COMP90050 Advanced Database Systems: Tutorial
Winter term, 2023 (Week 2)

Exercises

Part 1:

1. Which of the following RAID configurations that we saw in class has the lowest disk space utilization? Your answer needs to have explanations with calculations for each case.

   (a) RAID 0 with 2 disks

   (b) RAID 1 with 2 disks

   (c) RAID 3 with 3 disks

Where does this lack of utilization of space go, i.e., where we can use such a configuration as it has some benefits gained due to the loss of space utilization?

2. What is the Mean time to failure values of different RAID systems?
   a. RAID 0 with 2 disks
   b. RAID 2 with 2 disks
   c. RAID 1 with 2 disks
   d. RAID 1 with 3 disks
   e. RAID 3 with 3 disks
   f. RAID 4 with 3 disks
   g. RAID 5 with 3 disks
   h. RAID 6 with 5 disks

3. In a Failvote system, which of the following cases we can accept an action?

Total number of devices	Number of agreeing devices	Accept?
10	6	
10	5	
10	4	
5	3	
5	2	

4. In a Failfast system, which of the following cases we can accept an action?

Total number of devices	Number of working devices	Number of agreeing devices	Accept?
10	6	4	
10	6	3	
10	5	3	
5	5	3	
5	4	2	
5	2	2	
5	1	-	

5. There are two nodes in a network that use stable storage and acknowledgment message passing for reliable communication. The stable storage of Node A contains the following record - Received message (In6); Transmitted message(Out3); Out:3 Ack:3 In:6. The stable storage of Node B contains the following record - Received message (In3); Transmitted message(Out6); Out:6 Ack:6 In:3.

Now Node B sends a new message 7 to Node A. What will be in the stable storage of A and B if the message is received correctly, including a correctly received acknowledgement?

6. There are two nodes in a network that use stable storage and acknowledgment message passing for reliable communication. The stable storage of Node A contains the following record - Received message (In6); Transmitted message(Out3); Out:3 Ack:3 In:6. The stable storage of Node B contains the following record - Received message (In3); Transmitted message(Out6); Out:6 Ack:6 In:3.

Now Node B sends a new message 7 to Node A. **What will change in the stable storage of A and B if the message is received correctly, but the acknowledgement is not received?**

Part 2:

7. Discuss which query optimisation approach(es) (enumerating all plans, heuristic based, adaptive plans) can be suitable for the following scenarios:

- Scenario A: Given a table with 1000 tuples, run the following query:

```
SELECT customer
FROM Table
WHERE spend BETWEEN 100 AND 200
AND birth_year> 2000;
```

- Scenario B: Given 5 tables with 1000 tuples in each table, run a query:

```
SELECT T1.name, T2.salary, T3.qualification, T4.phone, T5.leader
FROM Table1 T1
   INNER JOIN Table2 T2 ON T2.id = T1.id
   INNER JOIN Table3 T3 ON T3.id = T1.id
   INNER JOIN Table4 T4 ON T4.id = T1.id
   INNER JOIN Table5 T5 ON T5.department = T1.department
WHERE T1.age > 50;
```

8. Review the examples on nested-loop join and block nested-loop join given in the lecture. Discuss and calculate why the later one can be more efficient.

9. A particular query on a table A used to run quite efficiently in a DBMS. After inserting many records and deleting many other records from table A, that same query is now taking more time to run, even when the total number of records has not changed. What can be the reason for that? What can you do as the user/database administrator of that DBMS to improve the performance of this query?

Part 3:

10. What indices are more suitable if a table is frequently used for finding records based on the following criteria: users' name, a range of users' birthday, and a spatial region covering users' residence?

11. Review the points on indexing with B+ trees. Assume a database table has 10,000,000 records and the index is built with a B+ tree. The maximum number of children of a node, is denoted as n. How many steps are needed to find a record if n=4? How many steps are needed to find a record if n=100?

12. Given the R-tree below please visit the nodes of the R-tree in a best-first manner as discussed in class to find the 1st nearest neighbour of query point "i". Is there anything peculiar that you notice while traversing an R-tree?