# DSAA 5012
# Advanced Data Management for Data Science

## LECTURE 10 EXERCISES
## RELATIONAL DATABASE DESIGN: NORMALIZATION

# EXERCISE 1

Given: R(A, B, C, D, E) $F = \{A \rightarrow BC\}$

Decomposition: $R_1$(A, B, C) and $R_2$(A, D, E)

a) Is the decomposition lossless? Why? (*iff* $R_1 \cap R_2 \rightarrow R_1$ or $R_1 \cap R_2 \rightarrow R_2$)

**Yes** The common attribute A is a key for $R_1$.

b) Is the decomposition dependency preserving? Why? (*iff* $(\cup F_i)^+ = F^+$)

**Yes** $A \rightarrow BC$ is preserved in $R_1$.

c) Is the decomposition $R_1$(A, B, C) and $R_2$(C, D, E) lossless? Why?

**No** C is not a key for any table.

# EXERCISE 2

Given: R(A, B, C, D, E) $F$ = {A→BC, CD→E, B→D, E→A}

Decomposition: $R_1$(A, B, C) and $R_2$(A, D, E)

a) Is the decomposition lossless? (*iff* $R_1 \cap R_2 \rightarrow R_1$ or $R_1 \cap R_2 \rightarrow R_2$)

**Yes** The common attribute A is a key for $R_1$.

b) Is the decomposition dependency preserving? (*iff* $(\cup F_i)^+ = F^+$)

**No** We lose CD→E and B→D.

# EXERCISE 3

**2NF**
R is in 2NF *if and only if*
For each FD: $X \rightarrow A$ in $F^+$:
- $A \in X$ (*trivial FD*) ***or***
- X is not a proper subset of a candidate key for R ***or***
- A is a prime attribute for R.

a) Given: R(A, B, C, D) $F = \{AB \rightarrow CD, B \rightarrow C\}$

Is R in 2NF? Why?

Key: AB $AB^+=\{A, B, C, D\}$ $B^+=\{B, C\}$

**No** For B→C, B *is* a proper subset of the key AB *and* C is non-prime. So, R is not in 2NF.

b) Given: R(A, B, C, D) $F = \{AB \rightarrow CD, C \rightarrow D\}$

Is R in 2NF? Why?

Key: AB $AB^+=\{A, B, C, D\}$ $C^+=\{C, D\}$

**Yes** For C→D, C *is not* a proper subset of the key, so R is in 2NF.

# EXERCISE 4

Identify the candidate key(s) and the current highest normal form for each of the following relation schemas given their corresponding FDs.

a) R(A, B, C, D, E)    $F$ = {A→B, C→D} = $F^+$

$A^+$={A, B}    $C^+$={C, D}

Candidate keys: ACE

⟹ For A→B and C→D

i. A and C *are* proper subsets of the candidate key ACE (both FDs fail 1st 2NF test).

ii. both B and D *are not* prime attributes of R (both FDs fail 2nd 2NF test).

☞ Both FDs violate 2NF.

Normal form: 1NF

**2NF**
R is in 2NF *if and only if*
For each FD: X→A in $F^+$:
A ∈ X (*trivial FD*) ***or***
X is not a proper subset of a candidate key for R ***or***
A is a prime attribute for R.

**3NF**
R is in 3NF *if and only if*
For each FD: X→A in $F^+$:
A ∈ X (*trivial FD*) ***or***
X is a superkey for R ***or***
A is a prime attribute for R.

# EXERCISE 4 (cont'd)

Identify the candidate key(s) and the current highest normal form for each of the following relation schemas given their corresponding FDs.

b) R(A, B, C) $F = \{AB \rightarrow C, C \rightarrow B\} = F^+$

$AB^+ = \{A, B, C\}$ $C^+ = \{C, B\}$

Candidate keys: AB, AC

⟹ For AB→C,
C *is* a prime attribute of R
(FD passes 2nd 2NF and 3NF tests).

⟹ For C→B,
B *is* a prime attribute of R
(FD passes 2nd 2NF and 3NF tests).

☞ Both FDs satisfy 3NF.

Normal form: 3NF

**2NF**
R is in 2NF *if and only if*
For each FD: X→A in $F^+$:
A ∈ X (*trivial FD*) ***or***
X is not a proper subset of a candidate key for R ***or***
A is a prime attribute for R.

**3NF**
R is in 3NF *if and only if*
For each FD: X→A in $F^+$:
A ∈ X (*trivial FD*) ***or***
X is a superkey for R ***or***
A is a prime attribute for R.

# EXERCISE 4 (cont'd)

Identify the candidate key(s) and the current highest normal form for each of the following relation schemas given their corresponding FDs.

c) R(A, B, C, F) $\mathcal{F}$ = {AB→C, C→F} = $\mathcal{F}^+$

$AB^+$={A, B, C, F} $C^+$={C, F}

Candidate keys: AB

⟹ For AB→C

i. AB *is not* a proper subset of a candidate key (FD passes 1st 2NF test);

ii. AB is a superkey for R (FD passes 1st 3NF test).

⟹ For C→F

i. C *is not* a proper subset of a candidate key (FD passes 1st 2NF test);

ii. C is not a superkey of R (FD fails 1st 3NF test);

iii. F *is not* a prime attribute (FD fails 2nd 3NF test).

☞ Both FDs satisfy 2NF.

Normal form: 2NF

**2NF**

R is in 2NF *if and only if*

For each FD: X→A in $\mathcal{F}^+$:

- A ∈ X (*trivial FD*) ***or***
- X is not a proper subset of a candidate key for R ***or***
- A is a prime attribute for R.

**3NF**

R is in 3NF *if and only if*

For each FD: X→A in $\mathcal{F}^+$:

- A ∈ X (*trivial FD*) ***or***
- X is a superkey for R ***or***
- A is a prime attribute for R.

# EXERCISES

**Exercise 5:** Decompose R(A, B, C, D, E, F, G) into 3NF relations for the FD set $F$ = {AB→CD, C→EF, G→A, G→F, CE→F}.

**Exercise 6:** Decompose R(A, B, C, D) into 3NF and BCNF relations for each of the following FD sets.

a) $F$ = {B→C, D→A)

b) $F$ = {ABC→D, D→A}

**Exercise 7:** Given: Sale(customer, store, product, price) and the constraints:

A customer buys from only one store.

There is a unique price for each product in a store, but the same product can have a different price in different stores.

a) What are the FDs implied by the above description?

b) What are the candidate keys?

c) Explain why Sale is not in 3NF.

d) Decompose Sale into 3NF relation schemas.

e) Is the decomposition dependency preserving? Why?

# EXERCISE 5

Decompose R(A, B, C, D, E, F, G) into 3NF relations for the FD set
$F$ = {AB→CD, C→EF, G→A, G→F, CE→F}.

Attribute closures

$AB^+$ = {A, B, C, D, E, F}    $C^+$ = {C, E, F}    $G^+$ = {G, A, F}    $CE^+$ = {C, E, F}

Candidate key: BG    From G→A we can infer BG→AB using **IR2**.

$BG^+$ = {B, G, A, C, D, E, F}    All FDs violate 3NF!

Canonical cover: $F_C$ = {AB→CD, C→EF, G→AF}

3NF decomposition: $R_1$(<u>A, B</u>, C, D)    $R_3$(<u>G,</u> A, F)

$R_2$(<u>C</u>, E, F)    $R_4$(<u>B, G</u>) ⟹ due to the candidate key

# EXERCISE 6

Decompose R(A, B, C, D) into 3NF and BCNF relations for each of the following FD sets.

a) $F = \{B \rightarrow C, D \rightarrow A\}$

Attribute closures: $B^+ = \{B, C\}$ $D^+ = \{D, A\}$

Candidate keys: BD

Canonical cover: $F_C = \{B \rightarrow C, D \rightarrow A\}$

<u>3NF Decomposition</u>

$R_1$(<u>B</u>, C)

$R_2$(<u>D</u>, A)

$R_3$(<u>B, D</u>) ⟹ due to the candidate key

<u>BCNF Decomposition</u>

$R_1$(<u>B</u>, C)

$R_2$(<u>D</u>, A)

$R_3$(<u>B, D</u>)

# EXERCISE 6 (cont'd)

Decompose R(A, B, C, D) into 3NF and BCNF relations for each of the following FD sets.

b) $F$ = {ABC→D, D→A}

Attribute closures: $ABC^+$ = {A, B, C, D}   $D^+$ = {D, A}

Candidate keys: ABC, BCD

**Using IR2:** BCD→ABC

Canonical cover: $F_C$ = {ABC→D, D→A}

3NF Decomposition

Already in 3NF

R(A, B, C, D)

Dependency preserving?

**Yes**

BCNF Decomposition

Only D→A violates BCNF

$R_1$(B, C, D)   $R_2$(D, A)

Dependency preserving?

**No** The FD ABC→D is lost.

Is this BCNF decomposition OK?

$R_1$(A, B, C) and $R_2$(D, A) **No Why?**

The FD ABC→D is lost, and we need to join $R_1$ and $R_2$ on A, but A is not a key of $R_2$.
⟹ **Not a lossless decomposition!**

# EXERCISE 7

Given: Sale(customer, store, product, price) and the constraints:

A customer buys from only one store.

There is a unique price for each product in a store, but the same product can have a different price in different stores.

a) What are the FDs implied by the above description?

customer → store

store, product → price

b) What are the candidate keys?

{customer, product}

Since customer, product → store, product (**IR2**)
and customer, product → price (**IR3**)

Also $\{customer, product\}^+$ = {customer, product, store, price}

# EXERCISE 7 (cont'd)

**3NF**
R is in 3NF *if and only if*
For each FD: X→A in $F^+$:
A ∈ X (*trivial FD*) ***or***
X is a superkey for R ***or***
A is a prime attribute for R.

Given: Sale(customer, store, product, price)

A customer buys from only one store.

There is a unique price for each product in a store, but the same product can have a different price in different stores.

c) Explain why Sale is not in 3NF.

candidate key: {customer, product}

Both FDs violate 3NF.

$F$ = {customer → store; store, product → price}

The LHS of the FDs are not superkeys; the RHS are not prime attributes of Sale.

d) Decompose Sale into 3NF relation schemas.

$R_1$(customer, store) $R_2$(store, product, price)

e) Is the decomposition dependency preserving? Why?

**Yes** Each FD is preserved in a relation schema, **BUT** … (next page).

# EXERCISE 7 (cont'd)

The decomposition $R_1$(customer, store), $R_2$(store, product, price) is lossy because the common attribute store is not a key of any table.

Sale

| customer | product | store | price |
|---|---|---|---|
| c1 | p1 | s1 | pr1 |
| c1 | p2 | s1 | pr2 |
| c2 | p1 | s1 | pr1 |

decompose

$R_1$

| customer | store |
|---|---|
| c1 | s1 |
| c2 | s1 |

$R_2$

| store | product | price |
|---|---|---|
| s1 | p1 | pr1 |
| s1 | p2 | pr2 |

The two decomposed relations do not generate the original one if joined (on the common store attribute). The join result contains 4 records instead of 3 as in the original relation.

**What is the problem?** None of the fragments contains the candidate key (customer, product).

**Solution?** Include an additional table $R_3$(customer, product) containing the candidate key in the decomposition.

$R_3$

| customer | product |
|---|---|
| c1 | p1 |
| c1 | p2 |
| c2 | p1 |