

Normalization

Université Grenoble Alpes

04/04/2023

Bahareh Afshinpour

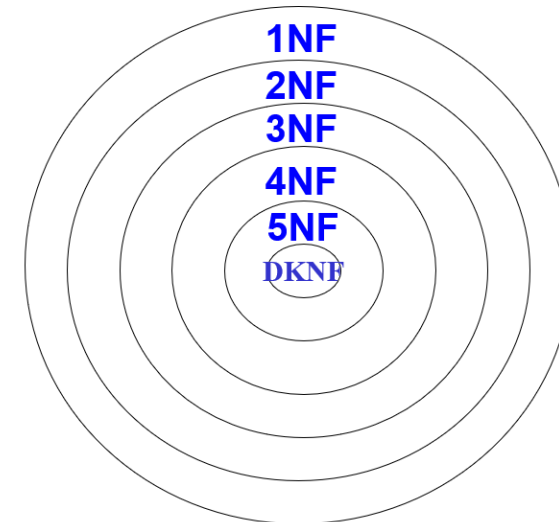
bahareh.afshinpour@univ-grenoble-alpes.fr

Main reference:

A First Course in Database Systems (and associated material) by
J. Ullman and J. Widom, Prentice-Hall

Levels of Normalization

- Levels of normalization based on the amount of redundancy in the database.
 - First Normal Form (1NF)
 - Second Normal Form (2NF)
 - Third Normal Form (3NF)
 - Boyce-Codd Normal Form (BCNF)
 - Fourth Normal Form (4NF)
 - Fifth Normal Form (5NF)
 - Domain Key Normal Form (DKNF)



Each higher level is a subset of the lower level

Most databases should be 3NF or BCNF in order to avoid the database anomalies.

First normal form

- A relation schema is in first normal form (1 NF) if any attribute has an atomic value.
- **1:Atomic value** : It cannot be decomposed into two or more component

- Create separate column for each member of composite attribute
- Make two or more different tuples for each multi-value attributes
- Define Fk

SID	Name	Add	Tel
1	x1	F1,a1,b1	T1,T2
2	x2	F1,a2,b5	T5
3	x3	F2,a3,b4	T6,T3

It is not atomic

- **2:In first normal form** : A column should contain values for the same domain.
- **3:Each column should have unique name**
- **4: NO ordering, No duplicate rows**

Closure of a set of FDs (F^+)

- The closure of F , said F^+ , is the set of all FD that can be derived from F
- Using **attributes closure** can help to answer , it is a **candidate** key or not
- Then by finding a candidate key we can solve the 2NF , 3NF ,

$R(A,B,C,D,E)$ $FD=\{A \rightarrow B, B \rightarrow C, C \rightarrow D, D \rightarrow E\}$

We can use rules and find more FD

$A \rightarrow B, B \rightarrow C \rightarrow A \rightarrow C$

$A \rightarrow A$ (Reflexivity)

$A \rightarrow C, C \rightarrow D \rightarrow A \rightarrow D$

$A \rightarrow D, D \rightarrow E \rightarrow A \rightarrow E$

$A \rightarrow ABCDE$ (splitting/merge)

x is a set of attribute

X^+ contains set of attributes determined by X

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





$R(A,B,C,D,E)$ $FD=\{A \rightarrow B, B \rightarrow C, C \rightarrow D, D \rightarrow E\}$

$AD^+=?$

$AD \rightarrow A$

$AD \rightarrow D$

$A \rightarrow B \rightarrow AD \rightarrow BD$

$AD \rightarrow BD \rightarrow AD \rightarrow B, AD \rightarrow D$

$AD \rightarrow B, B \rightarrow C \rightarrow AD \rightarrow C$

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

CD^+

$\{C,D, \dots\}$

$D \rightarrow E$

$CD^+=\{C,D, E\}$

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

Superkey

Set of attributes whose closure contains all attributes of a given relation

A^+ and AD^+ are SK

Candidate Key

- Is a superkey whose **proper subset** is **not** a superkey. (minimal super key)

SK= {A},{A,B},{A,C},{A,B,C}

{B} OR {C} NO

SK={B,C}

SK= {A},{A,B},{A,C},{A,B,C}, {B,C}

proper subset :

Suppose $X1=\{1,2,3\}$ and $X2=\{1,2\}$

$X2$ is **subset** of $x1$ if every member of $X2$ must be member of $X1$

$X2$ is **proper subset** of $x1$

First $x2$ is subset of $x1$

But $x1$ is not subset of $x2$

{A,B,C} : WHOSE proper subset are {A,B}, {B,C},{A,C},{A}, {B}, {C} **CK=NO** SOME ARE SUPERKEYS

{A,C}: WHOSE proper subset are {A},{C} **CK=NO** SOME ARE SUPERKEYS

{A}: **CK=YES** {B,C}: WHOSE proper subset are {B} ,{C} none of its proper subset is sk **CK=YES**

A	B	C
1	6	3
2	6	5
3	1	3
4	1	5

So every CK is a SK
But every SK is not a CK

Second Normal Form - 2NF

- First: It is in 1NF.
- Second: There would be **no partial dependency** present in the relation

 **What is partial functional dependency (PD) :**
- A proper subset of any candidate key → non-prim attributes

Well determine

← That should not be present in the relation

What is a prime attribute:

Prime attributes are those which are part of candidate keys.

For example, the candidate keys in R(A,B,C, D,E,F) are ADE,BC.

So the prime attributes are: A,B, C, D, E. and F is non-prime attributes

Example 1

R(A,B,C,D,E,F)

- A proper subset of any candidate key \rightarrow non-prim attributes

FD={ A \rightarrow B, B \rightarrow C,C \rightarrow D,D \rightarrow E}

- We need to find Candidate Key.  By taking the attribute closure

ABCDEF⁺ = {ABCDEF} Reflexivity properties

SK=YES

Superkey
Set of attributes whose closure contains all attributes of a given relation

Try to discard attributes by using the properties

ABCDEF⁺

A \rightarrow B ~~A~~BCDEF⁺

A \rightarrow C ~~A~~~~B~~CDEF⁺

A \rightarrow D ~~A~~~~B~~~~C~~DEF⁺

A \rightarrow E ~~A~~~~B~~~~C~~~~D~~E⁺  AF⁺

CK= Is a superkey whose proper subset is **not** a superkey. (minimal super key)

AF is candidate key?

So **AF** is a candidate key. **A** and **F** are **prime attributes** since they are part of CK.

A⁺ = {ABCDE} SK=NO

F⁺ = {} SK=NO

Short trick: If prime attributes are present on the **right-hand** side of any functional dependency, then there would be more candidate keys. If not, there would be no more candidate keys. ... so, in this relation, we have only one CK. (AF)

$R(A,B,C,D,E,F)$

$FD = \{ A \rightarrow B, B \rightarrow C, C \rightarrow D, D \rightarrow E \}$

- Non-prime = B,C,D,E
- Now you can check out this type of dependency:



- A proper subset of any candidate key \rightarrow non-prim attributes



That should not be present in the relation

$A \rightarrow B$

$F \rightarrow \dots$

B is in the non-prim attributes.

So, partial dependency exists.

R is not in the 2NF

Example2

$R(A,B,C,D)$

$FD=\{ AB \rightarrow CD, C \rightarrow A, D \rightarrow B \}$

- First: find out the candidate key

$ABCD^+ = \{A,B,C,D\}$ We have all the attributes and it is a super key.

We start to discard since the candidate key is minimal super key.

~~$ABCD^+$~~

$AB^+ = \{CDAB\} = \{ABCD\}$ Contains all the attributes \rightarrow **SK=YES**

To check if AB is the candidate key or not $\rightarrow A^+ = \{A\}$ $B^+ = \{B\}$ we do not have all the attributes of the relation so **CK=YES**

Prime attributes = $\{A,B\}$

To find if more candidate keys are present or not \rightarrow check the right side of the FD

$CK=AB$ $C \rightarrow A$, $CK=CB$, $C^+ = \{A,C\}$ **SK=NO** $B^+ = \{B\}$ **SK=NO** \rightarrow **CK=CB** AND **CK=AD**
 $D \rightarrow B$, $CK=AD$, $A^+ = \{A\}$ **SK=NO** $D^+ = \{B,D\}$ **SK=NO**

Prime attributes = $\{A,B,C,D\}$ all the attributes in R are prime- se we have no non-prime \rightarrow R would be second normal form