

University of Melbourne, School of Computing & Information Systems

COMP90050 Advanced Database Sys., Winter Sem. 2021 Final Exam

- This examination is worth of **50% of your final mark**. It is a 2 hour exam with 15 minutes reading time.
- There are 10 questions in the exam and a total of 4 pages including this cover page.
- You may access other materials during the exam. [Advise on accessing materials for the exam: You only need a text editor to answer this exam's questions in our view. No other materials are really needed. Calculators should not be needed as questions require simple calculations to answer. Referring to the book or other resources throughout this exam should not be of much help for this exam if you understood the contents of the subject and studied properly for the exam, and rather you may see that if you spend a lot of time looking through the pages of books, you may be losing time that you could have used answering questions properly. This is a short answer exam and all answers should be in your own words, calculations, and your own drawings. You should not put citations to other documents/papers/books/material, nor copy/paste any material into the exam paper.]
- Attempt all questions as partial marks will be available. No question requires writing lengthy answers. Please be clear and brief as you may lose points for unclear or redundant descriptions.
- The values in square brackets after questions show the marks allocated to each question.
- You are welcome to use the text editor of your choice to edit your answers. You do not need to repeat the questions themselves in your answers. Just make sure you use the right question number per answer. Type your answers: No handwritten answers! Drawings should also be done electronically. Note: no question requires extensive drawing effort!
- Start your answer document by writing your student ID on top of your document, e.g., "Student ID: ...". Then write "COMP90050 Final Exam Answers" in the next line. Then go to the line after that and start with your answers. Start each answer with a proper heading such as "Question 1 Answer: ..." and so on.
- Please answer questions in the given order and separate the answers with a few blank lines to give some space between the answers to different questions.
- Make sure to save your progress locally and regularly during the exam and at the end as a **PDF version** as well. Upload only that PDF version and when you are finished. We recommend not leaving the uploading of the PDF version to the last minute!

Question 1: [6 Marks]

We have a database server with a powerful CPU at a company that runs many queries for their business purposes. Unfortunately, this machine does not have a large memory. It has a cache that the CPU accesses while processing queries. When items are not found in the cache they are retrieved from the main memory. We heard that we also have to commonly go to the hard disk on this machine to get data as well i.e., most data is not found even in the main memory. We learn that the cache access time is C seconds for this machine while memory access time is 10 times slower than that. Only 2 out of 10 requests to the cache are in general successful. In addition, the hard disk access time is 20 times slower than the disk buffer access time where in this architecture the disk buffer is implemented as a part of the main memory. Finally, we also hear that the same hit ratio applies to the disk buffer as well, i.e., the hit ratio to the cache is the same as the hit ratio to the disk buffer. Given these please find the overall Effective Access time as a multiple of C for accessing data in this database using this server. Show your calculations.

Question 2: [6 Marks]

Recall the hash index concept we saw in class. If we were to use a database management system that implements a hash index on the *first_name* attribute of an employee table in the following way, then describe what would be the advantages and disadvantages that come with such an implementation. The hash function of the index uses the first character of the associated string attribute and finds the first bucket that has links to the records on a disk where that attribute's value starts with that same character. So "Adam" starts with "A" and our index takes you to the memory location where disk references to names that start with "A" are. In addition, in this hypothetical indexing scheme, the associated table is sorted on the disk with respect to *first_name* attribute in ascending order.

Question 3: [6 Marks]

We are given the following transaction history:

$H = \langle (T1, R, O1), (T2, W, O4), (T1, R, O3), (T3, W, O1), (T5, R, O3), (T3, W, O2), (T5, R, O4), (T4, R, O2), (T6, W, O4) \rangle$

please find the $DEP(H)$ and show it as a simple graph. Then using the concept of wormholes explain whether this history is equal to a serial history or not, i.e., if it is not give a wormhole example, and if it is then give a serial execution of these transactions that this history is equal to. Briefly explain your answers.

Question 4: [6 Marks]

We want to use the Cyclic Redundancy Check method for data storage as a means to deal with errors. Please compute the additional bits that we need to store with the original data, given the following information. Show your steps and calculations. Original data is 10100101 and your divisor polynomial is $x^4 + x^2 + x + 1$. [*Important Note:* When showing your calculations you can use the following notation, but also note that the numbers given in the notation example below have nothing to do with the question above and the operations there should not be taken as an indication how your solution would be done. This example below is there simply to give you an example means that you can use to easily enter your solution in electronic form with your editor, while avoiding complicated drawings/diagrams. You are welcome to use a table instead.]

Sample Notation:

1 0 0 0 0 1 1 1 1 0 0

1 0 0 1

0 0 0 1...

operation continues...

Question 5: [3 Marks]

Given the following operations for the two transactions T and U, give three concurrent execution of these transactions that are equal to a serial execution (do not assume any concurrency mechanism for this question) and also give which serial execution order each one equals to as well:

T has operations in the following order: {read(B); read(A); write(B); write(A);}

U has operations in the following order: {read(C); write(A); read(B); write(C);}

Question 6: [5 Marks]

Recall Distributed Transactions concepts we have seen in class. If the following two transactions are given to you with their operations:

T: {read(A); write(A); read(B); write(B);}

U: {read(B); write(B); read(A); write(A);}

and we know that there are two servers X and Y, create an execution order of these transaction operations that run concurrently using the two servers and in each server they seem to be running equal to a serial execution but globally there is no serial execution that the overall order is equal to. Do not assume any concurrency mechanism for this question and there is no replication in this particular system. You need to distribute objects A and B to servers X and Y and you can do this in any way you like as long as it adheres to the other parameters of this question given above.

Question 7: [5 Marks]

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. (1) RAID 0 with 2 disks, (2) RAID 1 with 2 disks, (3) 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?

Question 8: [6 Marks]

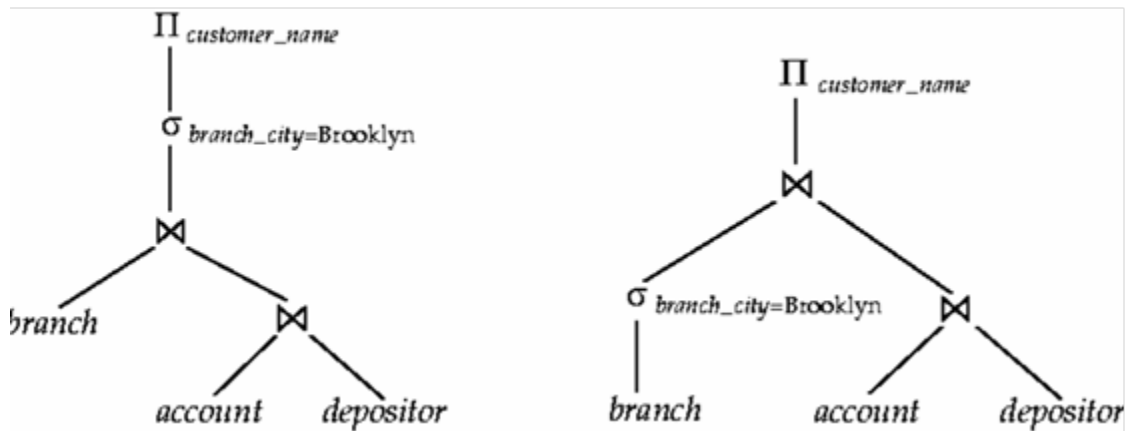
Shadow paging is a mechanism one can use in database recovery. First, briefly explain how this method works, then briefly discuss the benefits and disadvantages in comparison to log based systems, finally, give one realistic example that this type of mechanism can be preferred and used in comparison to logging.

Question 9: [3 Marks]

In a new database management system, we decide to introduce a simple locking mechanism that uses only XLOCKS and strict two-phase locking (for both reading and writing objects). What happens to concurrency control under this new regime? Discuss benefits and disadvantages? What transactions can run concurrency and which ones cannot in comparison to having SLOCKS as well?

Question 10: [4 Marks]

In the following figure we see two equal expressions a query optimizer is looking at to decide which one to run eventually:



Please describe which one of these plans the optimizer should choose and why. Is this choice true in general? If so, where in query optimization such an observation can be used? Briefly explain.

...END OF EXAM