COMPUTER SCIENCE

Database Management System

FD's & Normalization

Lecture_12

Vijay Agarwal sir

TOPICS TO BE COVERED

01 Normal Forms

(1) RDBMS Concept

(2) FD Concept

(3) FD types & Properties.

(4) Attribute closure.

(5) Key Concept
- → Super key
(6)
- → Candidate key
(7)
- → Multiple C.K

(8) membership set

(9) Equality b/w 2 FD set

(10) Minimal Cover

(11) Finding # Super key & C.K

(12) Closure of FD set

(13) Properties of Decomposition
- (i) Lossless Join
- (ii) Dependency Preserving

R(ABCDE)

Finding # Super key :

CK: [A, E, CD, CB]

- Venn Diagram
- More C.K
- every (n-1) attribute

Apply Mind.

Total # S.K $= 2^5 - 1 = 31$

$31 - 4 = 27$

D
C
B
BD

Not key

C
D
B
DB

2022 Q

C.K

(AB CDE)

C.K : AB

$\# S.K = 2^{5-2}$

$\Rightarrow 2^3$

$= 8$ S.K

Q

R(EFGH) E is C.K

$\# S.K = 2^{4-1} = 2^3$

$= 8$ S.K

Students Learning.

First Time

Normalization

Normal Forms

Normal Form is a Set of Rule is used to Reduce / eliminate the Redundancy

Redundancy is the Unnecessary Repeatition of Data.

Normal Forms

There are various Normal Form.

Codd.

1. 1NF (First Normal Form)
2. 2NF (Second Normal Form)
3. 3NF (Third Normal Form)
4. BCNF [Boyce Codd Normal Form]

× 4NF

× 5NF

$X \rightarrow\rightarrow y$ → Multivalued Functional Dependency.

Normal Forms

- Every Higher Normal Form Contain the Lower Normal Form.

- If a Relation R is in 2NF that means its already is in 1NF.

- If a Relation R is in 3NF that means its already is in 2NF & 1NF

- If a Relation R is in BCNF that means R already is in 3NF, 2NF & 1NF.

Normal Forms

First Normal Form [1NF]

A Relation R is in 1NF iff R does not Contain Any Multi-valued Attribute.

OR (Domain of attribute)

A Relation R is in 1NF iff all attribute of R are atomic.

Roll	Name	Course.
1	Vinod	C/JAVA

Multivalued Attribute

Not in 1NF

RollNo	Name	Course
1	Vinod	C
1	Vinod	JAVA

R is in 1NF.

Normal Forms

Note: 1NF Ensured by Candidate Key.

Note: Default RDBMS is in 1NF.

In 1NF Redundancy level is too High.

Redundancy Level:	1NF > 2NF > 3NF > BCNF.

Possible Non Trivial FD which create Redundancy

CASE I: Proper Subset of Candidate key $\rightarrow$ Non Prime/ Non key Attribute — Eliminated by 2NF

CASE II: Non key Attribute $\rightarrow$ Non key Attribute — El. By 3NF

CASE III: Proper Subset of One C.K $\rightarrow$ Proper Subset of another CK — El. By BCNF.

CASE I: Proper subset of Candidate key $\longrightarrow$ Non key/non Prime Attribute

eg R(ABCDEF) [AB → C, C → DF, B → E]

Candidate key = (AB)

Non Prime/Non key Attribute = [C, D, E, F]

$$B \longrightarrow E$$

Proper subset of Candidate key

Non key Attribute.

Violation of 2NF.

Not in 2NF.

CASE II

Non key Attribute $\longrightarrow$ Non key Attribute

eg R(ABC) [A → B, B → C]

Candidate key = [A]

Non Prime | Non key Attribute = [B, C]

B is Not Subset key

OR

C is Not key | Prime Attribute.

$$\underbrace{B}_{\text{Non key Attribute}} \longrightarrow \underbrace{C}_{\text{Non key Attribute}}$$

Not in 3NF

CASE III

Proper Subset of One C.K $\rightarrow$ Proper Subset of another C.K

eg R(ABCD) [AB → CD, D → A]

Candidate key = [AB, DB]

Non key Attribute = [C]

D → A

Proper subset of One C.K (D)

Proper Subset of another C.K (A)

Not in BCNF

CASE III

Proper Subset of One C.K → Proper Subset of another C.K

Subset of one CK → Subset of Same CK

eg R(ABCD) [AB → CD, D → A]

Candidate key = [AB, DB]

Non key Attribute = [C]

D → A

Proper subset of One C.K — Proper Subset of another C.K

B→A or A→B

D→B or B→D

eg B → A or A → B then AB is Not C.K either A or B is CK

eg D → B, or B → D then DB is Not CK D or B is CK

Q

FD I	FD II
$AB \rightarrow C$,	$B \rightarrow C$.

Identify which one is Partial Dependency?

(a) FD I

(b) FD II

(c) Both

(d) None.

~~$B \rightarrow C$ is Partial FD.~~

$AB \rightarrow C$ is Partial Dependency.

Q) FD I: $AB \rightarrow C$, FD II: $A \rightarrow C$.

Identify which one is Partial Dependency?

a) FD II

b) FD I

c) Both

d) None

~~$A \rightarrow C$ is Partial FD.~~

$AB \rightarrow C$ is Partial Dependency.

Partial Dependency.

eg $\boxed{AB \rightarrow C \text{ is Partial FD.}}$

if $A \rightarrow C$

OR

$B \rightarrow C$

$\boxed{X \rightarrow y \text{ is Partial FD}}$

If $A \in X$

$(X - A) \longrightarrow y$.

Normal Forms

Navathe

Ssn Pnumber → Hours
Full FD

Ssn ↛ Hours
Pnumber ↛ Hours.

15.3.5 Second Normal Form

Second normal form (2NF) is based on the concept of *full functional dependency*. A functional dependency $X \to Y$ 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 $X \to Y$ is a **partial dependency** if some attribute $A \varepsilon X$ can be removed from X and the dependency still holds; that is, for some $A \varepsilon X$, $(X - \{A\}) \to Y$. 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.

$X \to y$ is Full FD

Not determine
$(X-A) \not\to y$

AB → C is Full FD

if B ↛ C
or
A ↛ C

Act
$X \to y$ is Partial FD
iff $(X-A) \to y$

Ssn Pnumber → Ename is Partial

Ssn → Ename.

Normal Forms

Navathe

15.3.5 Second Normal Form

Second normal form (2NF) is based on the concept of *full functional dependency*. A functional dependency $X \to Y$ 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 $A \in X$, $(X - \{A\})$ does *not* functionally determine Y. A functional dependency $X \to Y$ is a **partial dependency** if some attribute $A \in X$ can be removed from X and the dependency still holds; that is, for some $A \in X$, $(X - \{A\}) \to Y$. 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.

$\boxed{AB \to C \text{ is Partial FD}}$

iff $A \to C$
or
$B \to C$.

(e) Ssn Pnumber → Ename is Partial FD

Ssn → Ename

$A \in X$

$\boxed{X \to y \text{ is Partial FD.}}$

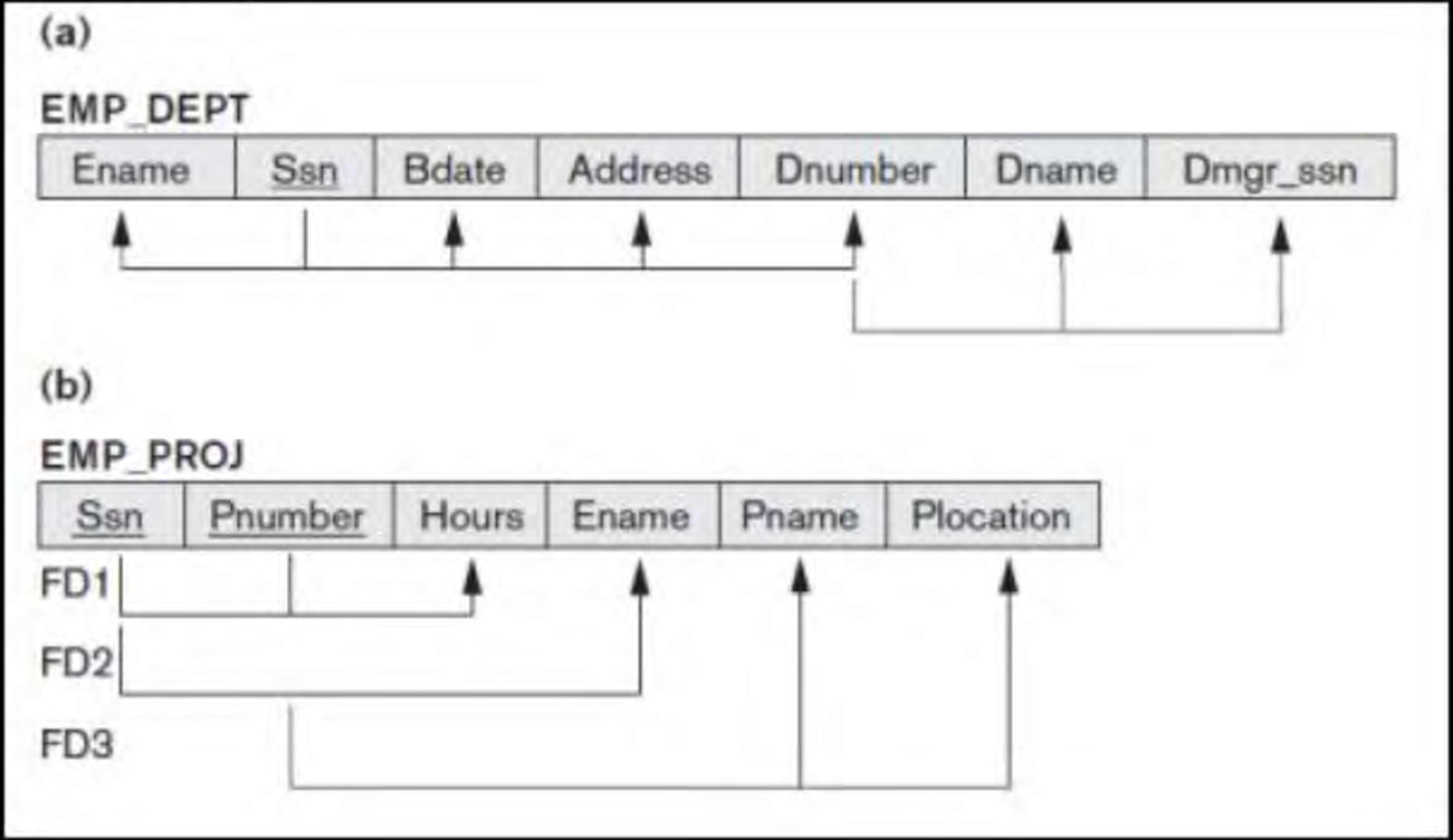
$(X - A) \to y.$

$AB \to C$ is P.F.D

$A \to C$

$B \to C$

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

Korth

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.

$\gamma \subset \alpha$

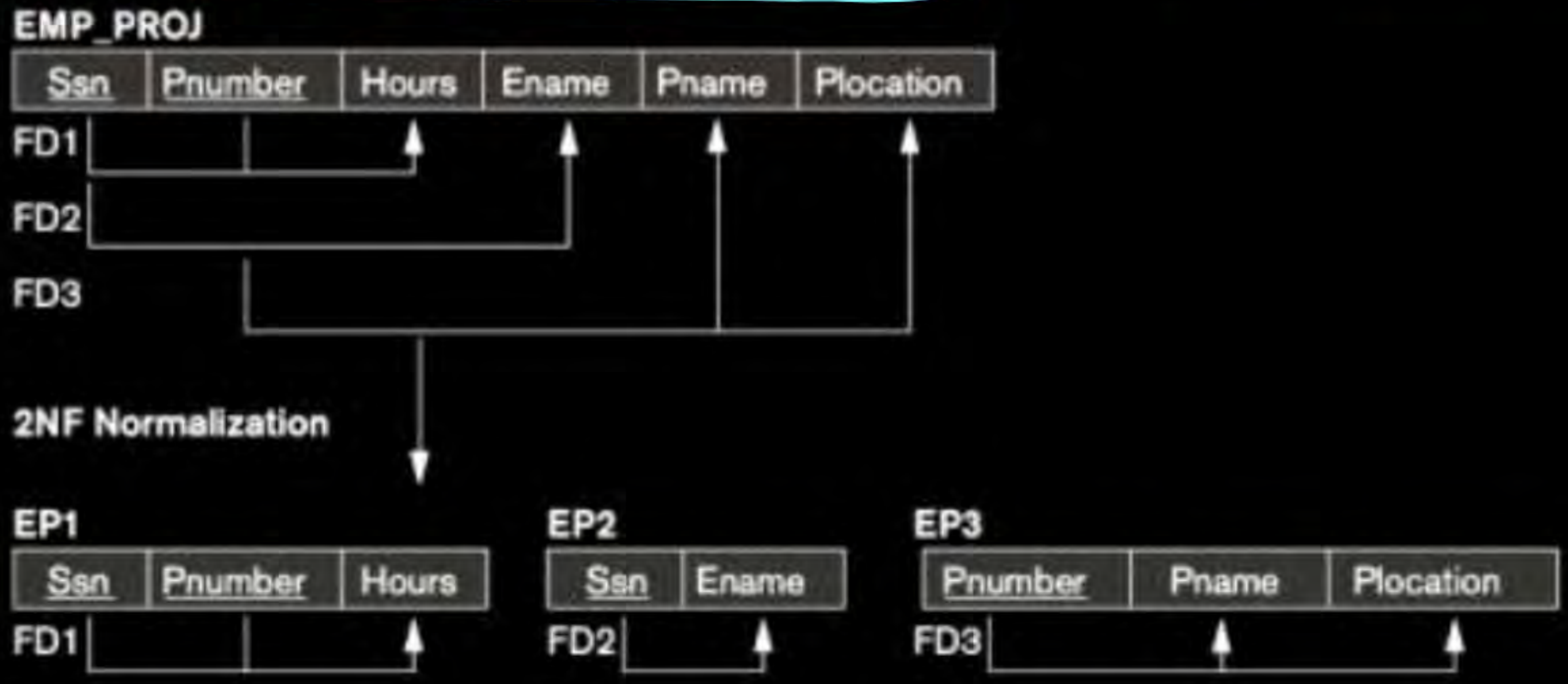
$\boxed{\alpha \rightarrow \beta \text{ is Partial FD}}$

if $\gamma \rightarrow \beta$

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.



Q) Identify which of the following is a Partial FD ?

$\boxed{X \rightarrow Y \text{ is Partial FD}}$

$A \in X$

$(X - A) \rightarrow Y$ Hold

$\boxed{AB \rightarrow C \text{ is Full FD}}$

if $A \not\rightarrow C$

$B \not\rightarrow C$

eg $\boxed{AB \rightarrow C \text{ is Partial FD}}$

if $A \rightarrow C$

OR

$B \rightarrow C$

R(ABCDEF) {ABC → DE, DE → ABC, AB → D, DE → F, E → C}

(i) ABC → D → P

(ii) AE → C → P

(iii) AF → D → F

(iv) AB → D

(v) AC → D

(vi) BC → D

(vii) DE → C

(viii) AB → F

(i) $\boxed{ABC \to D \text{ is Partial FD}}$

$\because$ AB → D

(ii) $\boxed{AE \to C \text{ is Partial FD}}$

$\because$ E → C Hold

(iii) AF → D

$[A]^+ = [A]$ $A \not\to D$

$[F]^+ = [F]$ $F \not\to D$

AF → D is Full FD.

AF → D is Partial FD

iff A → D $[A]^+ = [\ldots D]$

F → D $[F]^+ = [\ldots D]$

$\therefore$

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

(i) $ABC \to D \to P$

(ii) $AE \to C \to P$

(iii) $AF \to D \to F$

(iv) $AB \to D$: F

(v) $AC \to D$: F

(vi) $BC \to D$: F

(vii) $DE \to C$

(vi) $AB \to F$

(iv) $\underline{AB \to D}$

$(A)^+ = (A)$

$(B)^+ = (B)$

$\boxed{AB \to D \text{ is Full FD.}}$

(v) $AC \to D$

$(A)^+ = (A)$

$(C)^+ = (C)$

$\boxed{AC \to D \text{ is Full FD.}}$

(vi) $BC \to D$

$(B)^+ = (B)$

$(C)^+ = (C)$

$\boxed{BC \to D \text{ is Full FD.}}$

R(ABCDEF) {ABC → DE, DE → ABC, AB → D, DE → F, E → C}

(i) ABC → D : P

(ii) AE → C : P

(iii) AF → D : F

(iv) AB → D : F

(v) AC → D : F

(vi) BC → D : F

(vii) DE → C : P

(vii) AB → F : F

(ix) ABE → C

(vii) DE → C

$(D)^+ = [D]$

$(E)^+ = [E\underline{C}]$

E → C

$\boxed{\text{DE → C is Partial FD.}}$

(viii) AB → F

$(A)^+ = [A]$

$(B)^+ = [B]$

$\boxed{\text{AB → F is Full FD.}}$

(ix) ABE → C

$[A]^+ = [A]$

$(B)^+ = (B)$

$(E)^+ = [E\underline{C}]$

E → C

$\boxed{\text{ABE → C is Partial FD.}}$

Condition checking of Normal FORM.

OR

Only

WHAT is the Violation Condition of Normal Form.

second Normal Form. (Condition checking for 2NF)

A Relation Schema R is in 2NF iff

① R is in 1NF

② R does not Contain this below type FD.

2NF → Proper Subset of Candidate key → Non key Attribute — Violation of 2NF.

X

∴ Not in 2NF.

Third Normal Form : [Condition Checking for 3NF]

A Relation Schema R is in 3NF if every

$X \to Y$ Non Trivial FD Must Satisfy the following

Condition.

3NF →

$X \to Y$

X : Super key

OR

Y : key / Prime Attribute

BCNF [Condition checking in BCNF]

A Relational Schema R is in BCNF if every $X \to y$ Non Trivial FD Must satisfy the following Condition

BCNF → $\boxed{\begin{array}{c} X \to y \\ X : \text{Super key} \end{array}}$

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

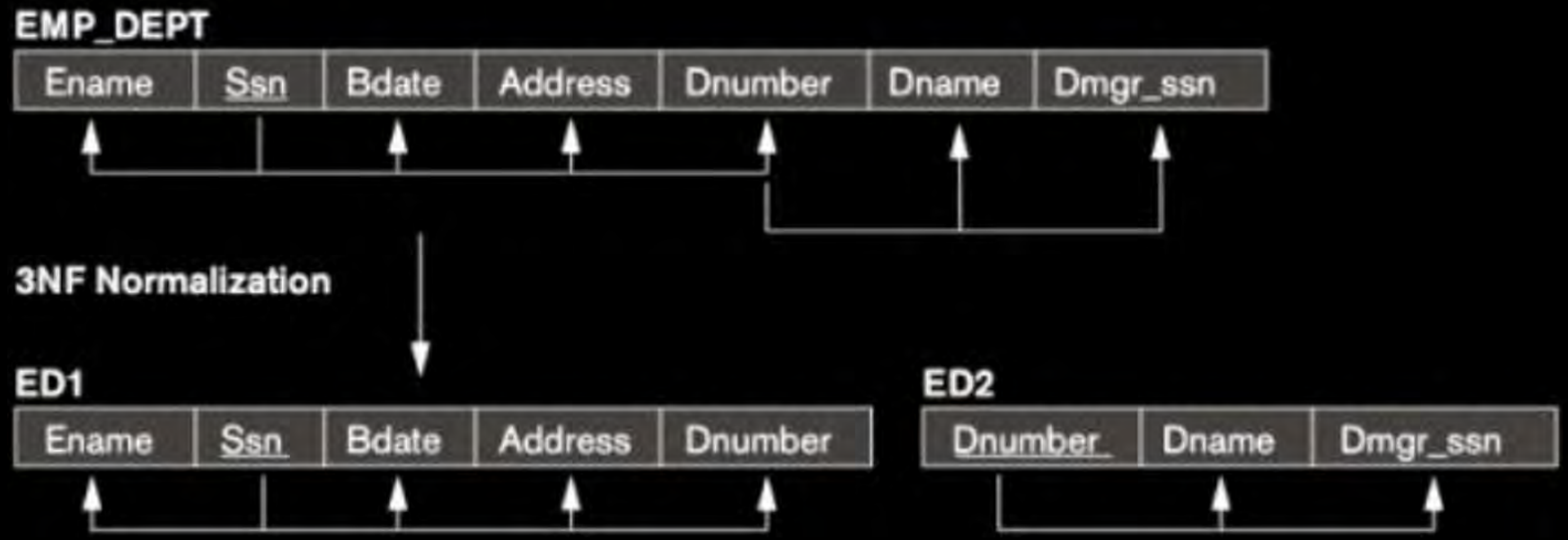
(D) Not in 2NF

Normal Forms

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 $X \rightarrow A$ holds in R either (a) X is a superkey of R, or (b) A is a prime attribute of R.


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.

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

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.

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 $\rightarrow$ courses

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

Non-trivial functional dependencies:

rollno, courseid $\rightarrow$ email

email $\rightarrow$ rollno

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

Non-trivial functional dependencies:

rollno, courseid $\rightarrow$ marks, grade

marks $\rightarrow$ grade

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

Non-trivial functional dependencies:

rollno, courseid $\rightarrow$ credit

courseid $\rightarrow$ 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

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

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

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

[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

Any Doubt ?

THANK YOU!