CSE 202, DBMS END-SEM EXAM, Winter 2024

Total points 35/48

Marks: 51 **Start Time: 2:30 PM**
Submission/End Time: 4:30 PM

(0) YOU MUST SHOW SUBMISSION PAGE TO THE INVIGILATOR BEFORE STANDING UP FROM YOUR SEAT. OTHERWISE, THE SUBMISSION WILL NOT BE CONSIDERED.

(1) Students should bring their own laptops fully charged and check for connection issues beforehand. No Excuses will be entertained regarding this.
(2) No other devices, like phone, Tablets, smart watches, etc are allowed. If caught, you will be treated as per institute plagiarism policy.
(3) No other applications/tabs should be open including the email tabs, except the google form tab (exam paper). If any tabs/apps found open at any point of time during the exam period, it will be considered cheating case and will be dealt as per institute's plagiarism policy.
(4) All rough work should be done in the blank sheets provided to you. You **MUST** write your name, Roll# and Section# on this sheet.
(5) No clarification are allowed with TAs/TF/instructors for any questions.
(6) Make sure you submit the form on time. We will NOT accept any submission after 4:32PM. No queries will be entertained regarding this.
(7) Once you have submitted the google form please stay seated in your place. Once you get the instruction from the invigilator, you can then stand up from your place and leave the room.

(8) This paper has 51 questions. Each question carries one mark.
(9) Each question may have more than one correct option. Please select all the correct answers from the options provided to get one mark. There is no partial marking for partially correct answers.
(10) IF ANY CORRECT OPTION SUBSUMES OTHER CORRECT OPTION(S), DO NOT SELECT THE SUBSUMED OPTION.

✓ Strict 2PL resolves the dirty read problem in schedule S: R2(X), R2(Y), W2(X), R1(X) as 1/1

- [x] T1 could not get a shared lock on X because T2 would be holding an exclusive lock on X. Thus, T1 would have to wait until T2 was finished. ✓
- [ ] T2 could not get an exclusive lock on X because T1 would be holding an exclusive lock on X. Thus, T1 would have to wait until T2 was finished.
- [ ] T1 could not get an exclusive lock on X because T2 would be holding an exclusive lock on X. Thus, T1 would have to wait until T2 was finished.
- [ ] T2 could not get a shared lock on X because T1 would be holding an exclusive lock on X. Thus, T1 would have to wait until T2 was finished.

✓ Consider the table Course(C-Name, Pre-Req-C-Name), where C-Name is the primary key
and Pre-Req-C-Name is the foreign key referencing to C-Name with on-delete cascade. 1/1

If a tuple (AP, IP) is deleted, which other tuples will be deleted for preserving the referential
integrity constraints?

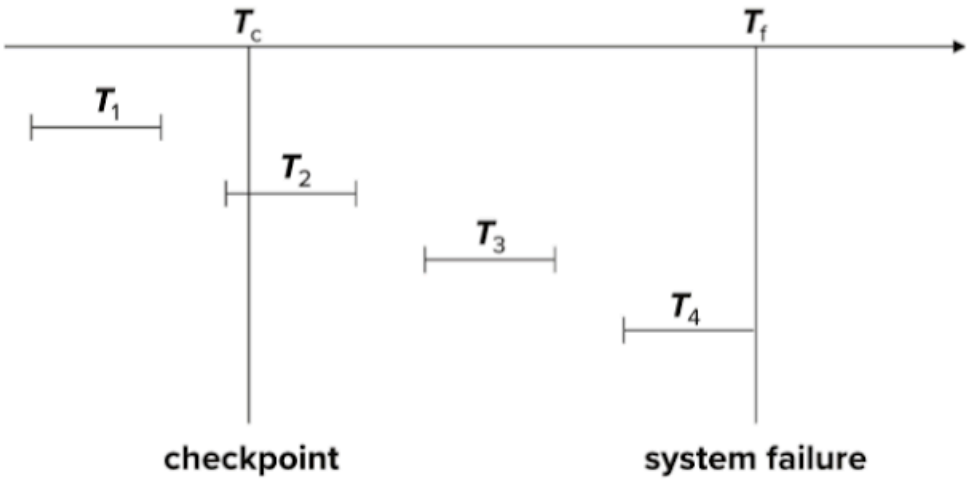
C-Name	Pre-Req-C-Name
DBMS	IP
AP	IP
IP	NULL
ADA	DS
DS	ADA
OS	NS
NS	AP

- [x] ((OS, NS), (NS, AP)} tuples will get deleted. ✓
- [ ] None, only (AP, IP) tuple will be deleted
- [ ] {(DBMS, IP) , (NS, AP)} tuples will be deleted
- [ ] (DBMS, IP) tuple will also be deleted

✓ Consider the below log sequence of four transactions in a banking environment. The first transaction increases the balance from Rs1000 to Rs2000. The second transaction calculate the interest at 10% and increase the balance amount by 10% as interest. The third transaction deducts Rs200 from the balance as the account service fee. The forth transaction deducts Rs1000 paid online. 1/1

<T1 start>
<T1 Balance, 1000, 2000>
<T1 commit>
<T2 start>
<T2 Balance 2000, 2200>
<T2 commit>
<T3 start>
<T3 Balance 2200, 2000>
<T3 Commit>
<T4 start>
<T4 Balance 2000, 1000>

What would be the recovery action?



- [ ] T 1 : Ignore, T 2 : undo, T 3 : undo, T 4 : redo
- [x] T 1 : Ignore, T 2 : redo, T 3 : redo, T 4 : undo ✓
- [ ] T 1 : redo, T 2 : redo, T 3 : redo, T 4 : undo

- [ ] T 1 : redo, T 2 : undo, T 3 : undo, T 4 : redo

✓ Which statement(s) is (are) true? 1/1

- [x] Sort-Merge Join algorithm gives better performance for the point queries. ✓
- [ ] In Index Nested Join, the outer relation is chosen that has an index table.
- [x] The Block Nested Loops Join is more sensitive to the buffer size, i.e. the I/O cost reduces as we increase the buffer size. ✓
- [x] Hash join is sensitive to data skew while sort-merge join is not. ✓

✓ Consider the database with the below parameters: 1/1
block size is 4096 bytes
header size is 100 bytes
record size is 100 bytes
where, a file storing variable length records. What would be the value of the block factor?

- [ ] 32
- [ ] 40
- [ ] 41
- [x] 39 ✓

✓ What would be the number of attributes of the resultant schema R NJ S, where NJ indicates natural join
operation. Here, let the schema of R and S be (A, B, C, D) and (A, C, D, F), respectively. 1/1

- [x] 5 ✓
- [ ] 6
- [ ] 4
- [ ] 8

✓ Which statement is (are) true w.r.t. the B+ tree? 1/1

- [x] It supports both equality and range queries. ✓
- [ ] Leaf nodes are not organized into doubly linked list.
- [x] All paths from the root to leaf nodes are of same length. ✓
- [ ] Its internal and leaf nodes direct the search.

✓ A magnetic disk with 32 surfaces having 512 tracks per platter with 128 sectors per track is
provided to you. You also know that this particular disk can only store 128 bytes in a sector.
How many number of bytes can be stored in this disk? 1/1

- [ ] 2^29 bytes
- [ ] 2^14 bytes
- [ ] 2 ^30 bytes
- [x] 2^28 bytes ✓

✗ Which one of the following statements is False? 0/1

- [ ] Lossless, dependency-preserving decomposition into BCNF is always possible
- [ ] If all attributes of a relation are prime attributes, then the relation is in BCNF.
- [x] Every relation has at least one non-prime attribute. ✓
- [ ] Any relation with two attributes is always in BCNF.

Correct answer

- [x] If all attributes of a relation are prime attributes, then the relation is in BCNF.
- [x] Every relation has at least one non-prime attribute.
- [x] Lossless, dependency-preserving decomposition into BCNF is always possible

✓ Consider the schedules S for transactions T1, T2 and T3 below. 1/1
S: r1(X); r2(Y); r3(Y); w2(Y); w1(X); w3(X); r2(X); w2(X)
This schedule is conflict serializable. What would be its equivalent serial schedule?

- [ ] T2 -> T3 ->T1
- [ ] T3 -> T2 -> T1
- [x] T1 ->T3 -> T2 ✓
- [ ] S1 is not conflict serializable

✗ Which statement(s) is (are) true? 0/1

- [x] Strict 2PL ensures the schedule recoverable. ✓
- [ ] Recoverable schedules do not avoid cascading aborts
- [x] For a schedule to be recoverable, transactions should commit only after all transactions whose changes they read commit. ✓
- [ ] For a schedule to be recoverable, transactions should abort only after all transactions whose changes they read commit.

Correct answer

- [x] For a schedule to be recoverable, transactions should commit only after all transactions whose changes they read commit.
- [x] Recoverable schedules do not avoid cascading aborts
- [x] Strict 2PL ensures the schedule recoverable.

✓ The magnetic disk rotates at a speed of 6000 rotations per minute and the average seek
time is 5ms. What would be the average access time (seek time + rotational time) for one
block? 1/1

- [ ] 25ms
- [x] 10ms ✓
- [ ] 12ms
- [ ] 15ms

✓ Let the schema of Employee and Branch relation be (E#, City, B#) and (B#, Location), 1/1
respectively. Here each E# and B# is the primary key in Employee and Branch relation, respectively, and Employee.B# is a foreign key referencing to Branch.B#, thus maintaining the
referential integrity constraints. Which of the following is(are) true?

$\pi_{B\#}(Branch) - \pi_{B\#}(Employee) \neq \emptyset$

☑ Option 4 ✓

$\pi_{B\#}(Employee) - \pi_{B\#}(Branch) = \emptyset$

☑ Option 1 ✓

$\pi_{B\#}(Employee) - \pi_{B\#}(Branch) \neq \emptyset$

☐ Option 3

(b)$\pi_{B\#}(Branch) - \pi_{B\#}(Employee) = \emptyset$

☐ Option 2

✓ Consider the schedules S below 1/1

S: r1(x), r2(x), w1(y), r2(y), r1(z), w2(x), T2(commit), w1(z), T1(commit)

This schedule S is:

- [ ] Conflict serializable, not recoverable and avoids cascading rollback
- [ ] Conflict serializable, recoverable and avoids cascading rollback
- [x] Conflict serializable, not recoverable and does not avoid cascading rollback ✓
- [ ] Conflict serializable, recoverable and does not avoid cascading rollback

✗ Let the schema of Employee and Branch relation be (Aadhar#, E#, City) and (E#, B#), 0/1

respectively. The functional dependencies (E# -> Aadhar#) and (Aadhar# -> City) hold for

Employee relation and its cardinality is 10000 and the cardinality of Branch relation is 500. What

would be the maximum number of tuples possible in (Employee NJ Branch)? Here, NJ means

Natural Join.

- [ ] 500
- [ ] 50000
- [x] 5000000 ✗
- [ ] 20

Correct answer

- [x] 500

✗ Consider the schedules S below. 0/1

S: r1(x), r2(x), w1(y), w1(x), w1(z), T1(commit), r2(y), w2(z), T2(commit)

This schedule S is:

- [x] Has no dirty reads ✓
- [x] Recoverable ✓
- [ ] Non-conflict serializable
- [x] Conflict serializable ✗

Correct answer

- [x] Non-conflict serializable
- [x] Has no dirty reads
- [x] Recoverable

✓ Which of the following statement(s) is(are) true? 1/1

- [ ] Lock downgrade reduces concurrency.
- [ ] A lock upgrade is to grant a transaction a shared lock of an object for which there is no lock holds on that object.
- [ ] Lock upgrade increases concurrency.
- [x] Lock downgrade violates the 2PL requirement because it reduces the locking privileges held by a transaction, and the transaction may go on to acquire other locks. ✓

✓ Consider the following two transactions with lock and unlock instructions obeying two-phase locking protocol. 1/1

T1: lock-S(A)
read(A);
lock-X(B)
read(B);
If A = 0, then B = B+1;
Write(B);
unlock(A)
unlock(B)

T2: lock-S(B)
read(B);
lock-X(A)
read(A);
If B = 0, then A = A+1;
Write(A);
unlock(B)
unlock(A)

Can there be a situation where the execution of these transactions in an interleaving operations schedule result in a deadlock?

- [ ] Never
- [ ] No
- [x] Yes ✓

✓ Consider the following BCNF schema of a database. 1/1

Emp (eid, ename, addr, sal, age, deptid)

Dept (did, dname, floor, budget)

Suppose you know that the following two queries are the most common queries in the workload

for this database and that both are roughly equivalent in frequency and importance:

I. List the id, name, and address of employees in a user-specified age range.

II. List the id, name, and address of employees who work in the department with a user-

specified department name.

Which of the following statement(s) is(are) true? Assume that B+ tree indexes are the only index

type supported by the DBMS and that both single- and multiple-attribute keys are permitted.

- [ ] Create an unclustered B+Tree index on deptid of the Emp relation and another clustered index on (dname, did) in the Dept relation, to answer the query II.
- [ ] Create a hash index on age of the Emp relation to answer query I.
- [x] Create a dense unclustered B+ tree index on age of the Emp relation to answer query I. ✓
- [x] Create an unclustered B+Tree index on deptid of the Emp relation and another unclustered index on (dname, did) in the Dept relation, to answer the query II. ✓

✓ The number of passes required for sorting 8 pages file using 3 buffer pages is 4. What 1/1

would be overall I/O cost?

- [ ] 16
- [ ] 32
- [x] 64 ✓
- [ ] 24

✓ Consider the following transactions with data items P and Q initialized to zero: 1/1

T 1 : read (Q) ;
read (P) ;
if P = 0 then Q : = Q + 1 ;
write (P).
T 2 : read (P) ;
read (Q)
if Q = 0 then P : = P + 1 ;
read (Q).
Any non-serial interleaving of T1 and T2 for concurrent execution leads to

- [x] A conflict serializable schedule ✓
- [ ] A schedule that is not conflict serializable
- [ ] A serial schedule
- [ ] None of the Above

✗ Which statement is (are) true w.r.t. “When a new page is to be placed in the main memory, a resident page should be evicted first (assuming there is not enough space in the memory)”. 0/1

- [ ] The DBMS administrator lets the buffer manager know to use the replacement policy.
- [x] All of the Above ✗
- [ ] The buffer manager capitalizes the different criterion to finalize the replacement policy.
- [ ] The buffer manager will determine the optimum uniform replacement policy.

Correct answer

- [x] The DBMS administrator lets the buffer manager know to use the replacement policy.
- [x] The buffer manager capitalizes the different criterion to finalize the replacement policy.

✓ What is the minimum space utilization for a B+ tree index? 1/1

- [ ] 100%, except with the root node
- [x] 50%, except with the root node ✓
- [ ] 75%, except with the root node
- [ ] None of the above

✗ Materialized Views can be incrementally maintained, whenever the database is updated, by 0/1

- [x] Using database pointers ✗
- [x] Executing change Management algorithm to maintain the materialized views incrementally (which runs periodically/on-demand/immediately) when database is updated ✓
- [x] Executing Triggers defined on database ✓
- [ ] None of the above

Correct answer

- [x] Executing change Management algorithm to maintain the materialized views incrementally (which runs periodically/on-demand/immediately) when database is updated
- [x] Executing Triggers defined on database

✓ In E-R model, the entity's life history (i.e. how its attributes and/or relationships change over time in response to events) can be shown as 1/1

- [ ] Entity enclosed by trapezium
- [x] None of the above ✓
- [ ] Entity enclosed by double rectangle
- [ ] Entity has an attribute named as, say, 'Life History'

✗ Which statement(s) is (are) false for the query optimizer 0/1

- [x] to leverage the data dictionary and statistics for computing the cost for each query plan ✗
- [x] it always executes the unary operations (selection, projection, renaming) as early as possible for reducing the I/O cost. ✓
- [x] to let the query processor know which algorithm to use for each relational operator ✗
- [ ] It always computes the cost of a single execution plan

Correct answer

- [x] It always computes the cost of a single execution plan
- [x] it always executes the unary operations (selection, projection, renaming) as early as possible for reducing the I/O cost.

✓ Which of the following best described the usage of metadata in DBMS? 1/1

- [x] It is used in analyzing the statitics of data, such as, number of records in each relation, selectivity factor, attributes' values distribution, attributes' data types, integrity constraints, etc. ✓
- [x] It is used for query processing. ✓
- [ ] It is used for data normalization and removing the data ambiguities.
- [ ] It is a backup of primary data, which is used for data recovery.

✗ Read-only transactions are not used for 0/1

- [ ] Business Intelligence reports
- [ ] Analysis applications
- [ ] Data backup
- [x] Detecting data anomalies ✗

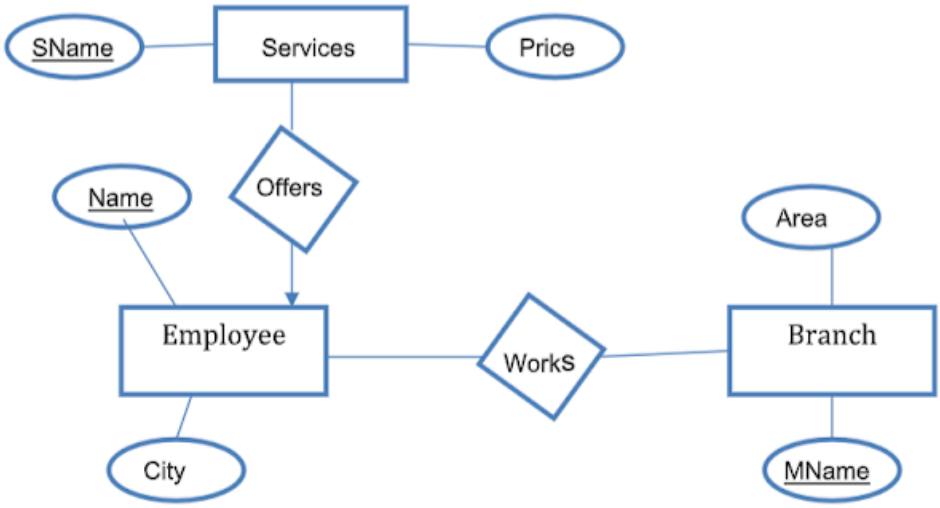
Correct answer

- [x] Data backup

✓ Which of the following option(s) satisfies(satisfy) that in relation schema R(A,B,C,D) with
primary key AB under which R is in 1NF but not in 2NF? 1/1

- [ ] A -> B
- [ ] C -> D
- [x] A-> C ✓
- [x] B -> C ✓

✓ In the following E-R diagram, how many minimum number of tables, adhering to 3NF,
would be generated? The Primary key of each entity is underlined. 1/1



- [ ] 3
- [ ] 5
- [ ] 6
- [x] 4 ✓

✓ Which of the statement(s) is(are) true wrt the RAID? 1/1

- [ ] RAID architecture is used for data backup solution.
- [x] RAID 5 is typically preferred when data update frequency is low. ✓
- [x] In RAID 5, You need at least 3 disks and writing the data on a disk is bit slower than RAID 1. ✓
- [ ] In RAID 1, if you loose a drive, you loose all data and therefore you cannot recover the data from other disks.

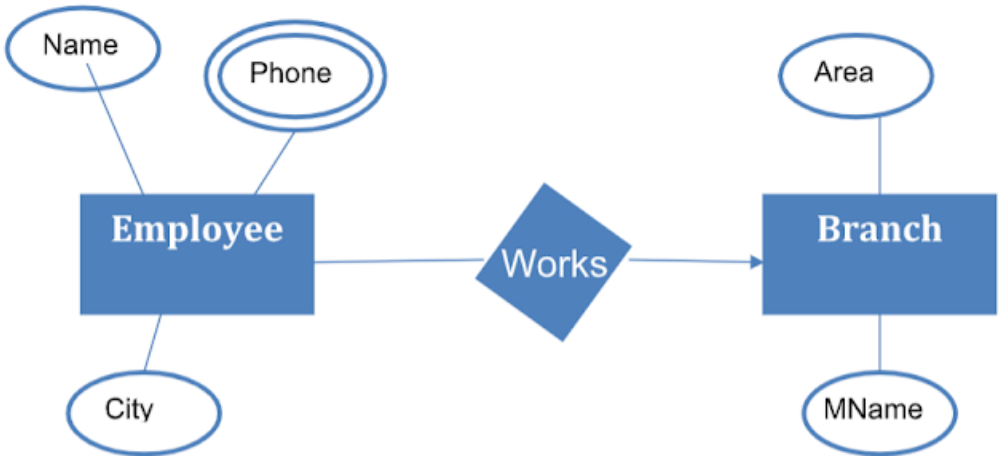
✓ Given a STUDENT(STD_ID, STD_Name, STD_Dept, STD-Hostel) table and a memory resident Index table on the primary key (STD_ID) of STUDENT pointing to the record position on the disk, if we need to answer the query “How many students are enrolled in 'CSE' departments?”, this query can be answered using 1/1

- [ ] taking the join of of Index Table and STUDENT Table
- [ ] Index Table
- [ ] cannot be answered
- [x] STUDENT Table ✓

✓ Which statement is (are) false? 1/1

- [x] The only advantage of using RAID is to achieve good performance ✓
- [ ] To minimize I/O time, it is necessary to store and locate data strategically
- [x] In RAID 5, the fault tolerance is achieved through distributed parity and 50% storage is used for data storage ✓
- [ ] I/O time for data transfer from disk to main memory is the summation of (seek time + rotational time+ transfer time)
- [ ] In RAID 1, data is fully replicated across the disks so less than 100% storage is used for data storage.

✗ In the following E-R diagram, how many minimum number of tables are required to convert
it into relational model? 0/1



- [ ] 5
- [ ] 4
- [x] 2 ✗
- [ ] 3

Correct answer

- [x] 3

✗ A phantom read is a situation where a transaction reads a set of records that satisfy a certain condition, but then another transaction inserts or deletes a record that also satisfies that condition, causing the first transaction to see a phantom record. What techniques can't be used to prevent Phantom reads? 0/1

- [ ] Undo and Redo
- [x] Locking ✗
- [ ] Multi-version concurrency control
- [ ] Serializability

Correct answer

- [x] Serializability
- [x] Undo and Redo

✓ Consider a relation R(A, B, C, D, E, F, G, H, I, J that satisfies the functional dependencies (FDs) {AB -> C}, {B -> EF}, {AD -> GH}, {G ->I}, {H ->J}. What would be the primary key of this relation? 1/1

- [ ] {BEF}
- [ ] {KLM}
- [x] {ABD} ✓
- [ ] {ABC}

✓ Consider the database with the below parameters: 1/1
Block Size = 500 bytes
Record Size = 50 bytes
Key Size = 10 Bytes
Blocking factor: 10
Pointer Size = 40 bytes
Given 28800 records, determine the number of dense index blocks?

- [ ] 4114
- [ ] 5760
- [x] 2880 ✓
- [ ] 1440

✓ A typical lock manager is implemented with a hash table, also called lock table, with the 1/1
data object identifier as the key. A lock table entry contains the following information:

- [ ] the number of transactions currently holding a lock on the object,
- [ ] a pointer to a queue of lock requests,
- [ ] the nature of the lock – shared or exclusive,
- [x] All of the above ✓

✗ Natural-Join operator is 0/1

- [ ] Distributive
- [ ] Associative
- [x] All of the Above ✗
- [ ] Commutative

Correct answer

- [x] Commutative
- [x] Associative

✓ Consider the schedules S below 1/1
S: r2(x), r1(x), w1(z), r2(z), r1(y), w1(y), T1(commit), w2(x), T2(commit)
This schedule S is:

- [x] Conflict serializable ✓
- [x] Has dirty reads ✓
- [ ] Avoids Cascading rollback
- [ ] Non-conflict serializable

✓ Consider the following schedule: 1/1

Here, R(X) stands for Read(X) and W(X) stands for Write(X). Replacing '?' by which of the
following options will make the above schedule serializable?

T1	T2
R(X)	
R(Y)	
	W(Y)
?	

- [x] R(X) ✓
- [ ] W(Y)
- [ ] R(Y)
- [ ] None of the above

✓ Which of the following concurrency control protocols ensure both conflict serializability and
freedom from deadlock? 1/1

- [ ] Strict Two-Phase Locking
- [x] Time-stamp ordering ✓
- [ ] All of the above
- [ ] Two-Phase Locking

✓ Which of the following schedules result(s) in a read-write conflict? 1/1

- [ ] R2(X), R2(Y), W2(X), R1(X), ...
- [x] R2(X), R2(Y), R1(X), R1(Y), W1(X), W2(X), ... ✓
- [x] R2(X), R2(Y), R1(X), R1(Y), W1(X), ... ✓
- [ ] None of the Above

✓ Let's consider the log as it appears at three instances of time 1/1

The recovery actions would be

(a)	(b)	(c)
$<T_0$ start>	$<T_0$ start>	$<T_0$ start>
$<T_0, A, 1000, 950>$	$<T_0, A, 1000, 950>$	$<T_0, A, 1000, 950>$
$<T_0, B, 2000, 2050>$	$<T_0, B, 2000, 2050>$	$<T_0, B, 2000, 2050>$
	$<T_0$ commit>	$<T_0$ commit>
	$<T_1$ start>	$<T_1$ start>
	$<T_1, C, 700, 600>$	$<T_1, C, 700, 600>$
		$<T_1$ commit>

- [x] In instance c: T 0 : redo, T 1 : redo ✓
- [ ] In instance c: T 0 : undo, T 1 : undo
- [x] In instance a: T 0 : undo, ✓
- [x] In instance b: T 0 : redo, T 1 : undo ✓

✗ When the E-R model is converted into a relational model, all the relations would be in 0/1

- [ ] 1NF
- [ ] 2NF
- [ ] 3NF
- [x] NNF (Non-Normal Forms) ✗

Correct answer

- [x] 1NF

✗ Let the schema of Employee and Branch relation be (E#, City, Salary, B#) and (B#, Location), respectively. Here Employee.B# is a foreign key referencing to Branch.B#, thus

0/1

maintaining the referential integrity constraints. The following two queries Q1 and Q2 are given below:

The ordered indexing or hashing can be done on attribute 'Salary'. Which of the following statement(s) is(are) true?

Q1: $\Pi_{E\#, City}(\sigma_{Salary=40000}(Employee))$

Q2: $\Pi_{E\#, City}(\sigma_{20000 \leq Salary \leq 40000}(Employee))$

- [ ] Hashing will outperform ordered indexing on Q1, but not on Q2.
- [ ] Hashing will outperform ordered indexing on Q2, but not on Q1.
- [ ] Hashing will always outperform ordered indexing for both queries.
- [x] Ordered indexing will always outperform hashing for both queries. ✗

Correct answer

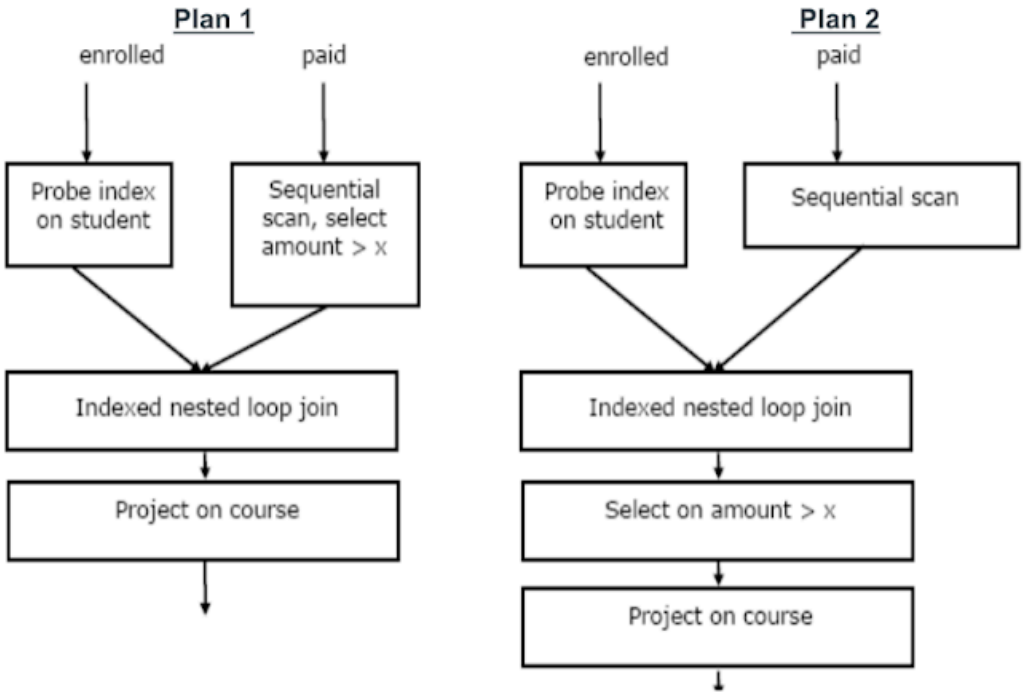
- [x] Hashing will outperform ordered indexing on Q1, but not on Q2.

✓ Consider the relation enrolled(student, course), and the relation paid(student, amount), here the primary key in eack relation is underlined and there is no foreign key. Assume the

1/1

selectivity factor of 'amount' attribute in paid relation is 20%. Consider the following query plans for the query "list all courses taken by students who have paid more than x".

A disk seek takes 4ms, disk data transfer bandwidth is 300 MB/s and checking a tuple to see if amount is greater than x takes 10 micro-seconds. Which of the following statements is correct, when the disk seek time is 4ms, data transfer bandwidth is 300MB, and checking a tuple to see if amount is greater than x takes 10 micro-seconds?



- [x] For x = 5000, Plan 1 executes faster than Plan 2 for all databases. ✓
- [ ] Plan 1 and Plan 2 will not output identical row sets for all databases.
- [ ] For x = 9000, Plan I executes slower than Plan 2 for all databases.

- [x] A course may be listed more than once in the output of Plan 1 for some databases ✓

✓ Which statement(s) is (are) true? 1/1

- [ ] Covering index table excludes all the columns required by a query and pointers to the data record.
- [x] It significantly improves query performance by reducing the number of I/O operations required to execute the queries which can be answered using covering index table. ✓
- [ ] There is no additional overhead for maintaining the covering index table.
- [x] Covering Index table can be created over materialized view. ✓