will the total number of columns in those partitions be less than, equal to, or greater than the number of columns in the original table before partitioning? Why? Justify.
 - b) In an execution plan, what do we gain by eliminating DISTINCT in the query? Explain.
9. We can also optimize performance by index tuning and log tuning.
 - a) Consider three types of indexes: clustered, non-clustered, and covering. Rank them in order of performance (from faster to slower) in answering a multi-point query. Justify your ranking.
 - b) If we put data and log on separate disks, this not only improves resilience but also performance. Why? Justify.
10. Finally, we can improve performance by hardware and operating system (OS) tuning.
 - a) If some transactions are more important than others, what kind of problem can occur if we assign them different priorities? Justify.
 - b) How does having more memory (RAM) help improve the hit ratio? Explain.