

COMP90050: Advanced Database Systems

Semester 1, 2023

Quiz 1: Solutions

*This document consists of the solutions for the Quiz 1 Multiple-choice questions. Note that, the **order** at which a question and the options appear for a particular student may **vary**. If the correct response option is chosen (as specified below), the mark will be awarded.*

1. Why solid-state drives do not have a seek time that is comparable to hard disk drives?

- SSDs are smaller in size
- Hard disk drives need to make backups more frequently
- SSDs do not have any moving arm
- SSDs have fast moving arm over the disk

Correct Response: SSDs do not have any moving arm

2. Machine A has a higher cache hit ratio than Machine B. The cache access time and the memory access time are two piece of information given for each machine that we do not know but we know that these two values are different for the two machines. Which machine has a faster effective memory access time?

- Machine B
- We cannot tell from the provided information in the question.
- Both machines have the same effective memory access time.
- Machine A

Correct Response: We cannot tell from the provided information in the question.

3. I decide to use a relational DBMS to store my data about my company. The company maintains customer, employee and supplier data in some tables. Which one of the reasons given below is a good reason not to store this data in some set of files that I manage myself, i.e. in comparison to keeping the data in a relational DBMS, and thus can be a good reason to choose a relational system.

- A friend will help me install the relational DBMS on my machine.
- I have many customers and employees accessing the data in a concurrent and complex manner over the web.
- I have specialized complex queries to run on this data.
- I want to save some storage space.

Correct Response: I have many customers and employees accessing the data in a concurrent and complex manner over the web.

4. A database stores the email address of customers who want to receive marketing information in the future. Customers can login to the database via a web interface. The customers can modify their email addresses after logging in. The database has one million records currently. Which of the following should be achieved?
- Different customers should not be modifying their own email addresses at the same time
 - A partial backup of the database is created every second and a full backup every minute
 - Data access should be faster than 10ms per customer to keep them happy
 - The data will not be lost if one of the database servers broke down

Correct Response: The data will not be lost if one of the database servers broke down

COMP90050: Advanced Database Systems

Semester 1, 2023

Quiz 2: Solutions

*This document consists of the solutions for the Quiz 2 Multiple-choice questions. Note that, the **order** at which a question and the options appear for a particular student may **vary**. If the correct response option is chosen (as specified below), the mark will be awarded.*

1. A restaurant database stores the name, cuisine type, GPS location, phone number as well as average expert rating of restaurants in a table. Customers can access the database via their web browsers and the database access website contains a map for querying as well. The main query type that is issued is a NN query by customers where given a point location on the map, the system shows the nearest restaurant. Which one of the following is the best index to create on this table for fast access?

- Hash index on phone numbers
- Hash index on GPS location
- R-tree index on GPS location
- Bitmap index on cuisine type

Correct Response: R-tree index on GPS location

2. Which one of the following statements is incorrect?

- Bitmap index is not commonly used in data analysis.
- A B+ tree index is a balanced index.
- Hash index can be used for fast access to individual records with id type attribute.
- Quadtree index is suitable for nearest neighbour queries on multi-dimensional data.

Correct Response: Bitmap index is not commonly used in data analysis.

3. Which of the following statements is true for transaction processing?

- Running transactions concurrently but not in isolation with each other is not desirable.
- SQL queries are the main atomic units of execution in relational DBMSs.
- Transactions are expected to be run obeying Chemical properties such as properties of acids and metals.
- Durability is the least important ACID property.

Correct Response: Running transactions concurrently but not in isolation with each other is not desirable.

4. If I have a hash index given as follows, what should we expect to happen during query processing using this index for accessing the associated table? The hash index is created on the "id" attribute of a table containing info about more than a billion social network users. Each "id" value is unique and generated per user at the time of creation of their social network account. This index is used to access the info associated with a user, i.e., given their id, e.g., by the system administrators. For example, administrators access a user's info during a service phone call from the customer taking the id value and entering to the system to retrieve the associated customer row from the customer info table. The DBMS uses a simple hash function which converts this big integer "id" number which has 20 decimal digits to binary form and takes the middle 4 bits out of the binary representation and uses it as the bucket number to access a bucket in the associated hash table. In this case we expect that:

- As this is a very simple hash function it will lead to slower than expected data access.
- The hash index finds the data location on the disk as quickly as it could be as there is an even distribution of ids to hash table buckets.
- The hash function is simple but elegant and thus saves time in data access and cannot be better.
- None of the above is relevant as we do not know which four bits are read per "id" instance at this time, i.e., 0000 or 1111 or etc.

Correct Response: As this is a very simple hash function it will lead to slower than expected data access.

COMP90050: Advanced Database Systems

Semester 1, 2023

Quiz 3: Solutions

*This document consists of the solutions for the Quiz 3 Multiple-choice questions. Note that, the **order** at which a question and the options appear for a particular student may **vary**. If the correct response option is chosen (as specified below), the mark will be awarded.*

1. Which statement is correct?
 - Spinlocks do not need any hardware support in general
 - It is efficient to do concurrency control with Dekker's algorithm because the algorithm relies on a lot of hardware support
 - Test and set implementation of spinlock requires memory locking support
 - OS supported interrupt calls are fast and cheap because they are done at OS level

Correct Response: Test and set implementation of spinlock requires memory locking support

2. Which action is not suitable for nested transactions?
 - Subtransactions do not have durability property until ancestors commit
 - Objects at parents can be made visible to children
 - All children as well as parent transaction can run in parallel together
 - Commit of a subtransaction makes data available to parents

Correct Response: All children as well as parent transaction can run in parallel together

3. If a transaction A is running concurrently with transaction B then the execution order should be equal to the outcome of which one of the following orders?
 - B running before A but not A running before B
 - A running before B or B running before A

Quiz 3 Solutions

- It does not really matter for these two specific transactions
- A running before B but not B running before A

Correct Response: A running before B or B running before A

4. Why do we really need to write code as Embedded or Dynamic SQL?

- Because only this way we can write our queries and save them for later use as well
- Because using another language in addition to SQL gives us more power to develop applications
- All of the choices given in this question are correct
- Because this is the main way we can learn about errors in our SQL queries

Correct Response: Because using another language in addition to SQL gives us more power to develop applications

COMP90050: Advanced Database Systems

Semester 1, 2023

Quiz 4: Solutions

*This document consists of the solutions for the Quiz 4 Multiple-choice questions. Note that the **order** at which a question and the options appear for a particular student may **vary**. If the correct response option is chosen (as specified below), the mark will be awarded.*

1. Which one of the following statements is correct?

- If $DEP(H1) > DEP(H2)$ that means if H1 has more wormholes than H2
- If $DEP(H1) = DEP(H2)$ then both histories H1 and H2 will have a wormhole transaction
- If $DEP(H1)$ is not equal to $DEP(H2)$ at least one of the histories is serializable
- If H1 is a serial execution of some transactions then $DEP(H1)$ cannot have a wormhole in it

Correct Response: If H1 is a serial execution of some transactions then $DEP(H1)$ cannot have a wormhole in it

2. Given the following which one of the following is not true?

$$T' \in Before(T) \cap After(T)$$

- If T and T' are in a dependency graph then the history that lead to that graph cannot be equal to any of the serial execution orders of the transactions it contains
- T' is a wormhole transaction
- T' is a wormhole transaction and the history that this analysis of two transactions comes from is a serializable history
- T is a wormhole transaction

Correct Response: T' is a wormhole transaction and the history that this analysis of two transactions comes from is a serializable history

3. Two-phase locking implies that:

- All locks in a transaction has to be taken during a transaction but can only be released at the end
- If you unlock an object then you cannot obtain another lock in that transaction
- All locks of a transaction has to be declared upfront
- All locks in a transaction has to be taken at the beginning of the transaction and be released at the end of the transaction

Correct Response: If you unlock an object then you cannot obtain another lock in that transaction

4. Degree 3 of isolation implies:

- Transactions are running serially and not concurrently
- None of the other answers are correct
- We do not use locking for concurrency control
- The associated transactions will have no lost updates, no dirty reads and have repeatable reads

Correct Response: The associated transactions will have no lost updates, no dirty reads and have repeatable reads

COMP90050: Advanced Database Systems

Semester 1, 2023

Quiz 5: Solutions

*This document consists of the solutions for the Quiz 5 Multiple-choice questions. Note that the **order** at which a question and the options appear for a particular student may **vary**. If the correct response option is chosen (as specified below), the mark will be awarded.*

1. Which one of the following is true regarding concurrency control mechanisms?
 - Snapshot isolation does not guarantee serializability
 - When there are many updates through transactions submitted to a DBMS, we should prefer optimistic concurrency control to others
 - Two-phase locking is the same thing as two-phase commit
 - Time-stamp based concurrency control is the worst among all mechanisms we have seen

Correct Response: Snapshot isolation does not guarantee serializability

2. Which one of the following is not true for DBMSs?
 - Two-phase commit protocol would still work if participants do not save their state before voting
 - Two-phase commit is a mechanism that we can use to help run distributed transactions in an atomic manner
 - For two-phase commit we need to update the optimistic concurrency mechanism to have a globally unique order of transactions if that is the concurrency mechanism in use
 - If we have replication in DBMSs, to achieve one-copy serializability, we need to take measures against servers coming and going into the system to avoid issues with isolation of transactions

Correct Response: Two-phase commit protocol would still work if participants do not save their state before voting

3. Which one of the following is not true for RAID level 4 system?
 - In RAID level 4 MTTF is better than RAID level 3
 - In RAID level 4 parity information is used
 - In RAID level 4 parity information creates bottlenecks in writes
 - In RAID level 4 parity is done through exclusive or operation

Correct Response: In RAID level 4 MTTF is better than RAID level 3

4. Which one of the following voting based systems could be deemed as giving the right answer?
- Failfast system with 10 modules, only 3 alive and 2 agrees
 - Failfast system with 5 modules, all alive but none agreeing with each other
 - Failfast system with 10 modules, 10 alive and 5 agreeing
 - Failfast system with 10 modules, all failed

Correct Response: Failfast system with 10 modules, only 3 alive and 2 agrees