

NORMAL FORMS

1) First Normal Form (1NF)

- * It is the first stage of database normalization. The domains of
- * A relation schema is in 1NF if all the attributes are atomic in nature (indivisible values)
- * It disallows multivalued attributes, ~~and~~ composite attributes and their combinations
- * There are only single valued attributes.

Eg:

Stud-roll	Sub-no	Sem	mark

ID	Name	Courses
1	A	C1, C2
2	E	C3
3	M	C2, C3

conversion
to 1NF →

ID	Name	Course
1	A	C1
1	A	C2
2	E	C3
3	M	C2
3	M	C3

2) Second Normal Form (2NF)

- * The second normal form is based on the concept of full functional dependency.
- * A functional dependency $X \rightarrow Y$ is a full functional dependency if removal of any attribute A from X means that the dependency does not hold any more. (A is subset of X)
i.e., for any attribute $A \in X$, $(X - \{A\})$ does not functionally determine Y.

$AB \rightarrow C$

* A functional dependency $X \rightarrow Y$ is a partial FD if $A \in X$ and ~~$A \rightarrow Y$ holds~~. $(X - \{A\}) \rightarrow Y$ holds

Eg: $R(A, B, C, D, E, F)$
 $F = \{ AB \rightarrow C, B \rightarrow D, D \rightarrow E, A \rightarrow C, AD \rightarrow F \}$
 \downarrow Partial FD as $A \rightarrow C$ holds
 \downarrow Full FD

Determination of Primary key

Let X be a primary key. Suppose X functionally determines attributes A_1, A_2, \dots, A_n .

$$X \rightarrow A_1 A_2 \dots A_n$$

Eg:

Example

Suppose we have the following set of FDs F :

$$F = \{ \text{roll} \rightarrow \text{name}, \text{roll subno} \rightarrow \text{mark}, \text{mark} \rightarrow \text{grade} \}$$

$$\text{roll}^+ = \text{roll, name}$$

$$\text{roll subno}^+ = \text{roll subno, mark, grade}$$

$$\text{mark}^+ = \text{mark, grade}$$

Prime Attributes = roll, subno

Non-Prime Attributes = name, mark, grade

So, roll subno^+ is the primary key.

NOTE: Let $R(A_1, A_2, \dots, A_n)$ be the relation schema. Then X is a primary key if

$$X \rightarrow A_1 A_2 \dots A_n \text{ holds on } R.$$

(all the attributes can be inferred or functionally determined from the primary key)

Definition of 2NF

A relation schema R is in 2NF if it is in 1NF and every non-prime attribute A in R is fully functionally dependent on the primary key of R .

or

There, does not exist any partial FD.

Eg: $R(A, B, C, D, E)$ be the relation schema.

Let the set of FDs be:

$$F = \{BC \rightarrow A, C \rightarrow DE, D \rightarrow E\}$$

Test whether R is in 2NF. If not make it 2NF.

Solⁿ:

First we determine the primary key of R .

$$BC^+ = ABCDE, C^+ = CDE, D^+ = DE$$

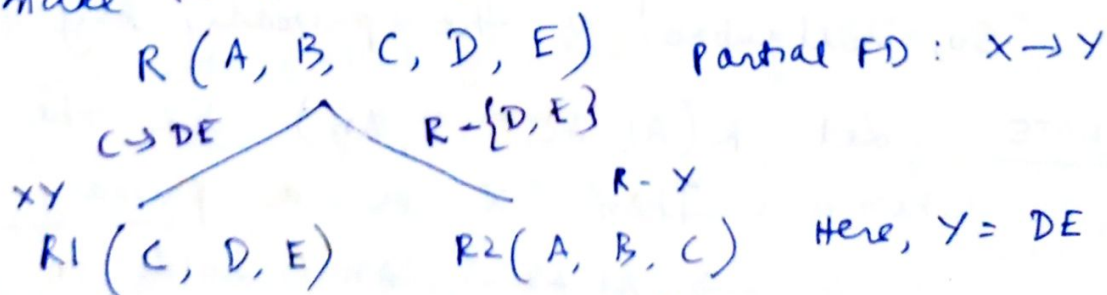
So, BC is the primary key.

But, partial FD exists : $C \rightarrow DE$

Non-prime attributes D and E are partially dependent on PK.

So, the given relation schema is not in 2NF.

To make it in 2NF, we decompose R as:



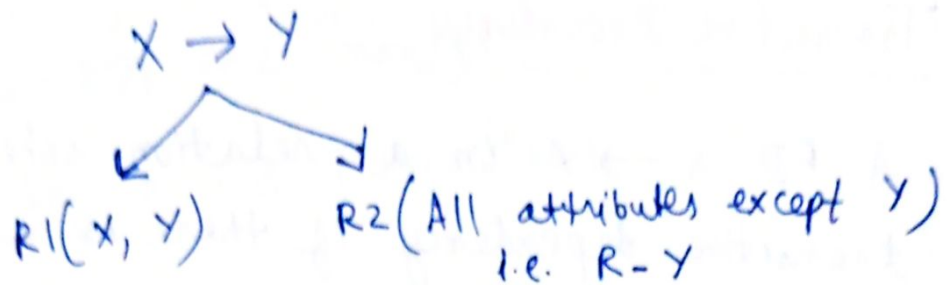
FDs: $C \rightarrow DE, D \rightarrow E$

$BC \rightarrow A$

No partial FD exists. So, R_1 and R_2 are in 2NF.

Decomposition Rule

9) If $X \rightarrow Y$ is the partial FD in R .
Then, R is decomposed as:



NOTE:

- The test for 2NF involves testing for FDs whose L.H.S attributes are part of the primary key ^(partial key)
- If the primary key contains a single attribute, the test is not required at all as by default it is in 2NF.
- * 2NF applies to relations with composite keys - relations with a P.K. composed of two or more attributes.
- * A relation with a single attribute primary key is automatically at least in 2NF.

3) Third Normal Form (3NF)

* Third normal form is based on the concept of transitive dependency

Transitive Dependency

A FD $X \rightarrow Y$ in a relation schema R is a transitive dependency if there is a set of ^{non-prime} attributes Z that is neither a candidate key nor a subset of any key of R and both $X \rightarrow Z$ and $Z \rightarrow Y$ hold.

Defⁿ of 3NF

not a member of any of the candidate keys

- A relation schema R is in 3NF if it is in 2NF and no non-prime attribute of R is transitively dependent on the primary key.
- This means all non-prime attributes should directly depend on the primary key.
- Let $R(A_1, A_2, \dots, A_n)$ be the relation schema and F be the set of FDs.
 R is in 3NF if for every non-trivial FD $X \rightarrow Y$ either of the following conditions hold:
(i) X is a superkey or
(ii) Y is a prime attribute

Example 1

Relation schema $R(A, B, C)$

FDs $F = \{ AB \rightarrow C, C \rightarrow B \}$

$$(AB)^+ = ABC$$

Both the FDs are non-trivial.

Primary key = AB (super-key)

For the FD, $AB \rightarrow C$

AB is the superkey but C is not prime attribute.

For the FD, $C \rightarrow B$

C is not superkey but B is a prime attribute.

Hence, R is in 3NF.

Example 2

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

$F = \{ A \rightarrow B, AC \rightarrow D \}$

Both the FDs are non-trivial.

$$A^+ = AB$$

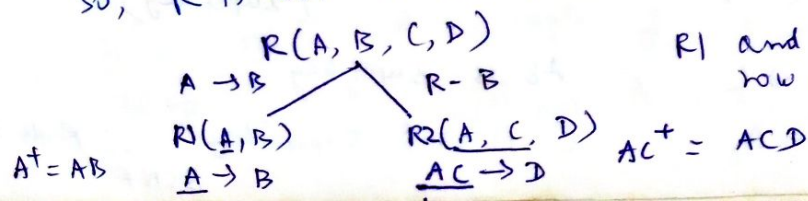
$$(AC)^+ = ACBD$$

$\therefore AC$ is the primary key.

For the FD $A \rightarrow B$,

A is not superkey and B is not a prime attribute.

So, R is not in 3NF, hence decomposed as:



4) Boyce-Codd Normal Form (BCNF)

* BCNF is a more stricter normal form than 3NF.

* Every relation in BCNF is also in 3NF, but a relation in 3NF is not necessarily in BCNF.

Defⁿ of BCNF

Let $R(A_1, A_2, \dots, A_n)$ be a relation schema and F be the set of FDs.

The relation R is in BCNF ~~iff~~ for all ~~non-trivial~~ ^{in F or of the form} $X \rightarrow Y$, either of the following conditions hold:

- X is a superkey of R .
- $X \rightarrow Y$ is a trivial FD (i.e. $Y \subseteq X$).

A relation schema R is in BCNF ~~iff~~ whenever a non-trivial FD $X \rightarrow Y$ holds in R , then X is a superkey of R .

* The only difference between the definitions of BCNF and 3NF is that condition (ii) of 3NF, which allows Y to be a prime attribute, is disallowed in BCNF.

Example

$R(A, B, C)$

$F = \{ AB \rightarrow C, C \rightarrow A \}$

Both the FD's are non-trivial.

$(AB)^+ = ABC \Rightarrow AB$ is the primary key (superkey)

$AB \rightarrow C$, AB is superkey

$C \rightarrow A$, C is not super-key $\therefore R$ is not in BCNF

But A is a prime attribute. Hence, R is in 3NF.

Lossless Join Decomposition into BCNF

The decomposition of a non-BCNF relation must be done by considering the lossless/non-additive decomposition requirement.

Example

For relation schema $R(A, B, C, D, E)$ and set of FDs $F = \{AB \rightarrow C, B \rightarrow D, A \rightarrow E\}$,

(i) Is R in BCNF?

(ii) If not decompose R into relations schemes such that they will be in BCNF.

(iii) Is the decomposition lossless join decomposition?

(iv) Test the decomposition to be dependency preserving.

Sol:

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

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

Determine the primary key of R.

$$(AB)^+ = ABCDE$$

$\Rightarrow AB$ is the primary key

(i) Test for BCNF

$AB \rightarrow C$

AB is a superkey

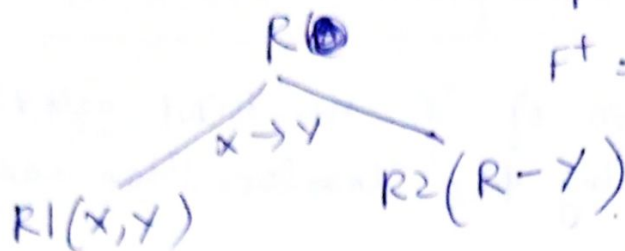
$B \rightarrow D$

B is not a superkey

~~Also~~

So, R is not in BCNF.

Let $X \rightarrow Y$ violates BCNF. Then, the relation schema R is decomposed as:



$$F = \{AB \rightarrow C, B \rightarrow D\}$$

$$F^+ = \{AB \rightarrow C, B \rightarrow D, AB \rightarrow ABCDE, B \rightarrow BCDE, A \rightarrow AE\}$$

Accordingly, R is decomposed into two relation schemas:

$$R_1(B, D)$$

$$F_1 = \{B \rightarrow D\}$$

Superkey $B^+ = BD$

R_1 is in BCNF

$$R_2(A, B, C, E)$$

$$F_2 = \{AB \rightarrow C, A \rightarrow E\}$$

superkey

$$AB^+ = ABCE$$

R_2 is not in BCNF

due to $A \rightarrow E$.

Test for lossless decomposition into R_1 and R_2

$$R_1 \cap R_2 = B, R_1 - R_2 = D$$

$B \rightarrow D$ is in F^+

Hence, the decomposition is lossless.

Test for BCNF

R_1 is in BCNF as B is a superkey.

But, R_2 is not in BCNF because:

For, $AB \rightarrow C$, AB is superkey (BCNF)

For, $A \rightarrow E$, A is not superkey. Hence, not in BCNF.

So, R_2 is further decomposed into:

$$R_3(A, E), R_4(A, B, C)$$

$$F_3 = \{A \rightarrow E\}, F_4 = \{AB \rightarrow C\}$$

$$A^+ = AE$$

$$AB^+ = ABC$$

Hence, R_3 and R_4 are in BCNF.

Test for Lossy Decomposition of R3 and R4

$$R3 \cap R4 = A, \quad R3 - R4 = E$$

$A \rightarrow E$ exists in F^+ .

Hence, the decomposition of $R2$ is lossy.

Test for Dependency Preserving $F = \{ B \rightarrow D, A \rightarrow E, AB \rightarrow C \}$

Finally, the decomposed relations are:

$$R1(B, D), \quad R3(A, E), \quad R4(A, B, C)$$

$$F1 = \{ B \rightarrow D \}, \quad F3 = \{ A \rightarrow E \}, \quad F4 = \{ AB \rightarrow C \}$$

$$(F1 \cup F3 \cup F4)^+ = (B \rightarrow D, A \rightarrow E, AB \rightarrow C)^+$$

closure of
Union of projections of
F on each R_i is same
as F^+

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

Hence, the decomposition is dependency-preserving.

$$\pi_{R1}(F) \cup \pi_{R3}(F) \cup \pi_{R4}(F) = F$$