Stanford University Computer Science Department
2005 Comprehensive Exam in Databases
SAMPLE SOLUTION

1. *Relational Algebra* **(12 points)**. In this question, we use $t[X]$ to represent the components of tuple t in attributes X. Let relation R have schema XY (i.e., the set of attributes $X \cup Y$) and let relation S have schema YZ. Assume X, Y, and Z are disjoint. The *strong join* of R by S is the set of tuples $t[X]$ such that t is a tuple of R, and for all tuples r in R (including t) such that $r[X] = t[X]$, there is some tuple s in S such that $r[Y] = s[Y]$. Write, as a sequence of assignments of relational-algebra expressions to variables, a program that computes the strong join of $R(A, B)$ by $S(B, C)$. **Note**: you should use only the basic relational-algebra operators: select, project, union, intersection, difference, product, natural join, and theta-join. Also, please explain what each step is doing, if you want to be considered for partial credit.

   T_1 := $R - \pi_{A,B}(R \bowtie S)$
   T_2 := $\pi_A(T_1) \bowtie R$
   Answer := $\pi_A(R) - \pi_A(T_2)$

2. *Functional Dependences* **(8 points)**. Armstrong's axioms for functional dependencies are:

   (a) *Reflexivity*: If $X \subseteq Y$, then $Y \rightarrow X$.
   (b) *Augmentation*: If $X \rightarrow Y$, then $XZ \rightarrow YZ$.
   (c) *Transitivity*: If $X \rightarrow Y$ and $Y \rightarrow Z$, then $X \rightarrow Z$.

   From these axioms, prove the *union rule*: If $X \rightarrow Y$ and $X \rightarrow Z$, then $X \rightarrow YZ$. Note that X, Y, and Z are sets of attributes, XY denotes $X \cup Y$, and nothing about disjointness is assumed.

   (1) From $X \rightarrow Y$ get $XZ \rightarrow YZ$ by Augmentation
   (2) From $X \rightarrow Z$ get $X \rightarrow XZ$ by Augmentation
   (3) From (1) and (2) get $X \rightarrow YZ$ by Transitivity

3. *ODL and E/R Models* **(6 points)**. Your answers to (a) and (b) will be judged not only on being truthful, but also on selecting among all options the things that are most significant.

   (a) **(4 points)** Name two things that object-oriented models such as ODL offer the designer and that are not found in the E/R model.

   - Methods
   - Complex data types

   (b) **(2 points)** Name one thing that the E/R model offers that ODL does not.

   - Multiway relationships

4. *XML* **(8 points)**. Here is a DTD:

```
<!DOCTYPE a [
   <!ELEMENT a (b+, c*)>
   <!ELEMENT b (c?)>
   <!ELEMENT c (#PCDATA)>
]>
```

   (a) **(3 points)** Give a shortest (fewest number of elements) example of a sequence of opening and closing tags that is valid for this DTD.

```
<a> <b> </b> </a>
```

   (b) **(5 points)** Give a shortest (fewest number of elements) example of a well-formed sequence of opening and closing tags where `a` is the root tag, `b` and `c` tags are also present, but the sequence is *not* valid for this DTD.

```
<a> <c> </c> <b> </b> </a>
```

5. *Multivalued Dependencies and SQL Constraints* **(10 points)**. Consider a table $R(A, B, C)$. Make no assumptions about keys.

   (a) **(6 points)** Using SQL Assertions, write an assertion stating that the multivalued dependency $A \twoheadrightarrow B$ holds on R.

```
create assertion MVD check (
  not exists (
    select * from R R1, R R2
    where R1.A = R2.A
    and (R1.A, R1.B, R2.C) not in (select * from R)))
```

(b) **(4 points)** Can you enforce the same dependency using SQL Tuple-Based `CHECK` constraints? If so, show the constraint(s). If not, explain why not.

Although the above general assertion can be translated directly to a tuple-based check constraint, it will only be checked on insertions and updates to R. Since deletions to R can cause an MVD violation, a tuple-based check cannot enforce an MVD.

6. *SQL Equivalence* **(8 points)**. Consider the following two SQL queries over tables $R(A, B)$ and $S(C)$. Make no assumptions about keys.

```
Q1: select A from R                      Q2: select distinct A from R, S
    where B in (select C from S)             where R.B = S.C
```

Are these two queries equivalent? That is, do they return the same answer on all possible instances of R and S? If so, justify why. If not, show a counterexample.

Not equivalent
Let $R = \{\langle 1, 2\rangle, \langle 1, 2\rangle\}$ and $S = \{2\}$
Q1 returns $\{1, 1\}$ while Q2 returns $\{1\}$

7. *Transactions* **(8 points)**. Consider a relation `Emp(ID,salary)` storing employee IDs and salaries, where ID is a key. Consider the following two transactions:

```
T1: begin transaction
    update Emp set salary = 2*salary where ID = 25
    update Emp set salary = 3*salary where ID = 25
    commit

T2: begin transaction
    update Emp set salary = 100 where salary > 100
    commit
```

Suppose the salary of the employee with ID=25 is 100 before either transaction executes.

(a) **(4 points)** If both transactions T1 and T2 execute to completion with isolation level *serializable*, what are the possible final salaries for the employee with ID=25?

100, 600

(b) **(4 points)** Now suppose transaction T1 executes with isolation level *read-committed*, transaction T2 executes with isolation level *read-uncommitted*, and both transactions execute to completion. What are the possible final salaries for the employee with ID=25?

100, 300, 600