*University of Melbourne, School of Computing & Information Systems*

# COMP90050 Advanced Database Systems

## Semester 1, 2023, 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.

**Continue to next page…**

**Question 1: [4 Marks]**

In one paragraph, discuss the importance of disks in DBMS design in its early years.

**Question 2: [6 Marks]**

We have seen two join algorithms in class. Describe in your own words how each of these algorithms work given two tables to join, with one paragraph to describe each algorithm. Then explain which one of these algorithms commonly works more efficiently than the other and why this is the case, explain with one paragraph.

**Question 3: [6 Marks]**

Given the following transaction history *h* we are told that a, b, c, d, e, f, and g 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 x, y, z, v, u, t. Here *h* is equal to

<(a,R,x),(c,W,x),(c,W,y),(f,R,t),(d,R,y),(a,W,z),(b,W,u),(e,R,z),(e,R,v),(g,W,v),(f,W,t)>

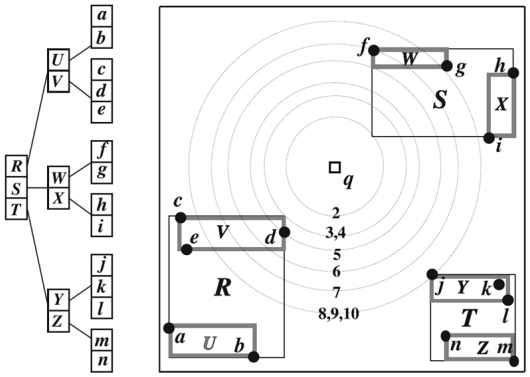
Please find the DEP(*h*) and draw as a simple graph version as well. Then using the concept of wormholes explain whether this history is equal to a serial history or not, i.e., if this history is not equal to a serial history then give a wormhole example, and if it is then give a serial execution of these transactions that this history is equal to. Briefly explain your steps with sentences while answering this question.

**Question 4: [5 Marks]**

What is the difference between Two-phase locking and a locking strategy where all locks are taken at the beginning of a transaction and released at the end of a transaction? Compare the two approaches in a few paragraphs, i.e. their benefits and disadvantages. Would there be any concurrency if the second method mentioned above is used. Briefly explain.

**Question 5: [6 Marks]**

Given the R-tree below with point data, where black labeled dots represent spatial coordinate data, and the tree structure is shown on the left of the figure, we also give a nearest neighbor query point "q" (and the circles represent distances of different entities to this query point). Run the nearest neighbor query that we saw in class with a priority queue step by step until the third nearest neighbor object is found. Explain briefly as well.




**Continue to next page…**

**Question 6: [4 Marks]**

What are the properties of an ideal hash function that we discussed in class. Describe each briefly. Then discuss why these properties are important for hash indices.

**Question 7: [4 Marks]**

Describe why a bank may prefer to refer to ACID properties than BASE properties in its database systems. At the same time why would an online shopping system may be built with mainly BASE properties in mind? Briefly explain each.

**Question 8: [3 Marks]**

Shadow paging is a mechanism one can use in database recovery. Despite its promise to be a fast recovery system discuss why it did not become a common recovery mechanism used by DBMS designers.

**Question 9: [3 Marks]**

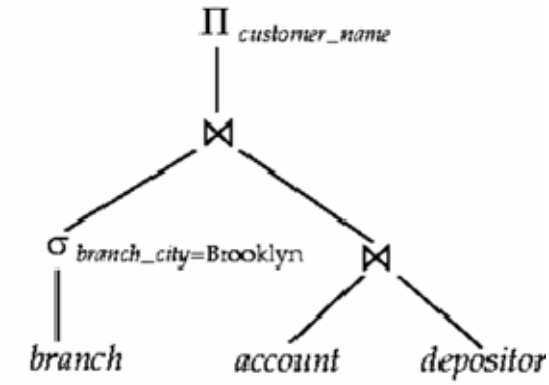
We have seen the supermodule concept in class with repairs. Explain in your own words even when we have disks where failures per disk may happen almost every other year, a supermodule using these disks may be fault tolerant at the level of many years or even hundreds of years. Briefly explain your rationale.

**Question 10: [6 Marks]**

Give an example concurrent execution of transactions, using three transactions of your choice/design, where the execution is an example for Degree 2 isolation and an unrepeatable read occurs. Show all the relevant details such as locks taken etc and briefly explain. Note: use a table where as time progresses transactions take their individual actions in time and they run concurrently.

**Question 11: [3 Marks]**

In the following figure we see an expression a DBMS is using to execute a query:




What happens if the DBMS does the selection operation later in the execution sequence given above, and in that case the last join operation given above gets to be done before the selection. Briefly explain.

***…END OF EXAM***