

# DSAA 5012

## Advanced Data Management for Data Science

### LECTURE 9 EXERCISES

### RELATIONAL DATABASE DESIGN: FUNCTIONAL DEPENDENCIES



# EXERCISE 1

Assume that this table contains the *only* set of tuples that may appear in a relation  $R(X, Y, V, W)$ . Which of the following FDs hold in  $R$ ?

tuple	X	Y	V	W
1	$x_1$	$y_1$	$v_1$	$w_1$
2	$x_1$	$y_1$	$v_2$	$w_2$
3	$x_2$	$y_1$	$v_1$	$w_3$
4	$x_2$	$y_1$	$v_3$	$w_4$

- $X \rightarrow X$       **Yes** – trivial (holds in any relation)
- $X \rightarrow Y$       **Yes** – for a given X value all Y values are identical
- $X \rightarrow V$       **No** – V values differ for same X value (e.g., tuples 1 & 2)
- $X \rightarrow W$       **No** – W values differ for same X value (e.g., tuples 1 & 2)
- $Y \rightarrow X$       **No** – X values differ for same Y value (e.g., tuples 2 & 3)
- $W \rightarrow X$       **Yes** – all W values are different
- $XV \rightarrow Y$     **Yes** – X alone determines Y
- $YV \rightarrow X$     **No** – X values differ for same YV value in tuples 1 and 3



## EXERCISE 2

In Exercise 1, we assumed that we know all possible records in the table, which is not usually true. In general, by looking at an instance of a relation, we can only tell FDs that are not satisfied.

List 5 FDs that are not satisfied in the table.

A	B	C
a <sub>1</sub>	b <sub>1</sub>	c <sub>1</sub>
a <sub>1</sub>	b <sub>1</sub>	c <sub>2</sub>
a <sub>2</sub>	b <sub>1</sub>	c <sub>1</sub>
a <sub>2</sub>	b <sub>1</sub>	c <sub>3</sub>

$A \rightarrow C$

$B \rightarrow A$

$B \rightarrow C$

$C \rightarrow A$

$AB \rightarrow C$

## EXERCISE 3

Given relation schema  $R(X, Y, U, V, W)$  and  $F = \{X \rightarrow Y, UV \rightarrow W, V \rightarrow X\}$

a) Determine the closure of each attribute.

$$X^+ = \{X, Y\} \quad (\text{Look for } X \text{ on LHS of FDs})$$

$$Y^+ = \{Y\}$$

$$U^+ = \{U\}$$

$$V^+ = \{V, X, Y\}$$

$$W^+ = \{W\}$$

b) What are the candidate keys of  $R$ ?

The candidate key is  $UV$  since  $UV^+ = \{X, Y, U, V, W\}$ .

## EXERCISE 4

Given relation schema  $R(A, B, C, G, H, I)$  and

$$F = \{A \rightarrow B, \quad A \rightarrow C, \quad CG \rightarrow H, \quad CG \rightarrow I, \quad B \rightarrow H\}$$

a) Is  $AG$  a (super)key of  $R$  given  $F$ ? **Yes** **Why?** Since  $AG \rightarrow R$

Compute  $AG^+$

$$AG^{(0)} = \{A, G\}$$

$$AG^{(1)} = \{A, G, B\}$$

$$(A \rightarrow B \text{ and } A \subseteq \{A, G\})$$

$$AG^{(2)} = \{A, G, B, C\}$$

$$(A \rightarrow C \text{ and } A \subseteq \{A, G\})$$

$$AG^{(3)} = \{A, G, B, C, H\}$$

$$(CG \rightarrow H \text{ and } CG \subseteq \{A, G, B, C\})$$

$$AG^{(4)} = \{A, G, B, C, H, I\}$$

$$(CG \rightarrow I \text{ and } CG \subseteq \{A, G, B, C, H\})$$

b) Is  $AG$  a candidate key? **Yes** **Why?** Cannot remove  $A$  or  $G$

c) Does  $A^+ \rightarrow R$  hold? **No** since  $A^+ = \{A, B, C, H\}$

d) Does  $G^+ \rightarrow R$  hold? **No** since  $G^+ = \{G\}$

# EXERCISES 5, 6, 7

**Complete Exercises 5, 6 and 7; upload your completed worksheet to Canvas by 11 p.m.**



## EXERCISE 5

Given relation schema  $R(A, B, C, D, E)$  and  $F = \{A \rightarrow B, AB \rightarrow C, D \rightarrow AC\}$

a) Determine the following attribute closures.

$$A^+ = \{A, B, C\} \quad B^+ = \{B\} \quad C^+ = \{C\} \quad D^+ = \{D, A, B, C\} \quad E^+ = \{E\}$$

b) What are the candidate keys of  $R$ ?

DE

c) Find a canonical cover of  $F$ .

$$F_c = \{A \rightarrow B, AB \rightarrow C, D \rightarrow AC\}$$

$$F_c = \{A \rightarrow B, A \rightarrow C, D \rightarrow AC\}$$

$$F_c = \{A \rightarrow BC, D \rightarrow AC\}$$

$$F_c = \{A \rightarrow BC, D \rightarrow A\}$$

B **extraneous** in  $AB \rightarrow C$  since  $A \rightarrow AB$  &  
 $AB \rightarrow C \Rightarrow A \rightarrow C$

use **union rule**

C **extraneous in**  $D \rightarrow AC$  since  $D \rightarrow A$  &  $A \rightarrow C$   
 $\Rightarrow D \rightarrow C$

# EXERCISE 6

We want to create the database for a bank that contains accounts (A), branches (B) and customers (C).

a) What are the functional dependencies implied by the following constraints?

i. An account cannot be shared by multiple customers.

Account  $\rightarrow$  Customer

A  $\rightarrow$  C

ii. Two different branches do not have the same account.

Account  $\rightarrow$  Branch

A  $\rightarrow$  B

iii. Each customer can have at most one account in a branch (but different accounts in different branches).

Branch, Customer  $\rightarrow$  Account

BC  $\rightarrow$  A

b) What are the candidate keys?

(Branch, Customer) and Account

BC and A





# EXERCISE 7

Given relation schema  $R(A, B, C)$ . Assume we do not know the keys of the relation. Write a valid SQL query whose result can be used to determine if  $A$  is a potential candidate key. Explain how to interpret the query result to determine if  $A$  is a potential candidate key.

```
select A
from R
group by A
having count(*)>1;
```

If the result is non-empty,  
then  $A$  is not a candidate key.

If the result is empty, we  
cannot be sure.

```
select A, count(*)
from R
group by A;
```

If the count is not 1 for  
all  $A$ , then  $A$  is not a  
candidate key.

If the count is 1 for all  
 $A$ , we cannot be sure.

```
select count(A), count(distinct A)
from R;
```

If the two counts do not match for  
all  $A$ , then  $A$  is not a candidate key.

If the counts match for all  $A$  we  
cannot be sure.

Write an SQL query to test if the FD  $A \rightarrow B$  holds in  $R$ ?

- Same as the first query above, but replace the last line with **having count(distinct B)>1**