

CS6890 - Final Exam

Name _____

Directions: Read each question carefully. Give your answer in the space provided. The points that can be earned for a question are listed. Partial credit will be given when appropriate. Please answer all questions.

Question 1 Create a schedule that obeys the practical two-phase locking protocol, but will result in cascading aborts. You may abort only one transaction with an explicit “abort” statement. Clearly state why each transaction will abort.

[6 points]

Question 2 Assume that the recovery log is stored in stable storage with a capacity of only 8MB (e.g., a small flash memory card). Which recovery strategy, e.g., undo/redo with no checkpointing, would be best to use? Explain why.

[3 points]

Question 3 Examine the schedule given below.

TA	TB	TC
READ tax		
WRITE tax		
	READ salary	
		READ rate
	READ rate	
READ profit		
		WRITE rate
	WRITE salary	
	READ tax	
		READ profit
	WRITE tax	
WRITE profit		
commit	commit	commit

1. Is this schedule serializable?

2. Assume that the transactions arrive in the following timestamp order TA, TB, then TC. Does this schedule obey the **practical** timestamp-ordering protocol? If no, state why.

3. Assume that the transactions arrive in the following timestamp order TA, TB, then TC. Does this schedule obey the **strict** timestamp-ordering protocol? If no, state why.

4. Does this schedule obey the optimistic concurrency control protocol? Assume that the transactions enter and finish their validation phase and their write phase immediately after the final READ or WRITE in each transaction.

[8 points]

Question 10 Assume that there exist relations with the following schemas.

$R(\underline{A}, B)$

$S(\underline{A}, C)$

$T(C, D)$

R and S each occupy 100 blocks. T occupies 5000 blocks. Assume the following disk organizations.

- R is sorted on attribute A.
- R has a sparse, single-level index on attribute A.
- S is sorted on attribute A.
- S has a sparse, single-level index built on attribute A.
- T has no indexes. T is organized as a heap file.

For each of the following queries, state the best query processing strategy (e.g., nested-loop join with R as outer and S as inner), and estimate the number of blocks read during evaluation of the operation assuming there are 10 blocks in the memory buffer.

1. $\pi_A(R)$

2. $R \bowtie_{R.A=S.A} S$

3. $S \bowtie_{S.C=T.C} T$

4. $\sigma_{C=32}(T)$

[8 points]

Question 11 Examine the schedule given below for transactions T1, T2, and T3. Initially, salary is 1 and tax is 2.

	T1	T2	T3
1			start
2			READ tax
3			tax := tax + 1
4	start		
5	READ salary		
6	salary := salary + 1		
7			WRITE tax
8			commit
9		start	
10		READ tax	
11		READ salary	
12		tax := tax + salary	
13		WRITE tax	
14		commit	
15		checkpoint	
16	READ tax		
17	tax := tax + salary		
18	WRITE salary		
19	commit		

1. Show the log file entries for steps 1 through 19, assume that the transaction starts in its first operation.

2. Assume that the undo/redo recovery algorithm with checkpointing is being used.

- (a) If the database crashes immediately after statement 10, which transactions would have to be undone? _____ redone? _____
- (b) If the databases crashes immediately after statement 17, which transactions would have to be undone? _____ redone? _____

[8 points]

Question 12 Examine the following query in the relational algebra.

$$\sigma_{S.Y=2}(\pi_{R.X,S.Y}(\sigma_{R.X < 23}(R \bowtie_{R.Y=S.Y} (S \bowtie_{S.Y=T.Y} T))))$$

Assuming that T occupies 1000000 blocks, but S and R occupy only 200 blocks each. Further assume that Y is the key of relation S . Give the most optimized, equivalent query to the one listed above. The query should be a single relational algebraic expression.

[6 points]

Question 13 What is the difference between a clustering index and a secondary index?

[3 points]

Question 14 Explain why the relational data model achieves physical data independence while the network data model does not. Also, does the XML or XQuery data model have physical data independence?

[3 points]