Sample Final Examination, Csci5708

Instructions

- Time Allocated: 120 minutes
- This is a CLOSED book examination. One 8.5x11 sheet with notes is allowed. Notes can’t be shared.
- This examination has 6 questions. Each question has its sub questions.
- Put your name, student university id, email address, class group on the coverage and blue book coverage.
- Please answer in the whitespace.
- Please state your assumptions clearly.

<table>
<thead>
<tr>
<th>Name</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Student ID</td>
<td></td>
</tr>
<tr>
<td>Email Address</td>
<td></td>
</tr>
<tr>
<td>Class Group</td>
<td></td>
</tr>
<tr>
<td>Course ID</td>
<td>CSci 5708</td>
</tr>
<tr>
<td>Course Name</td>
<td>Architecture and Implementation of DBMS</td>
</tr>
<tr>
<td>Semester</td>
<td>Spring</td>
</tr>
<tr>
<td>Year</td>
<td></td>
</tr>
</tbody>
</table>
Question 1. (20 points) Let $W_i(X)$ denote a write operation by Transaction $T_i$ on item $X$ in the database. Let $R_i(X)$ denote a read operation by Transaction $T_i$ on item $X$ in the database. Assume a database with item A, B, transactions T1, T2, T3 and schedule S1.

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

S1: R1(A), R2(A), W1(A), W2(A), R3(A), R3(B), R1(B)

Q1a. (6 points) Is S1 conflict serializable? Prove your answer by conflict graph.

Q1b. (6 points) Determine whether S1 is serializable by timestamp ordering (TO). Explain.

Q1c. (5 points) Does a strict schedule guarantees conflict serializability? If yes, provide brief proof. If not, give a brief counter example.

Q1d. (3 points) In Oracle, the default isolation level is read-committed. Which problem(s) does it have among dirty read, non-repeatable read and lose update?

<table>
<thead>
<tr>
<th>Problem</th>
<th>Dirty read</th>
<th>Non-repeatable read</th>
<th>Lost update</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presence</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Question 2. (10 points) Consider the following schedule consisting of 4 transactions T1, T2, T3 and T4. The system log is shown in textbox below. [read_item, Ti, X] represents log entry that Ti read item X. [write_item, Ti, X, A, B] states that transaction Ti writes item X from A to B. A, B, C and D are 0 initially. Suppose the immediate update protocol is followed.

<table>
<thead>
<tr>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
</tr>
</thead>
<tbody>
<tr>
<td>[CHECKPOINT]</td>
<td>[CHECKPOINT]</td>
<td>[CHECKPOINT]</td>
<td>[CHECKPOINT]</td>
</tr>
</tbody>
</table>
| Read_item(D) | Read_item(D) | Read_item(B); | Read_item(B);
| D=D+10 | B=B+10; | B=B+10; | B=B+10;
| Write_item(D) | Write_item(B); | Write_item(A); | Write_item(A);
| Commit; | Commit; | Commit; | Commit; |

System Log
[checkpoint]
[start transaction, T1]
[read_item, T1, D]
[write_item, T1, D, 0, 10]
[commit, T1]
[checkpoint]
[start transaction, T2]
[read_item, T2, B]
[write_item, T2, B, 0, 10]
[commit, T2]
[start transaction, T3]
[write_item, T3, A, 0, 10]
[read_item, T3, C]
[write_item, T3, C]
[read_item, T3, A]
[commit, T3]
[start transaction, T4]
[read_item, T4, B]
[commit, T4]

Q2a. (3 points) Which transactions are undone?

Q2b. (3 points) Which are redone?

Q2c. (4 points) Is cascaded rollback needed? If so, which transactions are rolled back?
Q3. (20 points)

Q3a. (5 points) Suppose user A is the owner of table T, A gives permission to user B to select all attributes in T and propagate it to other users. User B then grants this permission to user C and propagate it to others. User C, in turn, grants this permission to user A and user B. Now user A revoke permission from B. Could user B and user C still select T? Justify your answer with authorization graph.

Q3b. (10 points) The table below has multi-level access control. There are four security level: U (unclassified), C (classified), S (secret), TS (top secret).

<table>
<thead>
<tr>
<th>Eid</th>
<th>Ename</th>
<th>Esalary</th>
<th>TC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 U</td>
<td>John U</td>
<td>$10,000</td>
<td>C</td>
</tr>
<tr>
<td>2 C</td>
<td>Jill C</td>
<td>$12,000</td>
<td>S</td>
</tr>
</tbody>
</table>

(i) Suppose user A is in class S and user B is in class C. According to the simple security property, could they read Esalary attributes of both tuples respectively? According to the star property, could they write on Esalary attributes of both tuples respectively?

(ii) Why is the star property required? Provide a possible attack addressed by the star property.

Q3c. (5 points) Match security best practices (1) through (4) with security threat (a) through (d)

(1) mandatory access control (e.g. Bell LaPadulamodel, multi-level relations)  
(a) SQL injection attach  
(2) analyze DBMS logs for anomalous pattern  
(b) Unauthorized program (e.g. trojanhorse, virus) queries a sensitive field.  
(3) encryption  
(c) Insider Attack, DBA reads salaries of all executives.  
(4) use prepared (pre-compiled) parametric SQL statement in application  
(d) Dumpster diving: Unauthorized person finds a worn-out backup tape in garbage. He recovers credit-card numbers from readable portions of tape.
Q4. (15 points)

<table>
<thead>
<tr>
<th>SQL2 schema</th>
<th>SQL3 schema</th>
</tr>
</thead>
</table>
| **create table** employee (  
  eid integer,  
  name char(30),  
  age integer,  
  address char(30) );  
create table project (  
pid integer,  
title char(30),  
years integer,  
director_eid integer);  
create table work_on(  
eid integer,  
pid integer);  
| note: 1. primary keys are underlined.  
2. foreign keys are  
work_on.pid, work_on.eid and  
Project.director_eid.  
create type employee_type as (  
  name char(30),  
age integer,  
address char(30) );  
create table employee of employee_type (  
  ref is employeeid system generated,  
  primary key employeeid );  
create type project_type as (  
  title char(30),  
years integer,  
director ref(employee_type) scope employee );  
create table project of project_type (  
  ref is projectid system generated;  
  primary key projectid );  
create table workon (  
  employee ref(employee_type) scope employee,  
  project ref(project_type) scope project ); |

Q4a. (6 points) Match each Sql2 queries with corresponding Sql3 queries.

<table>
<thead>
<tr>
<th>SQL2 Queries</th>
<th>SQL3 Queries</th>
</tr>
</thead>
</table>
| (1) select e.name  
  from employee e, work_on w  
  where e.eid=w.eid;  
(2) select p.title  
  from project p, workon w  
  where p.pid=w.pid  
  and p.years>3;  
(3) select e.name  
  from employee e, works_on w, project p  
  where e.eid = w.eid and p.pid = w.pid  
  and p.director_Eid = e.eid | (a) select e.name  
  from employee e, workon w  
  where w.employee.employeeid=e.employeeid;  
(b) select p.title  
  from project p, workon w  
  where p.projectid=w.project.projectid  
  and p->years>3;  
(c) select w.employee->name  
  from workon w;  
(d) select w->project->title  
  from workon w  
  where w->project.years>3;  
(e) select w.employee->name  
  from works_on w  
  where w.employee->employeeid  
  = w.project->director->employeeid  
  (f) select e.name  
  from workon w, employee e, project p  
  where w.project.projectID=p.projectID  
  and e.employeeID=w.employee.employeeID; |

Answer:
Q4b. (3 points) Plot an ER diagram corresponding to the sql2 schema in Q4. List the main challenges in creating corresponding XML schema.

Q4c. (6 points) Provide an XML schema for the SQL2 relational schema in Q4. Briefly defy your answer.
Q5. (15 points)

Q5a. (7 points) Given the following points, use the K-means cluster method to group them into three clusters. (Hint: geometry reasoning may save time)

<table>
<thead>
<tr>
<th>Points</th>
<th>(1,1), (1,2), (2,1), (2,2), (3,5), (3,6), (4,5), (4,6), (5,1), (5,2), (6,1), (6,2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initials centroids</td>
<td>(3,5), (4,5), (5,1)</td>
</tr>
</tbody>
</table>

Answers:

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Centroid 1</th>
<th>Cluster 1 points</th>
<th>Centroid 2</th>
<th>Cluster 2 points</th>
<th>Centroid 3</th>
<th>Cluster 3 points</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(3,5)</td>
<td></td>
<td>(4,5)</td>
<td></td>
<td>(5,1)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Q5b. (8 points) Given the transactions with items below, use the Apriori algorithm to find all frequent item sets with the support \( \geq 0.5 \).

<table>
<thead>
<tr>
<th>TID</th>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A, C, D</td>
</tr>
<tr>
<td>2</td>
<td>B, C, E</td>
</tr>
<tr>
<td>3</td>
<td>A, B, C, E</td>
</tr>
<tr>
<td>4</td>
<td>B, E</td>
</tr>
</tbody>
</table>
Q6. (20 points) Consider the following relational schema in a distributed database, where the Project relation resides at site 1, the Employee relation resides at site 2 and the WorkOn relation resides at site 3.

Site 1:
- **Project** (Pno, Pbudget, Pname)
- 100 records
- Each record is 50 bytes.
- Pno field is 5 bytes.
- Pbudget field is 4 bytes.
- Pname field is 41 bytes.

Site 2:
- **Employee** (Eno, Ename, Esalary, Eage, Ehome_address, Eemail)
- 2000 records
- Each record is 200 bytes.
- Eno is 5 bytes.
- Ename is 20 bytes.
- Esalary is 4 bytes.
- Eage is 4 bytes.
- Eaddress is 167 bytes.

Site 3:
- **WorkOn** (Pno, Eno)
- 2000 records
- Each record is 10 bytes.
- Pno is 5 bytes.
- Eno is 5 bytes.

User query q at site 3: Retrieve the names (Pname) and budgets (Pbudget) of all projects that have employees between 40 years old and 50 years old, and the corresponding employee name (Ename). (Assume that 100 records have Eage attributes between 40 and 50 in Employee table.)

Q6a. (5 points) Suppose Project and Employee relations are transferred to site 3, query q is evaluated at site 3, and the result is sent to the user at site 3. How many bytes must be transferred?

Q6b. (5 points) Suppose Project and WorkOn relations are transferred to site 2, query q is evaluated at site 2 and the result is sent to the user at site 3. How many bytes must be transferred?

Q6c. (10 points) Given the SEMI JOIN strategy, which computes $\pi_{\text{name}, \text{eid}} (\text{Employee} \bowtie (\text{Employee}.\text{Eno} = \text{WorkOn}.\text{Eno}) \text{ WorkOn})$, followed by $\pi_{\text{pid}, \text{pname}, \text{pbudget}} (\text{Project} \bowtie (\text{Project}.\text{Pno} = \text{WorkOn}.\text{Pno}) \text{ WorkOn})$. Calculate the number of bytes transferred in your strategy. Assume the final result is to be presented at site 3.