Chennai Mathematical Institute

RDBMS AND SQL   90 MINUTES + 20 MINUTES FOR UPLOADING. 55 MARKS.

Instructions

- No negative marks.
- Submit your answer sheet as a single pdf file on moodle. For any reason if you cannot upload to moodle, email it to vvtesh.cmi@gmail.com.
- If you make multiple submissions, the last submission will be considered for grading.
- Late submissions: Two marks for first 5 minutes. 5 marks thereafter for every 5 minutes.

Each question carries 5 marks. Total = 11 x 5 = 55 Marks.

Question 1. We decided to split the relation R(A,B,C,D,E) into R1(A,D,E) and R2(A,B,C,D). Following functional dependencies exist: $\{A \rightarrow E, C \rightarrow BD, E \rightarrow AB, AB \rightarrow C\}$. Is the decomposition lossless? Why or why not? Explain in detail.

Question 2. Is the relation, R(A,B,C,D,E,F,G,H,I) in BCNF? The functional dependencies are given by F = $\{BC \rightarrow GHI, AD \rightarrow E, A \rightarrow H, E \rightarrow BCF, G \rightarrow H\}$. If not, decompose R into BCNF.

Question 3. a) Let A1, A2 and A3 be sets of attributes on relation r such that $A1 \subset A2 \subset A3$. F1 and F2 are boolean expressions over the attributes of r. Which of the following will produce the same output as $\pi_{A1}(\pi_{A2}(\pi_{A3}(\sigma_{F1}(\sigma_{F2}(r)))))$?

Choose all the correct answers:

(1) $\pi_{A1}(\sigma_{F1 \wedge F2}(r))$
(2) $\pi_{A1}(\sigma_{F1 \vee F2}(r))$
(3) $\pi_{A1}(\pi_{A2}(\sigma_{F1 \wedge F2}(r)))$
(4) $\pi_{A1}(\pi_{A2}(\sigma_{F1 \vee F2}(r)))$

b) Consider the relations r1(P, Q, R) and r2(R, S, T) with primary keys P and R respectively. The relation r1 contains 1500 tuples and r2 contains 1000 tuples. The maximum size of the join r1 ⋈ r2 is _____.

Question 4. Consider a file of 32768 records. Each record is 32 bytes long and its key field is of size 6 bytes. The file is ordered on a non-key field, and the file organization is unspanned. The file is stored in a file system with block size 2048 bytes, and the size of a block pointer is 10 bytes. If the secondary index is built on the key field of the file, and a multi-level index scheme is used to store the secondary index, the number of first-level and second-level blocks in the multi-level index are respectively _____.

Question 5.   Assume that search key occupies 12 bytes and a block pointer occupies 6 bytes. If the block size is 528 bytes, what is the maximum order of the B^+ tree?

Question 6.   Assume that an instance of relation R(A,B,C,D) is as given below.

A	B	C	D
3	11	d	4
2	11	b	6
1	11	c	6
4	11	a	7
5	11	e	8
6	11	f	9
7	11	g	9
8	11	g	10
9	11	g	11
10	11	h	12

The table is stored using sequential file organization sorted on the attribute A. For a fan-out factor of 3, draw the B^+ tree index over attribute C.

Question 7.   Consider the relation R(A,B,C,D,E). The set of functional dependencies are $F = \{A \rightarrow BC, CD \rightarrow E, B \rightarrow D, E \rightarrow A\}$. How many candidate keys exist for R? List the candidate keys.

Question 8.   Consider the following two schedules $S1$ and $S2$:
$S1:\ r1(A)\ r2(A)\ w1(A)\ r2(B)$
$S2:\ r1(A)\ w1(A)\ r2(A)\ w2(B)\ w2(B)$
Note that in this notation, $r1(A)$ means that transaction T1 reads the object A. Similarly, $w2(B)$ means that T2 writes the object B.
Which of these are conflict serializable? Justify your answer using precedence graphs.

Question 9.   CMI is tasked with setting up the database for the popular OTT platform hotstar. CMI is seeking your help for the same. For a part of their OTT system,

(1) Draw ER diagram with at least two entities. Show at least one weak entity.
(2) Convert that ER diagram to tables.

You may consider any use case such as storing the movie schedule, capturing user comments, etc.

Question 10.   Consider the instructor relation discussed in the class. Write a relational algebra expression to find the instructor with the least salary in the Biology department. Provide the corresponding TRC and DRC expressions. You may use extended and aggregate operations.

(Optional. Bonus 1 mark. Applies only if total is below 55.) Write the corresponding SQL query.

Question 11. We would like to capture CMI student and instructor data in a set of tables. You may make reasonable assumptions on the data items (i.e., columns of the table). Describe a relational schema which is in 3NF but not in 4NF for this purpose. Explain the reasons why you believe the schema you describe is in 3NF but not in 4NF.