Module - 3

Normalization : It is a technique to remove or reduce redundancy in a table.

Insertion Anomaly : It occurs when certain attributes cannot be inserted into the database due to missing additional data.

Ex. If a new course is introduced in University table. The student Id cannot remain empty as it is primary key.

Deletion Anomaly : It occurs when you delete a record that may contain attributes that should not be deleted.

ex. if we remove the information of the only student in a department, all of the department information vanishes.

Updation Anomaly : It occurs when the same data is repeated in multiple rows, and changes are made in some but not all instances.

ex. If an employees address changes & the update is made in one row but not in others, the database will contain inconsistent data

ow normalization helps in these cases?

It divides the original table into multiple tables from it.

irst Normal Form:

→ Table should not contain any Multivalued Attributes.

Ex

Rollno.	Name	course
1	Saw	C/C++
2	Harsh	Java
3	Omkar	DBMS, C/C++

multivalued

← Not in INF

INF form

Rollno.	Name	course
1	Saw	C
1	Saw	C++
2	Harsh	Java
3	Omkar	C
3	Omkar	DBMS

Primary key → (Rollno course)

Composite Primary key

or

Rollno.	Name	Cousel	Cous2
1	Saw	C	C++
2	Harsh	Java	Null
3	Omkar	C	DBMS

Dont do this as too many nulls

2 NF form

↓

Rollno.	Name
1	Saw
2	Harsh
3	Omkar

PK ⇒ Rollno.

Rollno.	course
1	C
1	C++
2	Java
3	C
3	DBMS

PK ⇒ ROLLNO. DBMS

FK ⇒ Rollno.

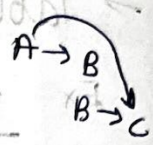
Functional dependency

→ Helps Find all the candidate keys of table

Ex. Table R with Attributes A B C D

R(ABCD)

FD { A → B , B → C, C → D }



∴ A^+ = BCDA

∴ A is candidate key

B^+ = BCD

B is not candidate key

C^+ = DC

Not CK

D^+ = D

Not CK.

$(AB)^+$ = ABCD

AB not CK as it is not minimal it is a super key

CK = {A}

Non PA = { }

PA = {A}

Ex. 2 > R(ABCD)

FD = { A → B , B → C, C → D, D → A }

A^+ = ABCD

B^+ = BCDA

C^+ = CDAB

D^+ = DABC

} All are CK

→ Prime attribute is the attribute which is used in making CK.

⇒ { A, B, C, D }

Non-PA ⇒ { }

$FD = \{A \to B, BC \to D, E \to C, D \to A\}$

Right Attributes (Elements after '→')

$\Rightarrow (B, D, C, A)$

The element which is not on right side, will be used for making candidate key.

$\therefore E = BDCAE$

$E^+ = EC$

$AE = ABECD$

$BE^+ = BECDA$
$CE^+ = CEBDA$
$DE^+ = DECBA$
} They all can be CK.

$\therefore (AE)$ is the candidate key.

So, to find their CK, in short

$CK = \{AE\}$ ⟶ check from this is 'A' or 'E' is present on the RHS of FD. If yes replace them.

$DE^+ = DEABC$

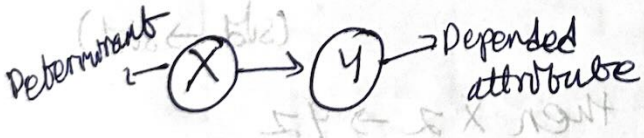
$BE^+ = BECDA$

$CE^+ = CEAX$

$CK = \{AE^*, DE^*, BE^*\}$

Prime Attributes $= \{A, D, B, E\}$

Non prime $= \{C\}$



X determines Y

or

Y is determined by X

✓ Sid → Sname

Sid	Sname
1	Ranjit
1	Ranjit

✓ Sid → Sname

Sid	Sname
1	Ranjit
2	Ranjit

✓ Sid → Sname

Sid	Sname
1	Ranjit
2	Varun

✗ Sid → Sname

Sid	Sname
1	Ranjit
1	Varun

✓ - valid

~~Reflexivity~~

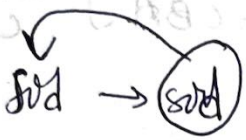
Trivial FD:

if

Y is subset of X

* The value to be determined is already a subset of determinant.

ex



Non-Trivial

$X \rightarrow Y \quad X \cap Y \neq \phi$ (Y is not a subset of X)

Fid → Fname
Fid → location

sid → sname
sid → semester
sid → phone

Properties of FD:

Reflexivity: If Y is subset of X, then X → Y
(sid → sid)

Augmentation: If X → Y, then XZ → YZ
(sid → sidname, sid phone → sname phone)

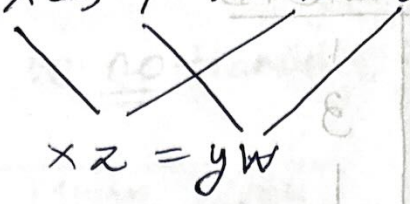
Transitive: if X → Y and Y → Z then X → Z
(sid → sname) & (sname → city)
[sid → city]

Union: If X → Y & X → Z then X → YZ

Decomposition: if X → YZ then X → Y & X → Z

Pseudo transitivity: If $x \to y$ & $wy \to z$ then $wx \to z$

composition: if $X \to Y$ & $Z \to W$ then $XZ \to YW$



→ Never break LHS //

2nd Normal form

→ Table or relation must be in 1st Normal form

→ All the non-prime attributes should be fully functional dependent on candidate key.

customer

customerID	storeID	Location
1	1	Delhi
1	3	Mumbai
2	1	Delhi
3	2	Banglore
4	3	Mumbai

Candidate key : customerID storeID

PA → CID, SID

NPA → Location

To make it into 2NF:

customer ID	store ID
1	1
1	3
2	1
3	2
4	3

CK ⇒ customer ID store ID

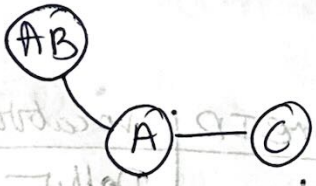
Store ID	Location
1	Delhi
2	Banglore
3	Mumbai

CK → Store ID

Hence both are in 2NF.

→ There should be no partial dependency on the relation

ex) Part of a candidate key, is determining another non-prime attribute 'C'.



R(A B C D E F)

FD = {C → F, E → A, EC → D, A → B}

CK: EC = FADB

EC⁺ = ECFADB

EC = candidate key

- Prime attributes ⇒ {E, C}
- non " " ⇒ {A, B, D, F}

This table is not in 2NF form as EC → F, E → A

3rd Normal Form

→ Table must be in 2nd Normal form

→ There should be no transitive dependency in table.

Roll no.	State	City
1	Pun JAB	Mohali
2	Haryana	Ambala
3	Punjab	mohali
4	Haryana	Ambala
5	Bihar	Patna

CK = Roll no.

FD: Roll no → state

state → city

PA ⇒ {Roll no}

NPA ⇒ {state, city}

R(A B C D)

FD: AB → CD, D → A

AB^+ = ABCD

DB^+ = DBAC

∴ CK ⇒ {AB, DB}

PA ⇒ {A, B, D}

NPA ⇒ {C}

To check for transitive:

for each FD ⇒ LHS must be CK or SK or RHS is a prime attribute

In the above example

FD: AB → CD, D → A ← RHS is a PA

LHS is a CK → valid FD

D is Not a CK, OR this is valid

∴ This table is in 3NF

Transitive occurs, only when A non-prime attribute is pointing to another NPA.

BCNF (Boyce Codd Normal Form)

→ It should be in 3NF

→ All the LHS in FD, must be a CK./super Key.

student

Rollno.	Name	VoterID	age
1	Ravi	K0123	20
2	Varun	M034	21
3	Ravi	K786	23
4	Rahul	B432	21

CK ⇒ {Rollno, voterID}

FD ⇒ { Rollno → Name
Roll no → voterID
voterID → age
voterID → name }

∴ Two table is in BCNF form

- Third normal form always ensures 'Dependency poeserving decomposition' but not in BCNF.
- Both third & BCNF ensure lossless decomposition

4th Normal Form

→ It should be in BCNF

– It should not contain Non-trivial multivalued dependency

Exc.

Course	Instructor	Textbook
Management	white	Drucker
	green	Peter
	Black	
Finance	Gray	weston
		Gilford

Course →→ Instructor
Course →→ Textbook

→ The above table is not in 4NF, as it has multiple values for the same column.
i.e course →→ Instructor etc.

Solution:

Course	Instructor
Mangement	white
Management	green
Management	Black
Finance	Gray

&

Course	Textbook
Management	Drucker
Management	Peter
Finance	weston
Finance	Gilford

→ 4NF is a level of database normalisation where there are non-trivial multivalued dependencies other than CK.

Database design & Implementation

1) Relational model concepts

Relations : Tables containing rows & columns

Tuples : Rows in a table, where each row represents a specific instance or data entry.

Attribute : Columns in a table, defining the properties of the data

Primary key : An attribute (or set of attributes) that uniquely identifies each tuple in a relation.

Domain : Defines the set of valid values for an attribute.

Schema : Describes the structure of relation, including its attributes & their domains

Relational Model Constraints:

Domain Constraints: Ensures that values in each attribute adhere to their defined domain

Key constraint: Ensures that each tuple can be uniquely identified by using the primary key.

Entity integrity: Requires that primary key attributes cannot contain null values.

Referential integrity: Ensures that relationships between tables are maintained, typically involving foreign keys referencing primary keys in other tables.

Referential Database Schema:

Database Schema: A set of relation schemas that make up a database.

Foreign key: An attribute in one table that references the primary key in another

Referential integrity constraint: Ensures that the foreign key values in one table are linked /referenced to another tables PK.

Semantic integrity constraint: Additional constraints based on application semantics, not directly expressible in model

Relational update operations:

Insertion
Deletion
Modification
} In tuples

Dealing with constraint violation:

Cancellation: Rejecting an operation would violate a constraint

Notification: Informing the user of the violation without performing the operation.

Correction: Automatically correcting violations, such as setting null values or propagating changes to maintain integrity.