2NF

Université Grenoble Alpes

23/03/2023

Bahareh Afshinpour

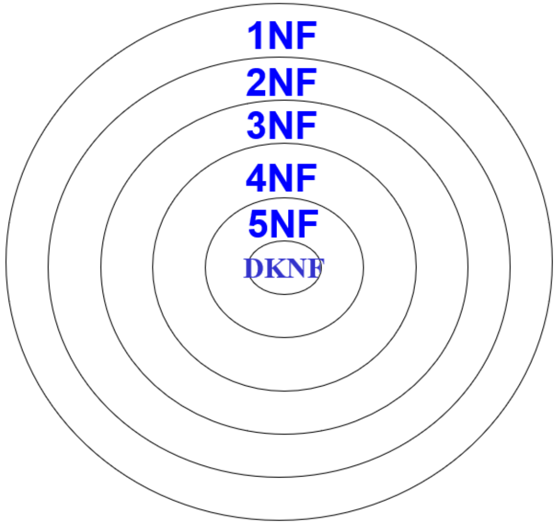
bahareh.afshinpour@univ-grenoble-alpes.fr

Université Grenoble Alpes

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.

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}

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

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

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

{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

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,F)   FD={A->B , B->C, C->D, D->E}

We can use rules and find more FD

A->B, B->C ➔ A->C

A->A (Reflexivity)

A->C, C->D ➔ A->D

A->D, D->E ➔ A->E

A->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->B , B->C, C->D, D->E}

AD^+=?

AD->A

AD->D

A->B ➔ AD->BD

AD->BD ➔ AD->B , AD->D

AD->B, B->C ➔ AD->C

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

CD^+

{C,D, ….}

D->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

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 <u>any</u> candidate key → non-prim attributes

That should not be present in the relation

Well determine

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

Example1

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

FD={ A->B, B->C,C->D,D->E}

- A proper subset of <u>any</u> candidate key → non-prim attributes

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

ABCDEF$^+$ ={ABCDEF} Reflexivity properties SK=YES

<u>Superkey</u>
Set of attributes whose closure contains all attributes of a given relation

Try to discard attributes by using the properties

	ABCDEF$^+$
A->B	A~~B~~CDEF$^+$
A->C	A~~BC~~DEF$^+$
A->D	A~~BCD~~EF$^+$
A->E	A~~BCDE~~F$^+$ → AF$^+$

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

AF is candidate key?

A$^+$ ={ABCDE} SK=NO

F$^+$ ={} SK=NO

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

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->B, B->C,C->D,D->E}

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

- A proper subset of <u>any</u> candidate key → non-prim attributes ← That should not be present in the relation

A-> B

F-> ---

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->CD, C->A,D->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.

AB~~CD~~$^+$

AB^+ ={CDAB}={ABCD} Contains all the attributes → SK=YES

To check if AB is the candidate key or not → 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 → check the right side of the FD

CK=AB C→A , CK=CB , C^+={A,C} SK=NO B^+={B} SK=NO

D→B , CK=AD , A^+={A} SK=NO D^+={B,D} SK=NO

→ CK=CB AND CK=AD

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