NOC24-CS75
Data Base Management System

Tutorial by: Adwait P. Parsodkar
Week - 8

Question - 1

Assume that immediate database modification scheme is followed in this question. Consider the following log records for transactions T0, T1, T2, T3 and T4:

steps	Details of log
1	⟨T0,start⟩
2	⟨T0,A,200,400⟩
3	⟨T1,start⟩
4	⟨T1,B,400,800⟩
5	⟨T1,commit⟩
6	⟨checkpoint{T0} ⟩
7	⟨T2,start⟩
8	⟨T3,start⟩
9	⟨T2,C,800,1600⟩
10	⟨T3,D,500,900⟩
11	⟨T3,commit⟩
12	⟨T4,start⟩
13	⟨T4,E,300,700⟩

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) After recovery completion, value of A will be 200.

b) After recovery completion, value of C will be 800.

c) After recovery completion, value of D will be 500.

d) After recovery completion, value of E will be 700.

Solution - 1

Answer: a), 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.

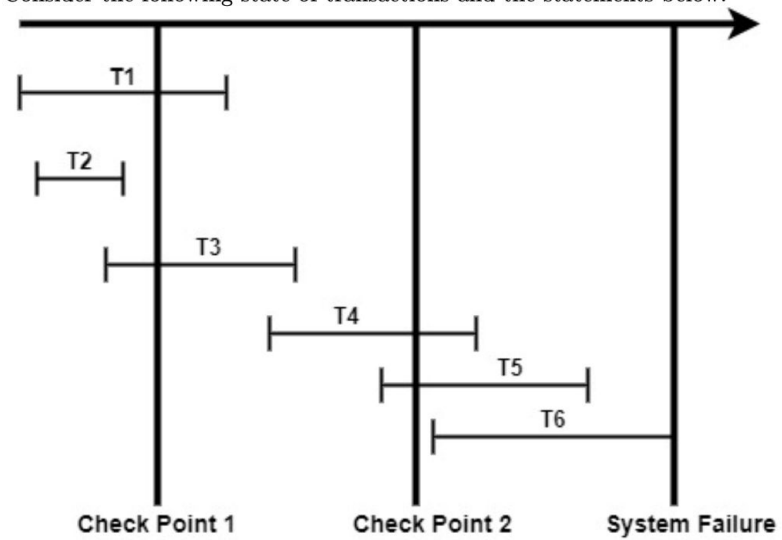
Any transactions that committed before the last checkpoint should be ignored(updates already output to disk due to the checkpoint).

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

As per the process of transaction recovery, options (a) and (b) are correct.

Question - 2

Consider the following state of transactions and the statements below.



1. T_1, T_2 and T_3 can be ignored.
2. T_2 and T_4 can be ignored.
3. T_3, T_4 and T_5 need to be redone.
4. T_4 and T_5 need to be redone.
5. Only T_6 needs to be undone.

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), 4), 5)

Solution - 2

Answer: c)
Explanation: Any transaction that is committed before the last checkpoint should be ignored. Therefore, T_1, T_2 and T_3 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_4 and T_5 are to be redone.
Any transaction that was running at the time of failure, needs to be undone and restarted. Hence, only T_6 is to be undone.
Hence, option (c) is correct.

Question - 3

Assume that `immediate database modification scheme` is followed in this question. Consider the following log records for transactions T0, T1, T2, T3 and T4:

steps	Details of log
1	⟨T0,start⟩
2	⟨T0,X,100,200⟩
3	⟨T1,start⟩
4	⟨T1,Y,200,300⟩
5	⟨T1,commit⟩
6	⟨checkpoint{T0} ⟩
7	⟨T2,start⟩
8	⟨T2,Z,300,500⟩
9	⟨T2,commit⟩
10	⟨T3,start⟩
11	⟨T3,W,800,1000⟩
12	⟨T3,commit⟩
13	⟨T4,start⟩
14	⟨T4,P,1000,1200⟩

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

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

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

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

Solution - 3

Answer: c)
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 {T2, T3} and undo list contain transaction {T0, T4} and for transaction {T1} no need any action because it is already committed before checkpoint.
Hence, option c) is the answer.

Question - 4

Identify the `cost estimation` of a `query evaluation plan`, if 6000 blocks are required to be `transferred from the disk` and the required number of `disk seeks` are 30.

- Time to transfer one block: $t_T = 2$ milliseconds.
- Time for one seek: $t_S = 0.2$ seconds.

a) 12.2 Seconds

b) 16 Seconds

c) 18 Seconds

d) 48 Seconds

Solution - 4

Answer: c)
Explanation: Cost for b block transfers plus S seeks will be $(b * t_T + S * t_S)$ seconds
$=(6000 * 2 * 10^{-3}) + (30 * 0.2)$ seconds
$= (12 + 6)$ Seconds
$= 18$ Seconds
For more details refer to 38.12 of lecture material.
Hence, option c) is the answer.

Question - 5

Let us consider the following statistics for two relations `Professor` and `Appointment_Details`:

- Number of records of `Professor`: $n_{\texttt{Professor}} = 4000$.
- Number of blocks of `Professor`: $b_{\texttt{Professor}} = 50$.
- Number of records of `Appointment_Details`: $n_{\texttt{Appointment\_Details}} = 2000$.
- Number of blocks of `Appointment_Details`: $b_{\texttt{Appointment\_Details}} = 20$.

Let us consider a natural join of `Professor` and `Appointment_Details` relations (`Professor` $\bowtie$ `Appointment_Details`).
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 `Professor` as the outer relation.

a) 120050 block transfers

b) 120000 block transfer

c) 80050 block transfers

d) 80000 block transfers

Solution - 5

Answer: c)
Explanation: Number of block transfers will be: $4000 \times 20 + 50 = 80050$, if `Professor` is taken as the outer relation.
For more details refer to 38.31 of lecture material.

Question - 6

Consider the following relational schema:

```
Professor(pid, pname, qualification, emailID)
Department(dname, building, budget)
Appointment_Details(pid, dname, date, time)
```

Two query trees are given below.

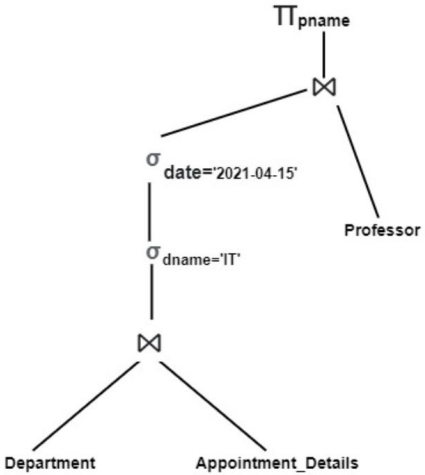


Figure 1:

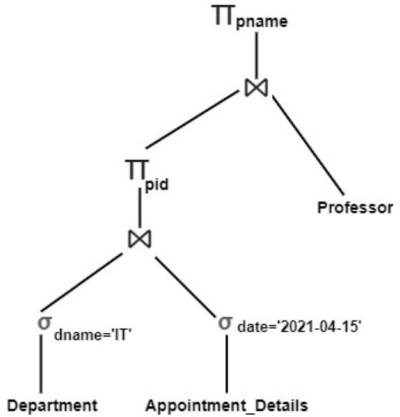


Figure 2:

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 selection or projection operation cannot be carried out before or after the natural join operation.

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.

Solution - 6

Answer: d)
Explanation: Two query tree is equivalent and Figure 2 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 (d) is correct.

Question - 7

Consider the following relational schema:

Professor(pid, pname, qualification, emailID)
Department(dname, building, budget)
Appointment_Details(pid, dname, date, time)

Four relational algebra queries are given below:

Q1: $(\sigma_{\texttt{building='Watson'}}(\sigma_{\texttt{date='2021-04-15'}}(\texttt{Department} \bowtie \texttt{Appointment\_Details})))$

Q2: $(\Pi_{\texttt{building, date}}(\texttt{Department} \bowtie \texttt{Appointment\_Details}))$

Q3: $\Pi_{\texttt{pname}}(\texttt{Professor} \bowtie \texttt{Appointment\_Details})$

Q4: $\Pi_{\texttt{pname}}(\sigma_{\texttt{Professor.pid=Appointment\_Details.pid}}(\texttt{Professor} \times \texttt{Appointment\_Details}))$

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.

Solution - 7

Answer: b), c)
Explanation: Q1 and Q2 will not give the same result because Q2 will print only building and date column over natural join of `Department` and `Appointment_Details` .
Q3 and Q4 will give the same result because $\sigma_{\texttt{Professor.pid=Appointment\_Details.pid}}$ is followed by (`Professor`$\times$`Appointment_Details`) is equivalent of (`Professor` $\bowtie$ `Appointment_Details`).
Hence, options b) and c) are correct.

Question - 8

Consider the following relational schema:
Professor(pid, pname, qualification, emailID)
Department(dname, building, budget)
Appointment(pid, dname, date, time)

A relational algebra expression given below:
$\Pi_{\text{pname}}(\Pi_{\text{pid,pname}}(\sigma_{\text{building='Packard'}}(\text{Professor} \bowtie \text{Appointment} \bowtie \text{Department})))$

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

a) $\Pi_{\text{pname,pid}}(\sigma_{\text{building='Packard'}}(\text{Professor} \bowtie \text{Appointment} \bowtie \text{Department}))$

b) $\Pi_{\text{pname,pid}}(\text{Professor} \bowtie (\text{Appointment} \bowtie (\sigma_{\text{building='Packard'}}(\text{Department}))))$

c) $\Pi_{\text{pname,pid}}(\text{Professor} \bowtie (\text{Appointment} \bowtie \Pi_{\text{dname}}(\sigma_{\text{building='Packard'}}(\text{Department}))))$

d) $\Pi_{\text{pname}}(\text{Professor} \bowtie (\text{Appointment} \bowtie (\sigma_{\text{building='Packard'}}(\text{Department}))))$

Solution - 8

Answer: d)
Explanation: According to the `transformation rules`, only the last in a sequence of projection operations is needed, the others can be omitted.
If we can do the selection operation as early as possible which reduces 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: $\sigma_{\theta 1}(\mathtt{P} \bowtie_{\theta 2} \mathtt{Q}) = (\mathtt{P} \bowtie_{\theta 1 \wedge \theta 2} \mathtt{Q})$

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

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

Identify the correct statement(s) from the followings.

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.

Solution - 9

Answer: a)
Explanation: Natural join operations are associative. Selections operations can be combined with Cartesian products and theta joins.
As per the equivalence rules of relational algebra expressions, option a) is correct.
For more details refer to Module 39, slides 11,12

Question - 11

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	⟨T1,start⟩
3	⟨T0,X,300,500⟩
4	⟨T1,Y,400,600⟩
5	⟨T0,commit⟩
6	⟨checkpoint{T1} ⟩
7	⟨T2,start⟩
8	⟨T1,commit⟩
9	⟨T3,start⟩
10	⟨T4,start⟩
11	⟨T2,Z,200,300⟩
12	⟨T3,W,300,800⟩
13	⟨T4,V,200,100⟩
14	⟨T2,commit⟩

If there is a crash just after step 14 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 300.

b) After recovery completion, value of Y will be 600.

c) After recovery completion, value of W will be 300.

d) After recovery completion, value of V will be 100.

Solution - 11

Answer: b), c)
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 contain transactions {T1, T2} and undo list contain transactions {T3, T4} and for transaction {T0} no need any action because it is already committed before checkpoint.
Hence, options (b) and (c) are correct.

Question - 12

Identify the `cost estimation` of a `query evaluation plan`, if 1000 blocks are required to `transferred from the disk` and the required number of `disk seeks` are 30.

- Time to transfer one block: $t_T = 4$ milliseconds.
- Time for one seek: $t_S = 0.3$ seconds.

a) 10 Seconds

b) 13 Seconds

c) 15 Seconds

d) 25 Seconds

Solution - 12

Answer: b)
Explanation: Cost for b block transfers plus S seeks will be $(b * t_T + S * t_S)$ seconds
$=(1000 * 4 * 10^{-3}) + (30 * 0.3)$ seconds
$= (4 + 9)$ Seconds
$= 13$ Seconds
For more details refer to 38.12 of lecture material.
Hence, option b) is the answer.

Question - 13

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

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

If there is a crash just after step 12 and the recovery of the system is successfully completed, identify the **correct** action for the above scenario.

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

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

c) No Action: T1; Redo: T2, T3; Undo: T4

d) No Action: T1; Redo: T2, T4; Undo: T3

Solution - 13

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

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

Hence, option b) is the answer.

Question - 14

Let us consider the following statistics for two relations of `Books` and `Authors`:

- Number of records of `Books`: $n_{Books} = 3000$.
- Number of blocks of `Books`: $b_{Books} = 300$.
- Number of records of `Authors`: $n_{Authors} = 1500$.
- Number of blocks of `Authors`: $b_{Authors} = 30$.

Identify the required number of `block transfers` and `seeks` if the smaller relation fits entirely in memory, use that as the inner relation using `Nested-loop join`.

a) 4830 block transfers and 2 seeks

b) 4500 block transfer and 2 seeks

c) 330 block transfers and 2 seeks

d) 300 block transfers and 2 seeks

Solution - 14

Answer: c)
Explanation: Number of block transfers will be: $300 + 30 = 330$, 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 - 15

Which of the following statement(s) is (are) **true**?

a) The `log` is a sequence of `log records`, which maintains information about update activities on the database.

b) The `log` is a sequence of `schedules`, which maintains information about each `schedules topological orderings`.

c) `Physical blocks` are those blocks which reside temporarily in the main memory.

d) `Physical blocks` are those blocks residing on the disk.

Solution - 15

Answer: (a), (d)

Question - 16

Consider the Relational Algebra expression given below:
$(\mathtt{P} \bowtie_\theta \mathtt{Q}) \cap (\mathtt{P} \bowtie_\theta \mathtt{R})$ where P, Q, and R are relational algebra expressions.

Identify the correct equivalent relational algebra expression of the above relational algebra expression.

a) $\mathtt{P} \bowtie_\theta (\mathtt{Q} \bowtie_\theta \mathtt{R})$

b) $((\mathtt{P} \bowtie_\theta \mathtt{Q}) - \mathtt{R})$

c) $(\mathtt{P} \bowtie_\theta \mathtt{Q}) - (\mathtt{P} \bowtie_\theta \mathtt{R})$

d) $\mathtt{P} \bowtie_\theta (\mathtt{Q} \cap \mathtt{R})$

Solution - 16

Solution: (d)

P(A,B,C) Q(C,D) → PQ(A,B,C,D)
P(A,B,C) R(B,C,D) → PR(A,B,C,D)

P(A,B,C) Q(C) → PQ(A,B,C)
P(A,B,C) R(B) → PR(A,B,C)

Question - 17

Consider the following relational schema: Students(student_id, student_name, address, email)
Courses(course_id, course_name, credits)
Registrations(student_id, course_id, semester, grade)
Two query trees are given below.

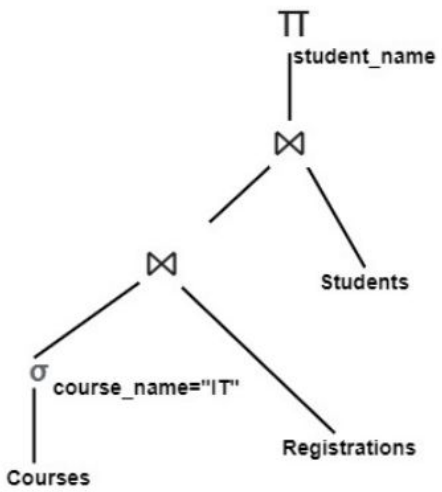


Figure 1:

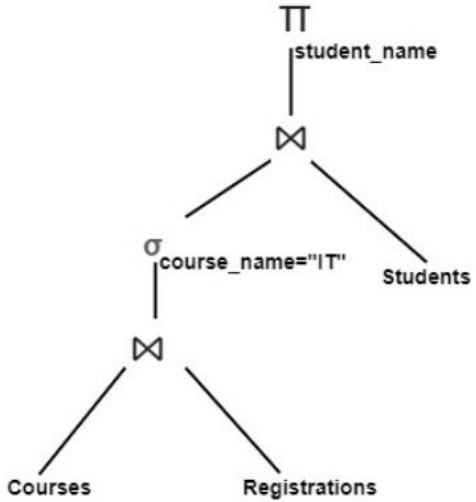


Figure 2:

Question - 17

Identify the correct statement for the above two query trees.

a) Two query trees are equivalent as selection operation can be carried out before or just after performing the natural join.

b) Two query trees are not equivalent as selection operation can not 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.

Solution - 17

Answer: a), c)
Explanation: Two query trees are equivalent, and the query tree of Figure 1 will lead to more efficient query processing. Performing the selection operation as early as possible reduces the size of the relation to be joined. If Selection with a condition is carried out before the join, query processing becomes efficient.
Hence, options (a) and (c) are correct.

Question - 18

Consider the following relational schema:
Students(student_id, student_name, address, email)
Courses(course_id, course_name, credits)
Registrations(student_id, course_id, semester, grade)
Four relational algebra queries are given below:

Q1: $\Pi_{\texttt{student\_name, email}}(\sigma_{\texttt{semester='Fall 2023'}}(\texttt{Students} \bowtie \texttt{Registrations}))$

Q2: $\Pi_{\texttt{student\_name, email}}(\texttt{Students} \bowtie_{\texttt{Students.student\_id = Registrations.student\_id}} (\sigma_{\texttt{semester='Fall 2023'}}(\texttt{Registrations})))$

Q3: $\Pi_{\texttt{course\_name}}(\sigma_{\texttt{semester='Spring 2023'}}(\texttt{Courses} \bowtie \texttt{Registrations}))$

Q4: $\Pi_{\texttt{course\_name}}(\texttt{Courses} \bowtie_{\texttt{Courses.course\_id = Registrations.course\_id}} (\Pi_{\texttt{course\_id}}(\sigma_{\texttt{semester='Spring 2023'}}(\texttt{Registrations}))))$

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.

Solution - 18

Answer: a), c)

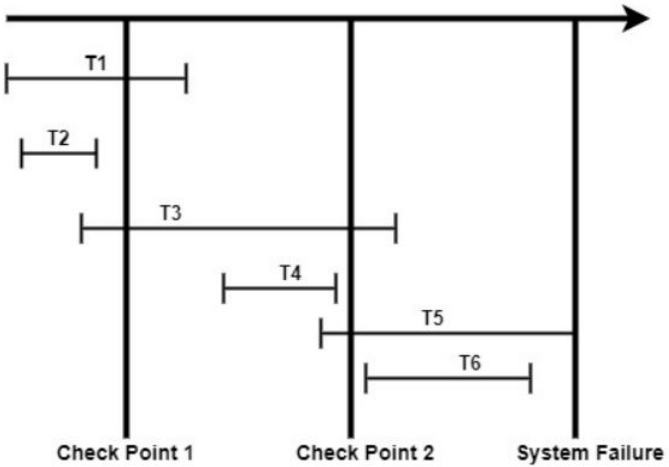
Explanation: Q1 and Q2 will give the same result because according to `transformation rules`, performing the selection as early as possible reduces the size of the relation to be joined.

Q3 and Q4 also give the same result because according to `transformation rules`, performing the selection as early as possible reduces the size of the relation to be joined. Then projection $\Pi_{\texttt{course\_id}}(\sigma_{\texttt{semester='Spring 2023'}}(\texttt{Registrations}))$ operation finds the required values for join operations $(\texttt{Courses} \bowtie \texttt{Registrations})$.

Hence, options a) and c) are correct.

Question - 19

Consider the following state of transactions and the statements below.



1. `T1, T2 and T4` can be ignored.
2. `T1, T2 and T3` can be ignored.
3. `T3 and T6` needs to be redone.
4. `T3, T4 and T6` needs to be redone.
5. Only `T5` need to be undone.

Identify the correct group of statements from the options below.

a) 1), 2), 3)

b) 1), 3), 5)

c) 2), 4), 5)

d) 2), 3), 5)

Solution - 19

Answer: b)
Explanation: Any transaction that is committed before the last checkpoint should be ignored. Therefore, `T1, T2 and T4` 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, `T3, T6` need to be redone.
Any transaction that was running at the time of failure, needs to be undone and restarted. So, `T5` need to be undone.
Hence, option (b) is the correct answer.