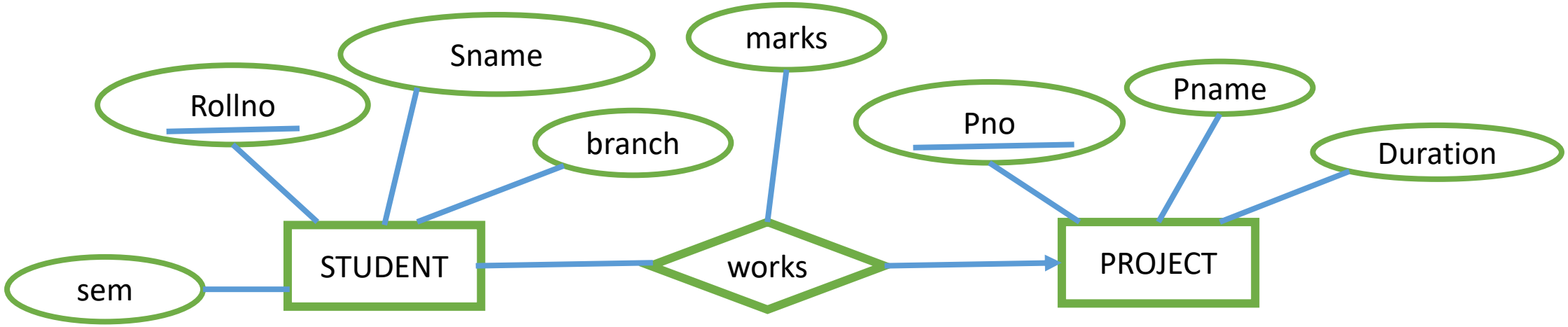
UNIT 2

Lecture 29

Normalization

Conversion of E R Diagram to Relational Model



STUDENT

Rollno	Sname	Sem	Branch	Marks	Pno
1	RAM	3	CSE	40	121
2	SHYAM	5	CSE	50	122
3	MOHAN	7	CSE	55	121
4	GOPAL	5	IT	65	121
5	RINKI	3	MECH	40	122
6	PINKI	3	ETC	90	123

PROJECT

Pno	Pname	Duration
121	P1	10
122	P2	20
123	P3	30

Insertion, Deletion and Update Anomalies

STUDENTPROJECT

Rollno	Sname	Sem	Branch	Marks	Pno	Pname	duration
1	RAM	3	CSE	40	121	P1	10
2	SHYAM	5	CSE	50	122	P2	20
3	MOHAN	7	CSE	55	121	P1	10
4	GOPAL	5	IT	65	121	P1	10
5	RINKI	3	MECH	40	122	P2	20
6	PINKI	3	ETC	90	123	P3	30

- **Insertion anomaly :** We cannot insert new project without a student.
- **Delete Anomaly :** If we delete a student then if that student is the only student who work for that project, then that project information is lost.
- **Update Anomaly :** If we update a project name then we have to update all the tuples that contain that project.

Normalization

- Normalization is the process of taking data from a problem and reducing it to a set of relations while *ensuring data integrity, eliminating data redundancy and minimizing the insertion, deletion and update anomalies.*
  - **Data Integrity** – All of the data in the database are consistent and satisfy all integrity constraints.
  - **Data redundancy** – If data in the database can be found in two different locations (*direct redundancy*) or if data can be calculated from other data items (*indirect redundancy*) then the data is said to contain redundancy.

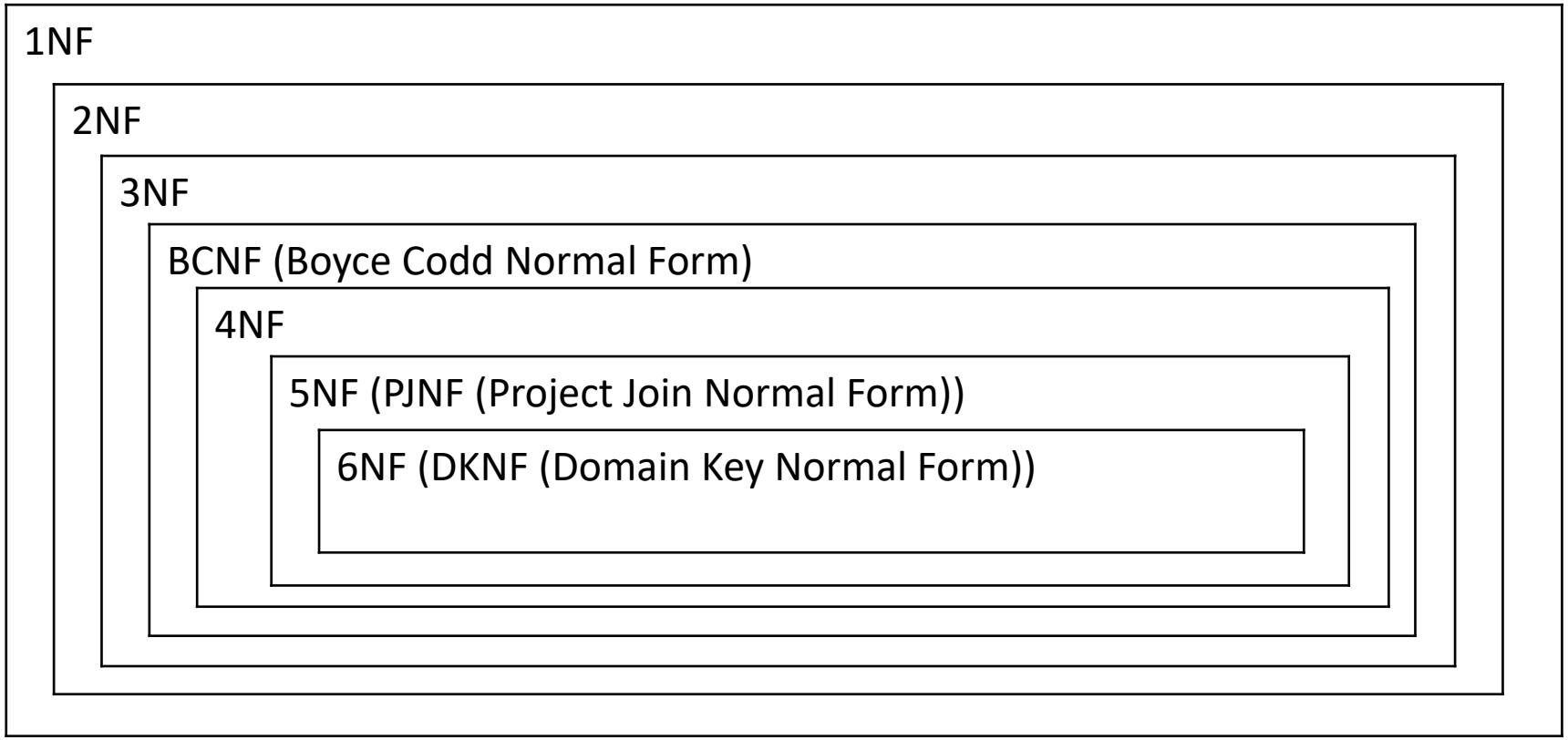
Goals of Normalization

1. Eliminate certain kinds of redundancy.
2. Avoid certain update anomalies.
3. Good representation of real world.
4. Simplify enforcement of Database Integrity.

Objective of Normalization

- *"to create relations where every dependency is on the key, the whole key, and nothing but the key".*

Levels of Normalization



Functional Dependency

A functional dependency, denoted by X→Y, between two sets of attributes X and Y that are subsets of R specifies a constraint on the possible tuples that can form a relation state r of R. The constraint is that, for any two tuples t_1 and t_2 in r that have $t_1[X] = t_2[X]$, they must also have $t_1[Y] = t_2[Y]$.

This means that the values of the Y component of a tuple in r depend on, or are ***determined by***, the values of X component; alternatively, the values of the X component of a tuple uniquely (or **functionally**) determine the values of the Y component. We also say that there is a functional dependency from X to Y, or that Y is functionally dependent on X.

Functional Dependency

Thus, X functionally determines Y in a relation schema R if, and only if, whenever two tuples of r(R) agree on their X value, they must necessarily agree on their Y value. Note the following:

1. If a constraint on R states that there cannot be more than one tuple with a given X value in any relation instance r(R), i.e. X is a **candidate key** of R – This implies that X→Y for any subset of attributes Y of R (because the key constraint implies that no two tuples in any legal state r(R) will have the same value of X).
2. If X→Y in R, this does not say whether or not Y→X in R.

Functional Dependency

STUDENT

Rollno	Sname	Sem	Branch	Marks	Pno
1	RAM	3	CSE	40	121
2	SHYAM	5	CSE	50	122
3	MOHAN	7	CSE	55	123
4	GOPAL	5	IT	65	121
5	RINKI	3	MECH	40	122
6	PINKI	3	ETC	90	123

If Rollno is primary key then, we can write

Rollno → Sname,

Rollno → Sem,

Rollno → Branch,

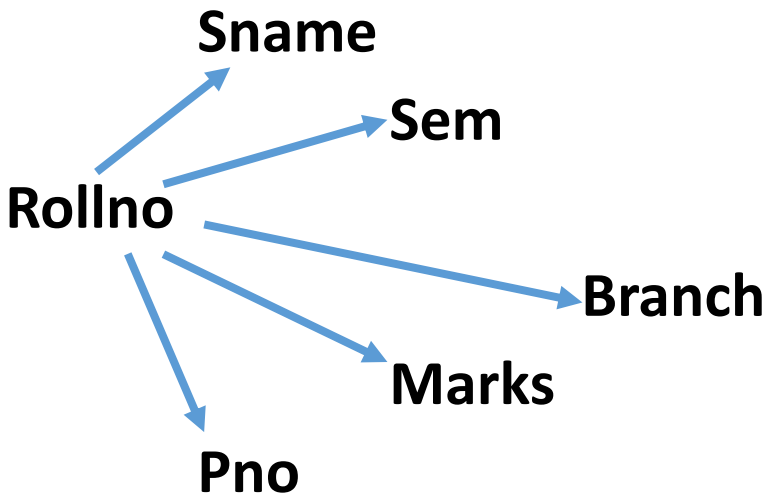
Rollno → Marks,

Rollno → Pno

4 ways to depict Functional Dependency

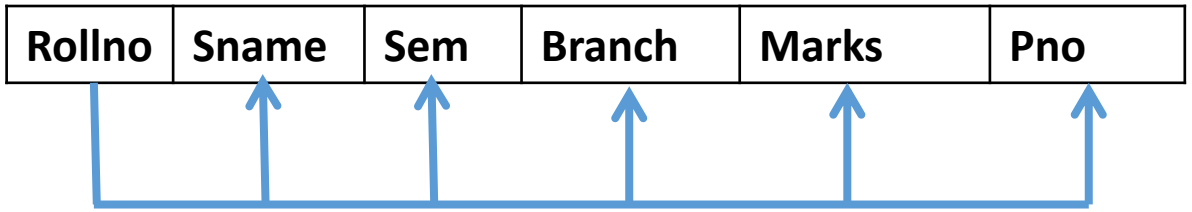
If Rollno is primary key then, we can write

1. Rollno → Sname,
   Rollno → Sem,
   Rollno → Branch,
   Rollno → Marks,
   Rollno → Pno

2. 

3. Rollno → {Sname, Sem, Branch, Marks, Pno}

4. Use a dependency diagram



Functional Dependency

R

A	B	C	D
1	1	2	4
2	3	4	5
3	4	2	4
4	5	5	7
5	3	1	5

A → B	B → A	C → A	D → A
A → C	B → C	C → B	D → B
A → D	B → D	C → D	D → C
AB → C	AC → B	AD → B	BC → D
AB → D	AC → D	AD → C	BC → A
BD → A	CD → A	ABC → D	BCD → A
BD → C	CD → B	ACD → B	

Super Keys of R can be : A, AB, AC, AD, BC, ABC, BCD, ACD, ABCD = 9

Candidate keys of R can be : A, BC = 2

Determinant

- In Functional Dependency

$$A \rightarrow B$$

Attribute on the L.H.S. of FD is known as ***determinant.***

- A is a determinant of B.
- A is a determinant of B if each value of A has precisely one (possibly null) associated value of B.

OR

- A is a determinant of B if and only if whenever two tuples agree on their A value they agree on their B value.

$$A \rightarrow B.$$

Inference Rules for functional dependencies

1. Reflexive Rule

   If $X \supset Y$ then $X \rightarrow Y$

2. Augmentation Rule

   If $X \rightarrow Y$ then $XZ \rightarrow YZ$

3. Transitive Rule

   If $X \rightarrow Y$ and $Y \rightarrow Z$ then $X \rightarrow Z$

4. Decomposition or Projective Rule

   If $X \rightarrow YZ$ then $X \rightarrow Y$, and $X \rightarrow Z$

5. Union or Additive Rule

   If $X \rightarrow Y$ and $X \rightarrow Z$ then $X \rightarrow YZ$

6. Pseudo transitive Rule

   If $X \rightarrow Y$ and $YZ \rightarrow W$ then $XZ \rightarrow W$

GATE Questions

Given the following relation instance.

X Y Z

1 4 2

1 5 3

1 6 3

3 2 2

Which of the following functional dependencies are satisfied by the instance?

(A) XY -> Z and Z -> Y

(B) YZ -> X and Y -> Z

(C) YZ -> X and X -> Z

(D) XZ -> Y and Y -> X

[GATE CS 2000]

GATE Questions

From the following instance of a relation schema R(A, B, C) we can conclude that :

A	B	C
1	1	1
1	1	0
2	3	2
2	3	2

(A) A functionally determine B and B functionally determine C.

(B) A functionally determine B and B does not functionally determine C.

(C) B does not functionally determine C.

(D) A does not functionally determine B and B does not functionally determine C.

[GATE 2002]

GATE Questions

Given an instance of the STUDENTS relation as shown below :

StudentD	StudentName	StudentEmail	StudentAge	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 (StudentName, StudentAge) to be a key for this instance, the value X should NOT be equal to_____.

[GATE 2014: 1 Mark]

For Video lecture on this topic please subscribe to my youtube channel.

The link for my youtube channel is

https://www.youtube.com/channel/UCRWGtE76JlTp1iim6aOTRuw?sub_confirmation=1