2018

COMPUTER SCIENCE

Paper : CSM-102

(Advances in Database Management System)

Full Marks : 70

The figures in the margin indicate full marks.

Candidates are required to give their answers in their own words as far as practicable.

Answer ***question nos.*** **1, 2** and ***any four*** questions from the rest.

1. Answer ***any five*** questions from the following : 2×5

(a) State the reason behind selecting "Number of Secondary memory access" as a metric for query optimization.

(b) Distinguish between Shared lock and Exclusive lock.

(c) What do you mean by un-recoverable schedule?

(d) Define conflict serializability.

(e) Why is cascading in a schedule not desirable?

(f) Why is the Conjunctive normal form preferable for selection operation?

(g) State the advantage of using rigorous two-phase locking concurrency control compared to the classical two-phase approach.

(h) What do you mean by data cleaning in context of data mining?

2. Answer ***any five*** questions from the following : 4×5

(a) How does multilevel indexing improve the efficiency of searching an index?

(b) State the difference between pipelining an materialization through an example.

(c) Discuss with an example, how redundancy is eliminated in snowflake schema.

(d) State the utility of dynamic locking in concurrency control protocol.

(e) Suppose that a B+-tree index on branch-city is available on relation branch, and that no other index is available. What would be the best way to handle the following selections that involve negation?

(i) $\sigma\neg$ (branch-city<"Brooklyn" V assets<5000)(branch)

(f) Derive the cost of joining two relations R and S using block nested loop join.

(g) Consider a schema Book (id, author_name, publisher, price)
having functional dependencies, id –> author and publisher –> price.
If Book schema is decomposed into A (id, author) and B(publisher, price), Comment on dependency preserving and Lossless join decomposition on
decomposition of Book schema.
If original Book schema also had another functional dependency id –> publisher, price.
then how does this FD affects your previous comment.

(h) Compute the closure of the following set F of functional dependencies for relation schema R = (A, B, C, D, E) with functional dependency f = {A → BC, CD → E, B → D, E → A}. List the candidate keys for R.

3. Describe the validation-based concurrency-control scheme. Also show that by choosing Validation (*Ti*), rather than Start (*Ti*), as the timestamp of transaction *Ti*, we can expect better response time provided that conflict rates among transactions are indeed low. 6+4

4. (a) Consider the following schema.
Hotel (hotelNo, hotelName, city)
Room (roomNo, hotelNo, type, price)
Booking (hotelNo, guestNo, dateFrom, dateTo, roomNo)
Guest (guestNo, guestName, guestAddress)
Draw a relational algebra tree for the following query and use the heuristic rules to transform the queries into a more efficient form. Show each step of transformation.

SELECT g.guestNo, g.guestName
FROM Room r, Hotel h, Booking b, Guest g
WHERE h.hotelNo = b.hotelNo **AND** g.guestNo = b.guestNo **AND** h.hotelNo = r.hotelNo **AND** h.hotelName = 'Grosvenor Hotel' **AND** dateFrom > '1-jan-08' **AND** dateTo < '31-Dec-08';

(b) "All interleaving manners of transactions within a concurrent schedule should not be accepted" – Comment and Justify your answer through an example. 5+5

5. (a) Why is the external sorting used for DBMS application? State the algorithm.

(b) Suppose that you have a file with 10,000 pages and that you have three buffer pages.
Answer the following questions for each of these scenarios, assuming that our most general external sorting algorithm is used :
A file with 2,000,000 pages and 17 available buffer pages.
(i) How many runs will you produce in the first pass?
(ii) How many passes will it take to sort the file completely?
(iii) What is the total I/O cost of sorting the file?
(iv) How many buffer pages do you need to sort the file completely in just two passes? (1+5)+4

6. (a) What is the need of maintaining multiple granularities within locking? Describe the utility of implicit and explicit mode locking through an example.

(b) State and justify the Thomas Write rule. (2+4)+4

7. (a) "Extendible Hasting adjusts gracefully to inserts and deletes"— Explain through an example.

(b) How does the recovery manager ensure atomicity of transactions? State the Write Ahead Log (WAL) protocol in this context. 6+(2+2)

8. (a) Consider the following graph-based locking protocol that allows only exclusive lock modes, and that operates on data graphs that are in the form of a rooted directed acyclic graph.

(i) A transaction can lock any vertex first.

(ii) To lock any other vertex, the transaction must have visited all the parents of that vertex, and must be holding a lock on one of the parents of the vertex.

Show that the protocol ensures serializability.

(b) Consider the relations r1(A,B,C), r2(C,D,E) and r3(E,F), with primary keys A, C and E, respectively. Assume that r1 has 1000 tuples, r2 has 1500 tuples, and r3 has 750 tuples. Estimate the size of r1 × r2 × r3, and give an efficient strategy for computing the join. 5+5

9. What is Dimensional Modelling? Discuss the utility of Fact table and Dimension table through an example. What is the 'Time variant' Property of DW? 3+5+2