

Final_Exam_COMP90050_WIN_2023

Started: Aug 1 at 12:00

Quiz Instructions

Total number of questions: 13

Total marks: 50

Authorised Materials:

This is an open book exam – the lecture slides, lecture recordings, textbooks, and study materials can be used. Note that there is a strict time limit for the exam, hence you should be mindful of the time spent using such resources.

While you are undertaking this exam you ***must not***:

- make use of any messaging or communications technology.
- act in any manner that could be regarded as providing assistance to another student who is undertaking this assessment, or will in the future be undertaking this assessment.

The work you submit ***must be based on your own knowledge and skills***, without assistance from any other person.

Instructions to Students:

This exam begins at the time mentioned in 'Available from' in Canvas (lms.unimelb.edu.au). The exam must be completed and submitted by the mentioned 'Due time'. This exam has a strict time limit.

- **Once submitted, you will NOT be able to reopen the quiz to change your answers. This means that you will be able to click the "Submit" button ONLY once.**
- Answers should only contain simple text. You do NOT need to upload any image. DO NOT click anything in the editing toolbar, e.g., changing format, uploading media, adding equation, etc.
- **Questions have unequal marks.** There is no negative marking, and we encourage you to **attempt all questions**.

Question 1

1 pts

Which of the following system has higher reliability than the other? (i) System A with 4 devices on failvote (ii) System B with 5 devices on failvote.

- ☐ System A has higher reliability than System B
- ☐ System B has higher reliability than System A

- ☐ Both system A and system B have the same reliability

Question 2**1 pts**

When a database needs to join two tables using page-oriented nested loop join algorithm, which one of the following strategies will provide better efficiency?

- ☐ Take the table with the smaller number of pages as the outer relation
- ☐ Take the table with the smaller number of pages as the inner relation

Question 3**2 pts**

A flat transaction with save-points has the following statements. If condition1 is true once and the final commit is successful, what will be printed as the value of the count?

BEGIN WORK

count = 10

SAVE WORK 1

count = count+10

SAVE WORK 2

count = count+5

SAVE WORK3

count = count+5

If (condition1) ROLLBACK WORK(3)

count = count-1


print count

COMMIT WORK

Question 4**2 pts**

A company has 20TBs of data. Only a particular 2GBs of that data gets frequently changed by its users, and the remaining data stays the same. This company has a limited budget for extra storage and hardware for backups of their data. What good backup strategies can they follow within their limited budget?

Edit View Insert Format Tools Table

12pt ▾ Paragraph ▾ | **B** *I* U A ▾  ▾ T^2 ▾ | ⋮

p

  | 0 words |   ⋮**Question 5****3 pts**

There are two versions of the same transaction as shown below. Here, the read and write operations in both versions are the same in the same order, both versions have Degree 3 isolation with all the necessary locks taken appropriately, but there are some differences in the locks and order of taking those locks. What are the advantages and disadvantages of each version compared to the other version? Answer with explanation.

Note that, the transaction on version 2 has updated its SLOCK on B to XLOCK before writing without releasing the lock on B.

(version 1)	(version 2)
SLOCK A	SLOCK A
XLOCK B	READ A
READ A	SLOCK B
READ B	READ B
WRITE B	XLOCK B
UNLOCK A	WRITE B
UNLOCK B	UNLOCK A
	UNLOCK B

Edit View Insert Format Tools Table

12pt Paragraph | **B** *I* U A |  | T² | ⋮

p



0 words



Question 6

3 pts

A particular database frequently has deadlock issues. Now, the database administrator has decided to run the following statement in their database before executing each transaction. Will there be any changes in the deadlock situation? If yes, what will be the changes and why? If not, explain why not.

*"SET TRANSACTION ISOLATION LEVEL Read committed with
READ_COMMITTED_SNAPSHOT on"*

Edit View Insert Format Tools Table

12pt ▾ Paragraph ▾ | **B** *I* U A ▾  ▾ T^2 ▾ | ⋮

p



0 words




Question 7

3 pts

Given the following transaction history H , where T_1 , T_2 , and T_3 are transactions, the operations are *Read* and *Write* that are labeled as R and W , and operations are done on the objects O_1 , O_2 , and O_3 , what are the dependencies in the following history? Answer as a sequence of tuples in the form (T_i, O_i, T_j) , and consider only the dependencies that we have focused on our subject.

$H = \langle (T_1, R, O_1), (T_3, W, O_3), (T_3, W, O_1), (T_2, R, O_3), (T_1, R, O_1), (T_2, W, O_2), (T_1, R, O_2), (T_1, R, O_3), (T_2, R, O_2) \rangle$

Edit View Insert Format Tools Table

12pt ▾ Paragraph ▾ | **B** *I* U A ▾  ▾ T^2 ▾ | ⋮

p

  | 0 words |   ⋮

Question 8

3 pts

In a distributed database transaction system, a transaction T only needs to read an object o1. This object o1 is replicated in the nodes D1, D2, and D3 of this distributed system. There is another node D4 in this distributed system, but D4 does not store any copies of object o1. Answer for the following scenarios:

Scenario 1: When T1 started its execution, all nodes (that is, D1, D2, D3, D4) are available, but when T1 commits, D4 is unavailable. Can T successfully commit? Why or why not, explain your answer.

Scenario 2: When T1 started its execution, the nodes D1, D2 are available, and the other nodes are unavailable. When T1 commits, the nodes D1, D2, D4 are available and D3 is unavailable. Can T successfully commit? Why or why not, explain your answer.

Edit View Insert Format Tools Table

12pt ▾ Paragraph ▾ | **B** *I* U A ▾  ▾ T^2 ▾ | ⋮

p



0 words

**Question 9****4 pts**

Consider a database storing the records of one million families, where each family is a row in a table. A unique ID, the location of the family's current residence, and the total household income per year of each family are stored in the database. The total household income of any family can be between 0 to 2 million, and it is an integer number (i.e., a whole number without any decimal value). For each of the following queries on this database, which index is the most suitable choice for efficiency? If there is any search key that needs to be specified for that index, mention the search key as well. If there's no index that can improve a query's performance, write 'no index' for that query. You need to explain the rationale of each of your answers.

Answer for each of the following queries -

A) find the IDs of the families where the total household income is less than 70,000. [2 marks]

B) find the household income of the families who live within 5km of Melbourne central station. [1 mark]

C) find the average household income of the families. [1 mark]

Question 10**6 pts**

Consider the following relational database table. The most common queries on this table are – (i) Finding all the postal addresses, and (ii) Finding all the email addresses. However, the company that is running these queries finds that the query time is not as fast as they would like, especially when this relational database table has many records. Can they use any different types of databases to improve the efficiency of these two queries (without adding any new hardware)? How do they need to store these data in those different types of databases to improve the efficiency of these two queries? Provide **three** such database solutions.

ID	Name	Postal address	... (many other columns)	Email address
1	Smith	10 Central Street, Disneyland	...	smith@email.com
2	Zhou	22 South Road, Someland	...	zhou@email.com
3	Lin	12 North Road, Lalaland	...	lin@email.com

Question 11**6 pts**

A researcher collected 10 Terabytes of data for her research. She needs to choose a RAID structure to store the data. She needs high reliability, high read throughput, and cheaper option for her work, but write throughput is not a concern for her. She also expects to collect 2-3 more TBs of additional data for her research in the near future. She is undecided between the following two options:

Option 1: RAID 1 with 3 disks

Option 2: RAID 4 with 3 disks.

Assuming each disk has 10 Terabyte capacity and the same mean time to failure, explain the advantages and disadvantages of each option for her. Is any of these two options clearly a more suitable choice over the other? Please explain why or why not.

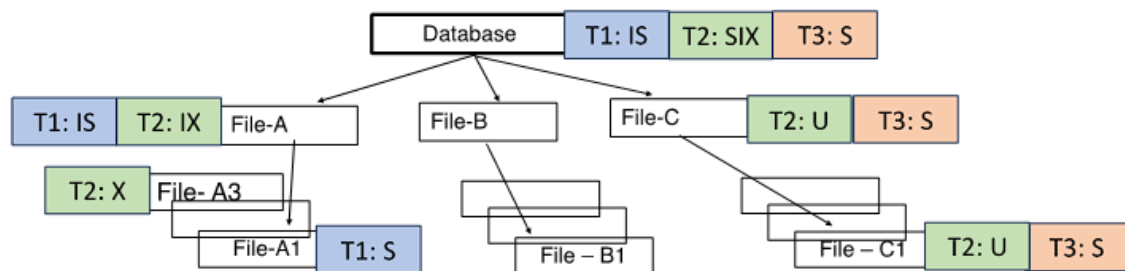
Question 12**8 pts**

Given the hierarchy of database objects and the corresponding granular locks in the following picture, which transactions can run concurrently from the beginning till the end (that is, all operations and locks are compatible to run concurrently with another one), and which ones need to be delayed, answer for both scenarios –

A) Scenario 1: If the transactions arrive in the order T1-T2-T3. Explain your answer.

B) Scenario 2: If the transactions arrive in the order T1-T3-T2. Explain your answer.

Note that all transactions need to take the locks when they start to run.



For your convenience, a granular lock compatibility matrix is as follows –

Compatibility Mode of Granular Locks							
Current	None	IS	IX	S	SIX	U	X
Request	+/- (Next mode) + granted / - delayed						
IS	+(IS)	+(IS)	+(IX)	+(S)	+(SIX)	-(U)	-(X)
IX	+(IX)	+(IX)	+(IX)	-(S)	-(SIX)	-(U)	-(X)
S	+(S)	+(S)	-(IX)	+(S)	-(SIX)	-(U)	-(X)
SIX	+(SIX)	+(SIX)	-(IX)	-(S)	-(SIX)	-(U)	-(X)
U	+(U)	+(U)	-(IX)	+(U)	-(SIX)	-(U)	-(X)
X	+(X)	-(IS)	-(IX)	-(S)	-(SIX)	-(U)	-(X)

Question 13**8 pts**

We have a simplified log at the time of a system crash. Assume that there is no log record before the checkpoint. The format of a log record is (LSN, Operation Details).

(00, begin checkpoint)

(05, end checkpoint)

(10, T1 write page1)

(20, T2 write page2)

(30, T1 write page5)

(40, T2 commit)

(45, T2 end)

(50, T3 write page5)

(60, T1 abort)

(70, CLR undo T1 LSN 30)

(80, CLR undo T1 LSN 10)

(85, T1 end)

(90, T3 write page3)

CRASH

The system recovery consists of three phases: analysis, redo, and undo. Please answer **each** of the following questions.

A) What information will be in the dirty page table after the analysis phase (write as a list of the format (Page id, LSN))? [1 mark]

B) What information will be in the transaction table after the analysis phase (write as a list of the format (Transaction ID, Status, LastLSN))? [1 mark]

C) If the pageLSN of Page5 stored in the database is found as 70, then what will be the order of the LSNs to be redone in the Redo phase? Assume that all the necessary pages are in the dirty page table, all LSNs in the log are greater than or equal to the corresponding page's recLSN, and all LSNs in the log are greater than the corresponding page's pageLSN (except for Page5). [2 marks]

D) What is the order of the LSNs in the 'ToUndo list' for the Undo phase? [2 marks]

E) Just right before the crash happened, what is the PageLSN of page1 and the PageLSN of page5? [2 marks]

Not saved

Submit Quiz