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 value)
* 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

$\xrightarrow{\text{conversion to 1NF}}$

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

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

Eg: $F = \{ \underset{\downarrow}{AB \to C}, B \to D, D \to E, A \to C, \underset{\downarrow}{AD \to F} \}$

Partial FD as $A \to C$ holds — Full FD

Determination of Primary Key

Let X be a primary key. Suppose X functionally determines attributes $A1, A2, \ldots, An$.

$$X \to A1\ A2 \ldots An$$

Eg:

~~F = { rollno → name, rollno → mark, rollno → address, rollno → grade }~~

Example

Suppose we have the following set of FDs F:

F = { roll → name, roll subno → mark, mark → grade }

$roll^+$ = roll, name

$rollsubno^+$ = roll subno, mark, grade

$mark^+$ = mark, grade

Prime Attributes = roll, subno

Non-Prime Attributes = name, mark, grade

So, $rollsubno^+$ is the primary key.

NOTE: Let $R(A1, A2, \ldots, An)$ be the relation schema. Then X is a primary key if

$$X \to A1\ A2 \ldots An$$

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.

Sol^n: First we determine the primary key of R.

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

So, BC is the primary key.

But, partial FD exists : $\underset{X}{C} \rightarrow \underset{Y}{DE}$ Non-prime attributes D and E are partially dependent on P.K.

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

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

R(A, B, C, D, E) Partial FD : $X \rightarrow Y$

$C \rightarrow DE$ (XY) R − {D, E} (R − Y)

R1(C, D, E) R2(A, B, C) Here, Y = DE

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

No partial FD exists. So, R1 and R2 are in 2NF.

Decomposition Rule

If $X \to Y$ is the partial FD in R.

Then, R is decomposed as:

$$X \to Y$$

$R1(X, Y)$ $R2$(All attributes except Y) i.e. $R - Y$

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)

Exa

* 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 attribute 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.

not a member of any of the candidate keys

Def^n of 3NF

- 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 (A1, A2, ..., An) be the relation schema and F be the set of FDs.

  R is in 3NF if for every non-trivial FD ($X \not\supseteq Y$) $X \rightarrow Y$ 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 $F = \{AB \rightarrow C,\ C \rightarrow B\}$ $\quad (AB)^+ = ABC$

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.

$A^+ = AB$

$(AC)^+ = \cancel{AB}\ ACBD$

∴ 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:

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

$A \rightarrow B$ $\qquad$ $R - B$

$R1(A, B)$ $\quad$ $\underline{A \rightarrow B}$ $\quad$ $A^+ = AB$

$R2(A, C, D)$ $\quad$ $\underline{AC \rightarrow D}$ $\quad$ $AC^+ = ACD$

R1 and R2 are now in 3NF.

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.

Def^n of BCNF

Let $R(A1, A2, \ldots, An)$ be a relation schema and F be the set of FDs.
~~The relation R is in BCNF if for all ~~non-trivial~~ FDs in F^+ of the form $X \to Y$, either of the following conditions hold :~~
~~- X is a superkey of R~~
~~- $X \to Y$ is a trivial FD (i.e $Y \subseteq X$)~~

> A relation schema R is in BCNF if whenever a nontrivial FD $X \to Y$ 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 $R(A, B, C)$

$F = \{ AB \to C, C \to A \}$ — Both the FD's are non-trivial.

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

$AB \to C$, AB is super-key

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

(i) Is R in BCNF?

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

(iii) Is the decomposition lossless join decomposition?

(iv) Test the decomposition to be dependency preserving.

Sol^n: R(A, B, C, D, E)

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

Determine the primary key of R.

$(AB)^+ = ABCDE$

$\Rightarrow$ AB is the primary key

(i) Test for BCNF

$AB \to C$

AB is a superkey

$B \to D$

B is not a superkey

So, R is not in BCNF.

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

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

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

R

$X \rightarrow Y$

R1(X, Y)   R2(R − Y)

Accordingly, R is decomposed into two relation schemas:

R1(B, D), R2(A, B, C, E) ← (R − D)

$F1 = \{B \rightarrow D\}$ $F2 = \{AB \rightarrow C, A \rightarrow E\}$

B: superkey, $B^{+} = BD$ AB: superkey, $AB^{+} = ABCE$

R1 is in BCNF   R2 is not in BCNF due to $A \rightarrow E$.

Test for lossless decomposition into R1 and R2

$R1 \cap R2 = B$, $R1 - R2 = D$

$B \rightarrow D$ is in F^{+}

Hence, the decomposition is lossless.

Test for BCNF

R1 is in BCNF as B is a superkey.

But, R2 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, R2 is further decomposed into:

R3(A, E), R4(A, B, C)

$F3 = \{A \rightarrow E\}$, $F4 = \{AB \rightarrow C\}$ Hence, R3 and R4 are in BCNF

$A^{+} = AE$ $AB^{+} = ABC$

Test for Lossless Decomposition of R3 and R4

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

$A \rightarrow E$ exists in F^+

Hence, the decomposition of R2 is lossless.

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

Finally, the decomposed relations are :

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

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

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

Closure of Union of projections of For each R_i is same as F^+

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

$= F^+$

Hence, the decomposition is dependency-preserving.

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