

Database Systems, CSCI 4380-01 Exam #3  
Thursday December 9, 2010 at 2 pm

1 (20)	2 (20)	3 (20)	4 (20)	5 (20)	6 (20)	TOTAL (100)

**Note.** The exam is open book and open notes. Use your own book and notes only, sharing is not allowed. Electronic gadgets are NOT allowed during the exam. Write your answers clearly, legibly and explain your reasoning as much as you can. If I cannot read or understand your answers, you will not get points.

There are 6 questions in this exam. Answer any 5 of them. If you answer all 6, I will only read the first 5 answers. Mark clearly which question you do not want me to grade.

In SQL queries in Question 5, do not use views (i.e. using CREATE VIEW statements), triggers and other procedural elements. Make sure you use DISTINCT only when you have to.

**Question 1 (20 points).** You are given the following schedule:

$r_1(X) \ r_2(Z) \ w_1(X) \ r_1(Z) \ r_3(X) \ r_3(Y) \ w_3(Y) \ r_2(Y) \ w_2(Z)$

- (a) List all conflicts in this schedule and draw the conflict diagram. Is this schedule is serializable? Why or why not? If this schedule is serializable, then find a serial schedule that is equal to this schedule.
- (b) Is this schedule possible under the two phase locking scheme if all items are locked with a single type of lock? Explain why or why not in detail.
- (c) Why serializable schedules are preferred? Explain with a single sentence.

**Question 2 (20 points each).** Answer the following questions:

- (a) Is it possible for recovery to erase the work done by a committed transaction? Explain with a single sentence.

- (b) Suppose you are using the following protocol for transaction management:

When a transaction modifies a page, the log containing the update to this page is written to disk first, then the updated page is written to disk before the transaction is allowed to continue.

When a transaction wants to commit, the commit record is written to log, the log is flushed to disk and then the transaction is allowed to commit.

Under this protocol, if a crash occurs, is there a need to do REDO and/or UNDO? Explain why.

- (c) If you are using REDO/UNDO recovery, and the following are the contents of the log and the disk after crash, which log entries should be redone, which should be undone and in which order?

LOG:		LSN	Entry
		10	T1 update P1
		11	T2 update P2
		12	T1 update P3
		13	T1 commit
		14	T3 update P3
		15	T2 update P4
		16	T3 commit

  

Data pages:		Data Page	LSN of Last recorded log entry
		P1	-
		P2	11
		P3	12
		P4	-

**Question 3 (20 points).** You are given the following statistics for  $R(A,B,C,D,E)$ .

$TUPLES(R) = 60,000$

$PAGES(R) = 500$

$VALUES(R.A) = 50,000$

$VALUES(R.B) = 1,000$

$VALUES(R.C) = 200$

$VALUES(R.D) = 10$

$VALUES(R.E) = 2$

and the following query:

```
SELECT E
FROM R
WHERE C=10 AND B > 500 AND D = 10
```

- (a) What is the cost of answering this query using sequential scan?
- (b) What is the cost of answering this query using an index  $I1$  on  $R(B,C,D)$  where  $I1$  is of height 1 (root and leaf) and has 80 leaf nodes.
- (c) What is the cost of answering this query using an index  $I2$  on  $R(C)$  where  $I2$  is of height 1 (root and leaf) and has 20 leaf nodes?
- (d) What would be the best index for this query? Explain your answer with a single sentence.

**Question 4 (20 points).** Answer the following given  $PAGES(R) = 2,000$ ,  $PAGES(S) = 5,000$ ,  $PAGES(R \bowtie S) = 3,000$ ,  $PAGES(T) = 10,000$ .

- (a) What is the cost of sorting relation  $R$  where  $PAGES(R) = 2,000$  with  $M = 50$  buffer pages?
- (b) What is the cost of joining  $R \bowtie S$  using block-nested-loop join with  $M = 1,001$  ( $R$  is the outer relation and  $S$  is the inner relation)?
- (c) What is the cost of computing the following join:  $(R \bowtie S) \bowtie T$  where  $R \bowtie S$  is computed with block-nested-loop join with  $M = 1,001$  as in the previous question above, and  $R \bowtie S$  and  $T$  are joined also using block-nested-loop join using an additional  $M = 1,001$  buffer pages?

**Data Model.**

Person(id, name, gender, birthday, fatherId, motherId)

Ancestor(personId, ancestorId)

In Person, fatherId and motherId are foreign keys to Person(id), but they can also be null if the values are not known.

In Ancestor, personId and ancestorId are both foreign keys to person. Ancestor stores all ancestors (parent, grandparent, great grandparent, etc.) of personId.

**Question 5 (20 points).** Suppose you are given the data model above. Write the following queries using SQL:

- (a) Find all twins in the database. Return tuples of the form (id1, id2) where id1 and id2 are the ids of the twins and for each twin, only return one tuple, i.e. one of (a,b) or (b,a).
- (b) Find people with eight or more children. Return their name and gender.

**Question 6 (20 points).** Using the data model from Question 5, write a trigger for the following using pl/pgsql syntax (though as long as syntax follows any embedded sql format, I'll accept):

Anytime a new person is entered into the database, if their mother id is not null, then enter all ancestors of this person from the mother side into the ancestor table (i.e. mother, grandmother, great grandmother, etc. only).

Do not assume that the ancestor table is correctly populated, simply compute all grandmothers of this person and insert them into the ancestor table.



Blank page for answers