CS482 – Test 2b

5pm – 6:10 pm

November 13, 2003

Note:

- There are 6 questions (100 points: 100 points for 20% of the total grade).

- Be short and precise in your answers.

Name:

Signature:

1. (15 points) Given two relation instances

<table>
<thead>
<tr><th colspan="3">R</th></tr>
<tr><th>P</th><th>Q</th><th>T</th></tr>
</thead>
<tbody>
<tr><td>3</td><td>1</td><td>1</td></tr>
<tr><td>1</td><td>1</td><td>2</td></tr>
<tr><td>2</td><td>1</td><td>3</td></tr>
</tbody>
</table>

<table>
<thead>
<tr><th colspan="3">S</th></tr>
<tr><th>A</th><th>B</th><th>C</th></tr>
</thead>
<tbody>
<tr><td>1</td><td>1</td><td>1</td></tr>
<tr><td>1</td><td>1</td><td>2</td></tr>
<tr><td>2</td><td>1</td><td>3</td></tr>
</tbody>
</table>

Compute the following:

1.1 $\pi_{A,B,C}(S)$ — it is S

<table>
<thead>
<tr><th>A</th><th>B</th><th>C</th></tr>
</thead>
<tbody>
<tr><td>1</td><td>1</td><td>1</td></tr>
<tr><td>1</td><td>1</td><td>2</td></tr>
<tr><td>2</td><td>1</td><td>3</td></tr>
</tbody>
</table>

1.2 $\sigma_{P=Q \ OR \ Q=T}(R)$

<table>
<thead>
<tr><th>P</th><th>Q</th><th>T</th></tr>
</thead>
<tbody>
<tr><td>3</td><td>1</td><td>1</td></tr>
<tr><td>1</td><td>1</td><td>2</td></tr>
</tbody>
</table>

1.3 $\pi_{P,Q}(R)[M, N] \cap \pi_{C,A}(S)[M, N]$

<table>
<thead>
<tr><th>M</th><th>N</th></tr>
</thead>
<tbody>
<tr><td>1</td><td>1</td></tr>
<tr><td>2</td><td>1</td></tr>
</tbody>
</table>

2. (10 points) Given the following relation schemas of a database about computers, printers, and producers:

PRODUCT(maker, model, type)

PC(model, speed, ram, hd, screen, price)

LAPTOP(model, speed, ram, hd, screen, price)

PRINTER(model, color, type, price)

(Sample data for these schemas are provided on Page 8). Write expressions of relational algebra to answer the following queries.

2.1 Find those manufacturers (i.e., makers) who produce Laptops but not PC's.

Several possible answers. Here is one:

$$\pi_{maker}(\sigma_{type='Laptop'}(Product) - \sigma_{type='PC'}(Product))$$

2.2 Find the model and price of all products made by manufacturer B (i.e., maker='B').

Several possible answers. Here is one:

$$\pi_{model,price}(\sigma_{maker='B'}(Product \bowtie PC)) \cup$$
$$\pi_{model,price}(\sigma_{maker='B'}(Product \bowtie Laptop)) \cup$$
$$\pi_{model,price}(\sigma_{maker='B'}(Product \bowtie Printer))$$

3. (15 points) Consider the schemas in the previous question. Write SQL queries to answer the following questions.

   3.1 Find those manufacturers (i.e., makers) who produce Laptops but not PC's.

               SELECT P.maker
               FROM PRODUCT P
               WHERE P.type='Laptop'
               EXCEPT
               SELECT P.maker
               FROM PRODUCT P
               WHERE P.type='PC'

   3.2 Find those hard-disk sizes that occur in two or more PC's.

               SELECT PC.hd
               FROM PC P
               WHERE
                    (SELECT COUNT(*) FROM PC P1
                    WHERE P.hd=P1.hd) > 1

4. (10 points) Consider the schema in the Student Database

```
STUDENT(Id, Name, Address, Status)
TRANSCRIPT(StudId, CrsCode, Semester, Grade)
```

and the query

```
SELECT S.Name
FROM STUDENT S
WHERE
    (SELECT COUNT(*) FROM TRANSCRIPT T
    WHERE S.Id=T.StudID AND T.Grade='A') > 2
```

What does this query mean? (Write one short English sentence to describe it).
This query gives us the names of students who get more than 2 (or at least 3) 'A'.

5. (25 points) Consider the relation schema $R = (ABCDE, \mathcal{F})$ where $\mathcal{F} = \{AB \rightarrow C, C \rightarrow D, D \rightarrow B, D \rightarrow E\}$. Answer the following.

5.1 Is the table

A	B	C	D	E
1	1	1	1	1
1	1	2	2	1
2	1	3	1	2

a valid instance of the schema R?

The answer is NO. This is because the FD $AB \rightarrow C$ is violated (the first two rows).

5.2 Find three FDs that violate the BCNF condition. Explain why they violate the conditon. Do not forget that there might be FDs that are entailed by $\mathcal{F}$, which violate the condition.

The key of the relation are AB and AC. All three FDs $C \rightarrow D, D \rightarrow B, D \rightarrow E$ violate the BCNF conditions since they are non-trivial FDs (RHS is not subset of LHS) and the left hand side is not a super key.

5.3 Compute a BCNF decomposition of R.

Decompose using $C \rightarrow D$, we get two shema $R_1 = (CD, \{C \rightarrow D\})$ and $R_2 = (ABCE, \{AB \rightarrow CE, C \rightarrow E\})$.

The first one is already in BCNF. The second one is not in BCNF since $C \rightarrow E$ violates the BCNF condition ($C^+ = CE$).

Decompose R_2 using $C \rightarrow E$, we get two shema $R_{21} = (CE, \{C \rightarrow E\})$ and $R_{22} = (ABC, \{AB \rightarrow C\})$. Both are in BCNF.

One BCNF decomposition is: $R_1 = (CD, \{C \rightarrow D\})$, $R_{21} = (CE, \{C \rightarrow E\})$, and $R_{22} = (ABC, \{AB \rightarrow C\})$.

6. (25 points) Consider the relation schema $R = (ABCDE, \mathcal{D})$ where $\mathcal{D} = \{AB \bowtie ACDE, AC \bowtie BCDE\}$. Answer the following.

6.1 Specify three MVD that violate the 4NF condition. Explain why they violate the conditon.

There is no FDs, so the key of the schema is $ABCDE$. So, both $AB \bowtie ACDE$ and $AC \bowtie BCDE$ violate the 4NF conditions because both MVDs are non-trivial (left-hand-side is not a subset of right-hand-side and right-hand-side is not a subset of left-hand-side) and the intersection of both side is not a super key (which is A or C, respectively). To find the third one, we can rewrite the first one to if $A \twoheadrightarrow B$ and the second one to $C \twoheadrightarrow A$. Using the transitivity, we get $C \twoheadrightarrow B$ and hence the MVD $CB \bowtie ACDE$. This MVD also violates the 4NF condition because $CB \not\subseteq ACDE$, $ACDE \not\subseteq BC$, and $CB \cap ACDE$ is not a superkey.

6.2 Compute a 4NF decomposition of R. Show the steps of your decomposition.

Using $AB \bowtie ACDE$ we get $R_1 = (AB, \{AB \bowtie A\})$ (in 4NF) and $R_2 = (ACDE, \{A \bowtie ACDE, AC \bowtie CDE\})$ (not in 4NF).

Decompose R_2 using $AC \bowtie CDE$ we get $R_{21} = (AC, \{AC \bowtie C\})$ (in 4NF) and $R_{22} = (CDE, \{C \bowtie CDE\})$ (in 4NF).

Final decomposition: $R_1 = (AB, \{AB \bowtie A\})$, $R_{21} = (AC, \{AC \bowtie C\})$, and $R_{22} = (CDE, \{C \bowtie CDE\})$.

6.2 Will your decomposition be changed if we add the FD $A \to CD$ into $\mathcal{D}$.

Adding the FD does not make A or C the key of the schema. The key of the relation is now ABE. Therefore, the violations to the 4NF condition stay. The above decomposition does not change since none of the decomposition steps uses a MVD that would be a 'non-violation' to the new schema.

Sample data for the tables in questions 2 and 3. (This is fictious number. The data is provided for you to test your queries/expressions. Your queries/expressions should work with any available data.)

Sample Data for Product

maker	model	type
A	1000	PC
A	1001	PC
B	1003	PC
B	1004	PC
D	1005	PC
A	2000	Laptop
B	2001	Laptop
D	2002	Laptop
A	3000	Printer
A	3001	Printer
C	3002	Printer

Sample Data for PC

model	speed	ram	hd	screen	price
1000	32	64	0.1	12	4000
1001	160	64	0.5	14	3500
1003	200	512	2	17	3000
1004	160	512	4	13	2500
1005	2800	1024	20	17	3100

Sample Data for Laptop

model	speed	ram	hd	screen	price
2000	64	128	0.1	10	5000
2001	64	128	0.5	12	6000
2002	128	128	1	12	6000

Sample Data for Printer

model	color	type	price
3000	color	ink-jet	1000
3001	bw	laser	1100
3002	color	laser	1500

8