CS 3743
Introduction to Database Systems
Midterm 2 Solutions

<table>
<thead>
<tr>
<th>Question</th>
<th>Mark</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NAME:_________________________

Instructions
1. Do all of the 5 problems
2. You have 60 minutes for the exam
4. Show all your work
5. Do not separate exam papers

<table>
<thead>
<tr>
<th>Easy</th>
<th>Difficulty Level</th>
<th>Difficult</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ 1</td>
<td>☐ 2</td>
<td>☐ 3</td>
</tr>
<tr>
<td>☐ 4</td>
<td>☐ 5</td>
<td>☐ 6</td>
</tr>
<tr>
<td>☐ 7</td>
<td>☐ 8</td>
<td>☐ 9</td>
</tr>
<tr>
<td>☐ 10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1. **SQL Queries**: (20 pts) Consider the following database schema that keeps track of Sailors, Boats and the boats reserved by sailors.

   - **Sailors**(*sid, sname, rating, age)
   - **Boats**(*bid, bname, color)
   - **Reserves**(*sid, *bid, *date)

Sample tuple for each relation is given below:

   - **Sailors**(22, Dustin, 7, 45)
   - **Boats**(101, Interlake, blue)
   - **Reserves**(22, 101, 1998-10-10)

Keys are denoted by *. Specify the following queries in SQL using above database schema.

   (a) find sailors whose rating is better than sailor named Horatio

   ```sql
   select s.sid
   from Sailors s
   where s.rating > ANY (select s2.rating
                           from Sailors s2
                           where s2.name = 'Horacio')
   ```

   (b) find the names of sailors who are older than the oldest sailor with a rating of 10

   ```sql
   select s.name
   from Sailors s
   where s.age > (select max(s2.age)
                  from Sailors s2
                  where s2.rating = 10)
   ```

   (c) find names of sailors who have reserved a red boat

   ```sql
   select s.name
   from Sailors s
   where s.sid in (select r.sid
                   from Reserves r
                   where r.bid in (select b.bid
                                    from boats b
                                    where b.color = 'red'))
   ```

   (d) find the average age of sailors for each rating level that has at least two sailors

   ```sql
   select s.rating, avg(s.age)
   from Sailors s
   group by s.rating
   having count(*) > 1
   ```
2. **Normalization:** (20 pts) Consider the universal relation \( R = \{A, B, C, D, E\} \) and the following functional dependencies

\[
\begin{align*}
BC & \rightarrow E \\
ED & \rightarrow A \\
A & \rightarrow B
\end{align*}
\]

Answer the following questions based on above information.

(a) List all keys for \( R \)

ACD, BCD, CDE

(b) is \( R \) in 3NF? Explain briefly

No, since it is not in 2NF (BC \( \rightarrow \) E violates 2NF)

(c) is \( R \) in BCNF? Explain briefly

No, since it is not in 2NF (BC \( \rightarrow \) E violates 2NF)
3. **Disk Storage:** (20 pts) Consider a disk with the following characteristics: block size B=512 bytes; interblock gap size G=32 bytes; number of blocks per track = 32; number of tracks per surface = 256. A disk pack consists of 64 double-sided disks. Answer the following based on above information.

(a) What are the total and useful capacity (excluding interblock gaps) of a track?

\[
\text{Total Track Size} = 32 \times (512 + 32) = 17408 \text{ bytes} \\
\text{Useful Capacity} = 32 \times 512 = 16384 \text{ bytes}
\]

(b) What are the total capacity and useful capacity of a cylinder?

\[
\text{Total Cylinder Capacity} = 64 \times 2 \times 32 \times (512 + 32) = 2228224 \text{ bytes} \\
\text{Useful Cylinder Capacity} = 64 \times 2 \times 32 \times 512 = 2097152 \text{ bytes}
\]

(c) What are the total capacity and useful capacity of a disk pack?

\[
\text{Total Capacity of a Disk Pack} = 64 \times 2 \times 256 \times 32 \times (512 + 32) = 570425344 \text{ bytes} \\
\text{Useful Capacity of a Disk Pack} = 64 \times 2 \times 256 \times 32 \times 512 = 536870912 \text{ bytes}
\]

(d) If the drive rotates at a speed of 6000 revolutions per minute, what is the block transfer time?

\[
\text{Transfer Rate} = \frac{17408}{(60 \times 1000/6000)} = 1740 \text{ bytes/msec} \\
\text{Block Transfer Time} = \frac{512}{1740} = 0.29425 \text{ msec}
\]
4. *ER-to-relational Mapping*: (20 pts) Draw an ER Diagram which leads to the following relational Mapping.

Student(*snum, sname, major, level, age)
Class(*cname, time, room, facultyid)
Enrolled(*snum, *cname)
Faculty(*facultyid, facultynname, deptid)

**Solution**: Er diagram is given below.

![ER Diagram](image.png)

Figure 1: ER Diagram
5. **Indexing Structures:** (20 pts) A `RangeQuery R(x, y)` returns all the items in the database that has value between x and y. \( x \leq \text{value} \leq y \). For example, consider the set \{1, 3, 4, 6, 7, 8, 9, 11, 12, 13\}. The query \( R(5, 10) \) will return 6, 7, 8, 9 since \( 5 \leq 6, 7, 8, 9 \leq 10 \). Your are given 10000 records. Each record is 200 bytes long. Each record includes a point plus some additional information. Block size is 1000 bytes. All the point coordinates are distinct and are integers in \{1, 2, 3, \ldots , 50000\}. You have full control over how the file is structured, stored on the disk. Assume each point is 4 bytes, block pointer is 6 bytes and record pointer is 7 bytes. Suggest an index structure that speeds up range queries. Note that based on the range specified, the query needs to access variable number of blocks. Answer the following based on your index.

(a) How is your file structured and stored on disk?

Sort the file based on the number.

(b) What is the number of disk blocks used to store the file?

Since block size is 1000 and record size is 200, 5 records fit in one block. For 10000 records, we need 2000 disk block.

(c) What is the number of index entries? and size of each index entry?

We can have one index entry for each disk block since data is sorted, so we need 2000 index entries. Each index entry requires 4 bytes for the point and 6 bytes for block pointer for a total of 10 bytes.

(d) What is the total size of the index?

Size of the index is 2000*10 = 20000 bytes.

(e) Draw a figure to show how a range query is executed using your index

For a range query \( R(a, b) \) find \( a \) in the index using binary search. Keep reading all the disk blocks in order until \( b \) is reached in the data.