DB Normalization

COMP3278B 2020

Overview

- Check if a decomposition is lossless-join
- Check if a decomposition is dependency preserving
- Check if a decomposition is in BCNF
- Perform lossless join decomposition into relations in BCNF

Testing lossless join

- Given relation R decomposed into $R1$ and $R2$.
- Let $Att(R)$ be the set of attributes in relation R. Decomposition is lossless if

$$Att(R1) \cap Att(R2) \rightarrow Att(R1)$$

Or

$$Att(R1) \cap Att(R2) \rightarrow Att(R2)$$

We test if the common attributes of R1 and R2 could be used to derive attributes in either R1 or R2.

Example

lossless join

Reflexivity – if $\beta \subseteq \alpha$, then $\alpha \rightarrow \beta$.
Transitivity – if $\alpha \rightarrow \beta$ and $\beta \rightarrow \gamma$, then $\alpha \rightarrow \gamma$
Augmentation – if $\alpha \rightarrow \beta$, then $\gamma\alpha \rightarrow \gamma\beta$
Union – if $\alpha \rightarrow \beta$ and $\alpha \rightarrow \gamma$, then $\alpha \rightarrow \beta\gamma$
Decomposition – if $\alpha \rightarrow \beta\gamma$, then $\alpha \rightarrow \beta$ and $\alpha \rightarrow \gamma$
Pseudo-transitivity – if $\alpha \rightarrow \beta$ and $\gamma\beta \rightarrow \delta$, then $\alpha\gamma \rightarrow \delta$

- Consider relation $R(A, B, C)$ with functional dependencies $\boldsymbol{F} = \{A \rightarrow B\}$.
  - Is $R1(A, B)$, $R2(A, C)$ a lossless join decomposition?
    - Common attribute is A
    - $A \rightarrow B \Rightarrow A \rightarrow AB$ (Augmentation)
    - It is a lossless join decomposition as common attribute A could derives attributes in $R1$.
  - Is $R3(A, B)$, $R4(B, C)$ a lossless join decomposition?
    - Common attribute is B
    - Attribute closure of B, $\{B\}^+ = \{B\}$.
    - It is not a lossless join decomposition as common attribute B could not derive $R1$ and $R2$.

We need to find attribute closure to show that a decomposition is not a lossless join.

Testing dependency preserving

- Given relation R with FD $\boldsymbol{F}$ decomposed into $R_1, R_2, \ldots$ let $\boldsymbol{F}_i$ be the set of FDs in $\boldsymbol{F}^+$ that include only attributes in R_i
- Decomposition is dependency preserving if

$$(\boldsymbol{F}_1 \cup \boldsymbol{F}_2 \cup \cdots)^+ = \boldsymbol{F}^+$$

We test if $\boldsymbol{F}^+$ could be constructed by combining the FDs in the decomposed relations.

Example

dependency preserving

Reflexivity – if $\beta \subseteq \alpha$, then $\alpha \rightarrow \beta$.
Transitivity – if $\alpha \rightarrow \beta$ and $\beta \rightarrow \gamma$, then $\alpha \rightarrow \gamma$
Augmentation – if $\alpha \rightarrow \beta$, then $\gamma\alpha \rightarrow \gamma\beta$
Union – if $\alpha \rightarrow \beta$ and $\alpha \rightarrow \gamma$, then $\alpha \rightarrow \beta\gamma$
Decomposition – if $\alpha \rightarrow \beta\gamma$, then $\alpha \rightarrow \beta$ and $\alpha \rightarrow \gamma$
Pseudo-transitivity – if $\alpha \rightarrow \beta$ and $\gamma\beta \rightarrow \delta$, then $\alpha\gamma \rightarrow \delta$

- Consider relation $R(A, B, C)$ with functional dependencies $\boldsymbol{F} = \{A \rightarrow B, B \rightarrow C\}$.
- $\boldsymbol{F}^+ = \{A \rightarrow B, B \rightarrow C, \text{and trivials/derived FDs}\}$
  - Is $R1(A, B)$, $R2(B, C)$ a dependency preserving decomposition?
    - $\boldsymbol{F}_1 = \{A \rightarrow B, \text{and trivials FD}s\}$
    - $\boldsymbol{F}_2 = \{B \rightarrow C, \text{and trivials FD}s\}$
    - $\boldsymbol{F}_1 \cup \boldsymbol{F}_2 = \boldsymbol{F} \Rightarrow \{\boldsymbol{F}_1 \cup \boldsymbol{F}_2\}^+ = \boldsymbol{F}^+$, it is a dependency preserving decomposition.
  - Is $R3(A, B)$, $R4(A, C)$ a dependency preserving decomposition?
    - $\boldsymbol{F}_3 = \{A \rightarrow B, \text{and trivials FD}s\}$
    - $\boldsymbol{F}_4 = \{A \rightarrow C, \text{and trivials FD}s\}$

      Note that $A \rightarrow C$ is in $\boldsymbol{F}^+$ due to transitivity.
    - $\boldsymbol{F}_3 \cup \boldsymbol{F}_4 = \{A \rightarrow B, A \rightarrow C, \text{and trivials FDs}\}$, we need to check if $B \rightarrow C$ is preserved.
    - In $\{\boldsymbol{F}_3 \cup \boldsymbol{F}_4\}^+$, $\{B\}^+ = \{B\}$ and so $B \rightarrow C$ is not in $\{\boldsymbol{F}_3 \cup \boldsymbol{F}_4\}^+$, it is not a dependency preserving decomposition.

Testing BCNF for relation

- Given relation R with FD $\boldsymbol{F}$
- R is in BCNF if for all non-trivial dependency $\alpha \rightarrow \beta$ in $\boldsymbol{F}$,

$$\{\alpha\}^+ = Att(R)$$

We test if LHS of each non-trivial dependencies is a superkey.

Similar to the case in dependency preserving, if R is decomposed, we need to construct the corresponding FDs in the decomposed relation by considering $\boldsymbol{F}^+$

Reflexivity – if $\beta \subseteq \alpha$, then $\alpha \to \beta$.
Transitivity – if $\alpha \to \beta$ and $\beta \to \gamma$, then $\alpha \to \gamma$
Augmentation – if $\alpha \to \beta$, then $\gamma\alpha \to \gamma\beta$
Union – if $\alpha \to \beta$ and $\alpha \to \gamma$, then $\alpha \to \beta\gamma$
Decomposition – if $\alpha \to \beta\gamma$, then $\alpha \to \beta$ and $\alpha \to \gamma$
Pseudo-transitivity – if $\alpha \to \beta$ and $\gamma\beta \to \delta$, then $\alpha\gamma \to \delta$

Example

BCNF (1)

- Consider relation $R(A, B, C)$ with functional dependencies $\boldsymbol{F} = \{A \to B, B \to C\}$.
  - Is R in BCNF?
    - $\{B\}^+ = \{B, C\} \neq \{A, B, C\}$ so it is not in BCNF.
- Consider relation $R(A, B, C)$ with functional dependencies $\boldsymbol{F} = \{A \to B\}$.
  - Is R in BCNF?
    - $\{A\}^+ = \{A, B\} \neq \{A, B, C\}$, so it is not in BCNF.
- Consider relation $R(A, B, C)$ with functional dependencies $\boldsymbol{F} = \{\}$ (trivial FDs only).
  - Is R in BCNF?
    - There is no non-trivial dependency. It is in BCNF.

No non-trivial dependency $\Rightarrow$ BCNF

Example

BCNF (2)

Reflexivity – if $\beta \subseteq \alpha$, then $\alpha \to \beta$.
Transitivity – if $\alpha \to \beta$ and $\beta \to \gamma$, then $\alpha \to \gamma$
Augmentation – if $\alpha \to \beta$, then $\gamma\alpha \to \gamma\beta$
Union – if $\alpha \to \beta$ and $\alpha \to \gamma$, then $\alpha \to \beta\gamma$
Decomposition – if $\alpha \to \beta\gamma$, then $\alpha \to \beta$ and $\alpha \to \gamma$
Pseudo-transitivity – if $\alpha \to \beta$ and $\gamma\beta \to \delta$, then $\alpha\gamma \to \delta$

- Consider relation $R(A, B, C, D)$ with functional dependencies $\boldsymbol{F} = \{A \to B, B \to C\}$.
  - Is R in BCNF?
    - $\{A\}^+ = \{A, B, C\} \neq \{A, B, C, D\}$, so it is not in BCNF.
  - Suppose R is decomposed into $R1(A, C, D)$, $R2(B, D)$, is $R1$ in BCNF?
    - $\boldsymbol{F}_1 = \{A \to C, \text{and trivials FD}s\}$
    - $\{A\}^+ = \{A, C\} \neq \{A, C, D\}$, it is not in BCNF

Again, $A \to C$ is in $\boldsymbol{F}^+$ due to transitivity.

  - Is $R2$ in BCNF?
    - $R2$ has no non-trivial FD, it is in BCNF

Decomposing relations

- Given relation R with FD $\boldsymbol{F}$. (where R is not in BCNF)
  - Pick a dependency $\alpha \rightarrow \beta$ in $\boldsymbol{F}$.
  - Split R into relation $R1 = \alpha \cup \beta$ and $R2 = \alpha \cup (\alpha \cup \beta)^c$.
  - Check that $R1$ is in BCNF, decompose $R1$ if not.
  - Check that $R2$ is in BCNF, decompose $R2$ if not.

We pick one rule for decomposition, repeat until all relations are in BCNF.

Example
BCNF (2)

Reflexivity – if $\beta \subseteq \alpha$, then $\alpha \rightarrow \beta$.
Transitivity – if $\alpha \rightarrow \beta$ and $\beta \rightarrow \gamma$, then $\alpha \rightarrow \gamma$
Augmentation – if $\alpha \rightarrow \beta$, then $\gamma\alpha \rightarrow \gamma\beta$
Union – if $\alpha \rightarrow \beta$ and $\alpha \rightarrow \gamma$, then $\alpha \rightarrow \beta\gamma$
Decomposition – if $\alpha \rightarrow \beta\gamma$, then $\alpha \rightarrow \beta$ and $\alpha \rightarrow \gamma$
Pseudo-transitivity – if $\alpha \rightarrow \beta$ and $\gamma\beta \rightarrow \delta$, then $\alpha\gamma \rightarrow \delta$

- Consider relation $R(A, B, C, D)$ with functional dependencies $\boldsymbol{F} = \{A \rightarrow B, B \rightarrow C\}$. Decompose R into relations in BCNF.
  - Pick dependency $A \rightarrow B$.
  - Decompose R into $R1 = \{A, B\}$ and $R2 = \{A, C, D\}$.
    - $R1$ is in BCNF as $\{A\}^+ = \{A, B\}$.
  - $\boldsymbol{F}_2 = \{A \rightarrow C, \text{ and trivials FDs}\}$, $\{A\}^+ = \{A, C\} \neq \{A, C, D\}$, it is not in BCNF.
  - Pick rule $A \rightarrow C$.
  - Decompose $R2$ into $R3 = \{A, C\}$ and $R4 = \{A, D\}$.
    - $R3$ is in BCNF as $\{A\}^+ = \{A, C\}$.
  - $R4$ has no non-trivial FD, it is in BCNF.

- R is decomposed into $R1 = \{A, B\}$, $R3 = \{A, C\}$, and $R4 = \{A, D\}$

Exercise

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

1. If R is decomposed into $R1(A, C, D)$ and $R2(B, C)$, is it a lossless join decomposition? If so, is it dependency preserving?
2. If R is decomposed into $R3(A, B, C)$ and $R4(B, C, D)$, is it a lossless join decomposition? If so, is it dependency preserving?
3. Show that R in not in BCNF.
4. Give a lossless join decomposition of R into relations in BCNF. Check if the decomposition is dependency preserving.
   - Try to repeat this question by picking different dependency to split.

An extra question if you have time

- Given $R(A, B, C, D, E), \boldsymbol{F} = \{A \rightarrow B, AD \rightarrow E, C \rightarrow ADE\}$.
  - Give a lossless join decomposition of R into relations in BCNF.
  - Check if the decomposition is dependency preserving.

Answer – 1

Given $R(A, B, C, D), \boldsymbol{F} = \{A \rightarrow B, BC \rightarrow D, D \rightarrow A\}$
If R is decomposed into $R1(A, C, D)$ and $R2(B, C)$, is it a lossless join decomposition? If so, is it dependency preserving?

- Common attributes: $\{C\}$
- $\{C\}^+ = \{C\}$, which doesn't cover $R1$ nor $R2$.
- It is not a lossless join decomposition.

Answer – 2

Given $R(A, B, C, D), \boldsymbol{F} = \{A \rightarrow B, BC \rightarrow D, D \rightarrow A\}$
If R is decomposed into $R3(A, B, C)$ and $R4(B, C, D)$, is it a lossless join decomposition? If so, is it dependency preserving?

- Common attributes: $\{B, C\}$
- $\{B, C\}^+ = \{B, C, D\}$, which covers $R4$.
- It is a lossless join decomposition.
- Consider FDs in $R3$ ($\boldsymbol{F}_3$) $= \{A \rightarrow B, BC \rightarrow A, \text{and trivials FDs}\}$
- Consider FDs in $R4$ ($\boldsymbol{F}_4$) $= \{BC \rightarrow D, \text{and trivials FDs}\}$
- Now consider $\{D\}^+$ for $\{\boldsymbol{F}_3 \cup \boldsymbol{F}_4\}^+$, $\{D\}^+ = \{D\}$. Dependency $D \rightarrow A$ is gone. Therefore, it is not dependency preserving.

Answer – 3

Given $R(A, B, C, D), \boldsymbol{F} = \{A \rightarrow B, BC \rightarrow D, D \rightarrow A\}$
Show that R in not in BCNF.

- $\{A\}$ is not a superkey as $\{A\}^+ = \{A, B\}$, so R is not in BCNF.
  - You can also check $\{D\}^+$.
  - Note that $\{B, C\}^+ = \{A, B, C, D\}$, so this could not be used here.

Answer – 4

Given $R(A, B, C, D), \boldsymbol{F} = \{A \to B, BC \to D, D \to A\}$
Give a lossless join decomposition of R into relations in BCNF.
Check if the decomposition is dependency preserving.

Pick $A \to B$

- Pick $A \to B$, R is decomposed into $R1(A, B)$ and $R2(A, C, D)$.
  - FDs in $R1$ (F_1) $= \{A \to B, \dots\}$, $\{A\}^+ = \{A, B\}$, $R1$ is in BCNF.
  - FDs in $R2$ (F_2) $= \{D \to A, \dots\}$, $\{D\}^+ = \{A, D\} \neq \{A, C, D\}$, $R2$ is not in BCNF.
- Pick $D \to A$ for $R2$, $R2$ is decomposed into $R3(A, D)$ and $R4(C, D)$.
  - FDs in $R3$ (F_3) $= \{D \to A, \dots\}$, $\{D\}^+ = \{A, D\}$, $R3$ is in BCNF.
  - There is no non-trivial FDs in $R4$, $R4$ is in BCNF.
- R is decomposed into $R1(A, B)$, $R3(A, D)$, and $R4(C, D)$.
- We can see that dependency $BC \to D$ is not preserved in the decomposition.

What if we pick $BC \to D$ first?

Answer – 4

Given $R(A, B, C, D), \boldsymbol{F} = \{A \rightarrow B, BC \rightarrow D, D \rightarrow A\}$
Give a lossless join decomposition of R into relations in BCNF.
Check if the decomposition is dependency preserving.

Pick $BC \rightarrow D$

- Pick $BC \rightarrow D$, R is decomposed into $R1(B, C, D)$ and $R2(A, B, C)$.
  - FDs in $R1$ (F_1) $= \{BC \rightarrow D, D \rightarrow B, \dots\}$, $\{D\}^+ = \{B, D\} \neq \{B, C, D\}$, $R1$ is not in BCNF.
  - FDs in $R2$ (F_2) $= \{A \rightarrow B, BC \rightarrow A, \dots\}$, $\{A\}^+ = \{A, B\} \neq \{A, B, C\}$, $R2$ is not in BCNF.
- Pick $A \rightarrow B$ for $R2$, $R2$ is decomposed into $R3(A, B)$ and $R4(A, C)$.
  - FDs in $R3$ (F_3) $= \{A \rightarrow B, \dots\}$, $\{A\}^+ = \{A, B\}$, $R3$ is in BCNF.
  - There is no non-trivial FDs in $R4$, $R4$ is in BCNF.
- Pick $D \rightarrow B$ for $R1$, $R1$ is decomposed into $R5(B, D)$ and $R6(C, D)$
  - Both are in BCNF.
- R is decomposed into, $R3(A, B), R4(A, C)$, $R5(B, D)$ and $R6(C, D)$.
- We can see that dependency $D \rightarrow A$ is not preserved in the decomposition. (Shown in Q2)

What if we pick $D \rightarrow A$ first?

Answer – 4

Given $R(A, B, C, D), \boldsymbol{F} = \{A \rightarrow B, BC \rightarrow D, D \rightarrow A\}$
Give a lossless join decomposition of R into relations in BCNF.
Check if the decomposition is dependency preserving.

Pick $D \rightarrow A$

- Pick $D \rightarrow A$, R is decomposed into $R1(A, D)$ and $R2(B, C, D)$.
  - FDs in $R1$ (F_1) $= \{D \rightarrow A, \dots\}$, $\{D\}^+ = \{A, D\}$, $R1$ is in BCNF.
  - FDs in $R2$ (F_2) $= \{BC \rightarrow D, D \rightarrow B, \dots\}$, $\{D\}^+ = \{B, D\}$, $R2$ is not in BCNF.
- Pick $D \rightarrow B$ for $R2$, $R2$ is decomposed into $R3(B, D)$ and $R4(C, D)$
  - Both are in BCNF.
- R is decomposed into $R1(A, D)$, $R3(B, D)$ and $R4(C, D)$.
- We can see that dependency $A \rightarrow B$ is not preserved in the decomposition.

Answer – 5

Pick $A \rightarrow B$

Given $R(A, B, C, D, E), \boldsymbol{F} = \{A \rightarrow B, AD \rightarrow E, C \rightarrow ADE\}$.
Give a lossless join decomposition of R into relations in BCNF.
Check if the decomposition is dependency preserving.

- Pick $A \rightarrow B$, decompose R into $R1(A, B)$ and $R2(A, C, D, E)$.
  - $R1$ is in BCNF. (proof skipped)
  - FDs in $R2$ (F_2) $= \{AD \rightarrow E, C \rightarrow ADE, \dots\}$, $\{A, D\}^+ = \{A, D, E\} \neq \{A, C, D, E\}$, $R2$ is not in BCNF.
- Pick $AD \rightarrow E$ for $R2$, $R2$ is decomposed into $R3(A, D, E)$ and $R4(A, C, D)$.
  - $R3$ is in BCNF. (proof skipped)
  - FDs in $R4$ (F_4) $= \{C \rightarrow AD, \dots\}$, $\{C\}^+ = \{A, C, D\}$, $R4$ is in BCNF.

Found using Armstrong's Axioms

- R is decomposed to $R1(A, B)$, $R3(A, D, E)$ and $R4(A, C, D)$.
- Is $C \rightarrow ADE$ preserved in the decomposition?

Answer – 5

Analysis

Given $R(A, B, C, D, E), \boldsymbol{F} = \{A \to B, AD \to E, C \to ADE\}$.
Give a lossless join decomposition of R into relations in BCNF.
Check if the decomposition is dependency preserving.

- Note that some of the dependencies could be reduced or transformed.
  - $C \to ADE \Leftrightarrow C \to A, C \to D, C \to E$

Must be two-way implications

- And some is redundant, e.g.,
  - $C \to A, C \to D, AD \to E \Rightarrow C \to E$
- Consider $\boldsymbol{F}' = \{A \to B, AD \to E, C \to A, C \to D\}$, then $\boldsymbol{F}^+ = \boldsymbol{F}'^+$.
  - Now if R is decomposed into $R1(A, B)$, $R3(A, D, E)$ and $R4(A, C, D)$, you can see that all dependencies are preserved.

In order to check if the decomposition is dependency preserving, you must consider $(\boldsymbol{F}_1 \cup \boldsymbol{F}_2 \cup \cdots)^+ = \boldsymbol{F}^+$ instead of just checking the given dependencies!

You can also find an equivalent (but simpler) set of FDs first before normalizing.