CS 452  
Fall 2002  
Midterm Exam I

Name: ____________________________________________________________________________

SSN: ____________________________________________________________________________

Closed Book  
Closed Notes  
No Calculators  
No Time Limit  
7 Questions

Instructor: David W. Embley
1. Explain how a database is similar to a type in programming languages as follows.

(a) (5 points) A type defines a set of values and a set of operations (e.g. `type integer` defines the integers in a certain range and defines operations on integers such as addition and multiplication). For an SQL-specified database, (1) describe the set of values a database scheme defines (i.e. describe the values the collection of `create` statements define), and (2) list the basic kinds of operations a database system provides (i.e. list the kinds of things you can do to a database using SQL statements).

(b) (3 points) A type hides its implementation. What do we call this principle of hiding for databases? What does this principle do for us if we want to change the physical data structures and indexes we use to implement the relations of a relational database?

(c) (2 points) Types may or may not have a closed set of operations. What do we mean when we say that the set of relational operators is closed?
(16 points) Produce appropriate relation schemes for this diagram. (“Appropriate” means to follow the procedure for generating schemes discussed in the book and in class.) Underline all candidate keys (be sure to make your underline continuous for composite candidate keys). Add a second underline for all primary keys, so that all primary keys are double underlined. Draw an arrow from each foreign key to the primary key it references.
3. Let $r$ and $s$ be relations as follows:

\[
\begin{array}{ccc}
A & B & D \\
1 & 2 & 3 \\
1 & 3 & 0 \\
4 & 7 & 0 \\
4 & 7 & 3 \\
\end{array}
\quad
\begin{array}{ccc}
B & C & D \\
2 & 3 & 3 \\
7 & 0 & 0 \\
7 & 0 & 3 \\
9 & 2 & 0 \\
0 & 5 & 5 \\
\end{array}
\]

Compute the relation for each of the following:

(a) (3 points) $\pi_D r$

(b) (3 points) $\pi_B r - \pi_B s$

(c) (3 points) $\pi_{BC\rho_D-C^r} \cap \pi_{BC}s$

(d) (3 points) $\pi_{AC}(r \bowtie s)$

(e) (3 points) $\pi_{C}s \bowtie (\pi_{E\rho_D-E^r})$
4. Consider the following database scheme.

- *yuppie*(SSN, Name, City, State, Salary)  \( key: SSN \)
- *car*(VIN, Make, Model, Year, Cost, Color)  \( key: VIN \)
- *owns*(SSN, VIN)  \( key: \{SSN, VIN\} \)

(a) (5 points) Write a relational-algebra expression to find the names and salaries of yuppies who live in Arizona and own a 2002 car that costs more than $20,000.

(b) (5 points) Write an SQL query to find the VIN of cars that were made neither in 1999 nor in 2000 (i.e. all but 1999 and 2000 cars).

(c) (5 points) Express the following SQL query in relational algebra.

```
select y1.SSN, y1.Name, y2.SSN, y2.Name
from yuppie y1, yuppie y2
where y1.City = y2.City and y1.SSN < y2.SSN
```
5. Consider relations \( r(ABC) \) and \( s(CDE) \) and suppose that \( r \) has 2,000 tuples stored 10 per block and that \( s \) has 8,000 tuples stored 20 per block. Assume that \( A \) is the primary key for \( r \) and that \( C \) is the primary key for \( s \). Assume that the average seek time is 8 milliseconds, the average rotational latency is 4 milliseconds, and the average block-transfer time is 68 microseconds.

(a) (4 points) Estimate the time required to execute \( \sigma_{C=1} r \) assuming that \( r \) is a sequential file sorted on its primary key and stored in contiguous blocks in disk-block read order.

(b) (4 points) Estimate the time required to execute \( \sigma_{A=1} r \) assuming that \( r \) has a B⁺-tree index on its primary key and that the entire index is on disk in a two-level B⁺ tree.

(c) (4 points) Assuming that that value 1 for \( A \) exists in \( r \), estimate the minimum and maximum time required to execute \( \sigma_{A=1} r \bowtie s \) assuming that both \( r \) and \( s \) are stored as sequential files that are not contiguous and not sorted.

(d) (4 points) Let the size of a block be 512 bytes, the size of a pointer be 2 bytes, and the size of a key be 18 bytes. (Assume also that the total size of the overhead information for the entire block does not exceed 2 bytes.) If a B⁺ tree’s nodes are the same size as a block (as they should be), how many possible pointers are in a node of the B⁺ tree (i.e. since a B⁺ tree is an \( m \)-way tree, what is \( m \))?
6. Let \( r_1(ABC), r_2(CDE), \) and \( r_3(ACF) \), and assume that the domain for all attributes is the set of integers. Rewrite the following queries so that they are optimal in the sense that \( \pi \) and \( \sigma \) operations are done before binary operators when possible.

(a) (6 points) \( \sigma_{A=4 \land C=6 \land E>F}(r_1 \bowtie r_2 \bowtie r_3) \)

(b) (6 points) \( \pi_A \sigma_{C>1}(\pi_{AC}r_1 - \pi_{AC}r_3) \)

(c) (6 points) \( \pi_{BCF}(r_1 \bowtie r_3) \)

7. Consider the following transaction execution schedule.

<table>
<thead>
<tr>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
</tr>
</thead>
<tbody>
<tr>
<td>LS(A)</td>
<td></td>
<td></td>
<td>LX(A)</td>
</tr>
<tr>
<td>UN(A)</td>
<td>LX(B)</td>
<td>LX(C)</td>
<td>UN(A)</td>
</tr>
<tr>
<td>LS(B)</td>
<td>UN(C)</td>
<td>UN(C)</td>
<td>UN(C)</td>
</tr>
<tr>
<td>UN(B)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) (4 points) For each of the four transactions, state whether the transaction does or does not satisfy the two-phase locking protocol.

(b) (6 points) The schedule has some locking violations (i.e. a database with a locking protocol would not allow the transactions to be executed as specified in the schedule). Explain all violations. (Note: these are simple locking violations, not 2PLP locking violations.)