University of Melbourne, School of Computing & Information Systems
COMP90050 Advanced Database Systems

Semester 1, 2022, Final Exam

- This examination is worth of **50 marks**. It is a **2 hour exam** with **15 min. reading** time. There are 11 questions in the exam and a total of 3 pages including this cover page. The values in square brackets after questions show the marks allocated to each.
- 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 mainly a short answer exam and all answers should be in your own words, own calculations, and own drawings. You should not copy/paste any material into the exam paper. All work should be your own work!]
- Attempt all questions as partial marks will be available. No question requires writing lengthy answers, be clear, as you may lose points for unclear/redundant descriptions.
- 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. Also answer questions in the given order. Type your answers: No handwritten or scanned 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 only 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! There will be some upload time at the end of the exam and submission server regularly slows down at the end with many people uploading!
- **In the exam**, if you have questions about the content, **use the LMS Exam Support**.
- **Monitor LMS Announcements** in case we make announcements about the exam.
- **Technical support** for this exam can be accessed at:
  https://students.unimelb.edu.au/your-course/manage-your-course/exams-assessments-and-results/exams/technical-support
  Information on **LMS Assignment exams** with some **troubleshooting** tips can be found: https://students.unimelb.edu.au/your-course/manage-your-course/exams-assessments-and-results/exams/how-do-i-take-my-exam/formats/LMS-assignments

Question 1: [4 Marks]

In one paragraph, compare Relational DB systems with Object Oriented DB Systems.

Question 2: [6 Marks]

We have seen the B+tree concept for indexing in class. These trees are deemed as an improvement over binary trees with their large fan out e.g., in the range of 100s. i) What is the benefit of making such a decision with fan outs, especially in the context of DBMSs? Briefly explain. ii) What are the benefits and disadvantages of using a B+tree in comparison to using hashing? Briefly explain.

Question 3: [6 Marks]

Given the following transaction history h we are told that k, l, m, n, o, and p are transactions, and operations are *Read* and *Write* operations which are labeled as R and W, and operations are done on the objects labeled as a, b, c, d, e:

$h = <$ (k,R,a),(m,W,a),(m,W,b),(n,R,b),(k,W,c),(l,W,e),(o,R,c),(o,R,d), (p,W,d) $>$

Please find the DEP(h) and draw as a simple graph version as well. Then using the concept of wormholes explain whether this history is equal to a serial history or not, i.e., if this history is not equal to a serial history then 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 steps with sentences while answering this question.

Question 4: [5 Marks]

What is cascading aborts? Does two-phase locking address them? If it does, explain how, but if it does not, then give a locking-based strategy that may address them, and then explain briefly how that strategy that you gave will address them.

Question 5: [4 Marks]

Given the k-d tree below with point data, where black labeled dots represent spatial coordinate data, and the rectangular area is divided into regions with the division order given with numbers: draw the associated k-d tree as a tree structure with leaves labeled as the data labels given below. Assume left subarea of a division goes to a left subtree, and lower subarea of a division also goes to a left subtree. The figure for the k-tree follows:
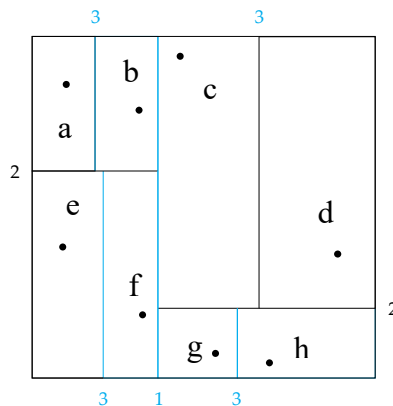
Question 6: [5 Marks]

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 11100001 and your divisor polynomial is $x^4 + x^2 + x + 1$.

Question 7: [4 Marks]

Compare ACID and BASE properties; what do they aim each; what are their differences, and advantages or disadvantages if a DBMS designer follows one or the other set. Briefly explain.

Question 8: [4 Marks]

Shadow paging is a mechanism one can use in database recovery. It is said that it is an approach where almost no recovery effort is needed after a crash. Explain why this may be the case.

Question 9: [4 Marks]

Given some data to be put on disks, which one of the following RAID configurations will have the best MTTF? Calculate each case and show your calculations with simple explanations. Assume individual MTTF of the disks in this question are the same. (1) RAID 0 with 3 disks, (2) RAID 1 with 4 disks (for this case first disk is mirrored 3 times), (3) RAID 3 with 3 disks, (4) RAID 4 with 3 disks.

Question 10: [4 Marks]

What is the difference between a classical Optimistic Concurrency control mechanism versus a Snapshot Isolation-based one. What is the implication of this difference? Briefly explain.

Question 11: [4 Marks]

Relation A has 1,000 records stored in 40 blocks that can be read consecutively. Relation B has 800 records stored in 50 blocks that can be read consecutively. For a SQL query with a join operation that joins these two relations, the query optimizer chooses to use block nested-loop join. Should the outer relation be A or B based on the costs that you may derive using the information above? Can we decide on this choice using the information provided? If so, show your calculations and then give your answer, if not briefly explain why not.

…END OF EXAM