Database Management System: Assignment 8

Total Marks : 100

October 3, 2022

Question 1

Assume an immediate database modification scheme. Consider the following log records for transactions T0, T1, T2, T3 and T4:   *Marks: 2* **MSQ**

steps	Details of log
1	⟨T0,start⟩
2	⟨T1,start⟩
3	⟨T1,X,100,200⟩
4	⟨T1,commit⟩
5	⟨checkpoint{T0} ⟩
6	⟨T0,Y,200,400⟩
7	⟨T2,start⟩
8	⟨T0,commit⟩
9	⟨T3,start⟩
10	⟨T4,start⟩
11	⟨T3,Z,400,600⟩
12	⟨T4,W,600,1200⟩
13	⟨T4,commit⟩

If there is a crash just after step 13 and the recovery of the system is successfully completed, identify the **correct** actions for the above scenario.

a) After recovery completion, value of X will be 100.

b) After recovery completion, value of Y will be 400.

c) After recovery completion, value of Z will be 600.

d) After recovery completion, value of W will be 1200.

Answer: b), d)

Explanation: In the `immediate database modification scheme`, during recovery after a crash, a transaction needs to be redone if and only if both ⟨T_i, `start`⟩, ⟨T_i, `commit`⟩ are present in the log. Otherwise, undo is required.

Any transactions that are committed before the last checkpoint should be ignored(updates already output to disk due to the checkpoint).

Redo list contains transactions {T0, T4} and undo list contains transactions {T2, T3} and for transaction {T1} no action is needed because it is already committed before checkpoint.

After recovery completion, for T0 the value of Y will be 400, for T3 the value of Z will be 400 and for T4 the value of W will be will be 1200. Hence, options (b) and (d) are correct.

Question 2

Identify the `cost estimation` of a `query evaluation plan`, if 2000 blocks are required to be `transferred from disk` and the required number of `disk seeks` are 50.

- Time to transfer one block: $t_T = 4$ milliseconds.

- Time for one seek: $t_S = 0.4$ seconds.

Marks:2 **MCQ**

a) 4.4 Seconds

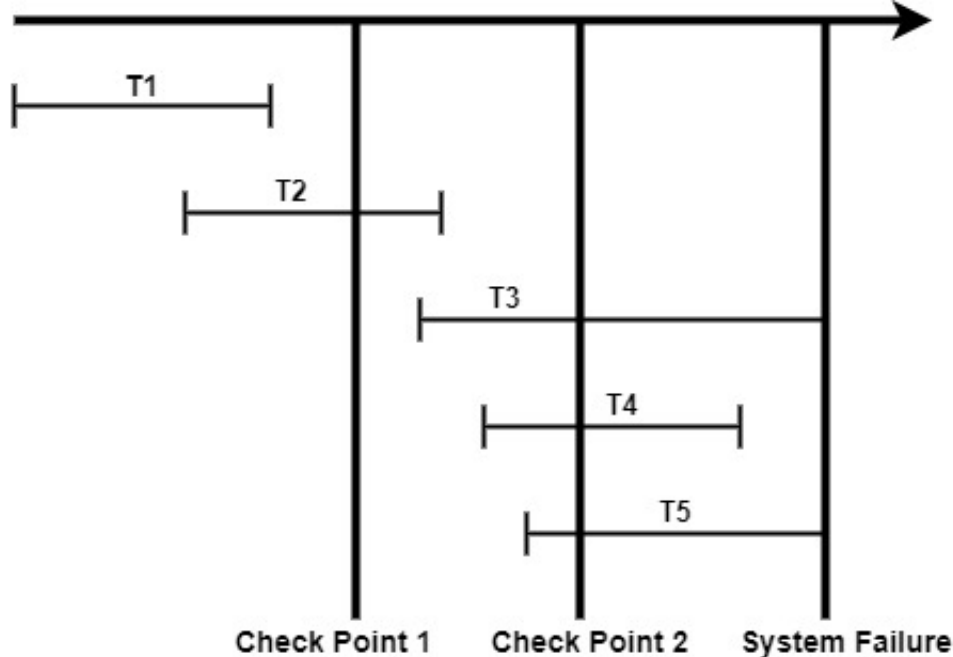b) 8 Seconds

c) 10 Seconds

d) 28 Seconds

Answer: d)
Explanation: Cost for b block transfers plus S seeks will be $(b * t_T + S * t_S)$ seconds
$=(2000 * 4 * 10^{-3}) + (50 * 0.4)$ seconds
$= (8 + 20)$ Seconds
$= 28$ Seconds
For more details refer to 38.12 of lecture material.
Hence, option d) is the answer.

Question 3

Consider the following state of transactions and the statements below.

1. `T1`, `T2` and `T4` can be ignored.

2. `T1` and `T2` can be ignored.

3. `T3` and `T5` need to be undone.

4. `T2`, `T3`, `T4` and `T5` need to be redone.

5. Only `T4` needs to be redone.

Identify the correct group of statements from the options below.   *Marks:2* **MCQ**

a) 1), 2), 3)

b) 1), 3), 4)

c) 2), 4), 5)

d) 2), 3), 5)

Answer: d)
Explanation: Any transaction that is committed before the last checkpoint should be ignored. Therefore, `T1 and T2` 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, only `T4` needs to be redone.
Any transaction that was running at the time of failure, needs to be undone and restarted. So, `T3 and T5` need to be undone.
Hence, option (d) is the correct answer.

Question 4

Let us consider the following statistics for two relations `Patients` and `Appointments`:

- Number of records of `Patients`: $n_{\texttt{Patients}} = 5000$.

- Number of blocks of `Patients`: $b_{\texttt{Patients}} = 200$.

- Number of records of `Appointments`: $n_{\texttt{Appointments}} = 1000$.

- Number of blocks of `Appointments`: $b_{\texttt{Appointments}} = 100$.

Identify the required number of `block transfers` and `seeks` for `Nested-loop join` if the smaller relation fits entirely in memory and the smaller relation is used as the inner relation in the join operation.

a) 500200 block transfers and 2 seeks

b) 200100 block transfers and 2 seeks

c) 6300 block transfers and 2 seeks

d) 300 block transfers and 2 seeks

Answer: d)
Explanation: Number of block transfers will be: $200 + 100 = 300$, and number of seeks will be :2, if the smaller relation fits entirely in memory.
For more details refer to 38.31 of lecture material.

Question 5

Assume an `immediate database modification scheme`. Consider the following log records for transactions T0, T1, T2, T3 and T4:   *Marks: 2* **MCQ**

steps	Details of log
1	$\langle$T0,start$\rangle$
2	$\langle$T0,X,200,400$\rangle$
3	$\langle$T1,start$\rangle$
4	$\langle$T1,Y,400,600$\rangle$
5	$\langle$T0,commit$\rangle$
6	$\langle$checkpoint$\{$T1$\}$ $\rangle$
7	$\langle$T2,start$\rangle$
8	$\langle$T2,Z,600,800$\rangle$
9	$\langle$T3,start$\rangle$
10	$\langle$T3,W,800,1000$\rangle$
11	$\langle$T3,commit$\rangle$
12	$\langle$T4,start$\rangle$
13	$\langle$T4,P,1000,1200$\rangle$

If there is a crash just after step 13 and the recovery of the system is successfully completed, identify the **correct** action for the above scenario.

a) No Action: T0, T1; Redo: T3 and Undo: T2, T4

b) No Action: T0; Redo: T3 and Undo: T1, T2, T4

c) Redo: T0, T3 and Undo: T1, T2, T4

d) Redo: T0, T1 and Undo: T2, T3, T4

Answer: b)

Explanation: In the `immediate database modification scheme`, during recovery after a crash, a transaction needs to be redone if and only if both $\langle$T$_i$, start$\rangle$, $\langle$T$_i$, commit$\rangle$ 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 $\{$T3$\}$ and undo list contain transaction $\{$T1, T2, T4$\}$ and for transaction $\{$T0$\}$ no need any action because it is already committed before checkpoint.

Hence, option b) is the answer.

Question 6

Consider the log record of transaction T0, transaction T1 with operation instances O1 and transaction T2 with operation instances O2 used in recovery system with early lock release, B+ tree based concurrency control.   *Marks: 2* **MCQ**

Step	Operation
1	$\langle$T0, start$\rangle$
2	$\langle$T0, X, 200, 400$\rangle$
3	$\langle$T0, commit$\rangle$
4	$\langle$T1, start$\rangle$
5	$\langle$T1, Y, 400, 600$\rangle$
6	$\langle$checkpoint$\{$T1$\}$ $\rangle$
7	$\langle$T1, O1, operation–begin$\rangle$
8	$\langle$T1, Z, 600, 800$\rangle$
9	$\langle$T1, O1, operation–end,(Z, -200)$\rangle$
10	$\langle$T2, start$\rangle$
11	$\langle$T2, O2, operation–begin$\rangle$
12	$\langle$T2, W, 800, 1000$\rangle$
13	crash or abort here

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

a)

$\langle$T2, W, 800$\rangle$
$\langle$T2, abort$\rangle$
$\langle$T1, Z, 800, 600$\rangle$
$\langle$T1, O1, operation–abort$\rangle$
$\langle$T1, Y, 600$\rangle$
$\langle$T1, abort$\rangle$
$\langle$T0, X, 200$\rangle$

b)

$\langle$T2, W, 1000$\rangle$
$\langle$T2, abort$\rangle$
$\langle$T1, Z, 800, 600$\rangle$
$\langle$T1, O1, operation–abort$\rangle$
$\langle$T1, Y, 600$\rangle$
$\langle$T1, abort$\rangle$
$\langle$T0, X, 400$\rangle$

c)

$\langle$T2, W, 1000$\rangle$
$\langle$T2, abort$\rangle$
$\langle$T1, Z, 800, 600$\rangle$
$\langle$T1, O1, operation–abort$\rangle$
$\langle$T1, Y, 600$\rangle$
$\langle$T1, abort$\rangle$

d)

$\langle$T2, W, 800$\rangle$
$\langle$T2, abort$\rangle$
$\langle$T1, Z, 800, 600$\rangle$
$\langle$T1, O1, operation–abort$\rangle$
$\langle$T1, Y, 400$\rangle$
$\langle$T1, abort$\rangle$

Answer: d)

Explanation: Step i: Scan the log records backward.

Step ii: For the step 12 update of W was part of O2, undone physically during recovery since O2 did not complete; so, physical undo is required for operation O2 on the variable W. That means, we have to add log ⟨T2, W, 800⟩. For step 10, add the log ⟨T2, abort ⟩

Step iii: For step 9 ⟨T1, O1, operation-end, (Z,-200)⟩ is found; so, logical undo is required for operation O1 on the variable Z using the information (subtraction of 200). That means, we have to delete the previous modifications on Z. So, add the following logs for steps 9, 8 and 7 respectively: ⟨T1, Z, 800, 600⟩ and ⟨T1, O1, operation–abort⟩.

Step iv: For step 5, add the log ⟨T1, Y, 400⟩.

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

For steps 3 to 1 no need any action because any transactions that committed before the last checkpoint should be ignored(updates already output to disk due to the checkpoint).

Hence, option (d) is correct.

Question 7

Consider the following relational schema: *Marks: 2* **MCQ**
Patients(pid, pname, age, p_email)
Doctors(did, dname, qualification, d_email)
Appointments(pid, did, date, time)
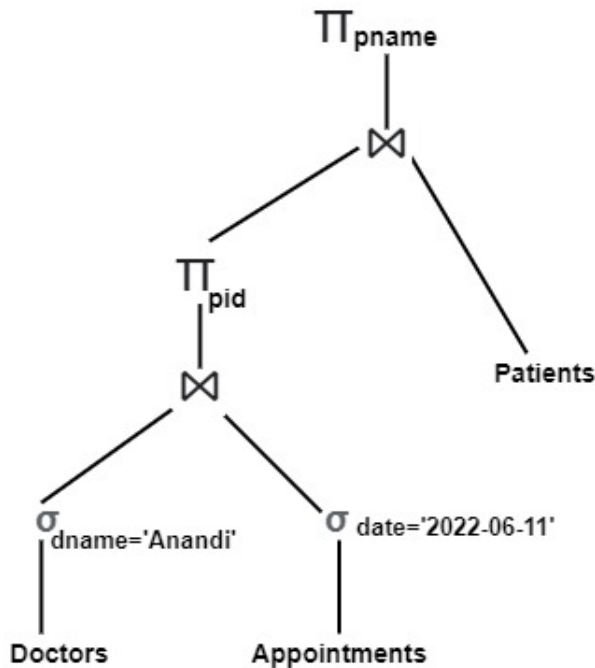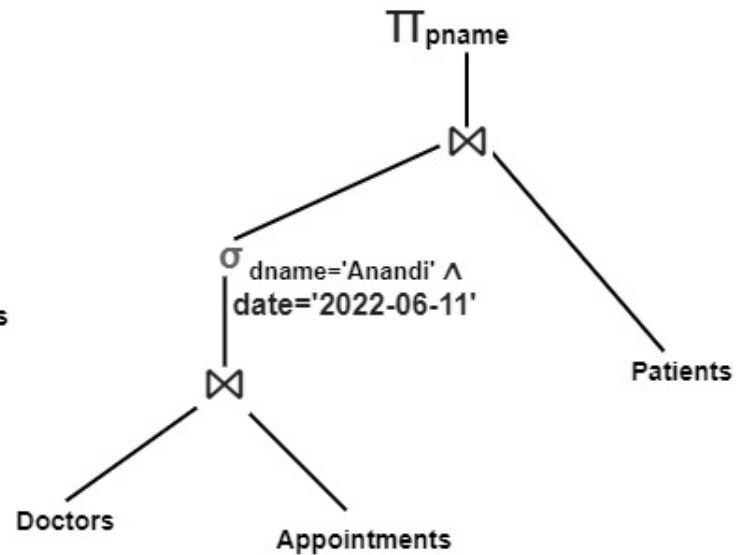
Two query trees are given below.

Figure 1:

Figure 2:

Identify the correct statement for the above two query trees.

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

b) Two query trees are not equivalent as neither selection nor projection operation can be carried out before performing the natural join.

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

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

Answer: c)
Explanation: Two query tree is 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.
If projection with needed attribute is carried out before join, query processing becomes efficient.
Hence, option (c) is correct.

Question 8

Consider the following relational schema: *Marks: 2* **MSQ**
```
Patients(pid, pname, age, p_email)
Doctors(did, dname, qualification, d_email)
Appointments(pid, did, date, time)
```

Four relational algebra queries are given below:

Q1: $(\sigma_{\text{age>20}}(\sigma_{\text{date='2022-06-12'}}(\text{Patients} \bowtie \text{Appointments})))$

Q2: $(\sigma_{\text{date='2022-06-12'}}(\sigma_{\text{age>20}}(\text{Patients} \bowtie \text{Appointments})))$

Q3: $\Pi_{\text{dname}}(\Pi_{\text{dname,d\_email}}(\text{Doctors} \bowtie \text{Appointments}))$

Q4: $\Pi_{\text{dname,d\_email}}(\text{Doctors} \bowtie \text{Appointments})$

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: a), d)
Explanation: Q1 and Q2 will give the same result because according to `transformation rules`, selection operations are commutative.
Q3 and Q4 will not give the same result because according to `transformation rules`, only the last in a sequence of projection operations is needed, the others can be omitted.
Hence, options a) and d) are correct.

Question 9

Consider the following relational schema: *Marks: 2* **MCQ**
```
Patients(pid, pname, age, p_email)
Doctors(did, dname, qualification, d_email)
Appointments(pid, did, date, time)
```

Consider the following query of the above relation:

```
Select Patients.pname, Doctors.dname
From Patients, Doctors, Appointments
Where Patients.pid=Appointments.pid and Doctors.did=Appointments.did and
Patients.age=50;
```

Identify the most optimized relational algebra for the above query using heuristic based query optimization technique.

a) $\Pi_{\texttt{Patients.pname,Doctors.dname}}$
   $(\sigma_{\texttt{Patients.pid=Appointments.pid} \land \texttt{Doctors.did=Appointments.did} \land \texttt{Patients.age=50}}$
   $(\texttt{Patients} \times (\texttt{Doctors} \times \texttt{Appointments})))$

b) $\Pi_{\texttt{Patients.pname,Doctors.dname}}((\Pi_{\texttt{Doctors.did,Doctors.dname}}(\texttt{Doctors}))$
   $\bowtie (\texttt{Appointments} \bowtie (\sigma_{\texttt{Patients.age=50}}$
   $(\Pi_{\texttt{Patients.pid,Patients.pname,Patients.age}}(\texttt{Patients})))))$

c) $\Pi_{\texttt{Patients.pname,Doctors.dname}}((\Pi_{\texttt{Doctors.dname}(Doctors)}) \bowtie (\texttt{Appointments} \bowtie$
   $(\sigma_{\texttt{Patients.age=50}}(\Pi_{\texttt{Patients.pid,Patients.pname,Patients.age}}(\texttt{Patients})))))$

d) $\Pi_{\texttt{Patients.pname,Doctors.dname}}(\sigma_{\texttt{Patients.age=50}}(\texttt{Patients} \bowtie (\texttt{Doctors} \bowtie \texttt{Appointments})))$

Answer: b)

Explanation: Performing selection and projection as early as possible reduces the size of the relations that are to be joined later. So, it is better to perform projection with required attribute and then perform selection.

Hence, option b) is correct.

Question 10

Consider the Relational Algebra expression given below:
P ⋈$_\theta$ (Q ∩ R) where P, Q, and R are relational algebra expressions.

Identify the correct equivalent relational algebra expression of the above relational algebra expression. *Marks:2* **MCQ**

a) P ⋈$_\theta$ (Q ⋈$_\theta$ R)

b) (P ⋈$_\theta$ Q) ∩ (P ⋈$_\theta$ R)

c) ((P ⋈$_\theta$ Q) − R)

d) (P ⋈$_\theta$ Q) − (P ⋈$_\theta$ R)

Answer: b)
Explanation: As per the equivalence rules of Relational Algebra expressions.