Total marks 15   **Maximum Time: 30 Minutes**

Name:_____   ID Number_____

Answers for the question must be written in the box below. Answers written anywhere won't be checked. Overwritten answers won't be considered for evaluation. (Don't ask Invigilator to sign.)

Ans 1	Ans 2	Ans 3	Ans 4	Ans 5	Ans 6	Ans 7	Ans 8

Only one answer is correct. Write the most appropriate answer. 0 Marks for No Attempt.

For Question 1-5: + 2 Marks for correct answer each, [-0.5 Marks for wrong answer for each question]

Ques1.   A B+ -tree index is to be built on the Name attribute of the relation STUDENT. Assume that all student names are of length 8 bytes, disk block are size 512 bytes, and index pointers are of size 4 bytes. Given this scenario, what would be the best choice of the degree (i.e. the number of pointers per node) of the B+ -tree?

a) 16
b) 42
c) 43
d) 45

Ques2.   Relation R is decomposed using a set of functional dependencies, F and relation S is decomposed using another set of functional dependencies G. One decomposition is definitely BCNF, the other is definitely 3NF, but it is not known which is which. To make a guaranteed identification, which one of the following tests should be used on the decompositions? (Assume that the closures of F and G are available).

a) Dependency-preservation
b) Lossless-join
c) BCNF definition
d) 3NF definition

Ques3.   Consider a schema R(A,B,C,D) and functional dependencies A->B and C->D. Then the decomposition of R into R1(AB) and R2(CD) is

a) dependency preserving and lossless join
b) lossless join but not dependency preserving
c) dependency preserving but not lossless join
d) not dependency preserving and not lossless join

Ques4.   Which one of the following statements is NOT correct about the B+ tree data structure used for creating an index of a relational database table?

a) B+ Tree is a height-balanced tree
b) Non-leaf nodes have pointers to data records
c) Key values in each node are kept in sorted order
d) Each leaf node has a pointer to the next leaf node

Ques5.   Let the set of functional dependencies F = {QR →S, R →P, S →Q} hold on a relation schema X = (PQRS). X is not in BCNF. Suppose X is decomposed into two schemas Y and Z, where Y = (PR) and Z = (QRS).

Consider the two statements given below.
   I.   Both Y and Z are in BCNF
   II.   Decomposition of X into Y and Z is dependency preserving and lossless.
Which of the above statements is correct? (choose the most relevant option)

a) Both I and II
b) I only
c) II only
d) Neither I nor II

Questions No 6-10: +1 Marks each [-0.25 Marks for wrong answer]

For Questions 6-8: Which basic file organization (heap, sorted, or hash) that is best for a large file where the most frequent operations are as follows
Ques6.   Search for records based on a range of field values.
a) Heap
b) Sorted
c) Hash

Ques7.   Perform inserts and scans where the order of records does not matter.
a) Heap
b) Sorted
c) Hash

Ques8.   Search for a record based on a particular field value.
a) Heap
b) Sorted
c) Hash

Ques9.   With respect to Two-Phase Locking. Select the FALSE statement.
a) Locks may be made at any point in the transaction schedule.
b) A transaction may start to release locks at any time.
c) The phases are "lock-acquisition" and "lock-release".
d) Locks are released on an ABORT.
e) Two-Phase Locking requires key pairs.

Ques10.   What is the main advantage of immediate update?
a) Changes can be stored on the disk before a commit.
b) All transaction data is held in memory making aborts faster.
c) The locking strategy used can be timestamp based.
d) Transactions that are long are given priority over short transactions.
e) Transactions that are short are given priority over long transactions.

Ques1. Consider the following schedules

S1: r7 (X); r8 (Z); r7 (Z); r8 (X); r9 (Y); w7 (X); c7; w9 (Y); c9; r8 (Y); w8 (Z); w8 (Y); c8;

S5: r1(X); w2(Q); w1(Q); w3(Q); c3;c1;c2;

S9: r6 (X); r8 (Z); r3 (X); r6 (Z); r8 (Y); r3 (Y); w6 (X); c6; w8 (Z); w3 (Y); w8 (Y); c3; c8;

a) Draw the graph and determine if these schedules are conflict serializable and view serializable. Write their order of serializability.

b) Are these schedules recoverable and cascadeless? Give proper justification. Answers without correct justification won't fetch any marks. **(12+ 3= 15 Marks)**

Ques2. Consider the following schedules

S1: r1 (X); r2 (Z); r3 (X); r1 (Z); r2 (Y); r3 (Y); w1 (X); c1; w2 (Z); w3 (Y); w2 (Y); c3; c2;

a) Implement the rigorous 2PL (2 phase locking) protocol in the above schedule

b) Do rigorous 2PL protocol ensures conflict serializability, view serializability, deadlock freedom and non-recoverable and cascadeless schedules. Justify your stand. **(5+ 5= 10 Marks)**

[start_transaction, T1]	
[start_transaction, T3]	
[read_item, T1, A]	
[read_item, T3, D]	
[write_item, T1, D, 20, 25]	
[start_transaction, T4]	
[read_item, T4, D]	
[Commit, T1]	
[Checkpoint]	
[start_transaction, T2]	
[read_item, T2, D]	
[start_transaction, T5]	
[read_item, T5, D]	
[read_item, T2, B]	
[write_item, T3, D, 25, 35]	
[write_item, T2, B, 12, 18]	
[read_item, T4, A]	
[Commit T3]	
[write_item, T4, D, 35, 15]	
[write_item, T5, C, 30, 40]	
[Commit T2]	
[write_item, T5, D, 15, 25]	
System Crash	

Fig: Ques 3

Time	Operation
T1	r1 (X);
T2	r2 (Z);
T3	r1 (Z);
T4	r3 (X);
T5	r3 (Y);
T6	w1 (X);
T7	w3 (Y);
T8	r2 (Y);
T9	w2 (Z);
T10	w2 (Y);
T11	c1;
T12	c2;
T13	c3;

Fig: Ques 5

Ques3. In the log file mentioned in Fig: Ques3

Suppose that we use the immediate update protocol with checkpointing.

a) In step by step manner elaborate the recovery process after the system crash.

b) Specify which transactions are rolled back, which operations in the log are redone and which (if any) are undone, and whether any cascading rollback takes place? (Ignore the logical errors and cascading)

(6+ 4 = 10 Marks)

Ques4. Construct a B+-tree index for the first 20 letters of the English alphabet (from "A" to "T"), with n = 4 (thus each leaf holds at most three keys, and each non-leaf has at most four children). Your final answer should be best full tree with min required number of levels. **(10 marks)**

Ques5.

a) Rewrite the schedule mentioned in Fig: Ques 5after Implementing the time stamp ordering protocol. Elaborate on the process of your implementation.

b) Whether the timestamp ordering protocol ensures conflict serializability, view serializability, deadlock freedom, cascadeless and non-recoverable schedules. Justify your stand. **(5+5 = 10 Marks)**

Ques6. Answer the following

(a) Suppose that you have a file that is already sorted in key order and you want to construct a dense, clustered B+ tree index on this file using <key, RID> pairs for data entries. A simple way to accomplish this is to create a B+tree, and then sequentially scan the file, inserting an index entry for each record using the normal B+tree insertion routine. What performance and storage utilization problems are there with this approach?

b) Briefly describe a change to the B+tree insertion routine that would solve the problems you identified in part (a).

(5+5=10 Marks)

Ques7. Suppose that a data file consists of the following data records (integers):

36, 44, 20, 5, 13, 17, 14, 18, 10, 30, 31, 35, 7, 11, 47, 34, 38.

a) Using the extendible hashing method starting with 4 buckets, show the final hashing results if the data records are hashed one by one according to the above ordering. Assume that each bucket can hold only 4 integers. (In your answer, please show the hashing index directory, local depth,and global depth.)

b) Using the linear hashing method starting with 4 buckets, show the final hashing results if the data records are hashed one by one according to the above ordering. Assume that each bucket can hold only 4 integers. [Hint : Take lower order bits]   **(5+5=10 marks)**

Ques8. Suppose that we have an ordered file with r = 20,000 STUDENT records of fixed length stored on a disk with block size B=512 bytes. Each record has the following fields: NAME (20 bytes), IDNO (12 bytes), ADDRESS (30 bytes), PHONE (10 bytes), DOB (8bytes), DISCIPLINE (10 bytes) and EMAIL (10 bytes).

a) Calculate the blocking factor and the number of file blocks assuming an unspanned organization. Also find number of block accesses to get data from file.

b) Assume a primary index is created on IDNO and record pointer is P = 8 bytes. Calculate (i) the index blocking factor (ii) the number of first-level index entries and the number of first-level index blocks (iii) number of block accesses to get data using primary index (iv) the number of levels needed if we make it into a multi-level index (v) total number of blocks required by the multi-level index (vi) number of block accesses to get data using multi-level index.   **(l3+7=20 Marks)**

Ques9.

a) Query processing takes time primarily in input/output and computation. The memory has limited size and also has direct implication on performance. List out 5 ways with example through which we can improve the performance of the SQL query.

b) Find the canonical/minimal cover of the following Relation. Show proper steps with the reasoning in each step

R = (A, B, C, G, H, I) F = {A → B, A → C, CG → H, CG → I, B → H}   **(5+5 = 10 Marks)**