

# CS482 – Test 2 Solution

5pm – 6:10 pm

November 3, 2003

**Note:**

- There are 5 questions (100 points: 100 points for 20% of the total grade).
- Be short and precise in your answers.

**Name:**

**Signature:**

Grades Points Total: .....(100 points)

1. ....(30 points)

1.1 .....(5 points)

1.2 .....(5 points)

1.3 .....(5 points)

1.4 .....(5 points)

1.5 .....(5 points)

1.6 .....(5 points)

2. ....(10 points)

3. ....(10 points)

4. ....(25 points)

4.1 .....(10 points)

4.2 .....(15 points)

5. ....(25 points)

5.1 .....(10 points)

5.2 .....(10 points)

5.3 .....(5 points)

1. (30 points) Given two relation instances

R		
A	B	C
1	0	1
1	1	2
2	1	3

S		
A	D	E
1	0	1
1	1	2
2	1	3

Compute the following:

1.1  $\pi_{A,B}(R)$ :  $\{(1, 0), (1, 1), (2, 1)\}$  or

A	B
1	0
1	1
2	1

1.2  $\sigma_{A>B}(R)$ :  $\{(1, 0, 1), (2, 1, 3)\}$

A	B	C
1	0	1
2	1	3

1.3  $\pi_{A,B}(R) \cup \pi_{D,A}(S)[A, B]$

$\pi_{A,B}(R)$

A	B
1	0
1	1
2	1

$\pi_{D,A}(S)$

D	A
0	1
1	1
1	2

$\pi_{D,A}(S)[A, B]$

A	B
0	1
1	1
1	2

Result

A	B
1	0
1	1
2	1
0	1
1	2

1.4  $\pi_{A,B}(R) - \pi_{D,A}(S)[A, B]$

A	B
1	0
2	1

1.5  $\pi_{A,B}(R) \cap \pi_{D,A}(S)[A, B]$

A	B
1	1

1.6  $R \bowtie S$

A	B	C	D	E
1	0	1	0	1
1	0	1	1	2
1	1	2	0	1
1	1	2	1	2
2	1	3	1	3

2. (10 points) Consider the schema `PERSON(Id, Name, Age)`. Write SQL query that finds the 10<sup>th</sup> oldest person in the relation. A 10<sup>th</sup> oldest person is one such that there are exactly 9 people who are strictly older. (There might be more than one person who have the same age; or there is none of them; see the example)

Id	Name	Age
001	Noah	10
002	Christine	20
003	Adam	20
004	Eva	30

In the above instance, there is no third oldest person since nobody is younger than *exactly* 2 persons (i.e., for everyone there are either 3 or 1 persons who are strictly older); However, there are two second oldest person (Christine and Adam) because there is only one person who is strictly older than them.

```
select T.Id, T.Name, T.Age
from   Person T
where  (select count(*) from Person P where P.Age > T.Age) = 9
```

3. (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, TRANSCRIPT T
```

```
WHERE S.Id = T.StudId
```

```
      AND (T.CrsCode='CS482' OR T.CrsCode='CS582')
```

```
      AND T.Grade='A'
```

What does this query mean? (Write one short English sentence to describe it).

This query asks for a list of students who get an A in either CS482 or CS582.

4. (25 points) Consider the relation schema  $R = (ABCD, \mathcal{F})$  where  $\mathcal{F} = \{AB \rightarrow C, C \rightarrow D, D \rightarrow A\}$ . Answer the following.

4.1 Specify three FDs that violate the BCNF condition. Explain why they violate the condition.

4.2 Compute a BCNF decomposition of  $R$ . Is your decomposition lossless?

4.1 The key of the relation is  $AB$  since  $AB^+ = \{A, B, C, D\}$ . There is no other key.

Since  $C^+ = \{C, D, A\}$  and  $D^+ = \{D, A\}$  we conclude that the two FDs  $C \rightarrow D$  and  $D \rightarrow A$  violate the BCNF condition.

The third FD that violates the BCNF condition is  $C \rightarrow A$ . This FD is also a FD of the relation due to the transitivity property of FDs and the two FDs  $C \rightarrow D$  and  $D \rightarrow A$  (or, due to the fact that  $A \in C^+$ ). Again, the reason is that  $C^+ = \{C, D, A\}$  (and hence  $C$  is not a superkey).

4.2. We follow the BCNF decomposition algorithm. Take  $C \rightarrow D$ , we get a decomposition of  $R$  into  $(CD, \{C \rightarrow D\})$  and  $(CAB, \{AB \rightarrow C, C \rightarrow A\})$ . The first one is already in BCNF but the second one is not due to the fact that  $C \rightarrow A$  violates the BCNF condition. We need to decompose the second one with  $C \rightarrow A$ . This gives us  $(CA, \{C \rightarrow A\})$  and  $(CB, \{\})$ . The final decomposition is given by  $CD, CA, CB$ .

The decomposition is lossless. There might be different answers for this question. One can simply say that the BCNF decomposition algorithm assures losslessness and therefore the decomposition is lossless. One can also check for losslessness using the criterion given in the book.

5. (25 points) Consider the relation schema  $R = (ABCD, \mathcal{D})$  where  $\mathcal{D} = \{AB \twoheadrightarrow ACD, AC \twoheadrightarrow ABD\}$ . Answer the following.

5.1 Specify three MVDs that violate the 4NF condition. Explain why they violate the condition. (It is helpful to think of  $AB \twoheadrightarrow ACD$  as  $A \twoheadrightarrow B$  and  $AC \twoheadrightarrow ABD$  as  $A \twoheadrightarrow C$ .)

5.2 Compute a 4NF decomposition of  $R$ .

5.3 What happens to your answer in (5.1) if  $\mathcal{D}$  is extended with the FD  $A \rightarrow BCD$  (i.e., we add  $A \rightarrow BCD$  to  $\mathcal{D}$ )?

5.1. The key of the relation is  $ABCD$ . Therefore, the two MVDs that belong to  $\mathcal{D}$  violate the 4NF condition because (i) the left-hand-side is not a proper subset of the right-hand-side; (ii) the right-hand-side is not a proper subset of the left-hand-side; and (iii) the intersection of both sides yield  $A$  which is not a superkey.

To find the third MVD that violates the 4NF condition, we can use the properties of the MVD. For example, the additivity rule says that if  $A \twoheadrightarrow B$  and  $A \twoheadrightarrow C$  then  $A \twoheadrightarrow BC$ . This gives us the MVD  $ABC \twoheadrightarrow AD$ . Checking again the 4NF condition, we can confirm that this one also violates the 4NF condition.

5.2. Decompose using  $AB \twoheadrightarrow ACD$  we get two schemes  $(AB, \{AB \twoheadrightarrow A\})$  and  $(ACD, \{A \twoheadrightarrow ACD, AC \twoheadrightarrow AD\})$ . The first one is in 4NF but the second one is not since  $AC \twoheadrightarrow AD$  violates the 4NF condition. Decomposing the second one with respect to this MVD gives us  $(AC, \{A \twoheadrightarrow AC\})$  and  $(AD, \{A \twoheadrightarrow AD\})$  both of them are in 4NF.

The final decomposition is  $AB, AC, AD$ .

5.3. If we add the FD  $A \rightarrow BCD$  then  $A$  is a key and there is no 4NF violation anymore.