Normalization

Data Anomalies refers to data incosistency:

- Due to data redundancy
- Improper database designing

Usually 3 types of anomalies:

1. Insertion
2. Deletion
3. Updation

Sno	Cno	Sname	Address	Cname
S21	9201	Jack	Patna	AI
S21	9267	Jack	Patna	DBDM
S24	9267	Sam	Goa	DBDM
S30	9201	Richa	Delhi	AI
S30	9322	Richa	Delhi	Math

Updation Anomaly:

One or more instances of duplicated data is updated, but not all

For example, if S30 updates the address: need to update all instances to avoid updation anomaly

Ritu

Sno	Cno	Sname	Address	Cname
S21	9201	Jack	Patna	AI
S21	9267	Jack	Patna	DBDM
S24	9267	Sam	Goa	DBDM
S30	9201	Richa	Delhi	AI
S30	9322	Richa	Delhi	Math

Consider, what happens if we want to remove the record of S30
cno: 9322 will be completely lost

Deletion Anomaly occurs when:
Certain attr. are lost due to deletion of other attributes

Sno	Cno	Sname	Address	Cname
S21	9201	Jack	Patna	AI
S21	9267	Jack	Patna	DBDM
S24	9267	Sam	Goa	DBDM
S30	9201	Richa	Delhi	AI
S30	9322	Richa	Delhi	Math
S31	9444	XYZ	Kolkata	ML

Insertion anomaly occurs when
certain attr. can't be inserted into DB without presence of other attr.

For example, we can't add a new course unless we have at leat ONE STUDENT ENROLLED on the course (provided NULL insertion is not allowed)

Nornalization:
A process of organizing data in DB to avoid:
- data redundancy,
- ins, del, upd. anomalies

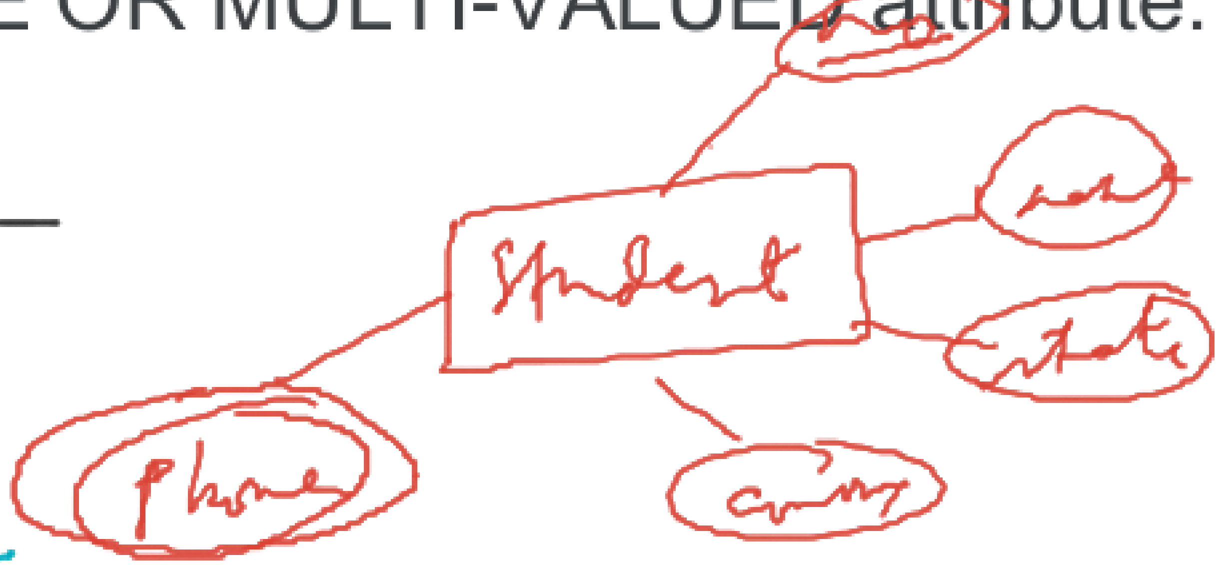
Helps to minimize the redundancy in relation

First Normal Form (1NF)

1

A relation is in 1NF if it ~~DOES~~ NOT contain any COMPOSITE OR MULTI-VALUED attribute.

NO	name	Phone	state	country
1	Ram	98750000, 7003721	Haryana	India
2	Ram	89104...	Bihar	India
3	Suresh	61003...	Punjab	India



not in 1NF

{No} : PK → A

A	B	C	D	E
no	name	Phone	state	country
1	Ram	9873	Har	Ind
1	Ram	70037..	Har	Ind
2	Ram	8910	Bihar	Ind
3	Suresh	6100	Punj	Ind

PK: {no, Phone} → AC

Student (no, name, Phone, State, Country)

r (A, B, C, D, E)

P.K. : {no} → not in 1NF since student can have multiple phones

PK : {no, Phone}

(2) Break the relation in

Student (no, name, state, country)

StudentContact (no, Phone)

no	name	state	country
1	Ram	—	-
2	Ram	—	-
3	Suresh		

no	phone
1	—
~~1~~	~~—~~
2	—
3	—

cname

Partial Dependency (PD): If a non-prime attribute (i.e., an attr. which is not part of ANY C.K) is dependent on any proper subset of any CK of a relation, the depedency is called PD

$\{cno\} \subset \{sid, cno\}$

A rel. is in 2NF:

1. It is in 1NF
2. Rel. must NOT contain any PD.
   or EVERY NON-PRIME (non-key) attribute is fully func. depen. on EACH PART OF CK

cs(sid, cno, cname, marks) || P.K. = {sid, cno}

sid	cno	cname	marks
S1	C1	DBMS	80
S1	C2	AI	70

$cno \longrightarrow cname$

$\Rightarrow$ P.D.

1. There is no MVA, $\Rightarrow$ in 1NF
2. P.D. found $\Rightarrow$ not in 2NF

We need to add a new course:
without student info, we can't add new course

r1

sid	cno	cname	Marks

r2 {cno, cname}

r3 (sid, cno, marks)

r2, r3 are in 1NF ∵ no MVA

r2 cno → cname ⟹ no P.D.

r3 {sid, cno} → marks ⟹ no P.D.

⟹ in 2NF

3NF

Transitive Dependency (TD): If $A \rightarrow B$ and $B \rightarrow C$ are two FDs then $A \rightarrow C$ is TD

A relation is in 3NF

1) the relⁿ is in 2NF

2) the relⁿ Must NOT contain any TD for non-prime attr.

OR

Every FD $\boxed{\alpha \rightarrow \beta}$ is in 3NF, it should satisfy either of the following

2. i) It is trivial dependency $(A \rightarrow A)$

or ii) α is a SK

roll no. $\rightarrow$ roll no.

or iii) β is a prime attribute (each element of β is part of some C.K) r1

Student (rno, name, state, contry, age)
R N S C A

in 2NF (check)

$FD_s\{\underbrace{R \to N}_{FD_1},\ \underbrace{R \to S}_{FD_2},\ \underbrace{R \to A}_{FD_3},\ \underbrace{S \to C}_{FD_4}\}$

r1 — X r2

CK: {R}

r2 (R, N, A, S)

r3 (S, C)

$S \to C$

$\alpha \to \beta$

$R \to S$

$S \to C$

$\Rightarrow$ T.D for non-prime attr.

C is tr. dep. on R

$\Rightarrow$ not in 3NF

FD1 : α is a S.K

FD_2 : "

FD_3 : "

$\Rightarrow$ not in 3NF

FD_4 : not trivial dep.
α is not SK
β is not prime attr.

1NF	2NF	3NF
no MVA	i) 1NF ii) no PD	i) 2NF ii) no T.D. of non-prime attr.

1NF < 2NF < 3NF < BCNF

Boyce Codd Normal Form (BCNF)

1) in 3NF
2) Every FD $\alpha \longrightarrow \beta$ in a relⁿ has to satisfy EITHER of the following
   i) It is trivial dependency
   ii) α is a S.K.

$R(\overset{S}{Sno}, \overset{C}{Cno}, \overset{T}{Tno})$ or $R(S, C, T)$

Studentno. Courseno Teacherno

FDs: $\{SC \rightarrow T,\ T \rightarrow C\}$

$T \rightarrow T$

key: $\{SC\}, \{ST\}$

1) R is in 3NF (checked)

$(SC)^+ = \{S, C, T\}$

$\cancel{(ST)^+ = \{S, T, C\}}$

2) i) ✓ ✗

ii)

$\alpha \rightarrow \beta$

$T \rightarrow C$

not in BCNF

NO

α is not S.K.

Decomposition: To convert R into BCNF, we have to decompose R

The decomposition should be LOSSLESS.

$R(S, C, T)$

$FD\{SC \rightarrow T,\ T \rightarrow C\}$

$C.K. = \{SC, ST\}$

1 $R_1 = \{S, C\}, R_2 = \{C, T\}$

OR

2 $R_3 = \{S, T\}, R_4 = \{T, C\}$

H.W.

Dependency Preserving Decomposition:

A decomposition $D = \{R_1, R_2, \cdots, R_n\}$ of a relation R is dep. pres. w.r.t. a FD set F of R if

$(F_1 \cup F_2 \cup \cdots \cup F_n)^+ = F^+$, where F_i's is FD set of R_i

F+ denotes closure of F, i.e., the set of all FDs can be derived from F

$FD \cong F$

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

$FDs: \{A \to B, C \to B, C \to D, B \to D\}$

The decomposition of R into $R_1(A, B)$, $R_2(B, C)$, $R_3(B, D)$

$FD_1: \{A \to B\}$, $FD_2 = \{C \to B\}$, $FD_3 = \{B \to D\}$

$$(FD_1 \cup FD_2 \cup FD_3)^+ = \{A \to B, \underbrace{C \to B, B \to D}_{C \to D}\}$$

$$= (FD)^+$$

Lossless Decomposition (LLD)

A decomposition of R into R_1 and R_2 is LLD w.r.t. a FD set F of R if $R_1 \bowtie R_2 = R$.

In other words, $R_1 \cap R_2 \Rightarrow R_1$ ✓

or $\underline{R_1 \cap R_2 \Rightarrow R_2}$ ✓

example

$R(A, B, C)$ has $F: \{A \to B, B \to C\}$

decomposition $R_1(A, B)$, $R_2(B, C)$

$F_1 := \{A \to B\}$, $F_2 : \{B \to C\}$

$R_1 \cap R_2 = \{B\} \Rightarrow R_2$

$\Rightarrow$ The decomp. is lossless

$B \to C$