CSE 190D Spring 2019 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>
Test:
1. When did the Pilgrims land at Plymouth Rock?

1620.

As you can see, I've memorized this utterly useless fact long enough to pass a test question. I now intend to forget it forever. I've taught me nothing except how to cynically manipulate the system. Congratulations.

They say the satisfaction of teaching makes up for the lousy pay.
Q 1. [12pts] Consider the following extendible hash index with 3 slots per bucket.

1. [6pts] Draw the index after the following sequence of update operations: delete 97*, insert 11*, delete 47*, and insert 34*.
2. [3pts] After performing the given sequence of updates, what is the minimum number of delete operations needed for any local 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 maximum number of insert operations possible before the global depth increases? Clearly circle the correct answer.

(a) 1   (b) 2   (c) 3   (d) 4   (e) 5   (f) 6
Q 2. [12pts] Suppose we are sorting a relation with 100 million 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 the naive 2-way EMS perform?
   (a) 22    (b) 24    (c) 26    (d) 28    (e) 30    (f) 32

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 5 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 5 pages and double buffering?
   (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. Redistribution of index keys among siblings following an overflow at a node is never preferable for reorganizing a B+ tree index after an insert operation.

   True

   False

2. Using variable-length record layout for fixed-length records leads to more metadata overhead than using the packed record layout.

   True

   False

3. Upon a new page request, if all buffer frames have pinned pages, a typical buffer manager will pick a frame with the lowest pin count for replacement.

   True

   False

4. Random accesses typically achieve peak data transfer throughput from hard disks.

   True

   False

5. Primary indexes are always unique indexes.

   True

   False

6. The delimiter-based record format can store records longer than a page.

   True

   False

7. The arm’s seek time for retrieving a disk block differs based on which cylinder that block is in.

   True

   False
Q 4. [12pts] Do you fly? Given the following relational database schema, translate each given (SQL) query into an equivalent relational algebra query (or a sequence of queries) with the correct notation used for the relational operators. Drawing the relational algebra tree is also acceptable.

Airports (Code, City, State)
Flights (ID, FromCode, ToCode, Time)

Code is the primary key of Airports. Both FromCode and ToCode are foreign keys referring to Airports.Code.

1. [3pts] SELECT DISTINCT City, State FROM Airports, Flights WHERE ID = 123 AND FromCode = Code

2. [3pts] SELECT City, State, AVG (Time) FROM Airports, Flights WHERE Code = ToCode GROUP BY City, State
3. [3pts] How many flights are there from San Diego (airport code SAN) to San Francisco (airport code SFO)?

4. [3pts] How many airports can be reached from San Diego’s with exactly two flights? (Hint: The rename operator could be helpful.)