COMPUTER SCIENCE

Database Management System

FD's & Normalization

Lecture_12

Vijay Agarwal sir

TOPICS TO BE COVERED

01 Normal Forms

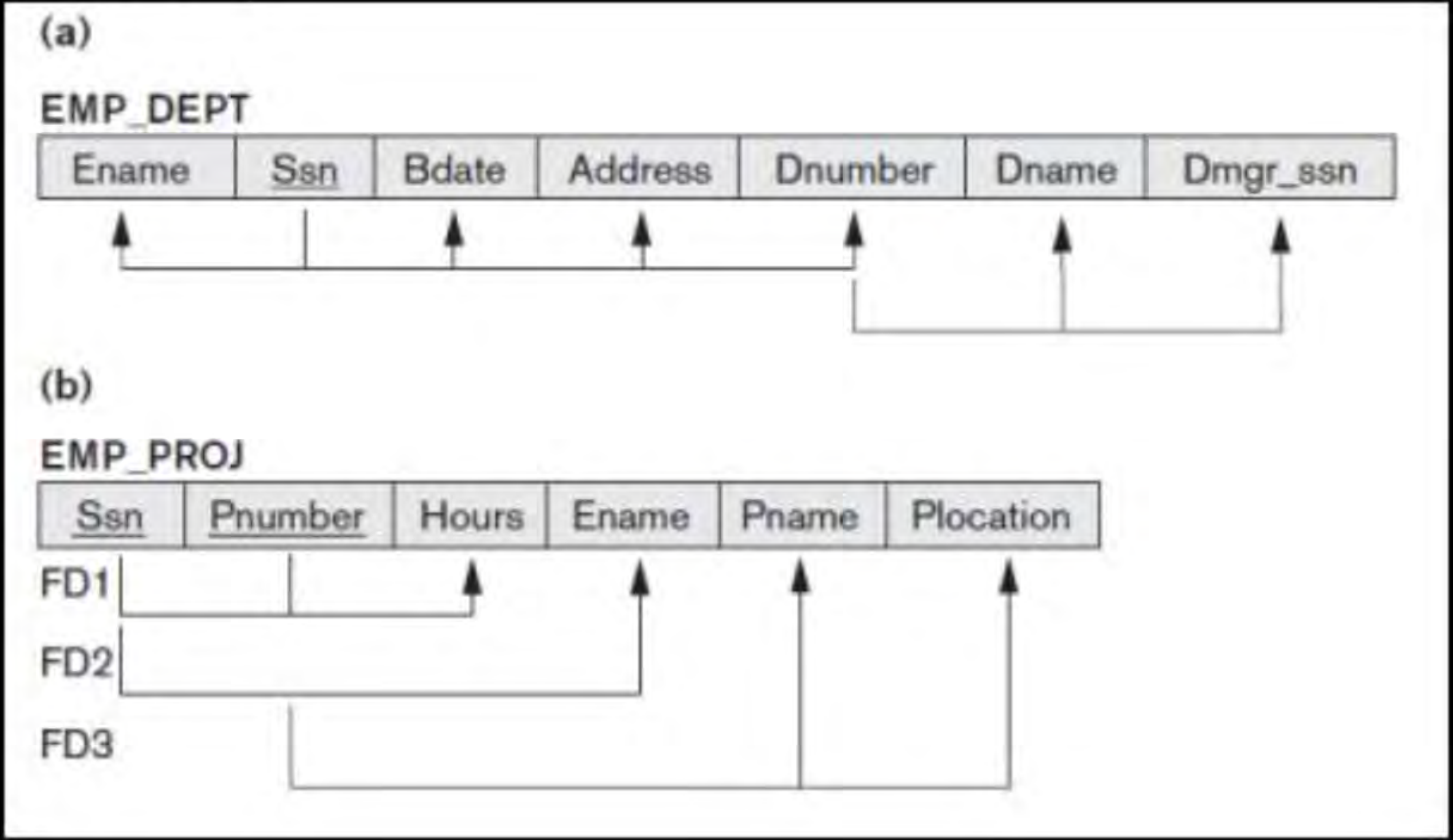
02 Normal Forms Decomposition

15.3.5 Second Normal Form

Second normal form (2NF) 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; that is, for any attribute $A \varepsilon X$, $(X - \{A\})$ does *not* functionally determine Y. A functional dependency $X \rightarrow Y$ is a **partial dependency** if some attribute $A \varepsilon X$ can be removed from X and the dependency still holds; that is, for some $A \varepsilon X$, $(X - \{A\}) \rightarrow Y$. In Figure 15.3(b), {Ssn, Pnumber} → Hours is a full dependency (neither Ssn → Hours nor Pnumber → Hours holds). However, the dependency {Ssn, Pnumber} → Ename is partial because Ssn → Ename holds.

Definition. A relation schema R is in 2NF if every nonprime attribute A in R is *fully functionally dependent* on the primary key of R.

Figure 15.3


(a)
EMP_DEPT
Ename
Ssn
Bdate
Address
Dnumber
Dname
Dmgr_ssn
(b)
EMP_PROJ
Ssn
Pnumber
Hours
Ename
Pname
Plocation
FD1
FD2
FD3

Normal Forms

8.17 A functional dependency $\alpha \rightarrow \beta$ is called a partial dependency if there is a proper subset γ of α such that $\gamma \rightarrow \beta$. We say that β is *partially dependent* on α. A relation schema R is in second normal form (2NF) 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.

Normal FORM

① 1NF

Possible Non Trivial FD which Cause Redundancy

CASE I, CASE II, CASE III

Partial FD &

FULL FD Concept

$R(ABCDEF)\ \{ABC \to DE,\ DE \to ABC,\ AB \to D,\ DE \to F,\ E \to C\}$

(i) $ABC \to D$ → Partial

(ii) $AE \to C$ → Partial

(iii) $AF \to D$ → FULL FD

(iv) $AB \to D$ → FULL FD

(v) $AC \to D$ → FULL FD

(vi) $BC \to D$ → FULL FD

(vii) $\underline{DE \to C}$ → Partial FD (Bcz $E \to C$)

(vi) $AB \to F$ → FULL FD.

$AF \to D$

$A \to D$ × $[A]^+ = [A]$

$F \to D$ × $[F]^+ = [F]$

∴ $AF \to D$ is FULL FD.

second Normal Form [2NF]

A Relation Schema R is in 2NF iff

(i) R is in 1NF

(ii) R does not Contain this Case

Proper subset of Candidate key $\longrightarrow$ Non Key Attribute

Only Violation of 2NF.

if this case exist then Not in 2NF

P.F.D

P.F.D

R(ABC) {AB → C , B → C}

Candidate key = AB

Non key Attribute = {C}

Not in 2NF

B → C

Proper subset of C.K

Non key Attribute

Only Violation of 2NF.

its Not Partial FD.

Not in 2NF.

Bcz AB → C is PFD.

2 + 2 = 4

2 * 2 = 4

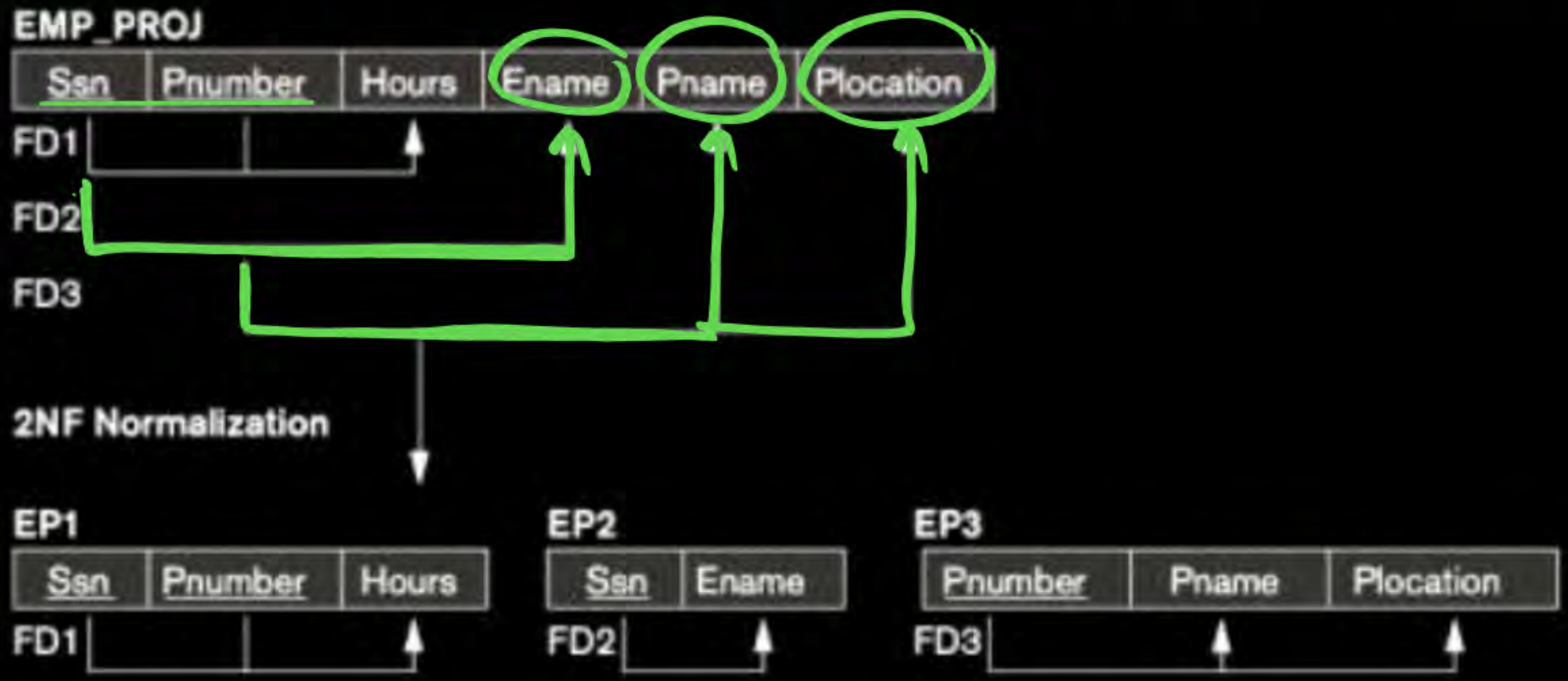
3 + 3 = 6

3 * 3 = 9

Normal Forms

Second Normal Form

Definition: A relation schema R is in 2NF if every nonprime attribute A in R is fully functionally dependent on the primary key of R.



$(AB)^+ = \{ABC\}$

Ck: AB

AB determines all attribute of Relation R.

$\boxed{AB \to C}$ is P.F.D

$\boxed{R(ABC) \quad A \to C,\ B \to C}$

Candidate key = AB

Non key Attribute = {C}

Check 2NF ?

$A \to C \quad B \to C$

Proper Subset of C.K → Non key Attribute

Violation of 2NF

Not in 2NF

$A \to C$

or

$B \to C$

then $AB \to C$ is Partial FD

Not fully Dep. on P.K

Non key Attribute

P.F.D

R(ABC) [AB → C , B → C]

AB → C

Partially Dep. on P.K

Non Prime Attribute

Not in 2NF

Due to this

$\boxed{B \to C}$

$R(ABCDEFG)$

P.F.D

$AB \rightarrow C$
$AB \rightarrow D$
$AB \rightarrow E$
$AB \rightarrow F$
$AB \rightarrow G$

$(AB)^+ = [ABCDEFG]$

[AB] is Candidate key.

AB

A B

Non Key Attribute = (C, D, E, F, G)

if

$A \rightarrow C$		$B \rightarrow C$
$A \rightarrow D$		$B \rightarrow D$
$A \rightarrow E$	OR	$B \rightarrow E$
$A \rightarrow F$		$B \rightarrow F$
$A \rightarrow G$		$B \rightarrow G$

Q. Let R(A, B, C, D, E, P, G) be a relational schema in which the following functional dependencies are known to hold:

$AB \rightarrow CD$, $DE \rightarrow P$, $C \rightarrow E$, $P \rightarrow C$ and $B \rightarrow G$.

The relational schema R is

Candidate key = [AB]

Non key Attribute = (C, D, E, P, G)

(A) In BCNF

(B) In 3NF, but not in BCNF

(C) In 2NF, but not in 3NF

(D) Not in 2NF

$B \rightarrow G$

Proper subset of C.K

Non key Attribute

Not in 2NF

2NF (Second Normal Form)

A Relation R is in 2NF iff

① R is in 1NF

② R does not Contain

Violation of 2NF

$$\text{Proper subset of C.K} \longrightarrow \text{Non key Attribute}$$

Not in 2NF.

Third Normal Form [3NF]

$X \longrightarrow y$

CASE I: Proper subset of CK $\longrightarrow$ Non key Attribute ✗

CASE II: Non key Attribute $\longrightarrow$ Non key Attribute ✗

A Relation Schema R is in 3NF iff every $X \rightarrow y$ Non Trivial FD Must Satisfy the following Condition

$X \longrightarrow Y$

X : Super key

OR

y : key/Prime Attribute

3NF

A Relation R is in 3NF iff

1. R is in 2NF.
2. R does not Contain Any Transitive Dependency.

R(ABC) [A → B, B → C]

C.K : A

Non key Attribute = [B, C]

Transitive Dep.

B → C
N.K   N.K

A → B & B → C

then A → C is Transitive Dep.

X → y is in 3NF

either

X : Super key

OR

y : key/Prime Attribute

R(ABC) [A → B, B → C]

C.K : A

Non key Attribute = [B, C]

Transitive Dep.

B → C

N.K N.K

$A \to B$ & $B \to C$

then $A \to C$ is Transitive Dep.

Check 3NF?

Not in 3NF

A → B ✓ 3NF

B → C X 3NF [B is Not Super key or C is Non key Attribute]

✓ $X \to y$ is in 3NF

either

X : Super key

OR

y : key/Prime Attribute

eg R(ABC) [A → B. A → C]

Candidate key = [A]

Non key Attribute = [B,C]

Check 3NF ?

A → B ✓ 3NF [A is Super key]

A → C ✓ 3NF [A is Super key]

✓

3NF

X → y

X: Super key

or

y: key | Prime Attribute

R(ABCD) [AB → CD, D → A]

Candidate key = [AB, DB]

Check 3NF ?

$X \to Y$

either

X: Super key

OR

Y: Prime/key Attribute

AB → CD ✓3NF [∵ AB is Super key]

D → A ✓3NF [Here D is Not Super key
But A is key/Prime Attribute]

$X \to Y$

Normal Forms

E. Codd

(a) YES

(b) NO

Third Normal Form

Definition: According to Codd's original definition, a relation schema R is in 3NF if it satisfies 2NF and no nonprime attribute of R is transitively dependent on the primary key.

Definition: A relation schema R is in third normal form (3NF) if, whenever a nontrivial functional dependency $X \to A$ holds in R either (a) X is a superkey of R, or (b) A is a prime attribute of R.

$X \to A$

(a) ⇒ X: superkey OR (b) A is key/Prime

R(ABC) [A → B, B → C]

Candidate key = [A]

Non key Attribute = [B, C]

Check 3NF ?

Best

(i)

B → C

N.K Attribute, N.K Attribute

Not in 3NF

(ii)

B → C

B is Not Super key

OR

C: Not key Attribute.

(iii)

A ——→ C

Here Non key Attribute (C) Transitive Determine by Primary key A.

A → B & B → C

then

A → C is transitive FD.

C (Non key) Attribute is T.D on P.K A.

R(ABC) [A → B, B → C]

C.K : A

Non key Attribute = [B, C]

<u>Transitive Dep.</u>

$$\underset{N.K}{B} \rightarrow \underset{N.K}{C}$$

A → B & B → C

then A → C is Transitive Dep.

<u>Not in 3NF</u>

A → B ✓ 3NF

B → C ✗ 3NF [B is Not Super key or C is Non key Attribute]

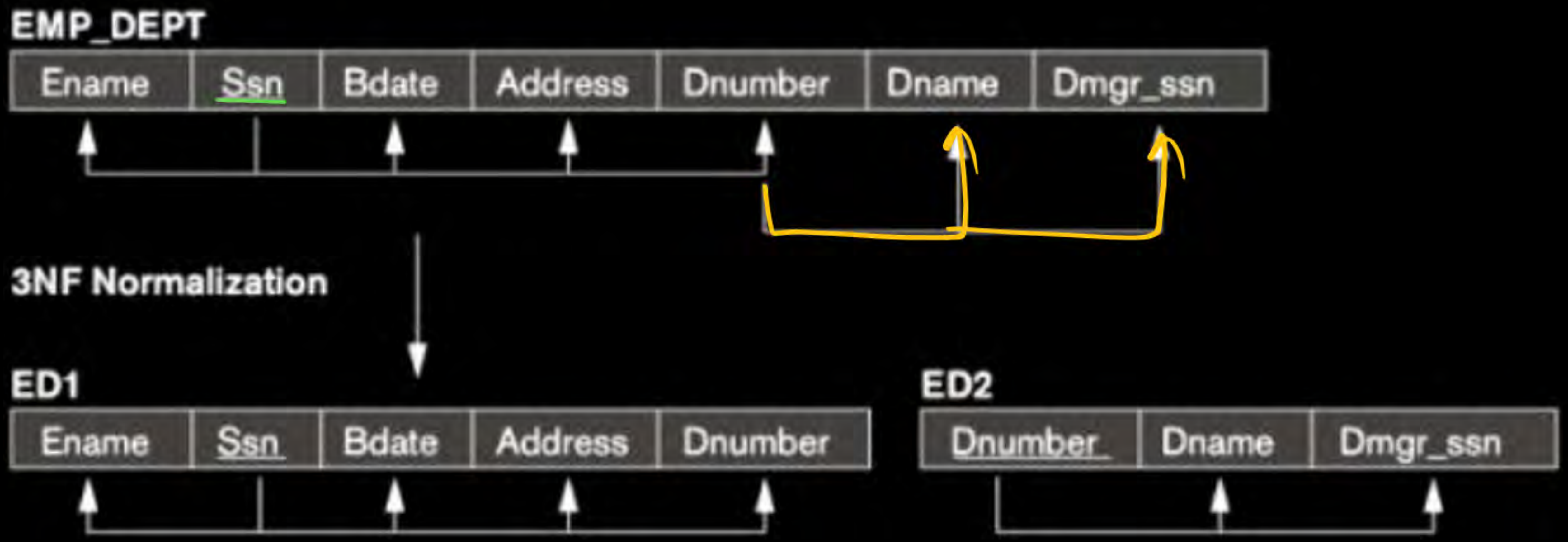
X → y is in 3NF

either

X : Super key

OR

y : key/Prime Attribute


EMP_DEPT
Ename
Ssn
Bdate
Address
Dnumber
Dname
Dmgr_ssn
3NF Normalization
ED1
Ename
Ssn
Bdate
Address
Dnumber
ED2
Dnumber
Dname
Dmgr_ssn

Boyce – Codd Normal Form

Definition: A relation schema R is in BCNF if whenever a nontrivial functional dependency $X \rightarrow A$ holds in R, then X is a superkey of R.

Every $X \rightarrow y$ Non Trivial FD Must Satisfy the following condition.

$\boxed{X : \text{Super key.}}$

2NF Checking

Proper subset of C.K $\longrightarrow$ Non key Attribute

Violation of 2NF

3NF Checking

Every Non Trivial FD $X \rightarrow y$ Must satisfy following Condition

either

X : Super key

OR

y: key/Prime Attribute

BCNF Checking

Every $X \rightarrow y$ Non Trivial FD is in BCNF iff

$X \longrightarrow y$

X: Super key.

Important Points : [1NF]

① If a Relation R has Only One Candidate key then R always is in 1NF But May or May Not in 2NF, 3NF & BCNF.

eg R(ABCDE) [AB→C, C→D, B→E]

Candidate key = (AB)

Non Key Attribute = (C, D, E)

B→E

violation of 2NF So Not in 2NF

Important Points [2NF]

If In a Relation R, all Candidate keys are Simple keys (No Composite key) then Relation R always is in 2NF But May or May Not in 3NF & BCNF

eg R(ABCDE) [A→B, B→C, C→D, D→E, C→A]

Candidate keys = [A, C, B]

Non Prime Non key Attribute = [D, E]

All keys are simple so R is in 2NF But May Not in 3NF.

D → E Not in 3NF.

D: is Not Super key

E is Not key / Prime Attribute.

Important Points [3NF]

3NF either X: super key
or
Y: key | Prime Attribute

If in a Relation all Attribute is a key | Prime Attribute then Relation R always is in 3NF But May or May Not in BCNF.

eg R(ABCDE) [A→B, B→C, C→D, D→A]

Candidate key = [AE, BE, CE, DE]

key | Prime Attribute = [A, B, C, D, E]

All Attribute are key | Prime Attribute then R is in 3NF But Not in BCNF

X → Y
A → B
B → C
C → D
D → A

X: Not Super key
Not in BCNF.

Important Points [BCNF]

① If R is in 3NF && All Candidate keys are Simple Candidate key (No Composite key) then R always is in BCNF.

eg R(ABCD) [A→B, B→C, C→D, D→A]

Candidate key = [A, B, C, D]

Prime|key Attribute = [A, B, C, D]

✓BCNF

A → B
B → C
C → D
D → A

X : Super key

Here all attributes are Prime|key Attribute so R is in 3NF && all keys are simple CK .... So R is in BCNF.

Transitive FD

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

then $A \rightarrow C$ is Transitive FD

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

$A \rightarrow B$ & $B \rightarrow C$

$[B]^+ = [B\ C\ D\ A]$

Important Points [BCNF]

② Binary Relation (Relation with 2 Attributes) is always in BCNF

R(AB) [A→B]	R(AB) [B→A]	R(AB) [A→B, B→A]
Candidate key : [A]	Candidate key = [B]	Candidate key = [A, B]
$A \rightarrow B$	$B \rightarrow A$	$A \rightarrow B$ & $B \rightarrow A$
Super key	Super key	Super key
∴ R is in BCNF	∴ R is in BCNF	∴ R is in BCNF

Important Points [BCNF]

③ A Relation R with No Non Trivial FD is in BCNF
(Trivial FD in BCNF)

eg $ABC \to AB$

⇓

Super key

eg $XYZ \to YZ$

⇓

Super key

BCNF.

Design Goal	1NF	2NF	3NF	BCNF
0% Redundancy	✗	✗	✗	✓
Lossless Join	✓	✓	✓	✓
Dependency Preserving	✓	✓	✓	May / May Not

But subset
M.V.F.D
$X \rightarrow\rightarrow Y$

Q. In a relational data model, which one of the following statements is TRUE?

(A) A relation with only two attributes is always in BCNF.

(B) If all attributes of a relation are prime attributes, then the relation is in BCNF.

(C) Every relation has at least one non-prime attribute.

(D) BCNF decompositions preserve functional dependencies.

Q. Let R(A, B, C, D, E, P, G) be a relational schema in which the following functional dependencies are known to hold:

$AB \rightarrow CD$, $DE \rightarrow P$, $C \rightarrow E$, $P \rightarrow C$ and $B \rightarrow G$.

The relational schema R is

(A) In BCNF

(B) In 3NF, but not in BCNF

(C) In 2NF, but not in 3NF

(D) Not in 2NF

Q

Consider the following statements: [MSQ]

S_1: If every attribute is prime attribute in R, then Relation R will always be in BCNF.

S_2: Any Relation with two Attribute is in 3 NF and 2 NF.

S_3: If every key of relation R is a simple candidate key (No composite key) then the relation R not always in NF.

S_4: In BCNF there is always a lossless join and Dependency Preserving Decomposition.

Which of the above statement are incorrect

(A) S_1

(B) S_2

(C) S_3

(D) S_4

Q

The relation scheme student Performance (name, courseNO, rollNo, grade) has the following functional dependencies:

[2004: 2 Marks]

name, courseNo $\rightarrow$ grade

RollNo, courseNo $\rightarrow$ grade

name $\rightarrow$ rollNo

rollNO $\rightarrow$ name

The highest normal form of this relation scheme is

(A) 2 NF

(B) 3 NF

(C) BCNF

(D) 4 NF

MCQ

In a relational data model, which one of the following statements is TRUE?

GATE-2022-CS: 1M]

A. A relation with only two attributes is always in BCNF.

B. If all attributes of a relation are prime attributes, then the relation is in BCNF.

C. Every relation has at least one non-prime attribute.

D. BCNF decompositions preserve functional dependencies.

NAT

Consider a relation R(A, B, C, D, E) with the following three functional dependencies.

$AB \rightarrow C$; $BC \rightarrow D$; $C \rightarrow E$;

The number of super keys in the relation R is ______.

[GATE-2022-CS: 1M]

MCQ

Consider a relational table R that is in 3 NF, but not in BCNF. Which one of the following statements is TRUE?

[GATE-2020-CS: 2M]

A. R has a non-trivial functional dependency $X \rightarrow A$, where X is not a superkey and A is a prime attribute.

B. R has a non-trivial functional dependency $X \rightarrow A$, where X is not a superkey and A is a non-prime attribute and X is not a proper subset of any key.

C. R has a non-trivial functional dependency $X \rightarrow A$, where X is not a superkey and A is a non-prime attribute and X is a proper subset of some key.

D. A cell in R holds a set instead of an atomic value.

MCQ

Let the set of functional dependencies F= {QR→S, R→P, S→Q} hold on a relation schema X=(PQRS) . X is not in BCNF. Suppose X is decomposed into two schemas Y and Z, where Y=(P R) and Z=(Q R S).

Consider the two statements given below.

I. Both Y and Z are in BCNF

II. Decomposition of X into Y and Z is dependency preserving and lossless

Which of the above statements is/are correct?

[GATE-2019-CS: 2M]

A. Both I and II

B. I only

C. II only

D. Neither I nor II

Q.

Consider the following four relational schemas. For each schema, all non-trivial functional dependencies are listed. The underlined attributes are the respective primary keys.

Schema I: Registration (<u>rollno</u>, courses)

Field 'courses' is a set-valued attribute containing the set of courses a student has registered for.

Non-trivial functional dependency:

rollno → courses

Schema II: Registration (<u>rollno, courseid</u>, email)

Non-trivial functional dependencies:

rollno, courseid → email

email → rollno

Schema III: Registration (<u>rollno, courseid</u>, marks, grade)

Non-trivial functional dependencies:

rollno, courseid → marks, grade

marks → grade

Schema IV: Registration (<u>rollno, courseid</u>, marks, credit)

Non-trivial functional dependencies:

rollno, courseid → credit

courseid → credit

Which one of the relational schemas above is in 3NF but not in BCNF?

[MCQ: 2018: 2M]

(A) Schema I

(B) Schema II

(C) Schema III

(D) Schema IV

Q.

A database of research articles in a journal uses the following schema.

(VOLUME, NUMBER, STARTPAGE, ENDPAGE, TITLE, YEAR, PRICE)

The primary key is (VOLUME, NUMBER, STARTPAGE, ENDPAGE) and the following functional dependencies exist in the schema.

(VOLUME, NUMBER, STARTPAGE, ENDPAGE) $\rightarrow$ TITLE

(VOLUME, NUMBER) $\rightarrow$ YEAR

(VOLUME, NUMBER, STARTPAGE, ENDPAGE) $\rightarrow$ PRICE.

The database is redesigned to use the following schemas.

(VOLUME, NUMBER, STARTPAGE, ENDPAGE, TITLE, PRICE)

(VOLUME, NUMBER, YEAR)

Which of the weakest normal form that the new database satisfies, but the old one does not? [MCQ: 2016: 1M]

A. 1NF

C. 2NF

B. 3NF

D. BCNF

NAT

Given an instance of the STUDENTS relation as shown below:

Student ID	Student Name	Student Email	Student Age	CPI
2345	Shankar	shankar@math	X	9.4
1287	Swati	swati@ee	19	9.5
7853	Shankar	shankar@cse	19	9.4
9876	Swati	swati@mech	18	9.3
8765	Ganesh	ganesh@civil	19	8.7

For (Student Name, Student Age) to be a key for this instance, the value X should NOT be equal to ____________.

[GATE-2014-CS: 1M]

NAT

The maximum number of superkeys for the relation schema R (E, F, G, H) with E as the key is ___________.

[GATE-2014-CS: 1M]

MCQ

Given the following two statements:

S1: Every table with two single-valued attributes is in 1 NF, 2 NF, 3 NF and BCNF.

S2: $AB \to C$, $D \to E$, $E \to C$ is a minimal cover for the set of functional dependencies $AB \to C$, $D \to E$, $AB \to E$, $E \to C$.

Which one of the following is CORRECT?

[GATE-2014-CS: 2M]

(A) S1 is TRUE and S2 is FALSE.

(B) Both S1 and S2 are TRUE.

(C) S1 is FALSE and S2 is TRUE

(D) Both S1 and S2 are FALSE.

MCQ

Relation R has eight attributes ABCDEFGH.
Fields of R contain only atomic values.
$F = \{CH \rightarrow G, A \rightarrow BC, B \rightarrow CFH, E \rightarrow A, F \rightarrow EG\}$ is a set of functional dependencies (FDs) so that F^+ is exactly the set of FDs that hold for R.
How many candidate keys does the relation R have?

[GATE-2013-CS: 2M]

(A) 3

(B) 4

(C) 5

(D) 6

MCQ

Relation R has eight attributes ABCDEFGH.
Fields of R contain only atomic values.
$F = \{CH \rightarrow G, A \rightarrow BC, B \rightarrow CFH, E \rightarrow A, F \rightarrow EG\}$ is a set of functional dependencies (FDs) so that F^+ is exactly the set of FDs that hold for R.
The relation R is

[GATE-2013-CS: 2M]

(A) in 1 NF, but not in 2 NF.

(B) in 2 NF, but not in 3 NF.

(C) in 3NF, but not in BCNF.

(D) in BCNF.

MCQ

Which of the following is TRUE?

[GATE-2012-CS: 1M]

A. Every relation in 3 NF is also in BCNF

B. A relation R is in 3 NF if every non-prime attribute of R is fully functionally dependent on every key of R

C. Every relation in BCNF is also in 3 NF

D. No relation can be in both BCNF and 3 NF

MCQ

Consider the following relational schemes for a library database:

Book (Title, Author, Catalog_no, Publisher, Year, price)

Collection (Title, Author, Catalog_no)

With the following functional dependencies:

I. TitleAuthor $\longrightarrow$ Catalog_no

II. Catalog_no $\longrightarrow$ Title Author Publisher Year

III. Publisher Title Year $\rightarrow$ Price

Assume { Author, Title} is the key for both schemes.

Which of the following statements is true?

[GATE-2008-CS: 2M]

A. Both Book and Collection are in BCNF

B. Both Book and Collection are in 3 NF only

C. Book is in 2 NF and Collection is in 3NF

D. Both Book and Collection are in 2 NF only

MCQ

Let R(A, B, C, D, E, P, G) be a relational schema in which the following functional dependencies are known to hold:

$AB \to CD$, $DE \to P$, $C \to E$, $P \to C$ and $B \to G$.

The relational schema R is

[GATE-2008-CS: 2M]

A. in BCNF

B. in 3NF, but not in BCNF

C. in 2 NF, but not in 3 NF

D. not in 2 NF

Q

Relation R is decomposed using a set of functional dependencies, F, and relation S is decomposed using another set of functional dependencies, G. One decomposition is definitely BCNF, the other is definitely 3NF, but it is not know which is which. To make a guaranteed identification, which one of the following tests should be used on the decompositions? (Assume that the closures of F and G are available).

[2002: 2 Marks]

A. Dependency-preservation

B. Lossless-join

C. BCNF definition

D. 3 NF definition

Q

Which of the following relational schema with given FD's follows is/are in BCNF?

A. R(ABCDE) and FD's are $\{A \to B, B \to C, C \to D, D \to E, C \to A\}$

B. R(ABCDE) and FD's are $\{A \to B, C \to D, D \to E\}$

C. R(ABCD) and FD's are $\{A \to B, B \to C, C \to D, D \to A\}$

D. R(ABCD) and FD's are $\{A \to B, B \to C, C \to A\}$

THANK YOU!