Final Examination, Csci5708, Spring 2010

INSTRUCTIONS

- **Time Allocated**: 120 minutes
- This is a CLOSED book examination. One 8.5 x 11 sheet with notes is allowed. Notes cannot be shared.
- This examination has 6 questions. Each question has its sub questions.
- Put your name, student id, email address, class group, course id, course name, semester and year on the cover page.
- Please answer questions in the spaces provided. You may request additional sheets of paper from the instructor or continue your answer on the back of each paper.
- Please state your assumptions clearly.

<table>
<thead>
<tr>
<th>Name:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student ID:</td>
</tr>
<tr>
<td>Email Address:</td>
</tr>
<tr>
<td>Class Group:</td>
</tr>
<tr>
<td>Course ID:</td>
</tr>
<tr>
<td>Course Name:</td>
</tr>
<tr>
<td>Semester:</td>
</tr>
<tr>
<td>Year:</td>
</tr>
</tbody>
</table>
Question 1. Let $Wi(j)$ denote a write operation by Transaction $Ti$ on item $j$ in the database. Let $Ri(j)$ denote a read operation by Transaction $Ti$ on item $j$ in the database. Assume a database with items A, B, C, transactions T1, T2, and T3 and schedules S1 and S2 where:

$T1$: $R1(B), R1(C), W1(B)$
$T2$: $R2(C), R2(B), W2(B), W2(A)$
$T3$: $R3(B), W3(B), R3(A), W3(A)$

$S1$: $R1(B), R1(C), R2(C), W1(B), R3(B), W3(B), R2(B), W2(B), W2(A), R3(A), W3(A)$
$S2$: $R1(B), R1(C), R2(C), R3(B), W3(B), W1(B), R2(B), W2(B), W2(A), R3(A), W3(A)$

Q1a. (5 points) Draw the conflict graphs for $S1$ and $S2$, and determine whether each schedule is (conflict) serializable or not.

Q1b. (5 points) Determine whether $S1$ and $S2$ are consistent with basic timestamp ordering (TO). Explain.

Q1c. (5 points) Consider schedule $S2'$ below:

$S2'$: $R1(B), R1(C), R2(C), R3(B), W3(B), W1(B), C1, R2(B), W2(B), W2(A), C2, R3(A), W3(A), C3$

Is $S2'$ recoverable? cascadeless? strict?
**Question 2. (15 points)** Consider the following schedule consisting of 4 transactions T1, T2, T3 and T4. Consider a system with immediate update and the following redo logs. \([\text{read\_item}, T_i, X]\) represents that transaction \(T_i\) reads item \(X\). \([\text{write\_item}, T_i, X, A, B]\) states that transaction \(T_i\) writes item \(X\) from \(A\) to \(B\). Which transactions are undone during recovery? Which transactions are redone? Is there any cascaded rollback?

<table>
<thead>
<tr>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
</tr>
</thead>
<tbody>
<tr>
<td>[CHECKPOINT]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[\text{read_item}(A)]</td>
<td>(A = A + 10)</td>
<td>[\text{write_item}(A)]</td>
<td>commit</td>
</tr>
<tr>
<td></td>
<td>[CHECKPOINT]</td>
<td>[\text{read_item}(B)]</td>
<td>(B = B + 10)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[\text{write_item}(B)]</td>
<td>commit</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[\text{read_item}(A)]</td>
<td>(A = A + 10)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[\text{write_item}(A)]</td>
<td>commit</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SYSTEM CRASH</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**System log**

- [checkpoint]
- [strat\_transaction, T1]
- [read\_item, T1, A]
- [write\_item, T1, A, 10, 20]
- [commit, T1]
- [checkpoint]
- [strat\_transaction, T2]
- [read\_item, T2, B]
- [write\_item, T2, B, 20, 30]
- [commit, T2]
- [strat\_transaction, T3]
- [read\_item, T3, A]
- [write\_item, T3, A, 20, 30]
- [strat\_transaction, T4]
- [read\_item, T4, A]
- [write\_item, T4, A, 30, 40]
- [commit, T4] ← system crash

---

**Question 3.** Consider the following relational schema created by user X:

Employee(SSN, Name, Salary, Mgrssn, Dno)
Works_on(ESSN, Pno, Hours)
Project(Pno, Pname, Plocation, Dno)

SSN is the primary key for Employee, ESSN and Pno together are the primary key for Works_on and Pno is the primary key for Project.

**Q3a. (10 points)** Write SQL statements for granting following privileges to account A, B and C.

- User A is a supervisor of department number(Dno) 5. User X gives permission to user A to select SSN, Name of employees in his/her department.

- User X gives permission to user B to select and update Works_on.Hours and propagate these to other users.
- User B gives permission to user C to select and update X.Works_on.Hours.

- User X revokes all privileges from user B on Works_on.

- User X gives permission to user A to view the total number of hours each employee has worked. Note that user A doesn’t have permission to view Project attributes, e.g. Pno, Pname, etc.

**Q3b. (5 points)** Can user C still select and update X.Works_on.Hours? Why? Briefly justify your answer by drawing the authorization graph.

**Q3c. (5 points)** Consider the updated Project relation shown below. How would it appear to a user with classification U? Suppose that a classification U user tries to update the Plocation of ‘ProductX’ to ‘Minneapolis’; what would be the result of this action?

<table>
<thead>
<tr>
<th>Pno</th>
<th>Pname</th>
<th>Plocation</th>
<th>Dno</th>
<th>TC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>U ProductX</td>
<td>U Bellaire</td>
<td>C 5</td>
<td>S</td>
</tr>
<tr>
<td>1</td>
<td>U ProductX</td>
<td>U Bellaire</td>
<td>C 5</td>
<td>C</td>
</tr>
<tr>
<td>2</td>
<td>C ProductY</td>
<td>C Sugarland</td>
<td>S 5</td>
<td>C</td>
</tr>
</tbody>
</table>

**Question 4.** Assume you are given the following schema:

```sql
CREATE TYPE MovieType AS (
    title CHAR(30),
    year INTEGER,
    inColor BOOLEAN
);
```
CREATE TABLE Movie OF MovieType (  
    REF IS MovieID SYSTEM GENERATED,  
    PRIMARY KEY (title, year)  
);  

CREATE TYPE AddressType AS(  
    street CHAR(30),  
    city CHAR(30),  
    zip INTEGER,  
    state CHAR(30)  
);  

CREATE TYPE PersonType AS (  
    name CHAR(30),  
    address AddressType,  
    bestMovie REF(MovieType) SCOPE Movie  
);  

CREATE TABLE Actor OF PersonType (  
    REF IS PersonID SYSTEM GENERATED;  
)  

CREATE TABLE StarsIn (  
    star REF(PersonType) SCOPE Actor,  
    movie REF(MovieType) SCOPE Movie  
);  

Q4a. (5 points) Using the StarsIn relation above, find the names of stars of Avatar. (Hint: The Movie and Actor relations are accessible through StarsIn).

Q4b. (10 points) Using the StarsIn relation above, find the titles and years of all movies in which at least one star lives in Malibu (Hint: The Movie and Actor relations are accessible through StarsIn).
**Question 5.** Consider the following dataset.

<table>
<thead>
<tr>
<th>Transaction ID</th>
<th>Items Purchased</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>A, B, C, D, E</td>
</tr>
<tr>
<td>T2</td>
<td>B, C, D</td>
</tr>
<tr>
<td>T3</td>
<td>A, C, D</td>
</tr>
<tr>
<td>T4</td>
<td>E, F</td>
</tr>
<tr>
<td>T5</td>
<td>A, D, E</td>
</tr>
<tr>
<td>T6</td>
<td>A, D, G</td>
</tr>
<tr>
<td>T7</td>
<td>A, B, C, D, E</td>
</tr>
<tr>
<td>T8</td>
<td>E, F, G</td>
</tr>
</tbody>
</table>

The set of items is \{A, B, C, D, E, F, G\}.

**Q5a. (10 points)** Identify all item-sets with support > 0.2 using the Apriori algorithm. Show intermediate results e.g. singletons, doubletons, etc.

**Q5b. (10 points)** Identify all item-sets with support > 0.2 using the frequent pattern tree algorithm. Show intermediate results e.g. FP-tree, result of each iteration, etc.
Question 6. Consider the following relational schema in a distributed database where the Project relation resides at site 1 and the Department relation resides at site 2:

**Site 1:**

**Project**\( (Pno, \text{Pname}, \text{Plocation}, \text{Dno}) \)
2000 records
each record is 100 bytes long
\(\text{Pno}\) field is 5 bytes long
\(\text{Pname}\) field is 15 bytes long
\(\text{Plocation}\) field is 76 bytes long

**Site 2:**

**Department**\( (\text{Dname, Dnumber, Mgr_ssn, Mgr_start_date}) \)
100 records
each record is 35 bytes long
\(\text{Dnumber}\) field is 4 bytes long
\(\text{Mgr_ssn}\) field is 9 bytes long
\(\text{Dname}\) field is 10 bytes long

**Query q:** For each project, retrieve the project name, project location and the name of the department that the project is associated with.

**Q6a. (5 points).** Suppose the Project relation is transferred to site 2, query q is executed, and the result is presented to the user at site 2. How many bytes must be transferred?

**Q6b. (5 points).** Suppose the Department relation is transferred to site 1, query q is executed at site 1 and the result is sent back to site 2. How many bytes must be transferred?

**Q6c. (5 points).** Suppose in executing query q we have Project \(\text{SEMI JOIN}_{\text{Dno} = \text{Dnumber}}\) Department where the join attributes of Department are projected at site 2 and transferred to site 1. The join then takes place at site 1 and the result is presented to the user at site 2. How many bytes must be transferred?