Database Technology

Teaching Session 1: Functional Dependencies and Normalization

Olaf Hartig

olaf.hartig@liu.se



Reasoning About FDs

- Logical implications can be derived by using inference rules called **Armstrong's rules**:
 - Reflexivity: If Y is a subset of X, then $X \rightarrow Y$
 - Augmentation: If $X \to Y$, then $XZ \to YZ$ (we use XY as a short form for $X \cup Y$)
 - Transitivity: If $X \to Y$ and $Y \to Z$, then $X \to Z$
- Additional rules can be derived:
 - Decomposition: If $X \rightarrow YZ$, then $X \rightarrow Y$
 - Union: If $X \rightarrow Y$ and $X \rightarrow Z$, then $X \rightarrow YZ$
 - Pseudo-transitivity: If $X \to Y$ and $WY \to Z$, then $WX \to Z$



Consider the relation R(A,B,C,D,E,F) with the following FDs:

FD1: A → BC

FD2: C → AD

FD3: **DE** → **F**

■ Use the Armstrong rules to derive the following FD: AC → D



Consider the relation R(A,B,C,D,E,F) with the following FDs:

FD1: A → BC

FD2: C → AD

- Use the Armstrong rules to derive the following FD: AC → D
 - FD4: AC → AD (Augmentation of FD2 with A)
 - FD5: AC → D (Decomposition of FD4)



Consider the relation R(A,B,C,D,E,F) with the following FDs:

FD1: A → BC

FD2: $\mathbf{C} \rightarrow \mathbf{AD}$

- Use the Armstrong rules to derive the following FD: AC → D
 - FD4: AC → AD (Augmentation of FD2 with A)
 - FD5: AC → D (Decomposition of FD4)
- Use the Armstrong rules to derive the following FD: A → D



Consider the relation R(A,B,C,D,E,F) with the following FDs:

FD1: **A** → **BC**

FD2: $\mathbf{C} \rightarrow \mathbf{AD}$

- Use the Armstrong rules to derive the following FD: AC → D
 - FD4: AC → AD (Augmentation of FD2 with A)
 - FD5: AC → D (Decomposition of FD4)
- Use the Armstrong rules to derive the following FD: A → D
 - FD6: A → C (Decomposition of FD1)
 - FD7: A → AD (Transitivity of FD6 and FD2)
 - FD8: A → D (Decomposition of FD7)



Warmup (cont'd)

Consider the relation R(A,B,C,D,E,F) with the following FDs:

FD1: A → BC

FD2: C → AD

FD3: **DE** → **F**

Use the Armstrong rules to derive the following FD: AE → ABCDEF



Warmup (cont'd)

Consider the relation R(A,B,C,D,E,F) with the following FDs:

FD1: **A** → **BC**

FD2: C → AD

- Use the Armstrong rules to derive the following FD: AE → ABCDEF
 - FD9: AE → BCE (Augmentation of FD1 with E)
 - FD10: AE → C (Decomposition of FD9)
 - FD11: **AE** → **AD** (Transitivity of FD10 and FD2)
 - FD12: **AE** → **ADE** (Augmentation of FD11 with E)
 - FD13: AE → DE (Decomposition of FD12)
 - FD14: AE → F (Transitivity of FD13 and FD3)
 - FD15: AE → ABCDEF (Union of FD9, FD11, and FD14)



Computing (Super)Keys

```
function ComputeAttrClosure(X, F)

begin
X^+ := X;

while F contains an FD Y \rightarrow Z such that

(i) Y is a subset of X^+, and

(ii) Z is not a subset of X^+ do

X^+ := X^+ \cup Z;

end while

return X^+;

end
```



Warmup (cont'd)

Consider the relation R(A,B,C,D,E,F) with the following FDs:

```
FD1: A \rightarrow BC
FD2: C \rightarrow AD
FD3: DE \rightarrow F
```

■ Compute the **attribute closure** of $X = \{ A, E \}$ w.r.t. $F = \{ FD1, FD2, FD3 \}$ to show that we have: $AE \rightarrow ABCDEF$

```
    Initially: X<sup>+</sup> = { A,E }
    By using FD1: X<sup>+</sup> = { A,E,B,C }
    By using FD2: X<sup>+</sup> = { A,E,B,C,D }
    By using FD3: X<sup>+</sup> = { A,E,B,C,D,F }
```

Your Turn

Consider the relation R(A,B,C,D) with the following set F of FDs:

FD1: AB → C

FD2: BC → D

- Compute the following attribute closures w.r.t. $F = \{FD1, FD2, FD3\}$
 - $\{B,C,D\}^{+} = ?$
 - $\{A,C,D\}^+ = ?$
 - $\{A,B,D\}^{+} = ?$
 - $\{A,B,C\}^{+} = ?$

Your Turn

Consider the relation R(A,B,C,D) with the following set F of FDs:

FD1: AB → C

FD2: BC → D

FD3: $D \rightarrow B$

- Compute the following attribute closures w.r.t. $F = \{FD1, FD2, FD3\}$
 - $\{B,C,D\}^+ = \{B,C,D\}$
 - $\{A,C,D\}^+ = \{A,C,D,B\}$
 - $\{A,B,D\}^+ = \{A,B,D,C\}$
 - $\{A,B,C\}^+ = \{A,B,C,D\}$

Consider the relation R(A,B,C,D) with the following set F of FDs:

FD1: AB → C

FD2: BC → D

FD3: **D** → **B**

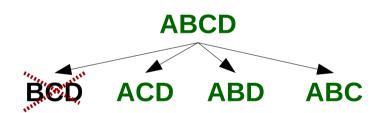
• A set X of attributes of R is a superkey if X^+ contains all the attributes of R

$$- \{B,C,D\}^+ = \{B,C,D\}$$

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

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

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



Consider the relation R(A,B,C,D) with the following set F of FDs:

FD1: AB → C

FD2: BC → D

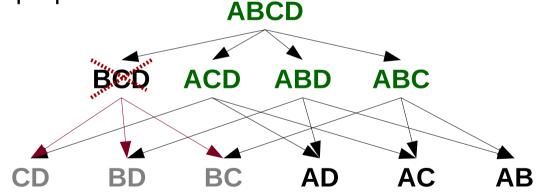
FD3: $D \rightarrow B$

- A set X of attributes of R is a superkey if X^+ contains all the attributes of R
- X is a candidate key (CK) if no proper subset X' of X is a superkey

$$- \{A,D\}^{+} = ?$$

$$- \{A,C\}^+ = ?$$

$$- \{A,B\}^{+} = ?$$



Consider the relation R(A,B,C,D) with the following set F of FDs:

FD1: **AB** → **C**

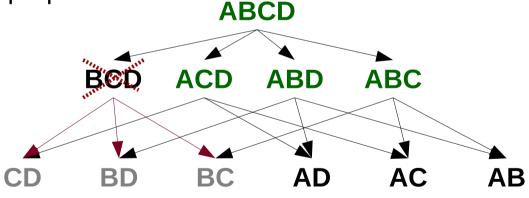
FD2: BC → D

- A set X of attributes of R is a superkey if X^+ contains all the attributes of R
- X is a candidate key (CK) if no proper subset X' of X is a superkey

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

$$- \{A,C\}^+ = \{A,C\}$$

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



Consider the relation R(A,B,C,D) with the following set F of FDs:

FD1: AB → C

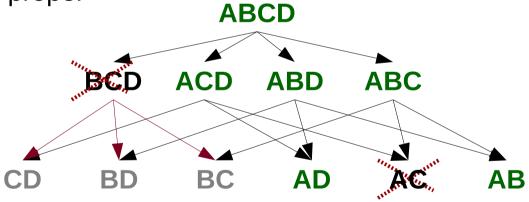
FD2: BC → D

- A set X of attributes of R is a superkey if X^+ contains all the attributes of R
- X is a candidate key (CK) if no proper subset X' of X is a superkey

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

$$- \{A,C\}^+ = \{A,C\}$$

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



Consider the relation R(A,B,C,D) with the following set F of FDs:

FD1: **AB** → **C**

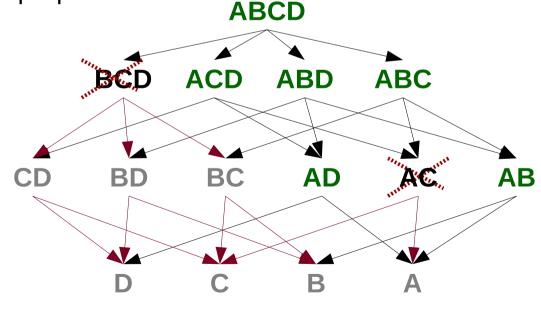
FD2: $BC \rightarrow D$

FD3: **D** → **B**

• A set X of attributes of R is a superkey if X^+ contains all the attributes of R

X is a candidate key (CK) if no proper

subset X' of X is a superkey





Consider the relation R(A,B,C,D) with the following set F of FDs:

FD1: **AB** → **C**

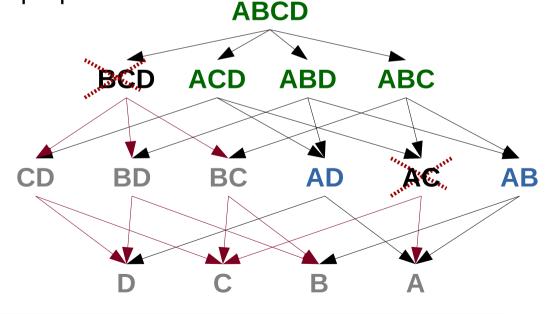
FD2: $BC \rightarrow D$

FD3: **D** → **B**

• A set X of attributes of R is a superkey if X^+ contains all the attributes of R

X is a candidate key (CK) if no proper

subset X' of X is a superkey





Consider the relation R(A,B,C,D) with the following set F of FDs:

FD1: AB → C

FD2: $BC \rightarrow D$

FD3: **D** → **B**

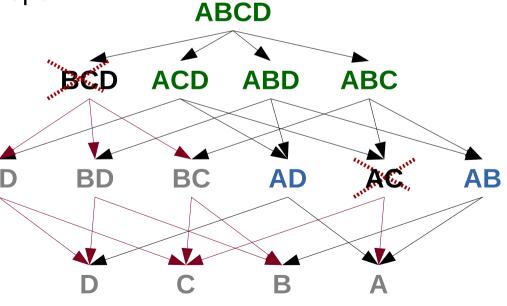
• A set X of attributes of R is a superkey if X^+ contains all the attributes of R

X is a candidate key (CK) if no proper subset X' of X is a superkey

Some steps may be skipped

 If an attribute is in the RHS of FDs but nowhere in the LHS, it cannot be part of any CK

 If an attribute is nowhere in the RHS, it must be part of any candidate key (such as A in the example)





Consider the relation R(A,B,C,D) with the following set F of FDs:

FD1: AB → C

FD2: BC → D

"FD3:"'**D**......B.

- A set X of attributes of R is a superkey if X^+ contains all the attributes of R
- X is a candidate key (CK) if no proper subset X' of X is a superkey
- Some steps may be skipped
 - If an attribute is in the RHS of FDs but nowhere in the LHS, it cannot be part of any CK
 - If an attribute is nowhere in the RHS, it must be part of every candidate key (such as A in the example)

- For instance, if we had only FD1 and FD2 (i.e., not FD3), we could immediately say that:
 - A and B must be in every CK
 - and, since {A,B}⁺ = {A,B,C,D},
 {A,B} is a superkey and thus
 the only CK



Your Turn

Consider the relation R(A,B,C,D,E,F) with the following FDs:

FD1: A → BC

FD2: C → AD

FD3: **DE** \rightarrow **F**

What are the candidate keys?



Your Turn

Consider the relation R(A,B,C,D,E,F) with the following FDs:

FD1: A → BC

FD2: C → AD

FD3: **DE** \rightarrow **F**

- What are the candidate keys?
 - {A,E} and {C,E}



BCNF

- Relation schema R with a set F of functional dependencies is in BCNF if for every non-trivial $FD X \rightarrow Y$ in F⁺ we have that X is a superkey
- Consider the relation R(A,B,C,D,E,F) with the following FDs:

FD1: A → BC

FD2: C → AD

- What are the candidate keys?
 - {A,E} and {C,E}
- Is the given relation in BCNF?
 - If not, identify the FDs that violate the BCNF condition.



Your Turn

- Relation schema R with a set F of functional dependencies is in BCNF if for **every** non-trivial **FD** $X \rightarrow Y$ in F⁺ we have that X is a superkey
- Consider the relation R(A,B,C,D) with the following set F of FDs:

FD1: AB → C

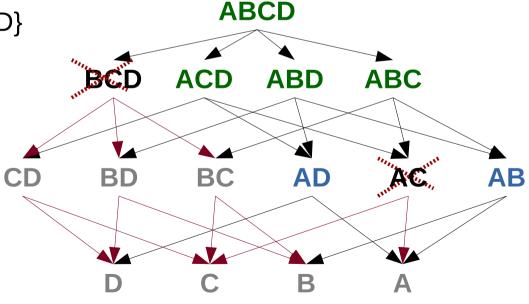
FD2: BC → D

FD3: **D** → **B**

Candidate keys: {A,B} and {A,D}

Is the given relation in BCNF?

 If not, identify the FDs that violate the BCNF condition.





- By using an FD $X \rightarrow Y$ that violates BCNF, decompose R into
 - R1 with all the attributes in X and in Y, and
 - R2 with all attributes from R except those that are in Y and not in X
- Consider the relation R(A,B,C,D) with the following set F of FDs:

FD1: AB → C

FD2: **BC** → **D**

FD3: **D** → **B**

Let's decompose based on FD2



- By using an FD $X \rightarrow Y$ that violates BCNF, decompose R into
 - R1 with all the attributes in X and in Y, and
 - R2 with all attributes from R except those that are in Y and not in X
- Consider the relation R(A,B,C,D) with the following set F of FDs:

FD1: AB → C

FD2: **BC** → **D**

FD3: **D** → **B**

Let's decompose based on FD2

- R1(B,C,D)
- R2(A,B,C)



- By using an FD $X \rightarrow Y$ that violates BCNF, decompose R into
 - R1 with all the attributes in X and in Y, and
 - R2 with all attributes from R except those that are in Y and not in X
- Consider the relation R(A,B,C,D) with the following set F of FDs:

FD1: AB → C

FD2: BC → D

- Let's decompose based on FD2 and don't forget to determine the FDs of the resulting relation schemas
 - R1(B,C,D)
 - R2(A,B,C)



- By using an FD $X \rightarrow Y$ that violates BCNF, decompose R into
 - R1 with all the attributes in X and in Y, and
 - R2 with all attributes from R except those that are in Y and not in X
- Consider the relation R(A,B,C,D) with the following set F of FDs:

FD1: AB → C

FD2: **BC** → **D**

- Let's decompose based on FD2 and don't forget to determine the FDs of the resulting relation schemas
 - R1(B,C,D) with FDs: FD2 and FD3
 - R2(A,B,C) with FDs: FD1



- By using an FD $X \rightarrow Y$ that violates BCNF, decompose R into
 - R1 with all the attributes in X and in Y, and
 - R2 with all attributes from R except those that are in Y and not in X
- Consider the relation R(A,B,C,D) with the following set F of FDs:

FD1: AB → C

FD2: BC → D

- Let's decompose based on FD2 and don't forget to determine the FDs of the resulting relation schemas, and the CKs
 - R1(B,C,D) with FDs: FD2 and FD3
 - R2(A,B,C) with FDs: FD1



- By using an FD $X \rightarrow Y$ that violates BCNF, decompose R into
 - R1 with all the attributes in X and in Y, and
 - R2 with all attributes from R except those that are in Y and not in X
- Consider the relation R(A,B,C,D) with the following set F of FDs:

FD1: AB → C

FD2: **BC** → **D**

- Let's decompose based on FD2 and don't forget to determine the FDs of the resulting relation schemas, and the CKs
 - R1(B,C,D) with FDs: FD2 and FD3, CKs: {B,C}, {C,D}
 - R2(A,B,C) with FDs: FD1, CK: {A,B}



- By using an FD $X \rightarrow Y$ that violates BCNF, decompose R into
 - R1 with all the attributes in X and in Y, and
 - R2 with all attributes from R except those that are in Y and not in X
- Consider the relation R(A,B,C,D) with the following set F of FDs:

FD1: AB → C

FD2: **BC** → **D**

- Let's decompose based on FD2 and don't forget to determine the FDs of the resulting relation schemas, and the CKs
 - R1(B,C,D) with FDs: FD2 and FD3, CKs: {B,C}, {C,D}
 - R2(A,B,C) with FDs: FD1, CK: {A,B}
 - Are they in BCNF?



- By using an FD $X \rightarrow Y$ that violates BCNF, decompose R into
 - R1 with all the attributes in X and in Y, and
 - R2 with all attributes from R except those that are in Y and not in X
- Consider the relation R(A,B,C,D) with the following set F of FDs:

FD1: AB → C

FD2: BC → D

- Let's decompose based on FD2 and don't forget to determine the FDs of the resulting relation schemas, and the CKs
 - R1(B,C,D) with FDs: FD2 and FD3, CKs: {B,C}, {C,D}
 - R2(A,B,C) with FDs: FD1, CK: {A,B}
 - Your turn: decompose R1 based on FD3 (and don't forget ...)



- By using an FD $X \rightarrow Y$ that violates BCNF, decompose R into
 - R1 with all the attributes in X and in Y, and
 - R2 with all attributes from R except those that are in Y and not in X
- Consider the relation R(A,B,C,D) with the following set F of FDs:

FD1: AB → C

FD2: **BC** → **D**

- Let's decompose based on FD2 and don't forget to determine the FDs of the resulting relation schemas, and the CKs
 - ..R.1(..B,.C,D.)-with--FD9::FD2:and:FD9;:"CK5::{B,C}; {C,D}"
 - R2(A,B,C) with FDs: FD1, CK: {A,B}
 - R3(D,B) with FD3, CK $\{D\}$ R4(C,D) only trivial FDs, CK: $\{C,D\}$



Your Turn

- By using an FD $X \rightarrow Y$ that violates BCNF, decompose R into
 - R1 with all the attributes in X and in Y, and
 - R2 with all attributes from R except those that are in Y and not in X
- Consider the relation R(A,B,C,D) with the following set F of FDs:

FD1: AB → C

FD2: BC → D

FD3: **D** → **B**

 Let's decompose based on FD3 – and don't forget to determine the FDs of the resulting relation schemas, and the CKs



Your Turn

- By using an FD $X \rightarrow Y$ that violates BCNF, decompose R into
 - R1 with all the attributes in X and in Y, and
 - R2 with all attributes from R except those that are in Y and not in X
- Consider the relation R(A,B,C,D) with the following set F of FDs:

FD1: AB → C

FD2: BC → D

FD3: **D** → **B**

- Let's decompose based on FD3 and don't forget to determine the FDs of the resulting relation schemas, and the CKs
 - R1(B,D) with FD3, CK: {D}
 - R2(A,C,D) with FD4: AD \rightarrow C, CK: {A,D}

R1 and R2 are in BCNE

can be derived from FD3 and FD1 using the augmentation rule and the transitivity rule



Different Example

- By using an FD $X \rightarrow Y$ that violates BCNF, decompose R into
 - R1 with all the attributes in X and in Y, and
 - R2 with all attributes from R except those that are in Y and not in X
- Consider the relation schema R(A,B,C,D,E,F) with the following FDs:

FD1: **AB** → **CDEF**

FD2: **E** → **F**

- Your turn:
 - Determine candidate key(s)
 - Is R in BCNF?
 - If not, normalize into a set of BCNF relation schemas (and don't forget to determine FDs and CKs along the way)



Different Example

- By using an FD $X \rightarrow Y$ that violates BCNF, decompose R into
 - R1 with all the attributes in X and in Y, and
 - R2 with all attributes from R except those that are in Y and not in X
- Consider the relation schema R(A,B,C,D,E,F) with the following FDs:

FD1: **AB** → **CDEF**

FD2: **E** → **F**

- Your turn:
 - Determine candidate key(s) {A,B}
 - Is R in BCNF? No, FD2 violates the BCNF condition.
 - If not, normalize into a set of BCNF relation schemas
 We decompose R based on FD2:
 - R1(E,F) with FD2; candidate key is {E}
 - R2(A,B,C,D,E) with a new FD: $AB \rightarrow CDE$; candidate key is $\{A,B\}$ R1 and R2 are in BCNF



can be derived from FD1 using the decomposition rule

One More

- By using an FD $X \rightarrow Y$ that violates BCNF, decompose R into
 - R1 with all the attributes in X and in Y, and
 - R2 with all attributes from R except those that are in Y and not in X
- Consider the relation schema R(A,B,C,D,E,F) with the following FDs:

FD1: **AB** → **CDEF**

FD2: **E** → **F**

FD3: **A** → **D**

- Your turn:
 - Determine candidate key(s)
 - Is R in BCNF?
 - If not, normalize into a set of BCNF relation schemas (and don't forget to determine FDs and CKs along the way)



One More

- By using an FD $X \rightarrow Y$ that violates BCNF, decompose R into
 - R1 with all the attributes in X and in Y, and
 - R2 with all attributes from R except those that are in Y and not in X
- Consider the relation schema R(A,B,C,D,E,F) with the following FDs:

FD1: **AB** → **CDEF**

FD2: **E** → **F**

FD3: **A** → **D**

- Solution: CK is {A,B}; R is not in BCNF because of FD2 and FD3.
 We decompose R based on FD2:
 - R1(E,F) with FD2; candidate key is {E}
 - R2(A,B,C,D,E) with FD3 and a new FD: AB → CDE; candidate key is {A,B}

R1 is in BCNF, but R2 is not because of FD3. So, we have to decompose R2 using FD3:

- R3(A,D) with FD3; candidate key is {A}
- R4(A,B,C,E) with new FD: $AB \rightarrow CE$; candidate key is $\{AB\}$

R3 and R4 are in BCNF. Hence, the result of the decomposition consists of R1, R3, and R4.



Back to the Earlier Running Example

- By using an FD $X \rightarrow Y$ that violates BCNF, decompose R into
 - R1 with all the attributes in X and in Y, and
 - R2 with all attributes from R except those that are in Y and not in X
- Consider the relation schema R(A,B,C,D,E,F) with the following FDs:

FD1: $A \rightarrow BC$ FD2: $C \rightarrow AD$ FD3: $DE \rightarrow F$

- Recall: CKs are {A,E}, {C,E}; all three FDs violate the BCNF condition
- Your turn: Normalize into a set of BCNF relation schemas
 (and don't forget to determine FDs and CKs along the way)



Back to the Earlier Running Example

- By using an FD $X \rightarrow Y$ that violates BCNF, decompose R into
 - R1 with all the attributes in X and in Y, and
 - R2 with all attributes from R except those that are in Y and not in X
- Consider the relation schema R(A,B,C,D,E,F) with the following FDs:

FD1: $A \rightarrow BC$ FD2: $C \rightarrow AD$ FD3: $DE \rightarrow F$

- Recall: CKs are {A,E}, {C,E}; all three FDs violate the BCNF condition
- A possible solution: We decompose R based on FD1:
 - R1(A,B,C) with FD1 and a new FD: $C \rightarrow A$; candidate keys are $\{A\}$ and $\{C\}$
 - R2(A,D,E,F) with FD3 and a new FD: $A \rightarrow D$; candidate key is $\{A,E\}$

R1 is in BCNF, but R2 is not because of FD3 and $A \rightarrow D$. Let's decompose R2 based on FD3:

- R3(D,E,F) with FD3; candidate key is {D,E}
- R4(A,D,E) with $A \rightarrow D$; candidate key is $\{A,E\}$

R3 is in BCNF, but R4 is not because of $A \rightarrow D$. Let's decompose R4 based on $A \rightarrow D$

- R5(A,D) with $A \rightarrow D$; candidate key is $\{A\}$
- R6(A,E) with only trivial FDs; candidate key is {A,E}

R5 and R6 are in BCNF. Hence, the result of the decomposition consists of R1, R3, R4, R5.



www.liu.se

