CSE 190A Spring 2018 Midterm Exam

Full Name:

Student ID:

Major:

INSTRUCTIONS

1. You have 50 minutes to complete this exam.

2. You can have up to one letter/A4-sized sheet of notes, formulae, etc. Apart from this, the exam is closed book/notes/electronics/peers.

3. Please wait until being told to start reading and working on the exam.

4. If you think a question is ambiguous, write down your assumptions, argue that they are reasonable, and then work on the problem using those assumptions.

5. Please ensure that your writing is clear and legible!

<table>
<thead>
<tr>
<th></th>
<th>Points</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q 1</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Q 2</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Q 3</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Q 4</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td></td>
</tr>
</tbody>
</table>
HI, THIS IS YOUR SON'S SCHOOL. WE'RE HAVING SOME COMPUTER TROUBLE.

OH, DEAR - DID HE BREAK SOMETHING? IN A WAY—

DID YOU REALLY NAME YOUR SON Robert? DROP TABLE Students;-- ?

~ OH, YES. LITTLE BOBBY TABLES, WE CALL HIM.

WELL, WE'VE LOST THIS YEAR'S STUDENT RECORDS. I HOPE YOU'RE HAPPY.

AND I HOPE YOU'VE LEARNED TO SANITIZE YOUR DATABASE INPUTS.
Q 1. [12pts] Are you on social media? Given the following relational database schema, translate each of the given SQL query over this relational database into an equivalent relational algebra query (or a sequence of queries) with the correct notation used for the relational operators.

Person (ID, Name, Age)
Friends (ID1, ID2)

Person.ID is the primary key of Person. Both Friends.ID1 and Friends.ID2 are foreign keys referring to Person.ID.

1. [2pts] SELECT ID FROM Person WHERE Age > 20 AND Name = "Neo"

2. [2pts] SELECT ID1, COUNT(*) FROM Friends GROUP BY ID1
3. [4pts] SELECT P2.Name FROM Person P1, Person P2, Friends F
   WHERE P1.ID = F.ID1 AND P2.ID = F.ID2 AND P1.Name = "Thanos"
   (Hint: The rename operator could be helpful.)
4. [4pts] SELECT AVG(P.Age) FROM Friends F1, Friends F2, Person P
    WHERE F1.ID2 = F2.ID1 AND F2.ID2 = P.ID AND F1.ID1 = 42
    (Hint: The rename operator could be helpful.)
Q 2. [12pts] Consider the following extendible hash index with 4 slots per bucket.

1. [6pts] Draw the index after the following sequence of update operations: delete 46*, insert 3*, insert 18*, delete 29*, and insert 46*. 
2. [3pts] After performing the given sequence of updates, what is the minimum number of delete operations needed for the global depth to decrease? Clearly circle the correct answer.

(a) 1  (b) 2  (c) 3  (d) 4  (e) 5  (f) 6

3. [3pts] Similarly, after performing the given sequence of updates, what is the minimum number of insert operations needed for the global depth to increase instead? Clearly circle the correct answer.

(a) 1  (b) 2  (c) 3  (d) 4  (e) 5  (f) 6
Q 3. **[14pts]** For the following questions, **clearly circle** True or False.

1. It is impossible to have a B+ tree index that is a clustered index with respect to more than one IndexKey.

   True  
   False

2. In the packed page layout for fixed-length records, deleting records could alter the record IDs of other records on the same page.

   True  
   False

3. The FIFO algorithm avoids the sequential flooding problem faced by LRU.

   True  
   False

4. Sequential data access is typically orders of magnitude faster than random data access when reading data from RAM.

   True  
   False

5. In the pointer-based record format, we need to scan the record from the start even to retrieve a single field in the middle.

   True  
   False

6. A typical magnetic hard disk has multiple arms.

   True  
   False

7. Redistribution of index keys among siblings following an underflow of a node is the first preference for reorganizing a B+ tree index after a delete operation.

   True  
   False
Q 4. [12pts] Suppose we are sorting a relation with 1 billion pages and we have 1000 buffer pages for the external merge sort (EMS). A "pass" over the relation is defined as one read and write of the whole file. In all of the following, you have to include both the sort and merge phases. Clearly circle the correct answer for each of the following questions.

1. [2pts] How many passes will a naive 2-way EMS perform?
   (a) 25  (b) 27  (c) 29  (d) 31  (e) 33  (f) 35

2. [2pts] How many passes will a standard multi-way EMS perform, assuming we do not use any of the three improvements discussed in class?
   (a) 2  (b) 3  (c) 4  (d) 5  (e) 6  (f) 7

3. [2pts] How many passes will a multi-way EMS perform, assuming we use replacement sort for internal sorting?
   (a) 2  (b) 3  (c) 4  (d) 5  (e) 6  (f) 7

4. [3pts] How many passes will a multi-way EMS perform, assuming we use replacement sort for internal sorting along with blocked I/O with block sizes of 10 pages but no double buffering?
   (a) 1  (b) 2  (c) 3  (d) 4  (e) 5  (f) 6

5. [3pts] How many passes will a multi-way EMS perform, assuming we use replacement sort for internal sorting along with blocked I/O with block sizes of 10 pages and double buffering?
   (a) 1  (b) 2  (c) 3  (d) 4  (e) 5  (f) 6