

M.Sc. Examination, 2019
Semester-IV
Computer Science
Course : MCSC-42
(Advanced DMBS)

Time : 3 Hours

Full Marks : 40

Questions are of value as indicated in the margin

Answer Question No. **1** and **any three** from the rest.

1. a) Distinguish between *ordered indices* and *hash indices*.
b) Estimate the cost of the disjunctive selection of the form : $\sigma_{\theta_1 \vee \theta_2 \vee \dots \vee \theta_n(r)}$.
c) How *strict-2PL* protocol is advantageous over *2PL*?
d) What is stable storage? 2×4=8
2. a) Answer with justification whether the following schedule of three transactions (T1, T2, and T3) is *conflict serializable*, *view serializable*, *recoverable* and *cascadeless*. Further answer with justification whether it may be generated with *Timestamp without the Thomas' Write Rule* and *Timestamp with the Thomas' Write Rule*.
T1: Read(X), Read(Y), Write(Y), T2: Read(X), Write(X), Read(Y), T3: Read(X), T1: Read(Z), T2: Read(Z), Write(Z), T3: Read(Y), T1: Write(Z), T1: Commit, T3: Read(Z), T2: Commit, T3: Commit.
b) How to determine *serializability* order from the precedence graph of transactions for conflict serializable schedule? 6+2=8
3. a) Does *timestamp-ordering* protocol ensure freedom from deadlock and starvation? Justify.
b) Show that the *2PL* protocol ensures *conflict serializability*.
c) Suppose there is a crash with the following log:
<T₀ start> T₀, B, 2000, 2050> <T₁ start> <checkpoint {T₀,T₁}> <T₁, C, 700, 600> T₁ commit> <T₂ start> <T₂, A, 500, 400> <T₀, B, 2000> Explain what would happen during recovery. 3+2+3=8
4. a) Briefly explain the steps involved in query processing.
b) Briefly explain the way of implementing the disjunctive selection of the form: $\sigma_{\theta_1 \vee \theta_2 \vee \dots \vee \theta_n(r)}$.
c) Suppose you have a file with 5,000 pages and three buffer pages. Answer the following questions using external merge sort algorithm. How many runs will be produced in the first pass? How many passes will it take to sort the file completely? What is the total I/O cost of sorting the file? 3+2+3=8
5. a) Briefly explain cost-based optimization and heuristic query optimization.
b) Consider the two relations: *depositor* (*customer-name*, *account-number*) and *customer* (*customer-name*, *customer-street*, *customer-city*). Suppose that customer has a primary B+ tree index on the join attribute *customer-name*, which contains 20 entries on an average in each index node. Number of records of *depositor* and *customer* are 5000 and 10,000, respectively. Number of blocks of *depositor* and *customer* are 200 and 500, respectively. Compute the cost of *depositor* ⋈ *customer*, with *depositor* as the outer relation using indexed nested-loop join.

(2)

- c) Consider the relations $r_1(A, B, C)$, $r_2(C, D, E)$, and $r_3(E, F)$, with primary keys A, C, and E, respectively. Assume that r_1 has 1000 tuples, r_2 has 2000 tuples, and r_3 has 750 tuples, Estimate the size of $r_1 \bowtie r_2 \bowtie r_3$ and give an efficient strategy for computing the join. $3+2+3=8$
6. a) What is the major problem of 2PC protocol in distributed transactions?
b) Briefly explain *semijoin* strategy in distributed query processing.
c) Briefly explain *Fragment-and-Replicate Join* in parallel database system. $2+2+4=8$
7. Write short notes on **any two** of the following : $4 \times 2 = 8$
- a) Data fragmentation
 - b) Equivalence rules for transformation of relational expressions
 - c) Fuzzy checkpoints in log-based recovery
 - d) OLTP vs OLAP
-