Final Examination, CSCI 4707, Spring 2008

Time Allocated: 120 minutes
Maximum Points: 100

1. This is a CLOSED book examination. Personal copies of textbook and class notes can not be used. However, each student may bring one 8.5inch by 11inch sheet of summary notes. Please do not share this sheet with other students. Laptop computers and calculators cannot be used.
2. There are 5 questions. Use a blue book to answer these questions. Start answer to each question on a fresh page. Put name, student id, email address, team group number, course id, course name, semester, and year on the cover page.
3. All questions about SQL refer to the SQL2 or SQL3 standard.
4. Clearly state any assumptions that you make regarding possible ambiguities.

Question 1.
Name is the unique candidate key in the table STUDENT. Each row in the table describes a different student. A student may only have two majors and one minor. To each major corresponds a unique minor.

<table>
<thead>
<tr>
<th>NAME</th>
<th>AGE</th>
<th>MAJOR</th>
<th>MINOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>GREEN</td>
<td>26</td>
<td>{MATH,CSE}</td>
<td>INFO</td>
</tr>
<tr>
<td>HONEY</td>
<td>18</td>
<td>{INFO,IT}</td>
<td>MATH</td>
</tr>
<tr>
<td>LAMPOC</td>
<td>33</td>
<td>{MKTG,BUS}</td>
<td>MGT</td>
</tr>
<tr>
<td>TAKKER</td>
<td>45</td>
<td>{MKTG,BUS}</td>
<td>MGT</td>
</tr>
<tr>
<td>YONAR</td>
<td>28</td>
<td>{INFO,IT}</td>
<td>MATH</td>
</tr>
</tbody>
</table>

Question 1a (5 Points).
Based on the STUDENT table, which of the following functional dependencies hold:
(A) Major -> Name
(B) Age -> Name
(C) Minor -> Name
(D) None of the above

For questions 1b and 1c consider the following description for the STUDENT table. In addition to the four columns (Name, Age, Major, Minor) in the STUDENT table, a new column for specialization is added for Major. For example, a student having CSE as a major may have DBMS or AI as a specialization. Below is a list of functional dependencies for the new table.
Name -> Age, Major, Specialization, Minor
Specialization -> Major

Question 1b (5 Points).
Provide the relational tables that satisfy 1NF but not 2NF.

Question 1c (5 Points).
Provide the relational tables that satisfy 3NF but not BCNF.
Question 2 (30 Points, 6 pts each).
Fill in the following TABLE within your blue book to summarize your answer in Question 2.

<table>
<thead>
<tr>
<th>English Query</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQL Expression(s)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Consider the following schema, in which the keys have been underlined.

EMPLOYEE (SSN, FNAME, LNAME, BIRTHDATE, ADDRESS, SUPERSSN, DNO)
WORKS_ON (ESSN, PNO, HOURS)

(Note: Foreign keys ESSN to Employee(SSN) and PNO to Project (PNUMBER))
PROJECT (PNUMBER, PNAME, PLOCATION, DNUM)

Below is a set of database queries written as English.

(a) Find the employees’ SSN who works on all the projects controlled by department number 6.
(b) Find the employees’ SSN who works on some projects controlled by department number 6.
(c) Find the employees’ SSN in department number 6 who worked on zero or more projects.
(d) Find the employees’ SSN where their supervisor works on some of the projects controlled by
    department number 6.
(e) Find the employees’ SSN who work only on projects controlled by department number 6.

For each query, find the SQL statement(s) below that produces the corresponding result. (Hint: it
is possible to have an English statement not matched to any SQL statement, and vice versa.
Moreover, more than one SQL statement may match the same English query statement.)

(1) select E.SSN
    from EMPLOYEE E
    where E.DNO = 6
    and E.SSN in (  
    select SSN
        from EMPLOYEE left join WORKS_ON
        on EMPLOYEE.SSN =  
        WORKS_ON.ESSN);  
(2) select E.SSN
    from EMPLOYEE E
    where exists (  
    select * from WORKS_ON W,
            PROJECT P
    where E.SSN=W.ESSN and
        W.PNO = P.PNUMBER and
        P.DNUM = 6);  
(3) select E.SSN
    from EMPLOYEE E, WORKS_ON W,  
    PROJECT P
    where E.SSN = W.ESSN and
        W.PNO = P.PNUMBER and
        P.DNUM = 6
    group by E.FNAME, E.LNAME
    having count(*) = (  
    select count(*) from PROJECT P2
    where P2.DNUM=6);  
(4) select E.SSN
    from EMPLOYEE E
    where exists (  
    select * from WORKS_ON W,
            PROJECT P
    where E.SUPERSSN = W.ESSN and
        W.PNO = P.PNUMBER and
        P.DNUM = 6);  
(5) select E.SSN
    from EMPLOYEE E
    where exists (  
    select * from WORKS_ON W,
            PROJECT P
    where E.SUPERSSN = W.ESSN or
        (W.PNO = P.PNUMBER and
        P.DNUM = 6) );  
(6) select E.SSN
    from EMPLOYEE E
    where exists (  
    select * from WORKS_ON W
    where E.SSN=W.ESSN and
    not exists (  
    select * from PROJECT P
    where W.PNO = P.PNUMBER and
    P.DNUM <> 6));
Question 3.
The following timeline shows the state of transactions with respect to the most recent backup, cache-consistent checkpoint, and failure. Use the timeline when solving each part in this question.

Question 3a (5 Points).
Describe the restart work for each transaction if a system failure occurs. Assume that the recovery manager uses the deferred update approach.

Question 3b (5 Points).
Describe the restart work for each transaction if a system failure occurs. Assume that the recovery manager uses the immediate update approach.

Question 3c (5 Points).
Describe the restart work if a device failure occurs affecting the database table. Assume that log-files are on a separate disk-drive and are not affected.
Question 4
Use the following data to perform the indicated calculations in questions 4a, 4b, and 4c.
Logical Record (i.e., Row) size = 100 bytes
Number of rows = 100,000
Primary key size = 6 bytes
Physical record size = 4,096 bytes
Pointer size = 4 bytes
Floor(X) is the largest integer less than or equal to X.
Ceil(X) is the smallest integer greater than or equal to X.

Question 4a (4 Points).
Which of the following is the number of rows that can completely fit in a physical record.
Assume that only complete rows can be stored (use the Floor Function).
(a) 20    (b) 30   (c) 40   (d) 50

Question 4b (4 Points).
Which of the following is the number of physical records necessary for a sequential file.
Assume that physical records are filled to capacity except for the last physical record (use
the Ceil function) and each logical record (i.e. row) is kept within a single physical record.
(a) 2100   (b) 2200   (c) 2300   (d) 2400

Question 4c (4 Points).
Which of the following is the maximum branching factor on a node in a Btree.  Assume
that each record in a Btree consists of <key value, pointer> pairs.
(a) 354   (b) 409   (c) 491   (d) 554

Question 4d (4 Points).
In distributed databases, what mechanism is used to ensure that the distributed
transactions are atomic despite site and communication failures?  Briefly describe this
mechanism, how it is used, and the pitfalls of this mechanism.

Question 4e (4 Points).
Which object-oriented concepts are usually not supported by object-relational databases?
(i) data abstraction, (b) encapsulation, (c) inheritance, (d) polymorphisms.
Briefly explain the reasons for the deviation.

Question 4f (5 Points).
Which relational database concepts are usually dropped in object-relational databases?
(i) Entity integrity constraints   (b) Referential integrity constraints
(c) Normal forms for table design   (d) ACID Properties for transactions
Briefly explain the reasons for the deviation.
**Question 5.**
Consider the following star schema:
Dimension tables:
- Customer(custKey, custName, custCity)
- Product(prodKey, name, type, busUnit)
- Date(dateKey, salesdate, month, year)
Fact table:
- Sales(custKey, prodKey, dateKey, amt, qty)

Write SQL expression for the following queries. Briefly justify your answer.

**Question 5a (5 Points).**
Provide a single block SQL expression to represent quantities of each product sold and the sales amount, by month and year using the SQL CUBE operation.

**Question 5b (5 Points).**
Repeat question 5a without using SQL operation of CUBE, ROLLUP, or GROUPING SETS.

**Question 5c (5 Points).**
Provide a single block SQL expression to represent sales amounts for the products starting with only product type first, followed by the business unit and nothing else. Do not use the SQL grouping set operator.