CS 315 – Introduction To Databases

Assignment 3: Normalization

Mridul Verma
Roll no. - 10415

1. Given the schema R= (A_1; A_2; A_3; A_4) and FDs A_1A_2->A_3, A_3->A_4 and A_4->A_1 answer the following:

(a) What non-trivial FDs follow from the given FDs?

<u>Sol:</u> Some of the non trivial functional dependencies which follow from the given FDs are:

A_2 A_3	**->**	A_1 A_2 A_3 A_4
A_2 A_4	**->**	A_1 A_2 A_3 A_4
A_3 A_4	**->**	A_1 A_3 A_4
A_1 A_2 A_3	**->**	A_1 A_2 A_3 A_4
A_1 A_2 A_4	**->**	A_1 A_2 A_3 A_4
A_2 A_3 A_4	**->**	A_1 A_2 A_3 A_4
A_3	**->**	A_1 A_3 A_4
A_4	**->**	A_1 A_4
A_1 A_2	**->**	A_1 A_2 A_3 A_4
A_1 A_3	**->**	A_1 A_3 A_4

(b) Find all candidate keys for R?

By calculating an attribute closure we can see the candidate keys are:
{A_1,A_2},{A_2,A_3}, and {A_2,A_4}

<u>As the Attribute closure of the written above attribute are:</u>

A_1 A_2	->	A_1 A_2 A_3 A_4
A_2 A_3	->	A_1 A_2 A_3 A_4
A_2 A_4	->	A_1 A_2 A_3 A_4

(c) Find all super keys for R that are not candidate keys.

Sol: The Super keys for R would be the superset of the candidate keys that are

$\{A_1, A_2\}, \{A_2, A_3\}$, and $\{A_2, A_4\}$

So the Super keys would be

Super keys: $\{A_1, A_2, A_3\}, \{A_2, A_3, A_4\}$, $\{A_1, A_2, A_4\}$ and $\{A_1, A_3, A_2, A_4\}$

(d) Find all minimal bases for R .

Sol: The minimal bases for R are:

$$\{ A_1A_2 \rightarrow A_3,\ A_3 \rightarrow A_4 \text{ and } A_4 \rightarrow A_1 \}$$

The method to find out the minimal bases is to basically:

Iterate over the given two rules:

a) First remove the dependencies of the form ($a \rightarrow b$ and $a \rightarrow c$) by replacing them from ($a \rightarrow bc$)

b) By removing the extraneous attributes from the dependencies.

Now in the given dependencies we cannot apply the above two rules as the given dependencies do not have the form ($a \rightarrow b$ and $a \rightarrow c$) and also these dependencies do not have the extraneous attributes in them.

2. Let X be a set of attributes. We say X is closed with respect to a set of FDs F if $X^+ = X$. Given R= (A1; A2; A3; A4) and information about which subsets of R are closed we can find F. Find F in the following cases:

(a) All subsets of R including R are closed.

Sol: F = {**Y -> Z| Y ⊆ R, Z ⊆ Y**} i.e. all the trivial FDs.

(b) Only R and ϕ are closed.

Sol: For instance F = {A_1->A_2, A_2->A_3, A_3->A_4, A_4->A_1} or

F = {A_1->A_3, A_3->A_2, A_2->A_4, A_4->A_1}

Or any **cyclic Dependencies** can lead to the solution

(c) R, ϕ and {A1; A2} are closed.

Sol: For instance F = {A_1->A_2, A_2->A_1, A_3->A_1A_4, A_4->A_2A_3}

---

3. In the past we have shown that certain rules about FDs are sound using Armstrong's axioms. For the following rules show that the rules do not hold by giving counter examples:

(a) If $A_1 \rightarrow A_2$ then $A_2 \rightarrow A_1$

A_1	A_2
1	4
2	4
1	4
4	10

As you can see from the above example that the $A_1 \rightarrow A_2$ is satisfied but not $A_2 \rightarrow A_1$.

(b) If $A_1A_2 \rightarrow A_3$ and $A_1 \rightarrow A_3$ then $A_2 \rightarrow A_3$

A_1	A_2	A_3
1	2	4
1	2	4
1	3	4
3	2	5

As you can see from the above example that the given table satisfies $A_1A_2 \rightarrow A_3$ and $A_1 \rightarrow A_3$ but not $A_2 \rightarrow A_3$.

(c) If $A_1A_2 \rightarrow A_3$ then $A_1 \rightarrow A_3$ or $A_2 \rightarrow A_3$

A_1	A_2	A_3
1	3	4
1	3	4
1	5	6
2	3	8

As you can see from the above table that the table satisfies $A_1A_2 \rightarrow A_3$ but not $A_1 \rightarrow A_3$ or $A_2 \rightarrow A_3$.

---

4. A functional dependency $\boldsymbol{\alpha} \rightarrow \boldsymbol{\beta}$ is called a partial dependency if there is a proper subset $\boldsymbol{\gamma}$ of $\boldsymbol{\alpha}$ such that $\boldsymbol{\gamma} \rightarrow \boldsymbol{\beta}$. We say that $\boldsymbol{\beta}$ is partially dependent on $\boldsymbol{\alpha}$. A relation schema R is in second normal form if each attribute A in R meets one of the following criteria:

- it appears in a candidate key;

- it is not partially dependent on a candidate key.

Show that every 3NF schema is in 2NF.

Sol:

An Attribute that does not occur in any of the candidate keys in R is called a Non Prime Attribute.

A relation schema R is said to be in 3NF if there is no attribute A(that does not occur in any candidate key in R) for which A is transitively dependent on a key for R.

Now we can also rewrite the definition of 2NF given here as:

"A relation schema R is in 2NF if no attribute A(that does not occur in any candidate key) is partially dependent on any candidate key for R."

To prove that every 3NF schema is in 2NF, it suffices to show that a attribute A(that does not occur in any candidate key) is partially dependent on a candidate key **α**, then A is also transitively dependent on the key **α.**

Let A be attribute that does not occur in any of the candidate keys in R. Let **α** be a candidate key for R. Suppose A is partially dependent on **α**.

- from the definition of a partial dependency, we know that for some proper subset beta of **α**, **β** → **α** .
- since beta in **α** , **α** → **β** . Also **β** → **α** does not hold, since **α** is a candidate key.
- finally, since A does not occur in any candidate key, it cannot be in either **β** or **α** .

Thus we conclude that **α** → **A** is transitive dependency. Hence we have proved that every 3NF schema is also in 2NF.

---

5. Define a prime attribute as one that appears in at least one candidate key. Let **α,β** be attribute sets such that the FD

$\alpha \rightarrow \beta$ holds but **$\beta \rightarrow \alpha$** does not hold. Let attribute A be such that A $\notin$ **α** ,A $\notin$ **β** and **$\beta \rightarrow$ A.** Then A is said to be transitively dependent on **α** **.**A relation schema R is in 3NF with respect to a set of FDs F if there is no non-prime attribute A $\in$ R that is transitively dependent on a key of R.

Argue that the above definition of 3NF is equivalent to the definition discussed in class.

<u>Sol:</u> Again describing non prime attribute, non prime attribute are those attributes that does not occur in any of the candidate keys in R.

Suppose R is in 3NF according to the definition discussed in class. We have to show that it is in 3NF according to the definition in the question given above. Let A be a non-prime attribute in R that is transitively dependent on a key α for R. Then there exists $\beta \subseteq R$ such that $\beta \rightarrow A$, $\alpha \rightarrow \beta$, A $\notin \alpha$, A $\notin \beta$, and $\beta \rightarrow \alpha$ does not hold. But then $\beta \rightarrow A$ violates the definition told in the class of 3NF since

• A $\notin \beta$ implies $\beta \rightarrow A$ is nontrivial.

• Since $\beta \rightarrow \alpha$ does not hold, β is not a super-key

• A is not any candidate key, since A is non-prime.

Now, we show that if R is in 3NF according to this question definition , it is in 3NF according to the definition told in class . Suppose R is not in 3NF according to the definition told in class. Then there is an FD $\alpha \rightarrow \beta$ that fails all three conditions. Thus

• $\alpha \rightarrow \beta$ is nontrivial.

• α is not a super key for R.

• Some A in $\beta - \alpha$ is not in any candidate key.

This implies that A is non-prime and $\alpha \rightarrow A$. Let γ be a candidate key for R. Then $\gamma \rightarrow \alpha$, $\alpha \rightarrow \gamma$ does not hold (since α is not a super-key), A $\notin \alpha$, and A $\notin \gamma$ (since A is non-prime).

Thus A is transitively dependent on γ, violating the exercise definition.

6. Let schema R be decomposed into $R_1, \ldots, R_n$, let r(R) be a relation and $r_i = \pi_{R_i}(r)$ where $i \in 1..n$.
Show:

$$\mathbf{r \subseteq r_1 \bowtie r_2 \ldots \bowtie r_n}$$

<u>Sol:</u> Consider some tuple **t** in **r**

Note that $r_i = \pi_{R_i}(r)$ implies that $t[R_i] \in r_i$, $1 \le i \le n$.
Thus

$$\mathbf{t[R_1] \bowtie t[R_2] \bowtie \ldots \bowtie t[R_n] \in r_1 \bowtie r_2 \bowtie \ldots \bowtie r_n}$$

By the definition of natural join,

$$\mathbf{t[R_1] \bowtie t[R_2] \bowtie \ldots \bowtie t[R_n] = \pi_\alpha(\sigma_\beta(t[R_1] \times t[R_2] \times \ldots \times t[R_n]))}$$

where the condition β is satisfied if values of attributes with the same name in a tuple are equal and where $\alpha = U$. The Cartesian product of single tuples generates one tuple. The selection process is satisfied because all attributes with the same name must have the same value since they are projections from the same tuple.

Finally, the projection clause removes duplicate attribute names.

By the definition of decomposition,

$$\mathbf{R = R_1 \cup R_2 \cup \ldots \cup R_n} ,$$

which means that all attributes of t are in

$$\mathbf{t[R_1] \bowtie t[R_2] \bowtie \ldots \bowtie t[R_n]}$$

That is, **t** is equal to the result of this join. Since **t** is any arbitrary tuple in **r**,

$$\mathbf{r \subseteq r_1 \bowtie r_2 \bowtie \ldots \bowtie r_n}$$

---

7. Give an example of a schema R and set of FDs F such that there are at least three distinct lossless join decompositions of R into BCNF.

Sol: Consider the schema $R = (A_1, A_2, A_3, A_4)$, the set of FDs

$F = \{ A_1 \rightarrow A_4 , A_2 \rightarrow A_4 , A_3 \rightarrow A_4 \}$

Candidate key for R is for given FD F is $\{A_1, A_2, A_3\}$

a) Decomposition using $A_1 \rightarrow A_4$

$R_1 = \{A_1, A_4\}$
$R_2 = \{A_1, A_2, A_3\}$

FD's

$A_1 \rightarrow A_4$

and $A_1A_2A_3 \rightarrow A_1A_2A_3$

Both FD's are in BCNF

b) Decomposition using $A_2 \rightarrow A_4$

$R_1 = \{A_2, A_4\}$
$R_2 = \{A_1, A_2, A_3\}$

FD's

$A_2 \rightarrow A_4$

and $A_1A_2A_3 \rightarrow A_1A_2A_3$

Both FD's are in BCNF

c) Decomposition using $A_3 \rightarrow A_4$

$R_1 = \{A_3, A_4\}$
$R_2 = \{A_1, A_2, A_3\}$

FD's

$A_3 \rightarrow A_4$

and $A_1A_2A_3 \rightarrow A_1A_2A_3$

Both FD's are in BCNF