Course Name: Database Management System

Assignment 8 - Week 8 (Jan 2025)

TYPE OF QUESTION: MCQ/MSQ

Number of questions: 10 Total mark: 10 X 2 = 20

Question 1

Identify the `cost estimation` of a `query evaluation plan`, if 9000 blocks are required to `transferred from the disk` and the required number of `disk seeks` are 25.

- Time to transfer one block: $t_T = 5$ milliseconds.
- Time for one seek: $t_S = 0.4$ seconds.

a) 40 Seconds

b) 45 Seconds

c) 50 Seconds

d) 55 Seconds

Answer: d)
Explanation: Cost for b block transfers plus S seeks will be $(b * t_T + S * t_S)$ seconds
$=(9000 * 5 * 10^{-3}) + (25 * 0.4)$ seconds
$= (45 + 10)$ Seconds
$= 55$ Seconds
For more details refer to 38.12 of lecture material.
Hence, option d) is the answer.

Question 2

Assume an immediate database modification scheme. Consider the following log records for transactions T0, T1, T2, T3 and T4:

steps	Details of log
1	⟨T0,start⟩
2	⟨T0,A,500,600⟩
3	⟨T1,start⟩
4	⟨T1,B,300,500⟩
5	⟨T1,commit⟩
6	⟨T2,start⟩
7	⟨checkpoint{T0, T2} ⟩
8	⟨T3,start⟩
9	⟨T2,C,200,400⟩
10	⟨T3,D,700,900⟩
11	⟨T2,commit⟩
12	⟨T3,commit⟩
13	⟨T4,start⟩
14	⟨T4,E,300,700⟩

If there is a crash just after step 14 and the recovery of the system is successfully completed, identify the **correct** action for the above scenario.

a) After recovery completion, value of A will be 600.

b) After recovery completion, value of C will be 200.

c) After recovery completion, value of D will be 900.

d) After recovery completion, value of E will be 300.

Answer: c), d)
Explanation: In the `immediate database modification scheme`, during recovery after a crash, a transaction needs to be redone if and only if both ⟨$\texttt{T}_i$, start⟩, ⟨$\texttt{T}_i$, commit⟩ are present in the log. otherwise undo is required.
Any transactions that committed before the last checkpoint should be ignored(updates already output to disk due to the checkpoint).
Redo list contains transaction {T2, T3} and undo list contain transactions {T0, T4} and for transaction {T1} no need any action because it is already committed before checkpoint.
As per the process of transaction recovery, options (c) and (d) are correct.

Question 3

Let us consider the following statistics for two relations `Instructor` and `Job_Assignments`:

- Number of records of `Instructor`: $n_{\texttt{Instructor}} = 5050$.
- Number of blocks of `Instructor`: $b_{\texttt{Instructor}} = 30$.
- Number of records of `Job_Assignments`: $n_{\texttt{Job\_Assignments}} = 1050$.
- Number of blocks of `Job_Assignments`: $b_{\texttt{Job\_Assignments}} = 10$.

Let us consider a natural join of `Instructor` and `Job_Assignments` relations (`Instructor` $\bowtie$ `Job_Assignments`). Identify the required number of `block transfers` in the worst case (enough memory only to hold one block of each relation) using `Nested-loop join` and assume `Instructor` as the outer relation.

a) 40000 block transfers

b) 40030 block transfer

c) 50030 block transfers

d) 50530 block transfers

Answer: d)
Explanation: Number of block transfers will be: $5050 \times 10 + 30 = 50530$, if `Instructor` is taken as the outer relation.
For more details refer to 38.31 of lecture material.

Question 4

Consider the following relational schema:
INSTRUCTOR(InstructorID, Name, HireDate)
COURSE(CourseID, CourseName, Credits)
JOB_ASSIGNMENTS(InstructorID, CourseID, JobTitle, StartDate, EndDate)

Two query trees are given below.

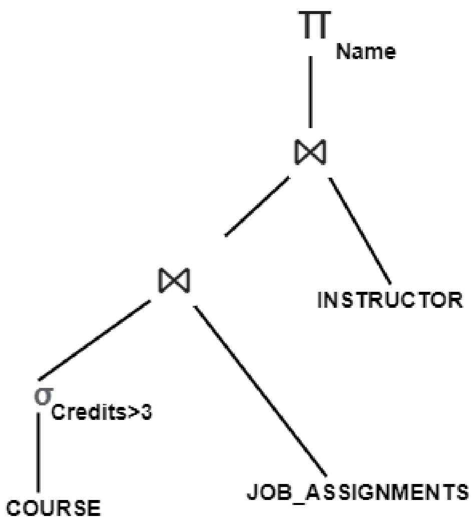


Figure 1:

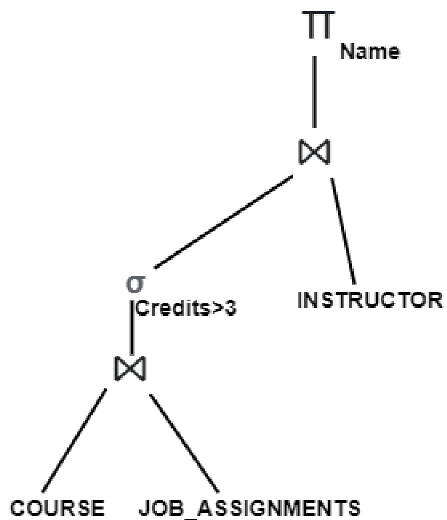


Figure 2:

Identify the correct statement for the above two query trees.

a) Two query trees are equivalent, and the query tree of Figure 2 will lead to more efficient query processing.

b) Two query trees are equivalent, and the query tree of Figure 1 will lead to more efficient query processing.

c) Two query trees are equivalent, as identical operations (irrespective of their positions) are used in both trees.

d) Two query trees are not equivalent as selection or projection operation cannot be carried out before or after the natural join operation.

Answer: b)
Explanation: Two query trees are equivalent, and Figure 1 will lead to more efficient query processing because performing the selection operation as early as possible reduces the size of the relation to be joined.
Hence, option (b) is correct.

Question 5

Assume deferred database modification scheme. Consider the following log records for transactions T1, T2, T3 and T4:

steps	Details of log
1	⟨T1,start⟩
2	⟨T1,A,500,800⟩
3	⟨T2,start⟩
4	⟨checkpoint{T1, T2} ⟩
5	⟨T1,commit⟩
6	⟨T3,start⟩
7	⟨T2,B,200,400⟩
8	⟨T2,commit⟩
9	⟨T4,start⟩
10	⟨T3, C, 700, 300⟩

If there is a crash just after step 10 and the recovery of the system is successfully completed, identify the **correct** action(s) for the above scenario.

a) After recovery completion, the value of B is 400.

b) After recovery completion, the value of C is 300.

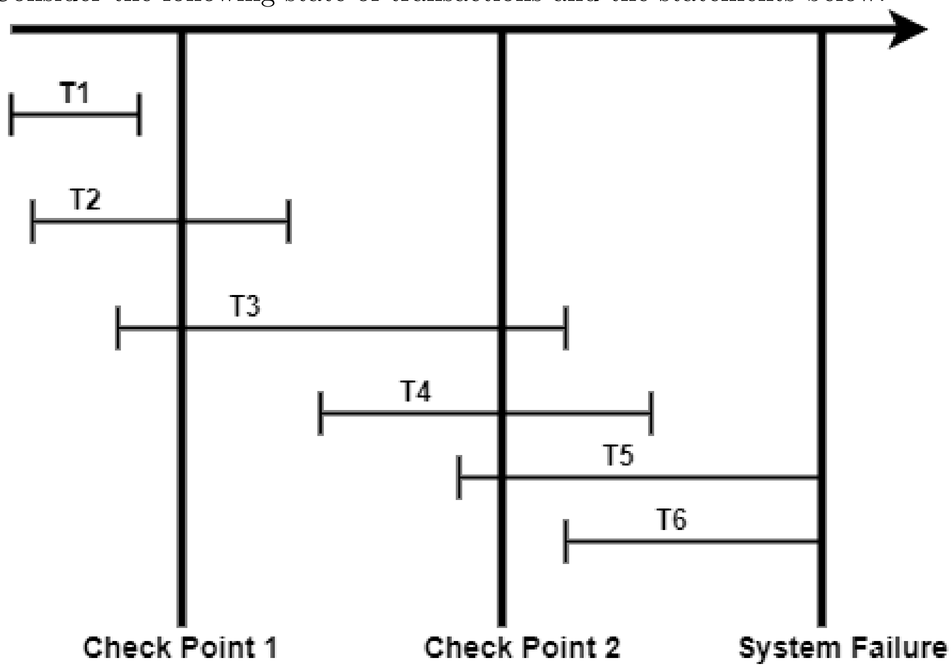
c) Redo list contain transactions {T1, T2} and undo list contains {T3, T4}.

d) Redo list contain transactions {T1, T2} and undo list contains { }.

Answer: a), d)
Explanation: In the deferred database modification scheme, updates are not written to the database until the transaction commits. If a transaction commits, its updates are redone from the log. If a transaction does not commit, it is ignored (no undo is needed, since no changes were written to the database).
As per the process of transaction recovery, options a) and d) are correct.

Question 6

Consider the following state of transactions and the statements below.



1. $\mathtt{T}_1$, $\mathtt{T}_2$ and $\mathtt{T}_3$ can be ignored.
2. $\mathtt{T}_5$ and $\mathtt{T}_6$ need to be undone.
3. $\mathtt{T}_1$ and $\mathtt{T}_2$ can be ignored.
4. $\mathtt{T}_3$, $\mathtt{T}_4$ and $\mathtt{T}_5$ need to be redone.
5. $\mathtt{T}_3$ and $\mathtt{T}_4$ need to be redone.

Identify the correct group of statements from the options below.

a) 1), 2), 3), 5)

b) 1), 3), 4), 5)

c) 1), 4), 5)

d) 2), 3), 5)

Answer: d)
Explanation: Any transaction that is committed before the last checkpoint should be ignored. Therefore, T_1, and T_2 can be ignored (updates already output to disk due to the last checkpoint).
Any transaction that is committed since the last checkpoint, needs to be redone. Hence, T_3 and T_4 are to be redone.
Any transaction that was running at the time of failure, needs to be undone and restarted. Hence, T_5 and T_6 are to be undone.
Hence, option (d) is correct.

Question 7

Consider the following relational schema:
`INSTRUCTOR(InstructorID, Name, HireDate)`
`COURSE(CourseID, CourseName, Credits)`
`JOB_ASSIGNMENTS(InstructorID, CourseID, JobTitle, StartDate, EndDate)`

Four relational algebra queries are given below:

Q1: $\sigma_{\texttt{StartDate='2023-01-01'}}(\texttt{INSTRUCTOR} \bowtie \texttt{JOB\_ASSIGNMENTS})$

Q2: $\pi_{\texttt{StartDate, Name}}(\texttt{INSTRUCTOR} \bowtie \texttt{JOB\_ASSIGNMENTS})$

Q3: $\pi_{\texttt{Name}}(\texttt{INSTRUCTOR} \bowtie \texttt{JOB\_ASSIGNMENTS})$

Q4: $\pi_{\texttt{Name}}(\sigma_{\texttt{INSTRUCTOR.InstructorID=JOB\_ASSIGNMENTS.InstructorID}}(\texttt{INSTRUCTOR} \times \texttt{JOB\_ASSIGNMENTS}))$

Identify the correct options from the options given below.

a) Q1 is equivalent to Q2.

b) Q1 is not equivalent to Q2.

c) Q3 is equivalent to Q4.

d) Q3 is not equivalent to Q4.

Answer: b), c)
Explanation: Q1 and Q2 will not give the same result because Q2 will project only the StartDate and Name columns over the natural join of `INSTRUCTOR` and `JOB_ASSIGNMENTS`, whereas Q1 selects records where `StartDate` is '2023-01-01'.
Q3 and Q4 will give the same result because $\sigma_{\texttt{INSTRUCTOR.InstructorID=JOB\_ASSIGNMENTS.InstructorID}}$ followed by $(\texttt{INSTRUCTOR} \times \texttt{JOB\_ASSIGNMENTS})$ is equivalent to $(\texttt{INSTRUCTOR} \bowtie \texttt{JOB\_ASSIGNMENTS})$.
Hence, options b) and c) are correct.

Question 8

Consider the following relational schema:
`INSTRUCTOR(InstructorID, Name, HireDate)`
`COURSE(CourseID, CourseName, Credits)`
`JOB_ASSIGNMENTS(InstructorID, CourseID, JobTitle, StartDate, EndDate)`

A relational algebra expression is given below:
$\Pi_{\texttt{Name}}(\Pi_{\texttt{InstructorID, Name}}(\sigma_{\texttt{Credits=3}}(\texttt{INSTRUCTOR} \bowtie \texttt{JOB\_ASSIGNMENTS} \bowtie \texttt{COURSE})))$

Identify the most optimized relational algebra expression equivalent to the above relational algebra expression.

a) $\Pi_{\texttt{Name, InstructorID}}(\sigma_{\texttt{Credits=3}}(\texttt{INSTRUCTOR} \bowtie \texttt{JOB\_ASSIGNMENTS} \bowtie \texttt{COURSE}))$

b) $\Pi_{\texttt{Name, InstructorID}}(\texttt{INSTRUCTOR} \bowtie (\texttt{JOB\_ASSIGNMENTS} \bowtie (\sigma_{\texttt{Credits=3}}(\texttt{COURSE}))))$

c) $\Pi_{\texttt{Name, InstructorID}}(\texttt{INSTRUCTOR} \bowtie (\texttt{JOB\_ASSIGNMENTS} \bowtie \Pi_{\texttt{CourseID}}(\sigma_{\texttt{Credits=3}}(\texttt{COURSE}))))$

d) $\Pi_{\texttt{Name}}(\texttt{INSTRUCTOR} \bowtie (\texttt{JOB\_ASSIGNMENTS} \bowtie (\sigma_{\texttt{Credits=3}}(\texttt{COURSE}))))$

Answer: d)
Explanation: According to the `transformation rules`, only the last in a sequence of projection operations is needed, so the others can be omitted.
It is optimal to perform the selection operation as early as possible to reduce the size of the relation to be joined.
Hence, option (d) is correct.

Question 9

Consider the following two relational algebra expressions (RA) given below:

RA I: $\Pi_{A,B}(\texttt{P} \cup \texttt{Q}) = \Pi_{A,B}(\texttt{P}) \cup \Pi_{A,B}(\texttt{Q})$

RA II: $((\texttt{P} \bowtie \texttt{Q}) \bowtie \texttt{R}) = (\texttt{P} \bowtie (\texttt{Q} \bowtie \texttt{R}))$

where `P`, `Q`, and `R` are relational algebra expressions.

Identify the correct statement(s) from the following.

a) Both RA I and RA II are true.

b) Both RA I and RA II are false.

c) RA I is true but RA II is false.

d) RA I is false but RA II is true.

Answer: a)
Explanation: RA I follows the projection distribution property over union: projecting after a union is equivalent to taking projections first and then performing a union. Thus, RA I is valid.
RA II is correct because natural join operations are associative.
Thus, option a) is correct.

Question 10

Consider the log record of Transaction `T1` with two operation instances `O1` and `O2` used in recovery system with early lock release, B+ tree based concurrency control.

Step	Operation
1	⟨T1, start⟩
2	⟨T1,X, 200, 400⟩
3	⟨T1, O1, operation–begin⟩
4	⟨T1, Y, 100, 500⟩
5	⟨T1, O1, operation–end,(Y, -400)⟩
6	⟨T1, O2, operation–begin⟩
7	⟨T1, Z, 500, 800⟩
8	crash or abort here

Choose the correct set of log entries for the recovery of transactions.

a)

⟨T1,Z,500⟩
⟨T1,Y,500,100⟩
⟨T1,O1,operation–abort⟩
⟨T1,X,400⟩
⟨T1,abort⟩

b)

⟨T1,Z,500⟩
⟨T1,Y,500,100⟩
⟨T1,O1,operation–abort⟩
⟨T1,X,200⟩
⟨T1,abort⟩

c)

⟨T1,Z,500⟩
⟨T1,Y,100,500⟩
⟨T1,O1,operation–abort⟩
⟨T1,X,200⟩
⟨T1,abort⟩

d)

⟨T1,Z,800⟩
⟨T1,Y,500,100⟩
⟨T1,O1,operation–abort⟩
⟨T1,X,400⟩
⟨T1,abort⟩

Answer: b)

Explanation: Step i: Scan the log records backward.

Step ii: For the step 7 update of Z was part of O2, undone physically during recovery since O2 did not complete; so, physical undo is required for operation O2 on the variable Z. That means, we have to add log ⟨`T2, Z, 500`⟩. Step iii: For the step 5 ⟨`T1, O1, operation-end, (Y,-400)`⟩ is found; so, logical undo is required for operation O1 on the variable Y using the information (subtraction of 400). That means, we have to delete the previous modifications on Y. So, add the following logs for steps 5, 4 and 3 respectively: ⟨`T1, Y, 500, 100`⟩ and ⟨`T1, O1, operation–abort`⟩.

Step iv: For step 2, add the log ⟨`T1, X, 200`⟩.

Step v: For step 4, add the log ⟨`T1, abort` ⟩.

Hence, option (b) is correct.