Normalization

What is Normalizarion? The technique to minimize **data redundancy**.

What is data redundancy? Duplicated data is present in many tables which is not required.

Why is it important to minimize? Eats up a lot of space and creates issues: Insertion, Deletion, and Updation these are normally called anormalise.

Problem :

rollno	name	branch	hod	office_tel
1	Akon	CSE	Mr. X	53337
2	Bkon	CSE	Mr. X	53337
3	Ckon	CSE	Mr. X	53337
4	Dkon	CSE	Mr. X	53337

Issues :

1. Insertion Anomaly: while inserting data every time we need to provide this repeated data which is not required and is just a waste of time and space.

2. Deletion Anomaly: Now while deleting data we are not only deleting the student information but also deleting the branch information.As soon as all records are deleted all student is deleted but along with that branch info is also deleted.

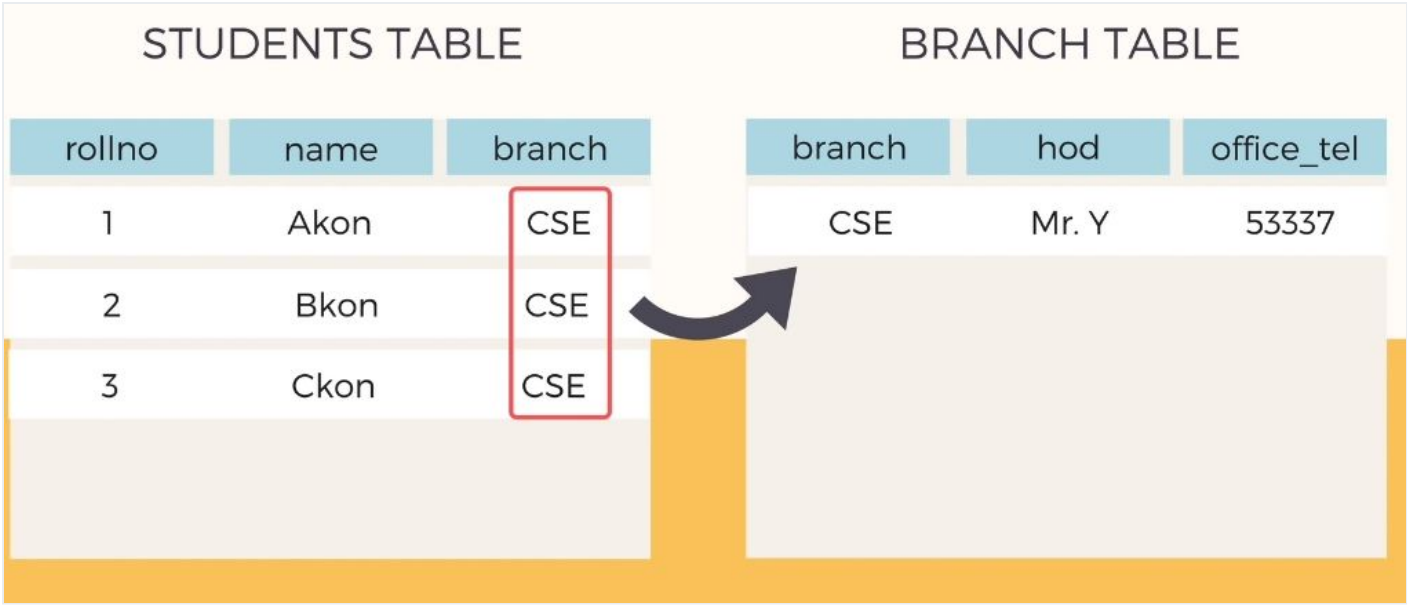
3. Updation Anomaly: If one of the HODs is changed in a department then we need to update all the records taking into care that the student belongs to that department only.

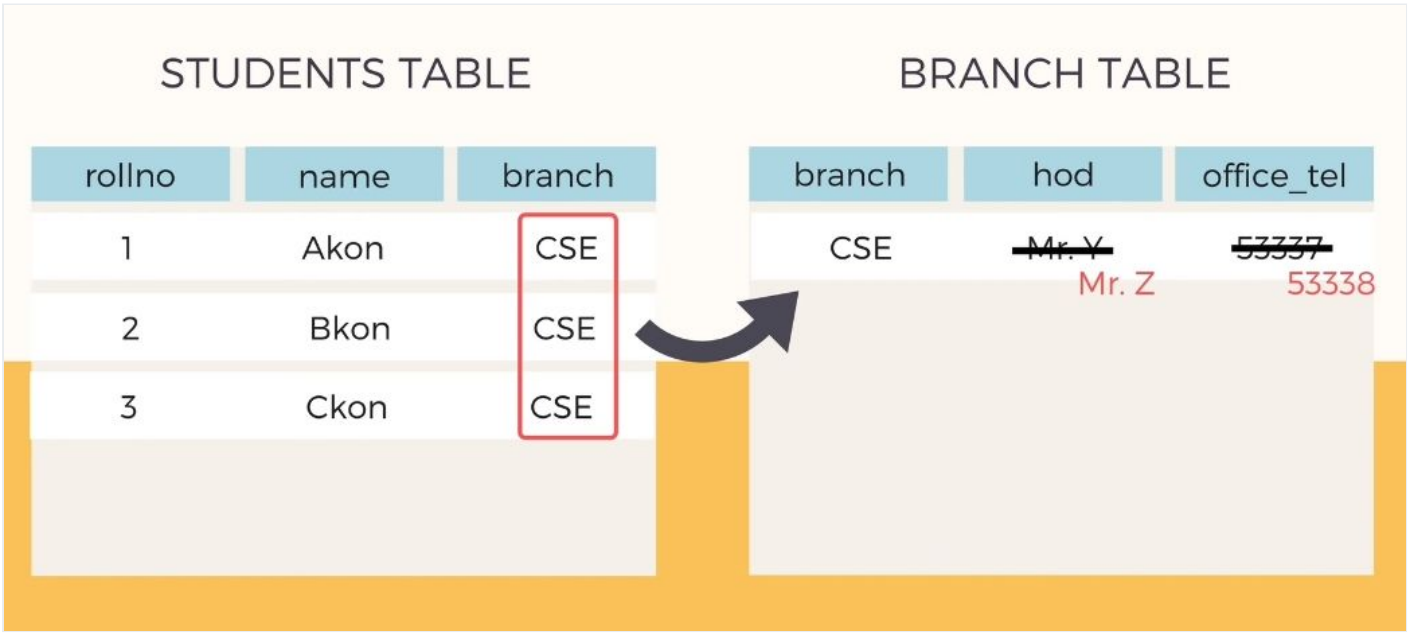
Howwill Normalisation solve this problem:

1. Split the table into two tables : student table and branch table

2. Student will only have roll no , name , branch

3. Branch will have branch , hod name and office_tel





You might think that still branch name is still repeated but it has minimized the data redundancy to a large extent.

Normalization is not about eliminating data redundancy but about reducing data redundancy.

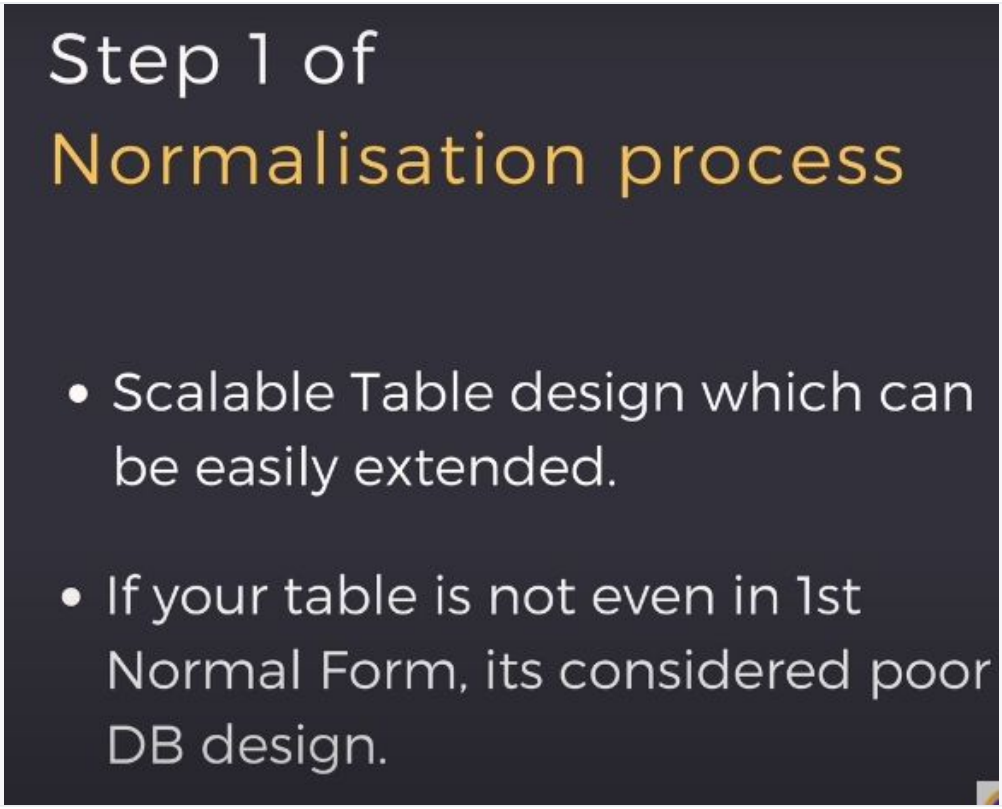
Problem is solved as

1. If we want to insert data only important or relevant data is provided and the branch info is just saved once.
2. Deletion can also be done by the student name without affecting the branch infomation
3. Updating of branch info will be in one place and to only one record which leads to fewer mistakes and less space is used up.

Types of Normalization

1. 1st Normal Form: simple teams the step 1 of normalization .

1NF


Step 1 of
Normalisation process
• Scalable Table design which can be easily extended.
• If your table is not even in 1st Normal Form, its considered poor DB design.

4 rules :

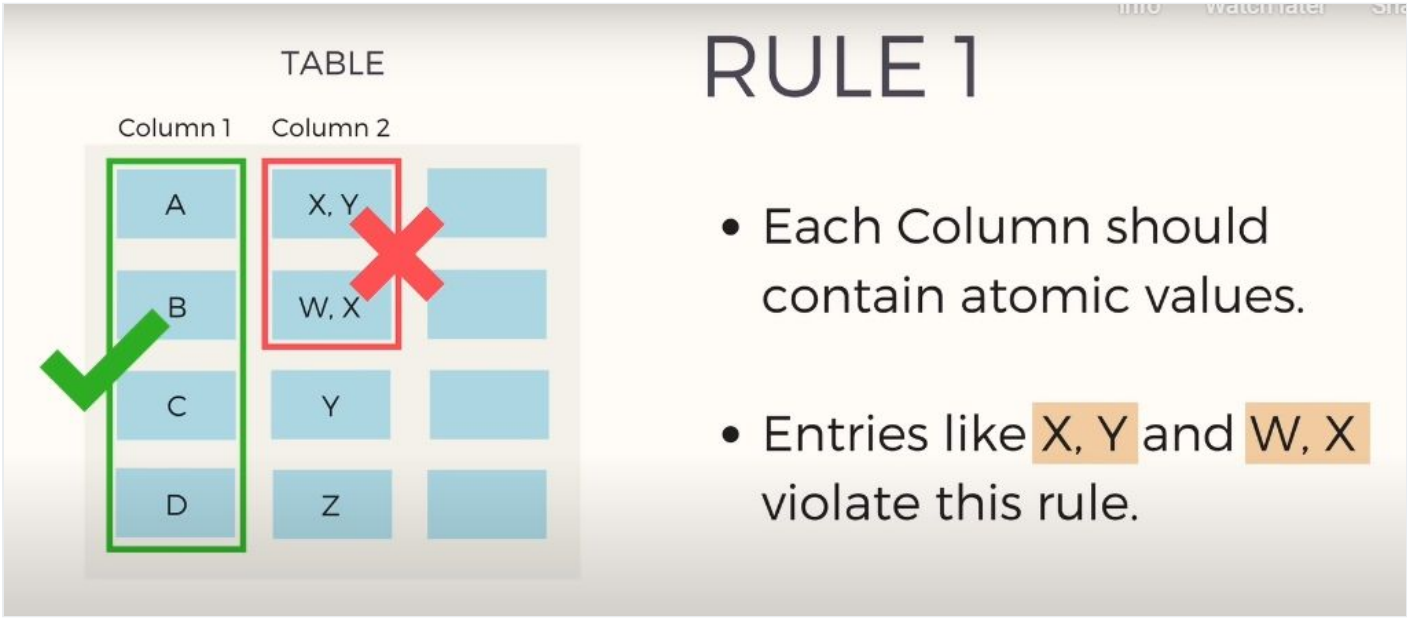


TABLE
Column 1
Column 2
A
B
C
D
X, Y
W, X
Y
Z
RULE 1
• Each Column should contain atomic values.
• Entries like X, Y and W, X violate this rule.

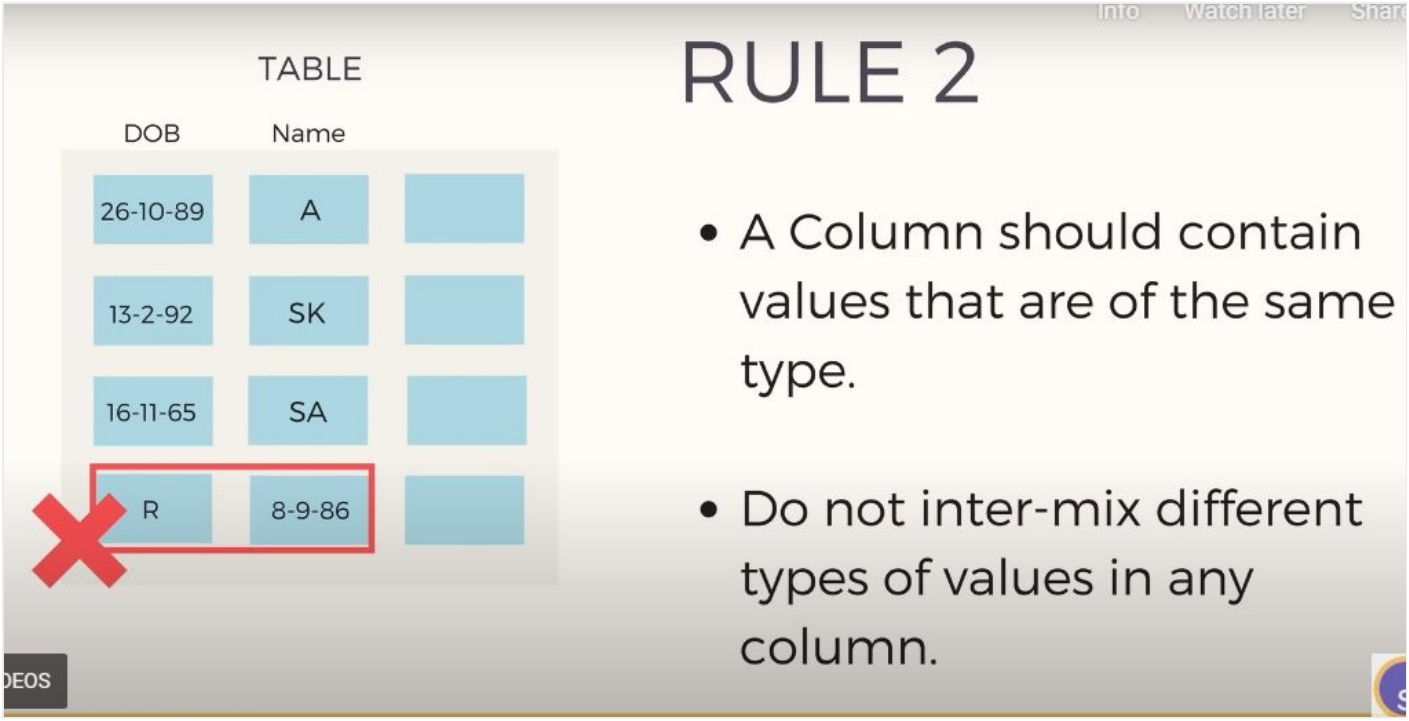


TABLE
DOB
Name
26-10-89
A
13-2-92
SK
16-11-65
SA
R
8-9-86
RULE 2
• A Column should contain values that are of the same type.
• Do not inter-mix different types of values in any column.

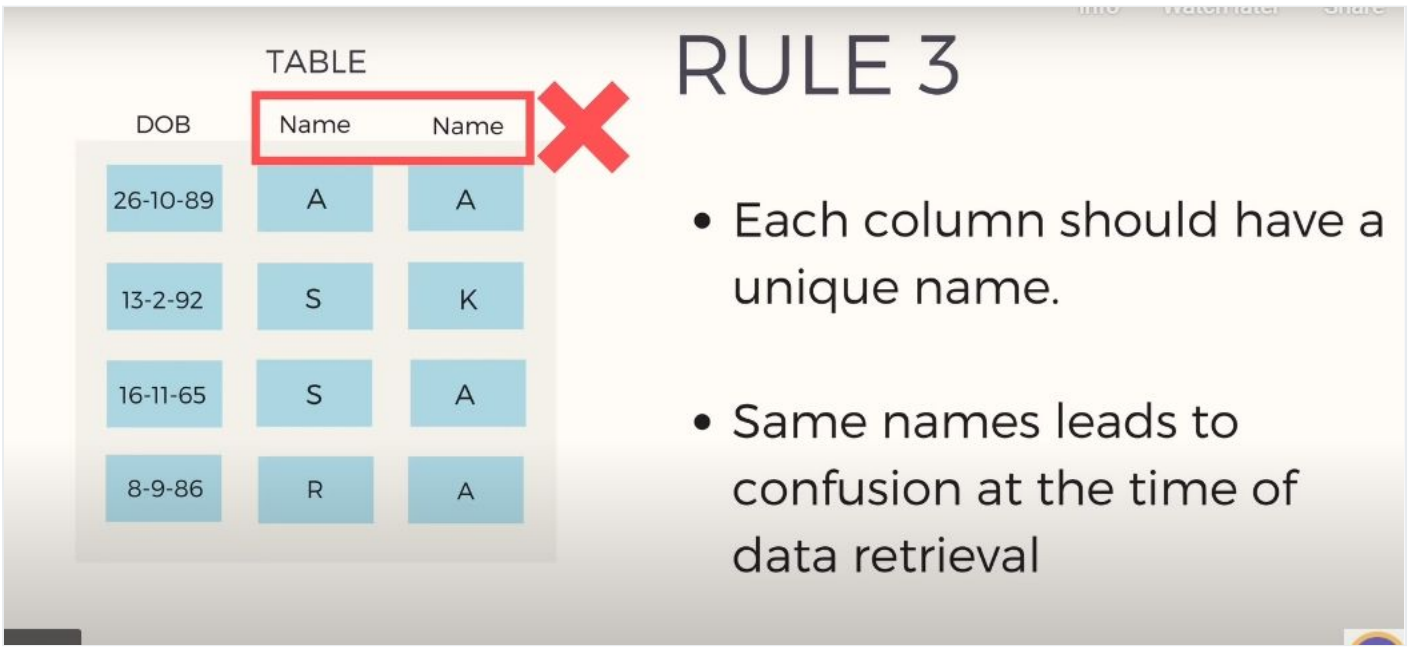


TABLE
DOB
Name
Name
26-10-89
A
A
13-2-92
S
K
16-11-65
S
A
8-9-86
R
A
RULE 3
• Each column should have a unique name.
• Same names leads to confusion at the time of data retrieval

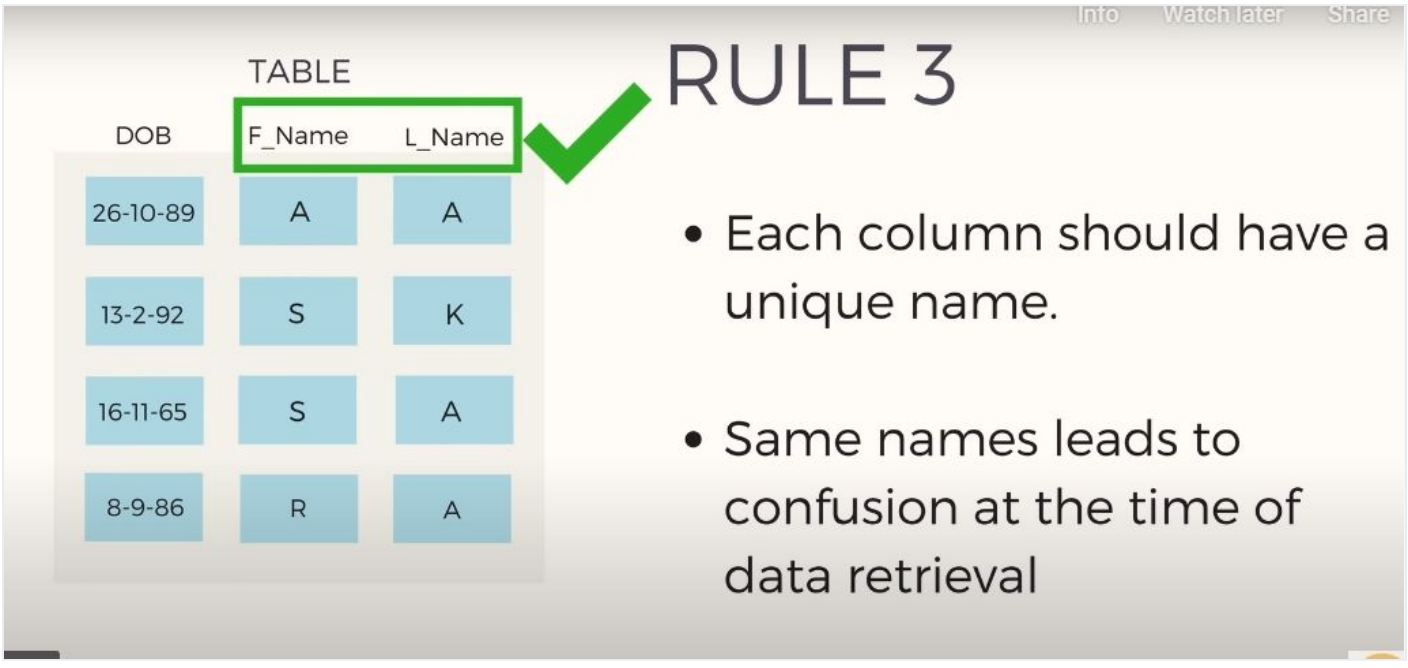


TABLE
DOB
F_Name
L_Name
26-10-89
A
A
13-2-92
S
K
16-11-65
S
A
8-9-86
R
A
RULE 3
• Each column should have a unique name.
• Same names leads to confusion at the time of data retrieval

TABLE

Roll_no	F_Name	L_Name
3	A	A
4	S	K
1	S	A
2	R	A

RULE 4

- Order in which data is saved doesn't matter.
- Using SQL query, you can easily fetch data in any order from a table.

STUDENTS TABLE

rollno	name	subject
101	Akon	OS, CN
103	Ckon	JAVA
102	Bkon	C, C++

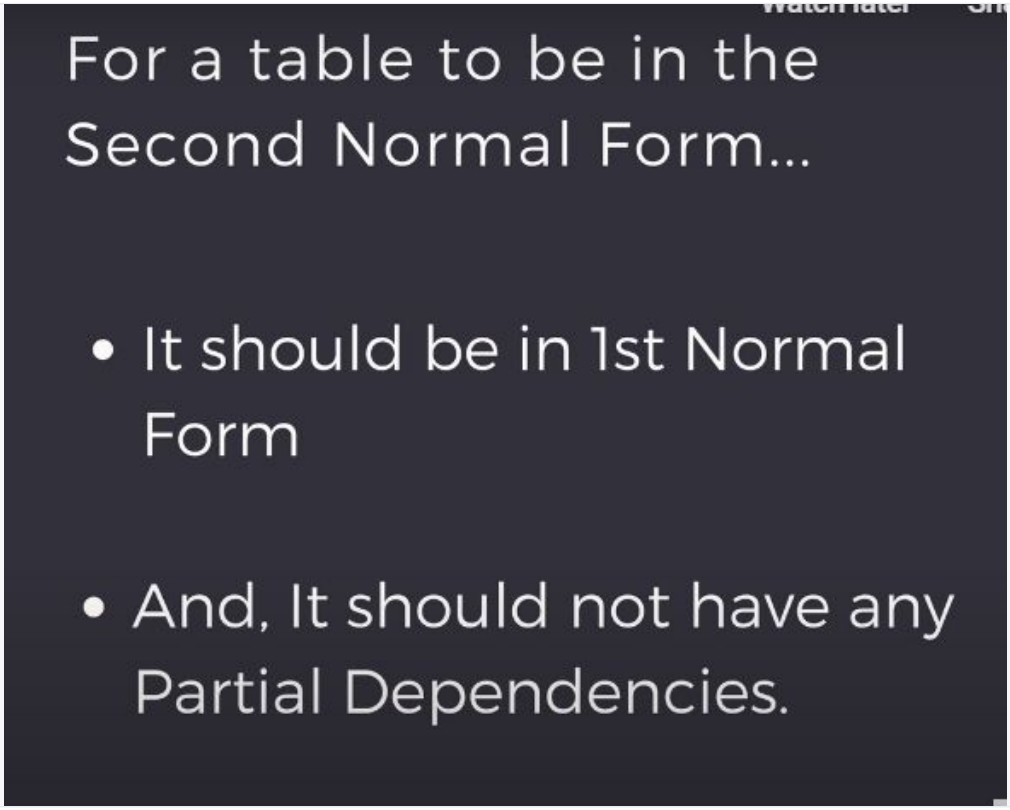
Violation of 1 NF

BY 1NF the final result is :

STUDENTS TABLE

rollno	name	subject
101	Akon	OS
101	Akon	CN
103	Ckon	JAVA
102	Bkon	C
102	Bkon	C++

2NF

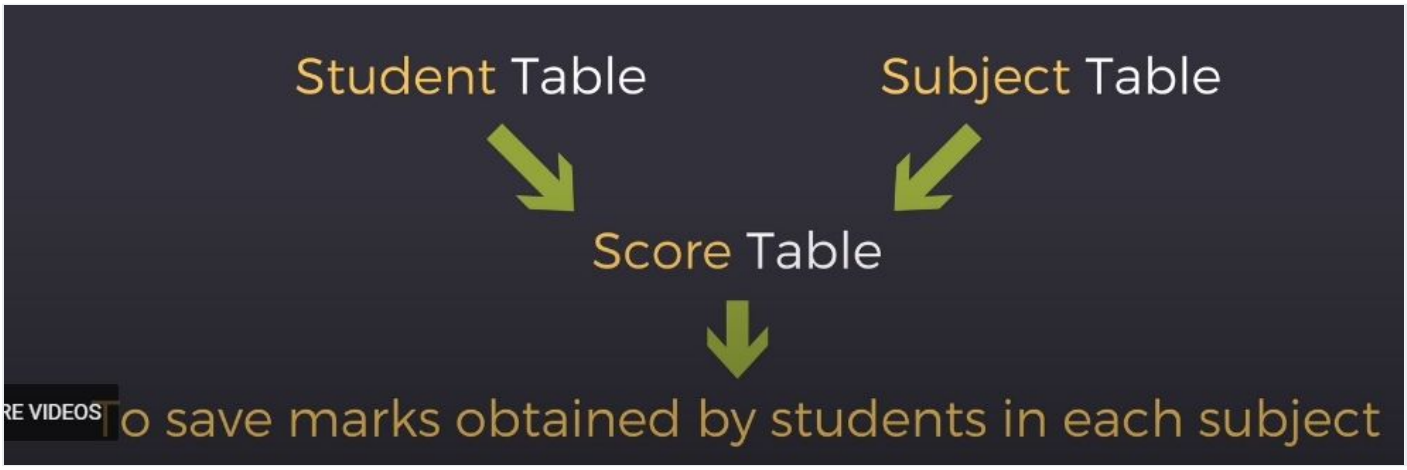


What is Partial Dependency?

Before knowing partial dependency, first, we should know what is dependency or functional dependency in the table i.e. primary key.

STUDENTS TABLE

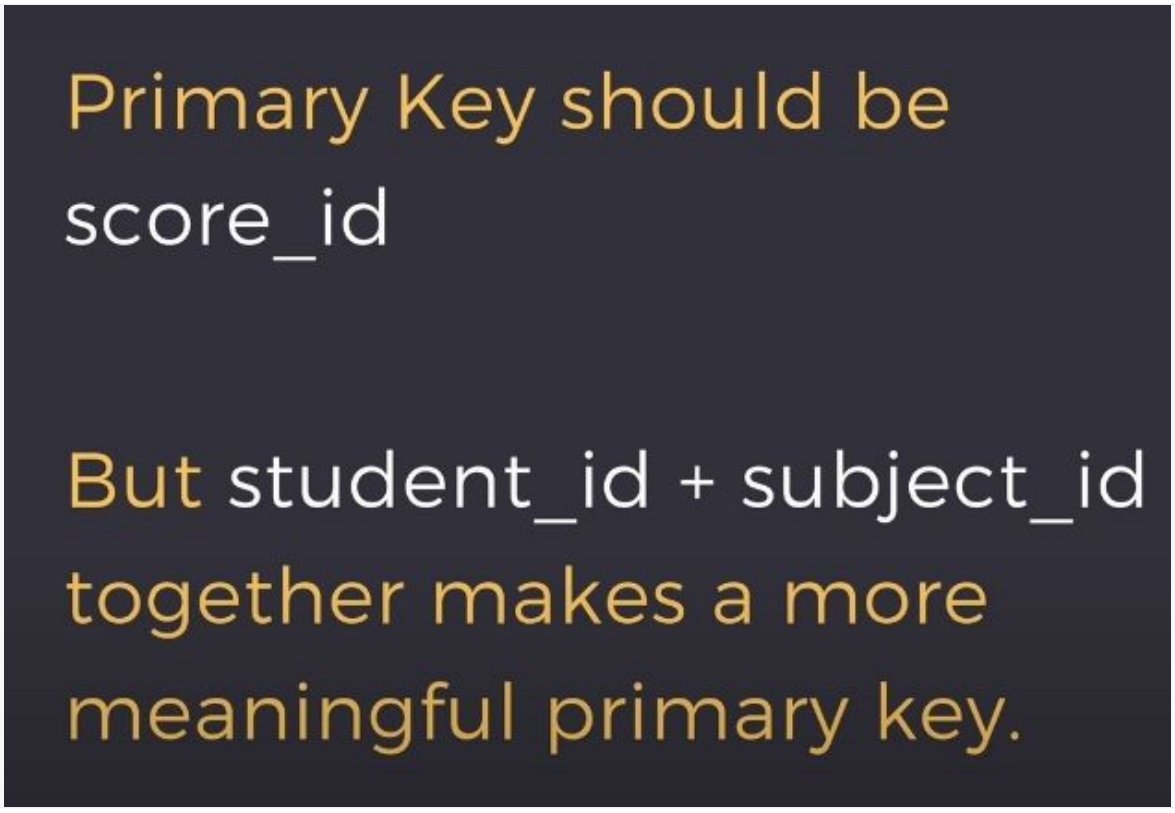
student_id	name	reg_no	branch	address
1	Akon	CSE-18	CSE	TN
2	Akon	IT-18	IT	AP
3	Bkon	CSE-18	CSE	HR
4	Ckon	CSE-18	CSE	MH



In this case we can see Many To Many relationship

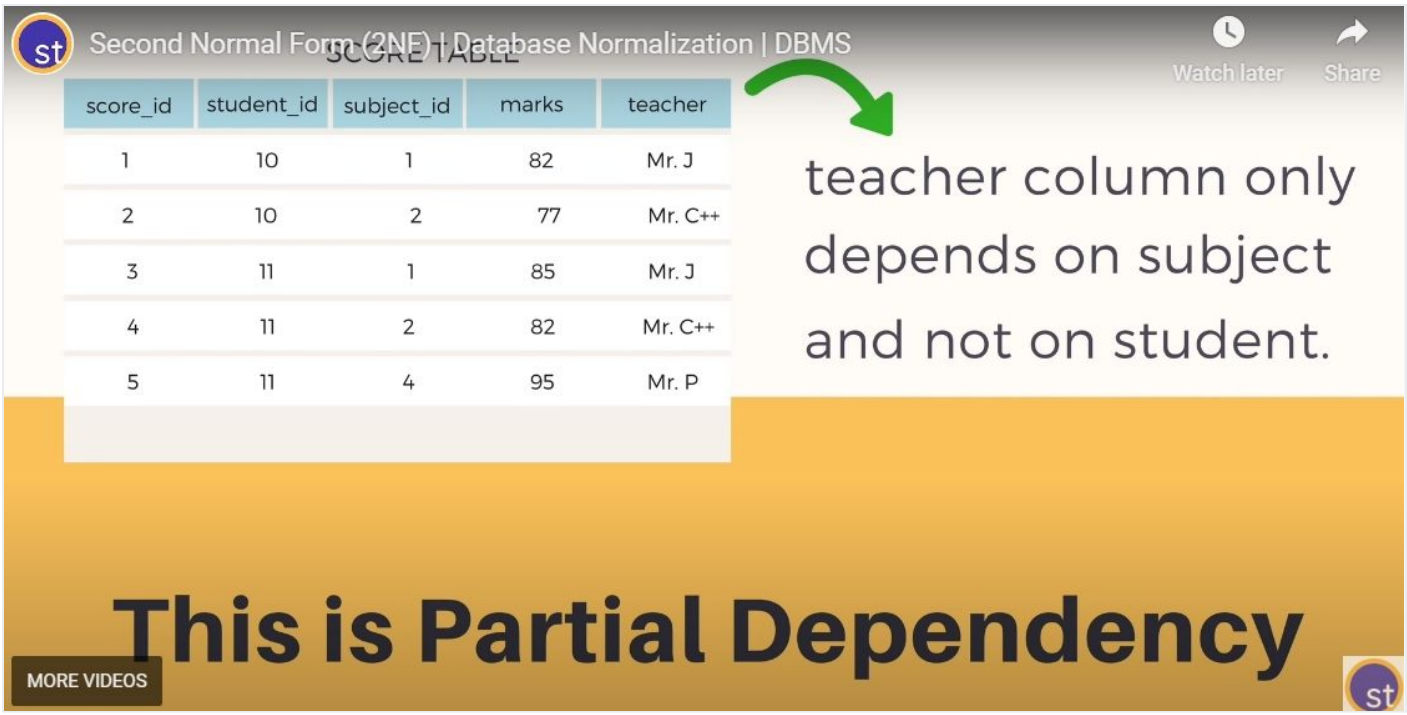
SCORE TABLE

score_id	student_id	subject_id	marks	teacher
1	1	1	82	Mr. J
2	1	2	77	Mr. C++
3	2	1	85	Mr. J
4	2	2	82	Mr. C++
5	2	4	95	Mr. P



Here the primary key in score table is combination of student_id and subject_id

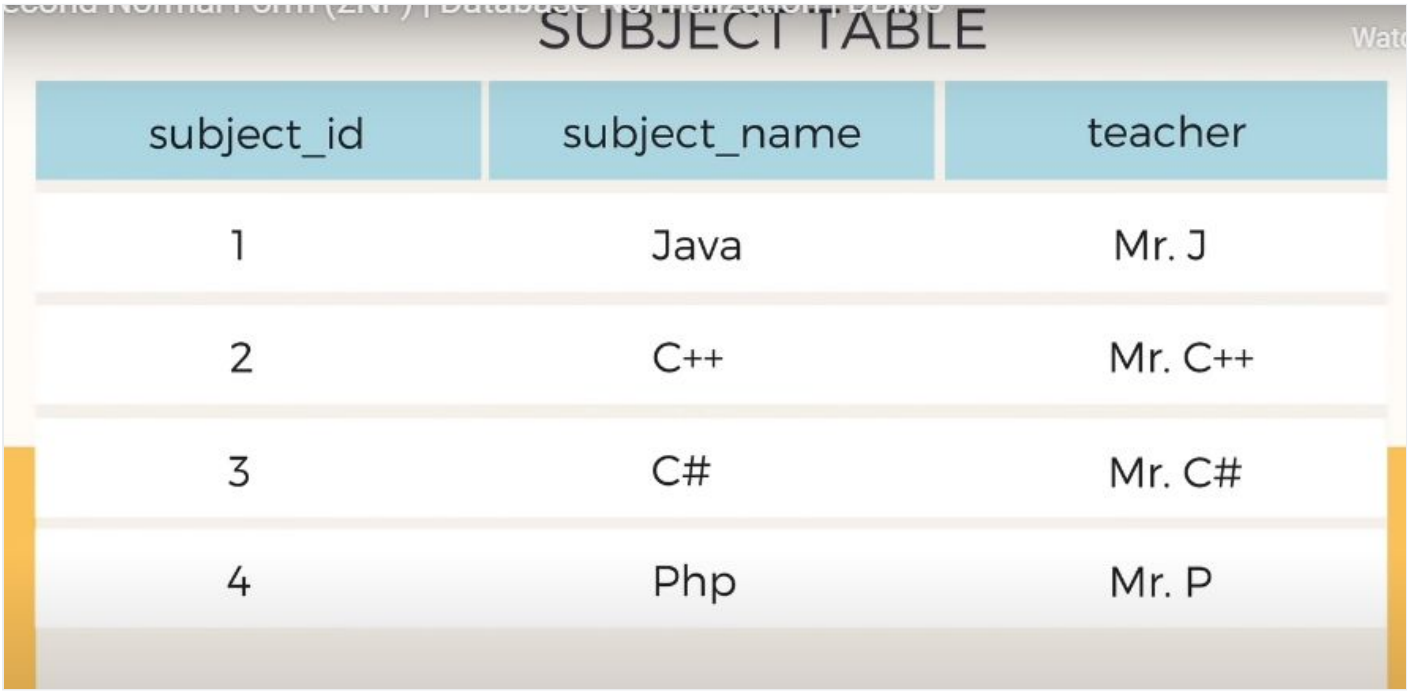
But to fetch teacher details only subject id is required. **This is called Partial dependency.**



The table to be in second normal form this should not exist .

How to remove the partial dependency?

1. To move teacher name with subject table
2. Or we can make a new table for teachers

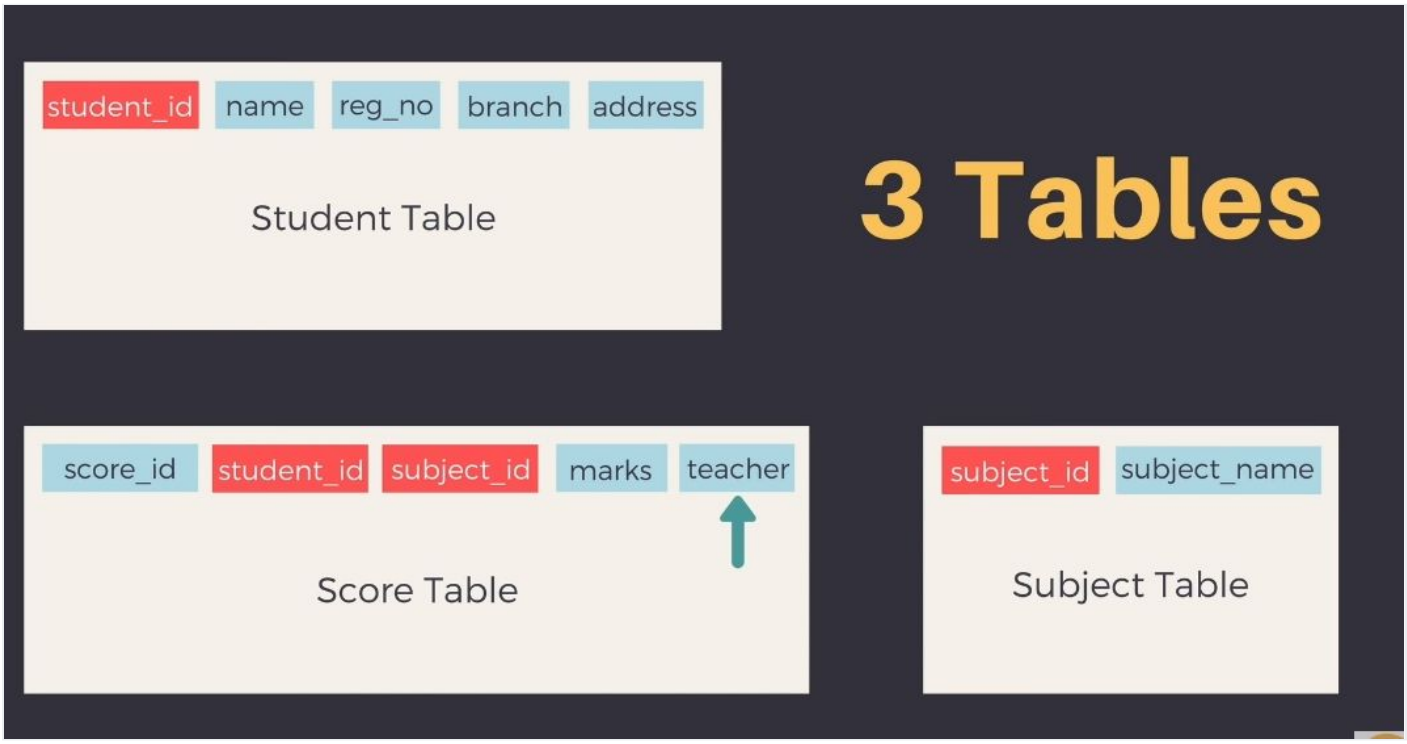


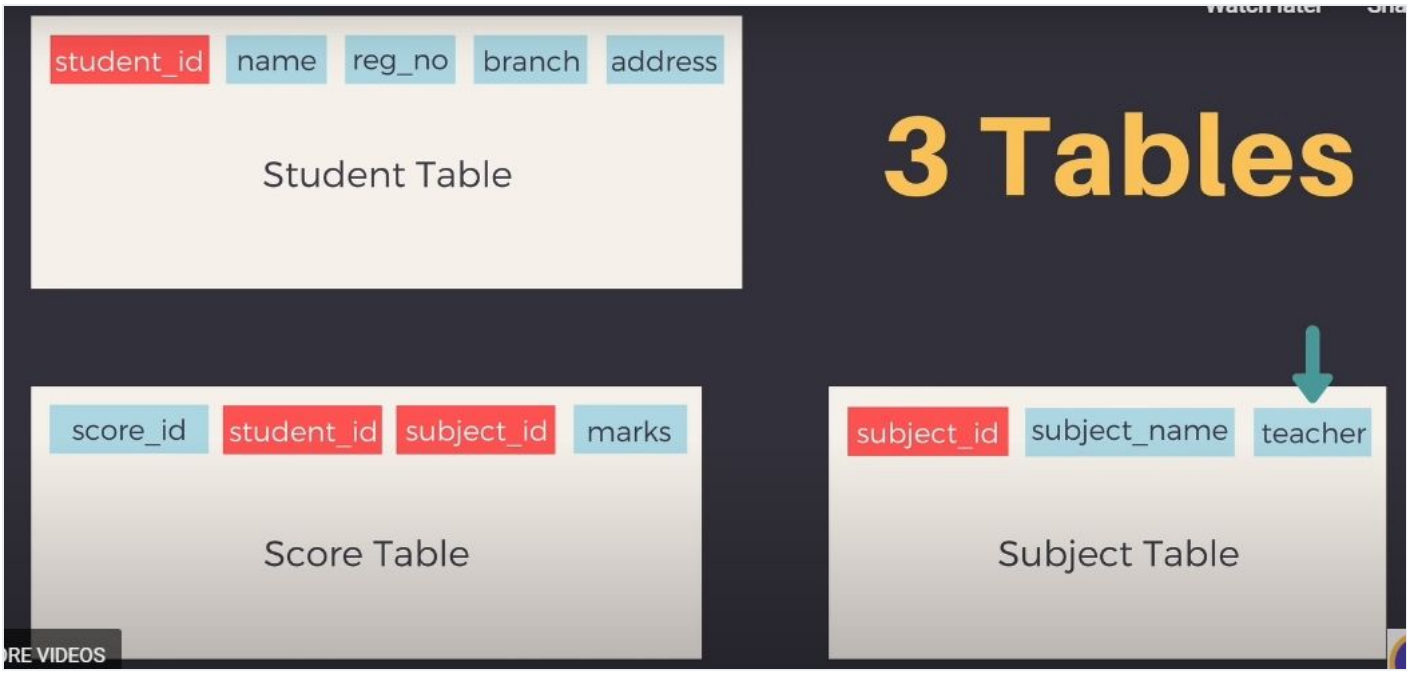
Teacher TABLE

teacher_id	teacher_name
1	Mr. J
2	Mr. C++
3	Mr. C#
4	Mr. P

Can even add more info. related to teachers like date of joining, salary etc.

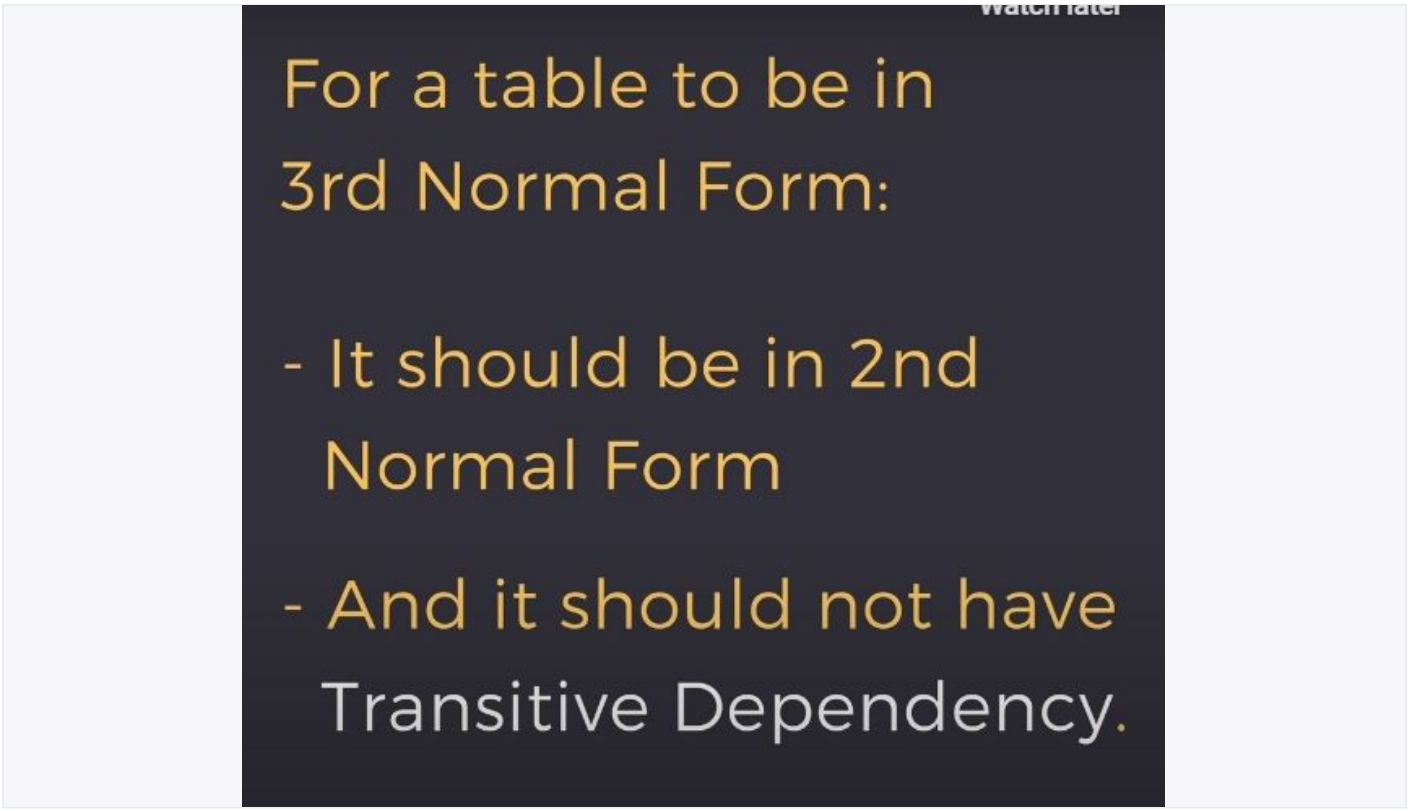
3NF





We also have the total marks and exam name after that the table would look this this :

IN 3NF

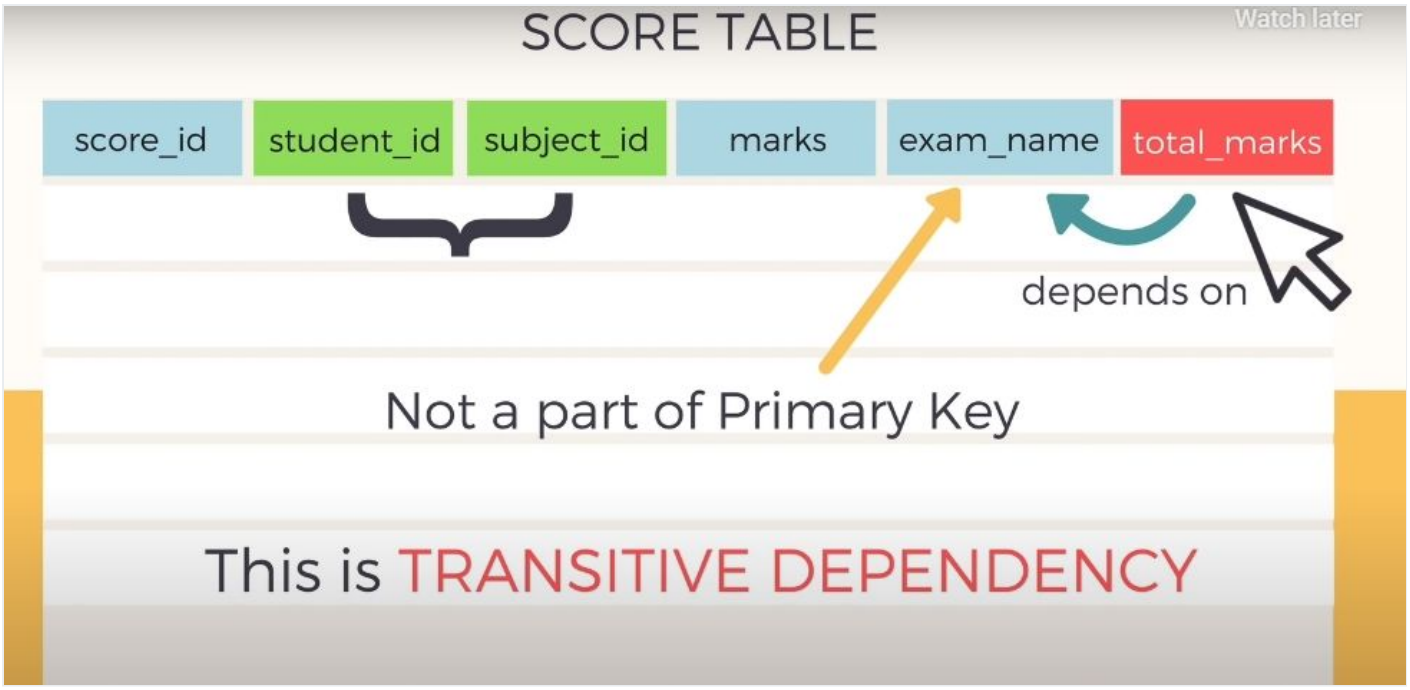


score_id = student_id + subject_id

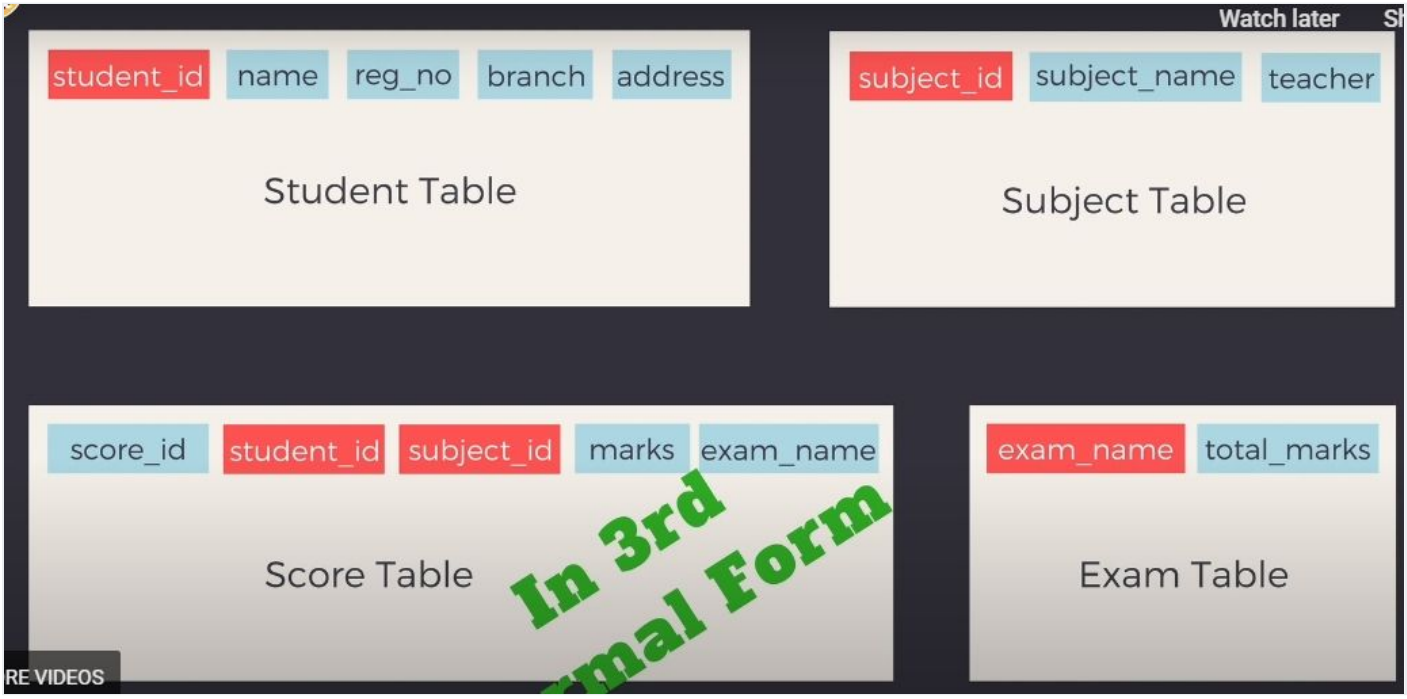
exam_name : depend on student name and subject both

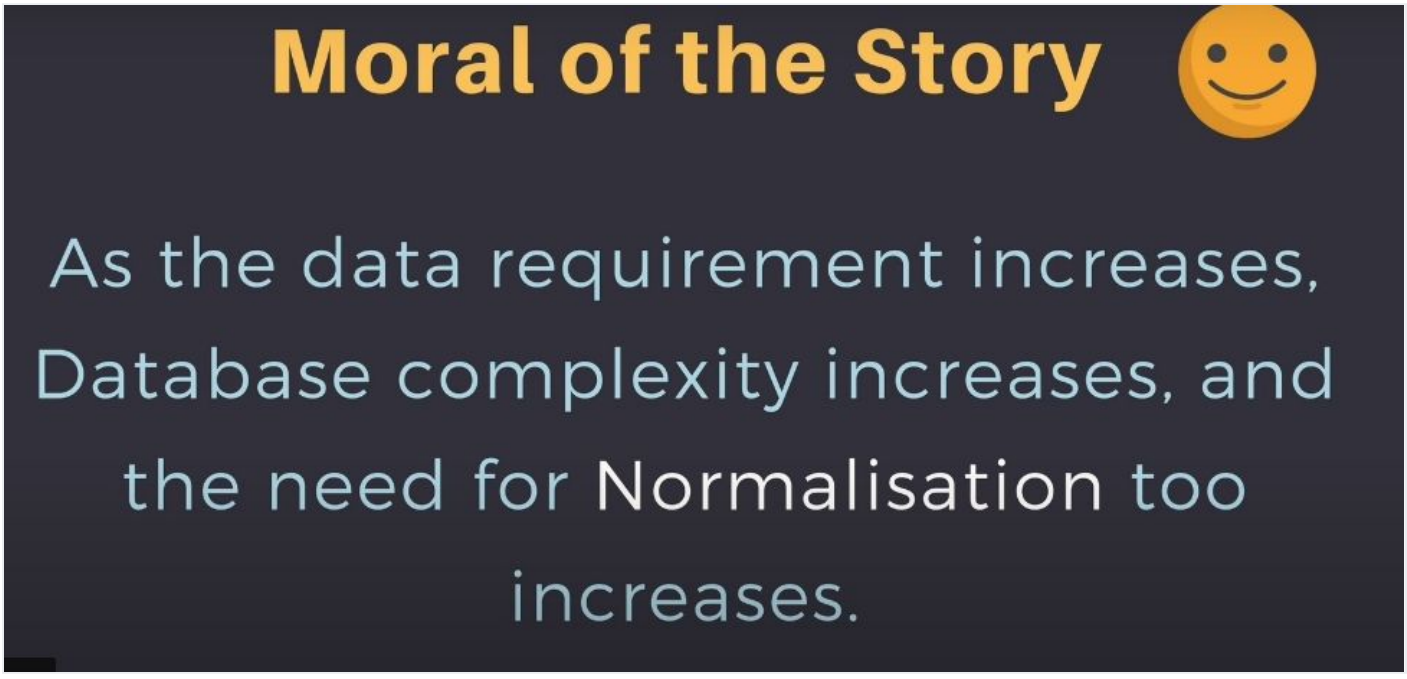
total marks: depends on the exam name

but the exam name is not our primary key this is a **transitive dependency**



so we need to put the exam name and total marks in **new** exam table .





BCNF or 3.5NF (Boyce - Codd Normal Form)

Let's recap 2NF AND 3NF first,

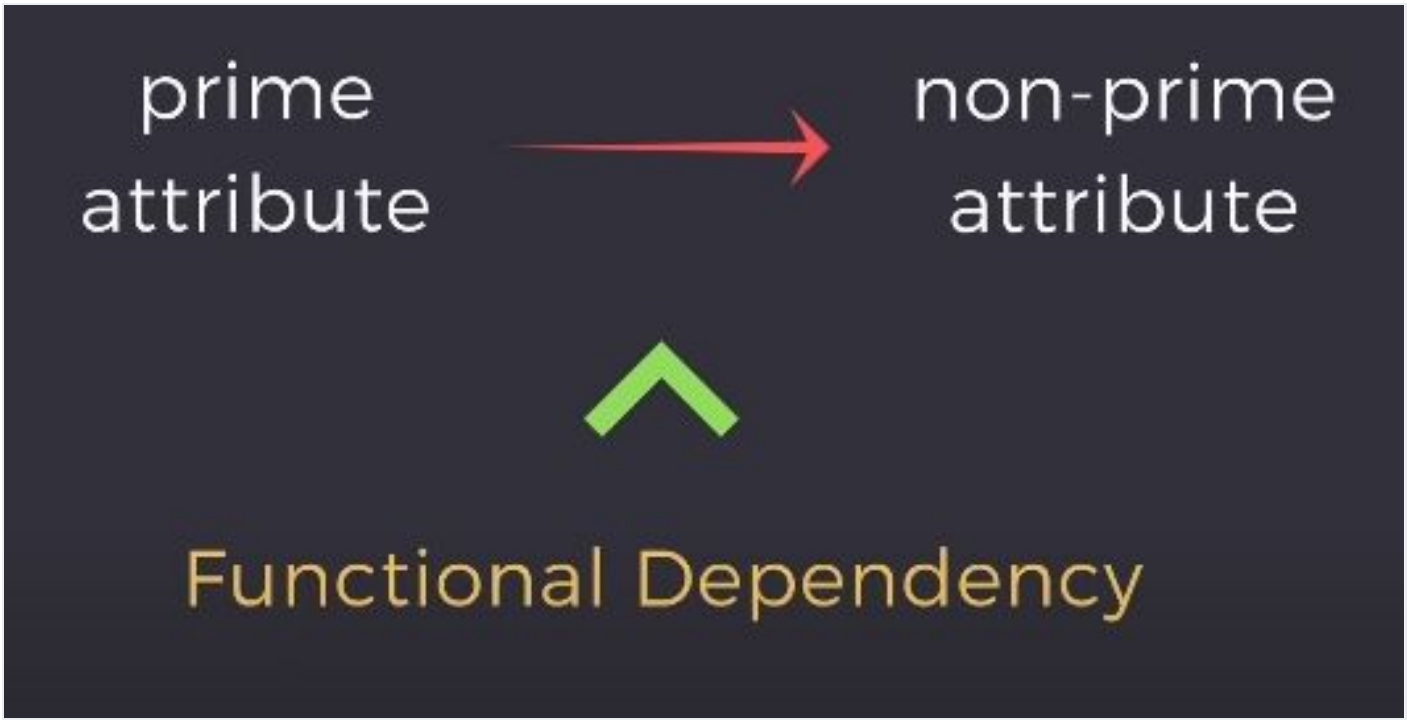
1. 2NF : A table to be in 2nd normal form , it should be in 1NF and should not follow partial dependency.

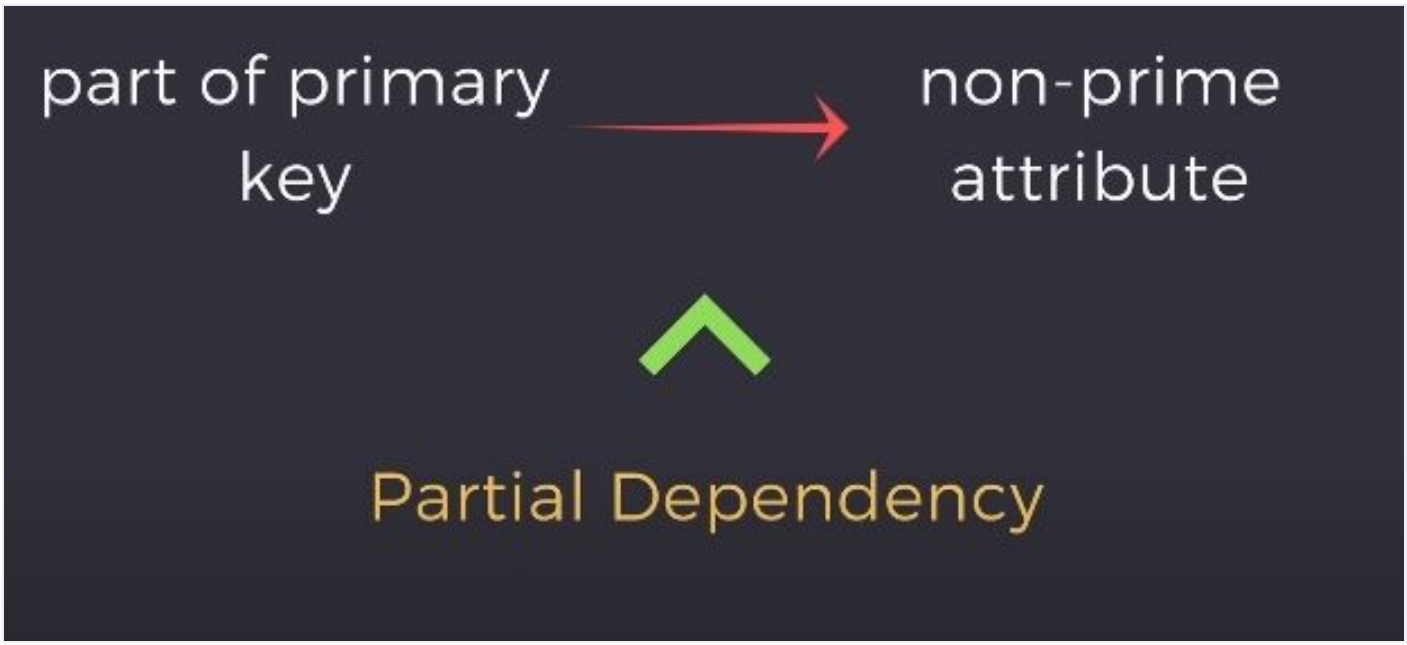
2. Functional Dependency: A is prime attribute or primary key and B is not prime attribute , B is derived from A .

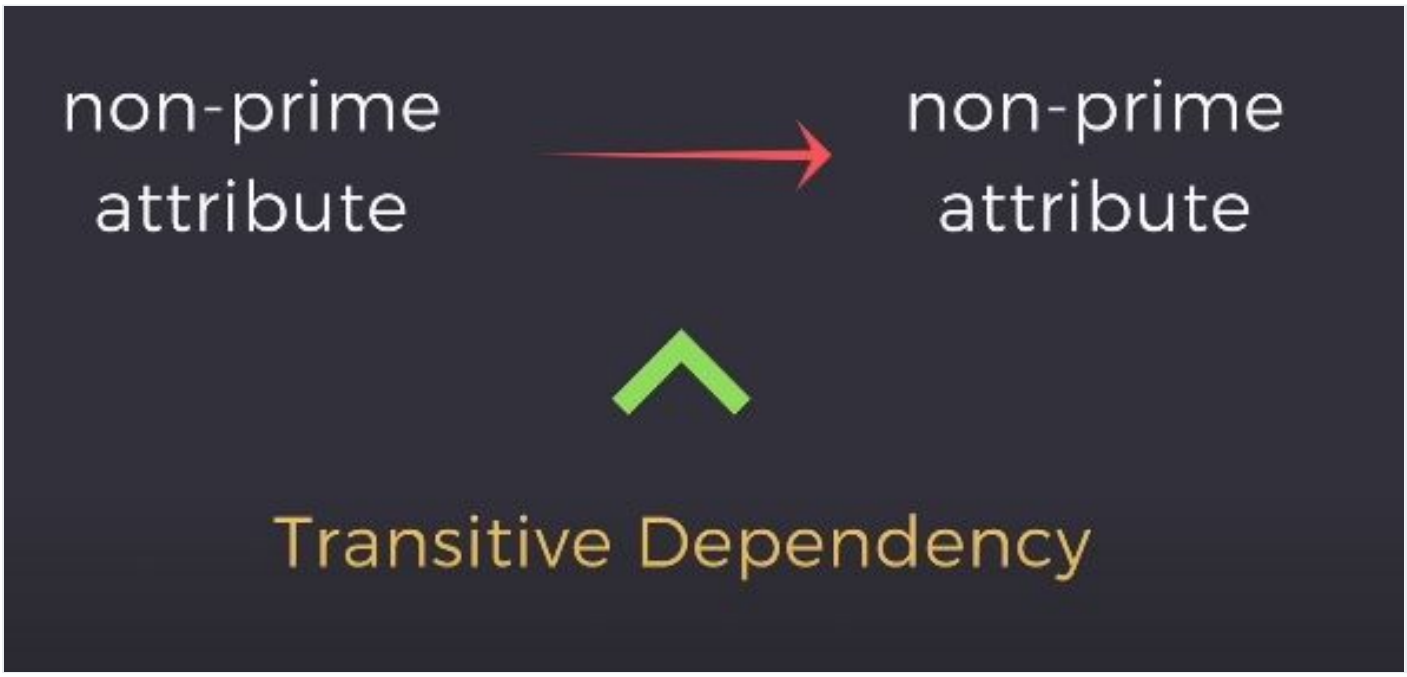
3. If we have AXY as candidate key where B only depends on A rather than depending on all the 3 keys then it's said to be partial dependency.

4. For a table to be in 3rd Normal form it should be in 2NF and should not have transitive dependency.

5. **When a non-prime attribute depends on another non-prime attribute then this is called a Transitive dependency.**

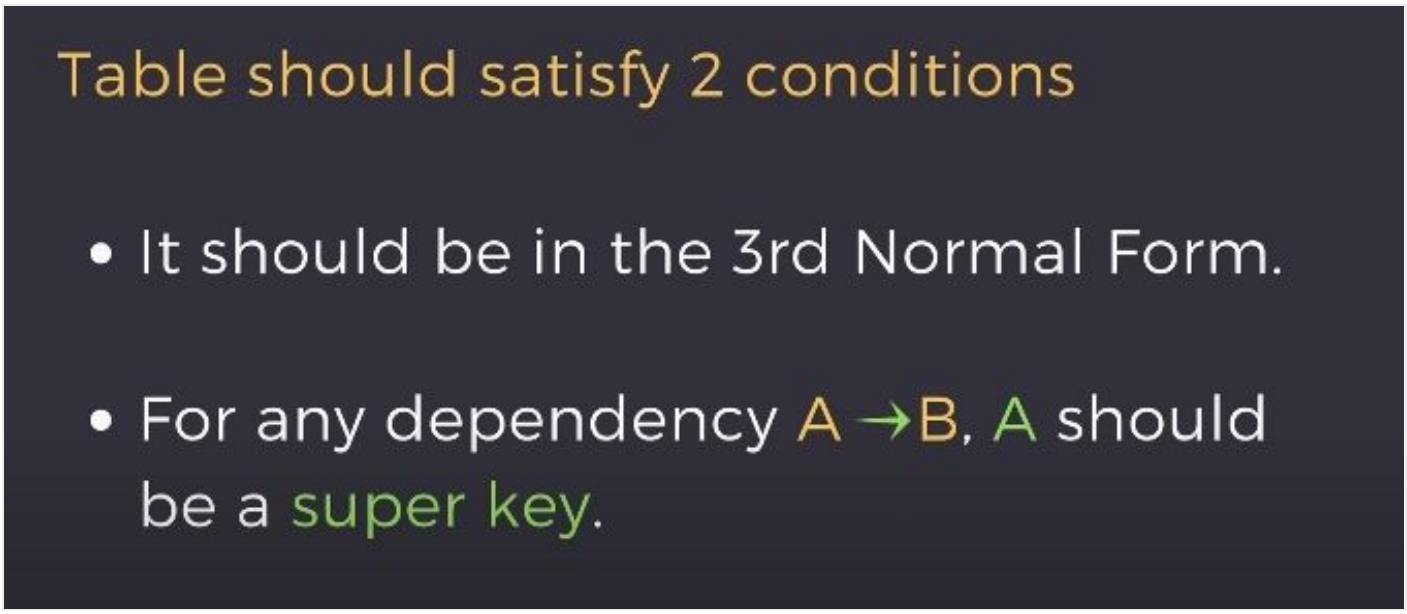






BCNF

1. First it should be in 3NF
2. For any dependency A→ B, **A should be a super key .**

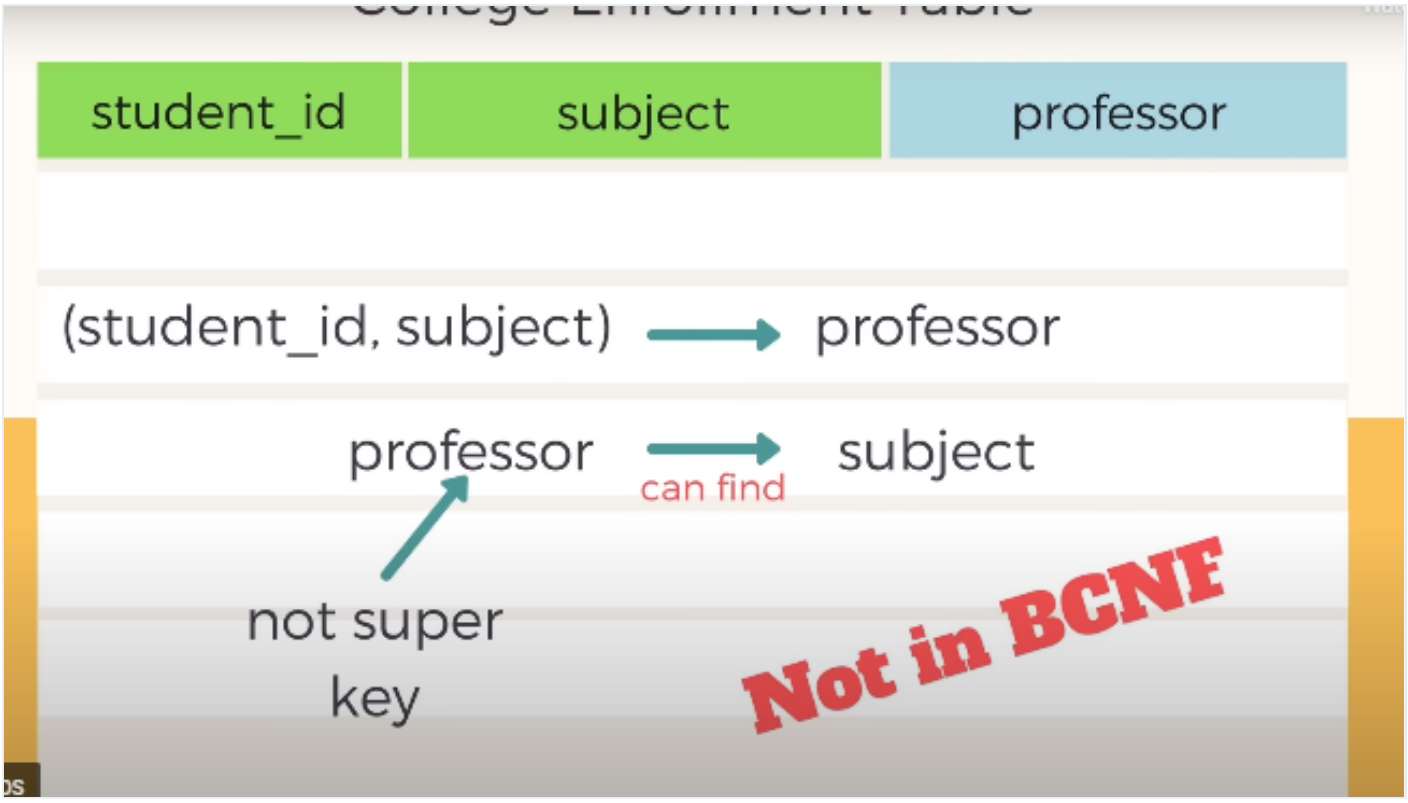


A **can not be** a non -prime attribute where B is a prime attribute.

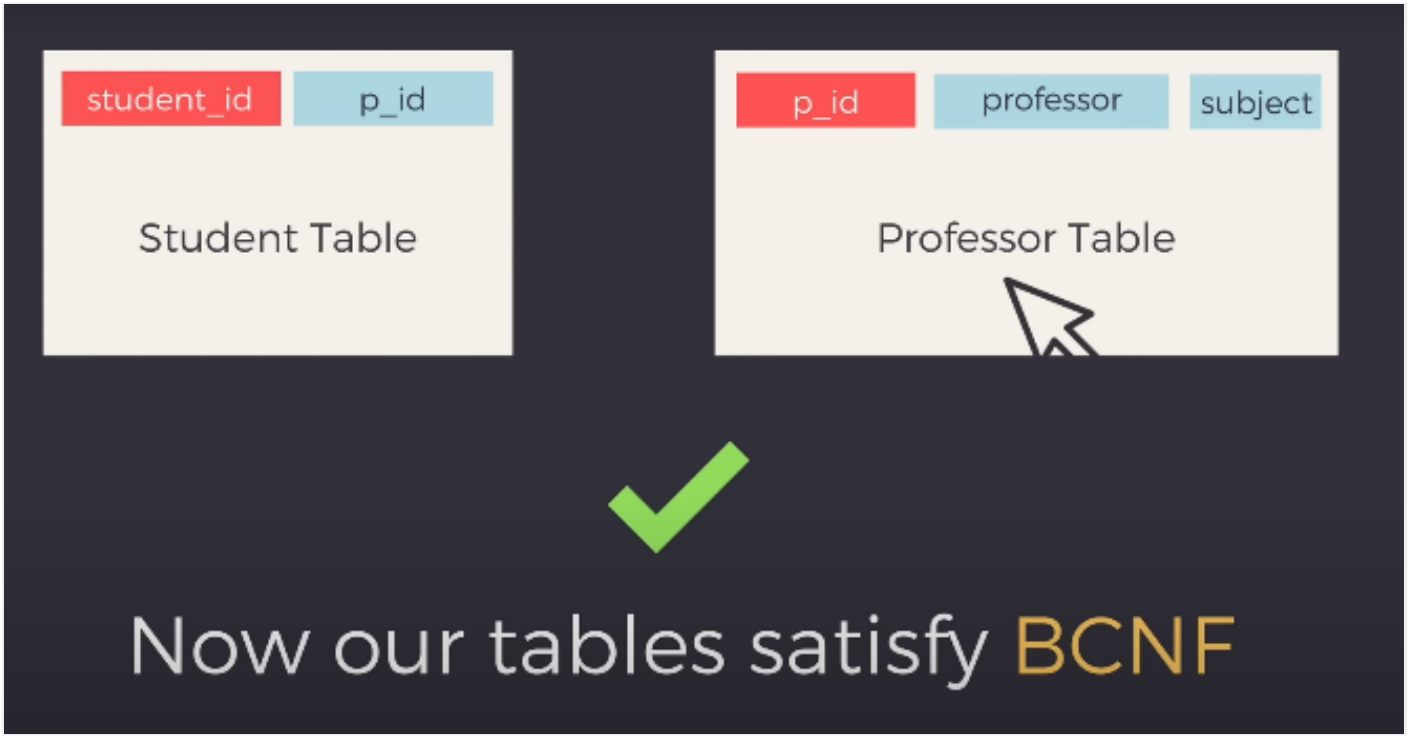
College Enrollment Table

student_id	subject	professor
101	Java	P. Java
101	C++	P. Cpp
102	Java	P. Java2
103	C#	P. Chash
104	Java	P. Java

in this case student_id and subject ⇒ primary key .



To solve this problem we can do this.



4NF

1. It should be in BCNF

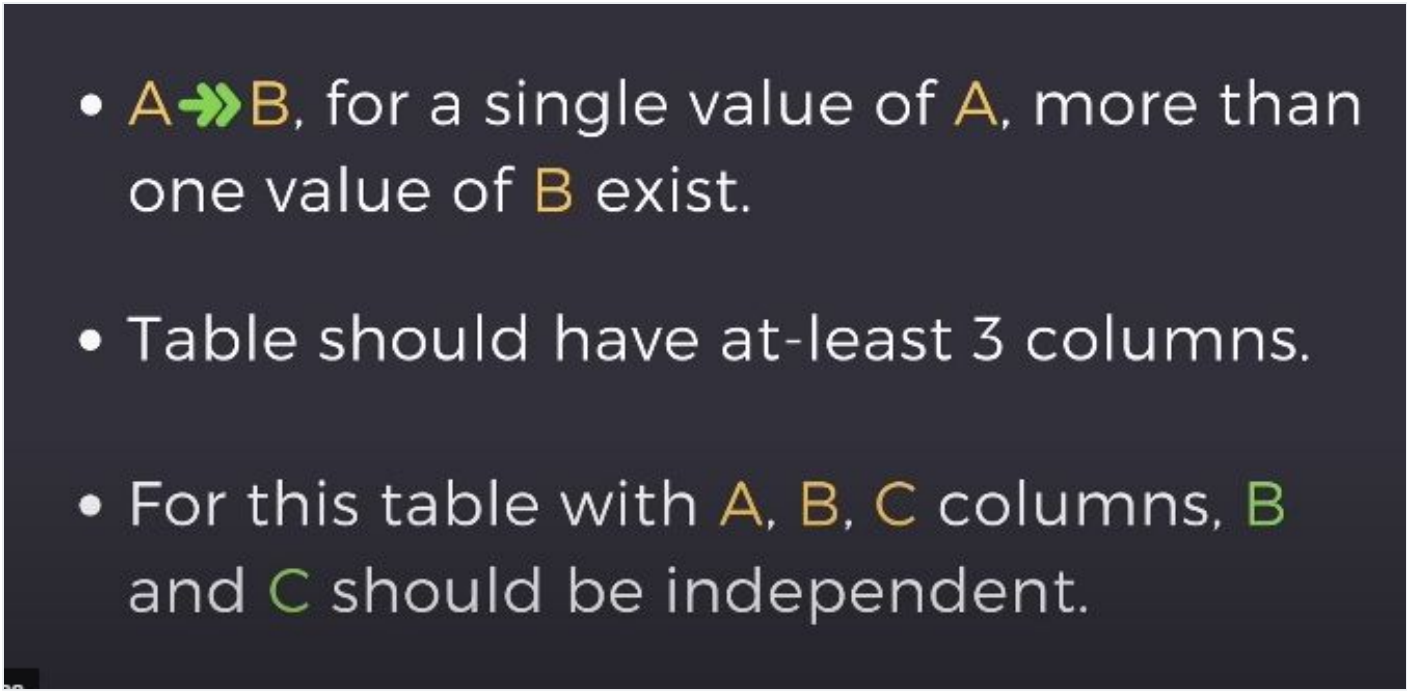
2. There should not be multi-valued dependency.

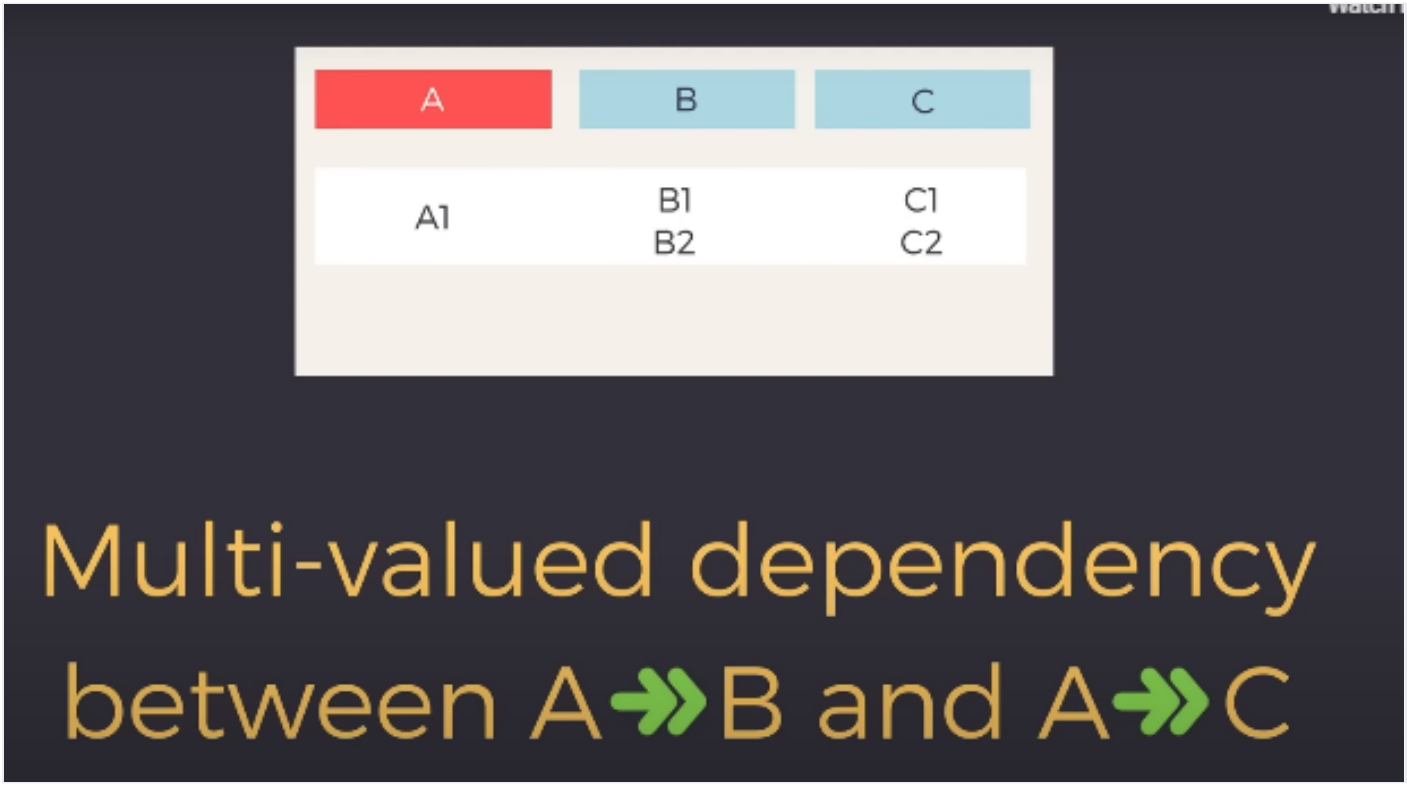
3. **Multi-valued dependency**:

4. when for a single value of A more than one value of B exists. A is a prime attribute and B is a non-prime attribute.

5. For multivalued dependency, there should at least be 3 columns.

6. For a table A , B ,C where B and C should be independent.





ENROLMENT TABLE

s_id	course	hobby
1	Science	Cricket
1	Maths	Hockey
2	C#	Cricket
2	Php	Hockey

ENROLMENT TABLE

s_id	course	hobby
1	Science	Cricket
1	Maths	Hockey
1	Science	Hockey
1	Maths	Cricket

ENROLMENT TABLE

No relationship

s_id	course	hobby
1	Science	Cricket
1	Maths	Hockey
1	Science	Hockey
1	Maths	Cricket

CourseOpted TABLE

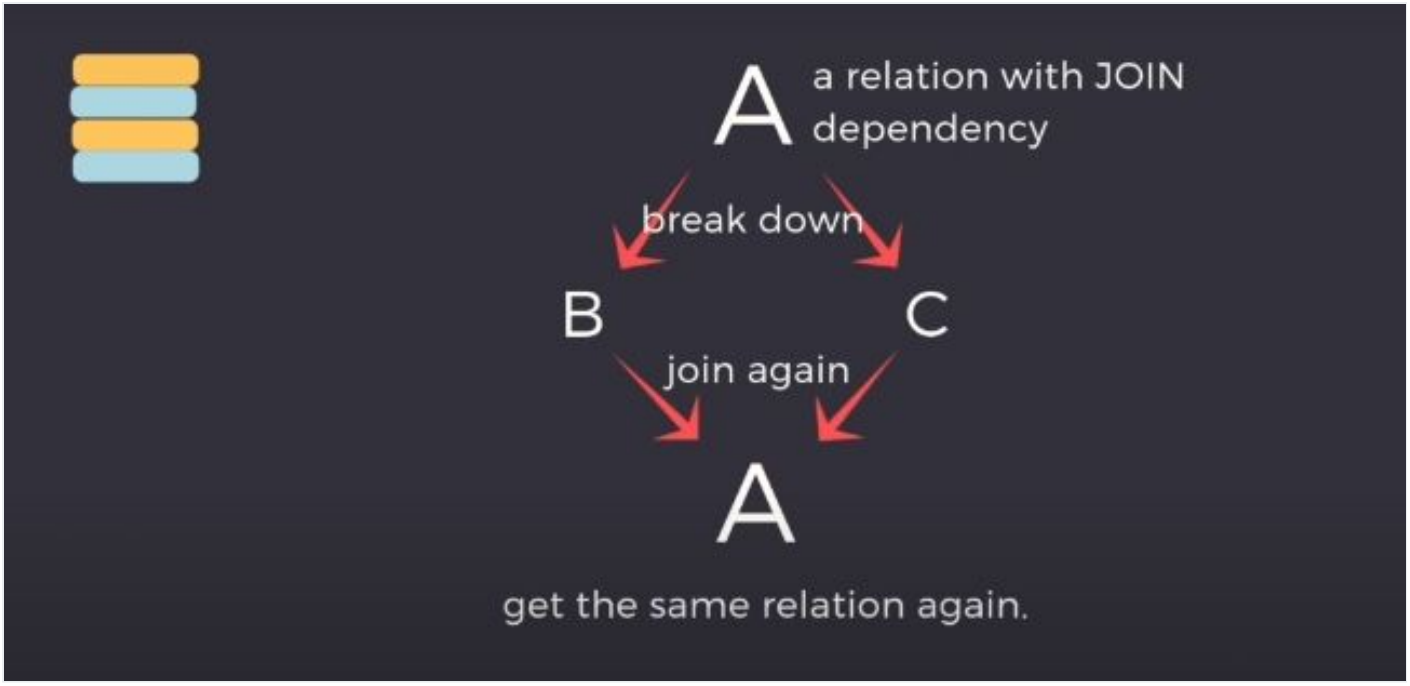
s_id	course
1	Science
1	Maths
2	C#
2	Php

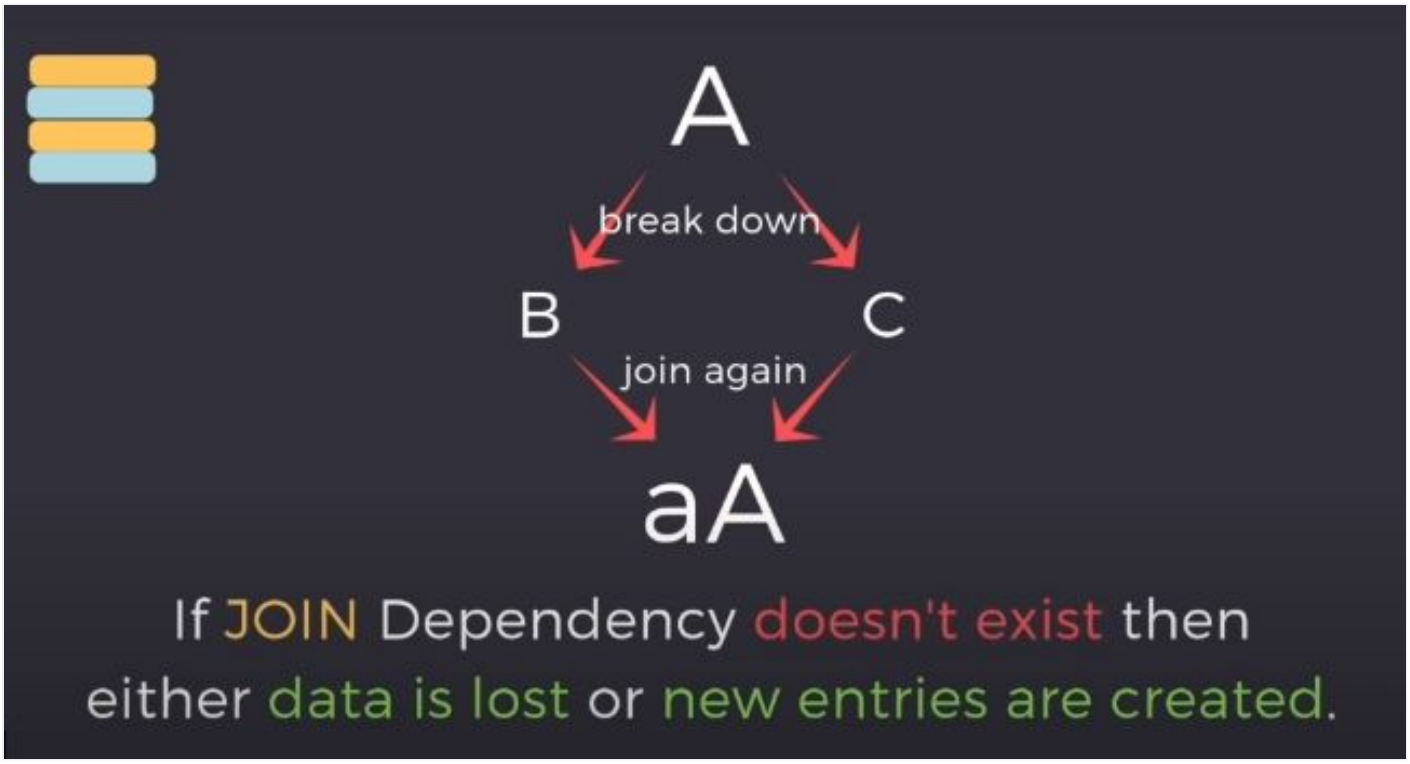
Hobbies TABLE

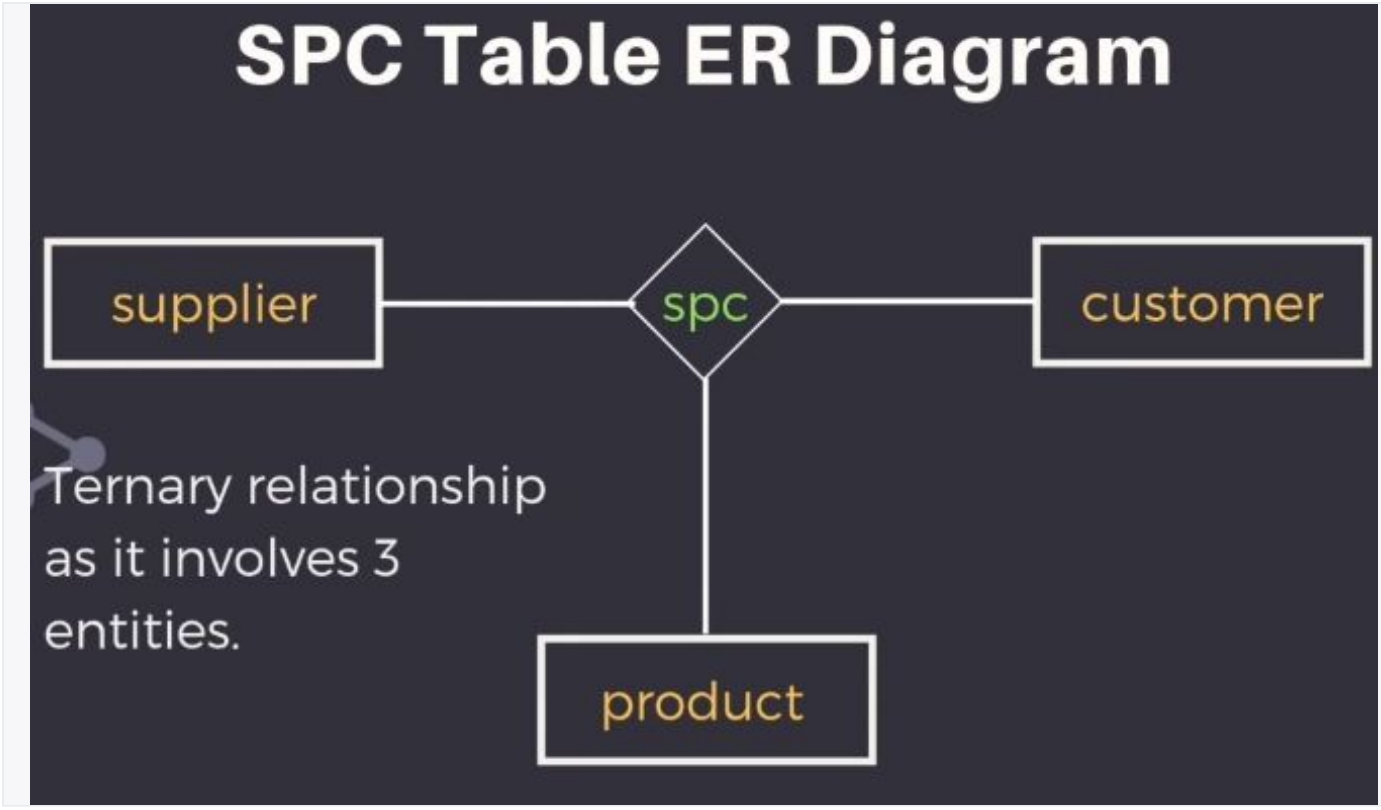
s_id	hobby
1	Cricket
1	Hockey
2	Cricket
2	Hockey

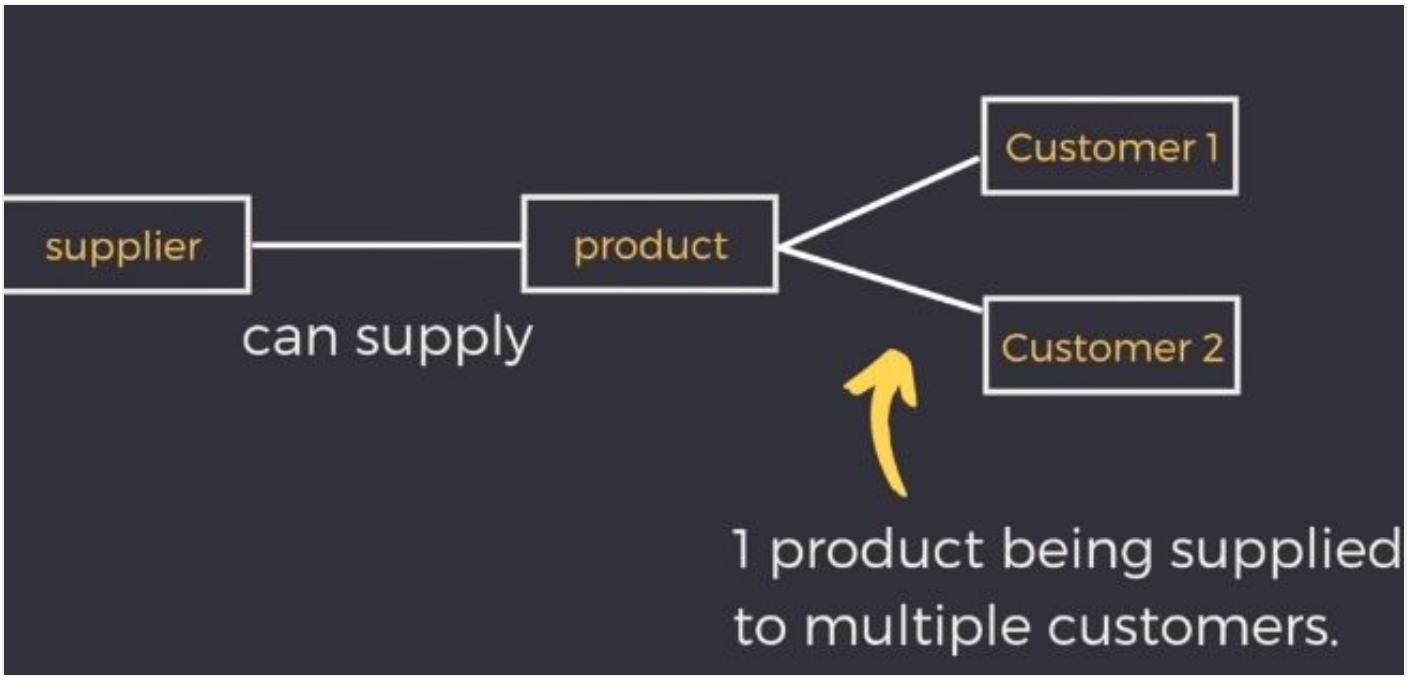
5NF or PJNF (Project Join Normal Form)

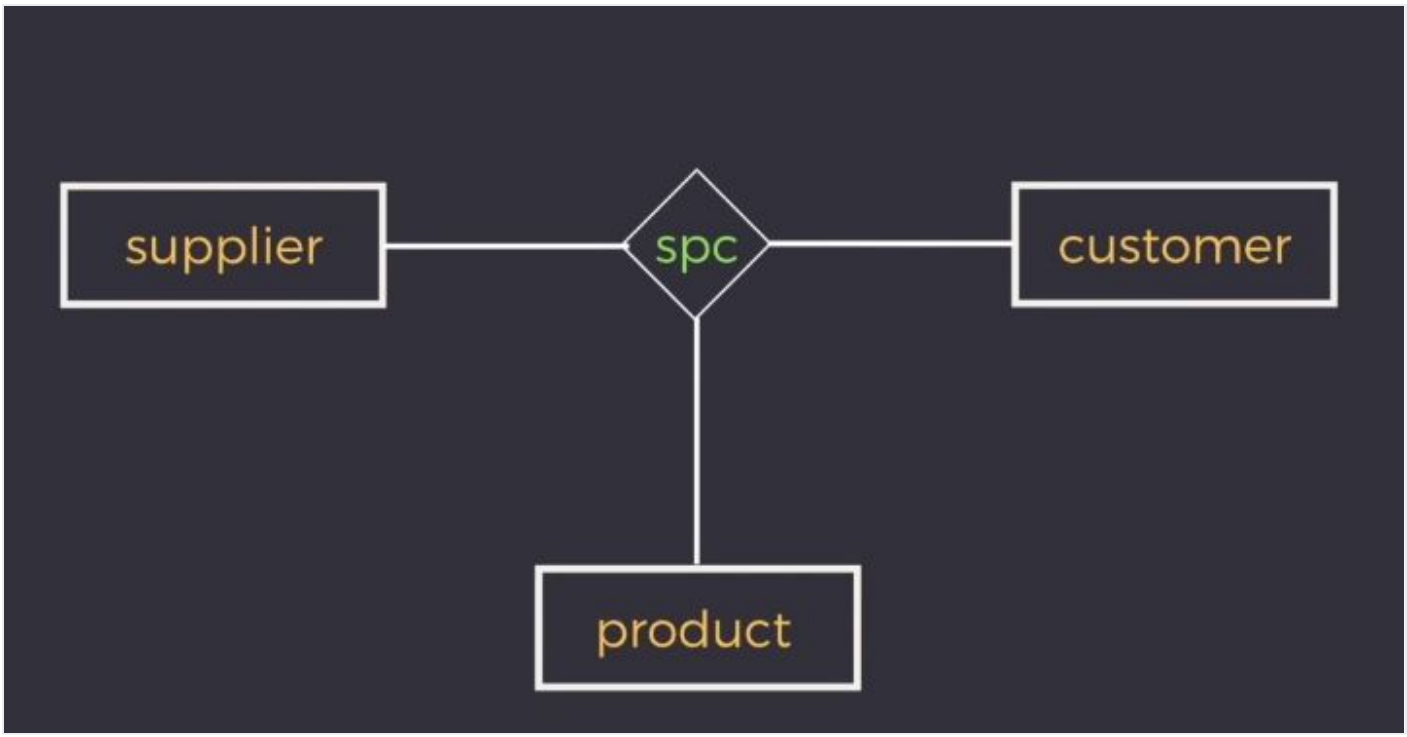
1. It should be in 4NF
2. It should not have **Join dependency , decompose the table.**

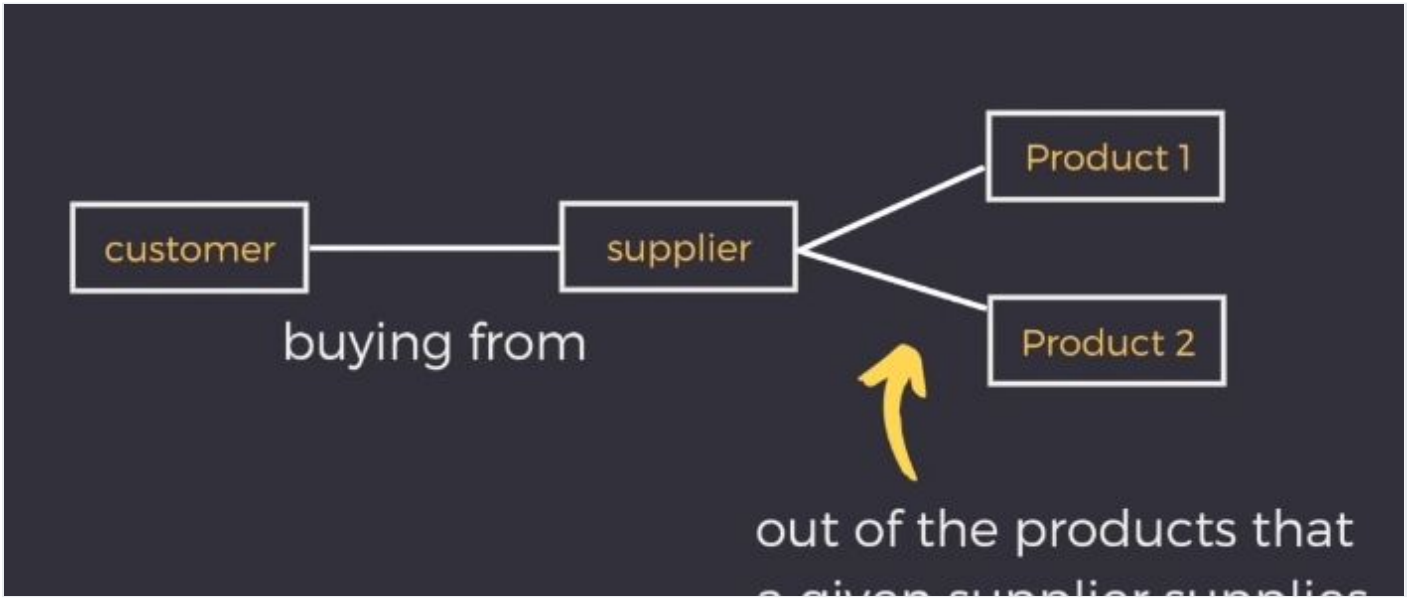

A
a relation with JOIN dependency
break down
B
C
join again
A
get the same relation again.

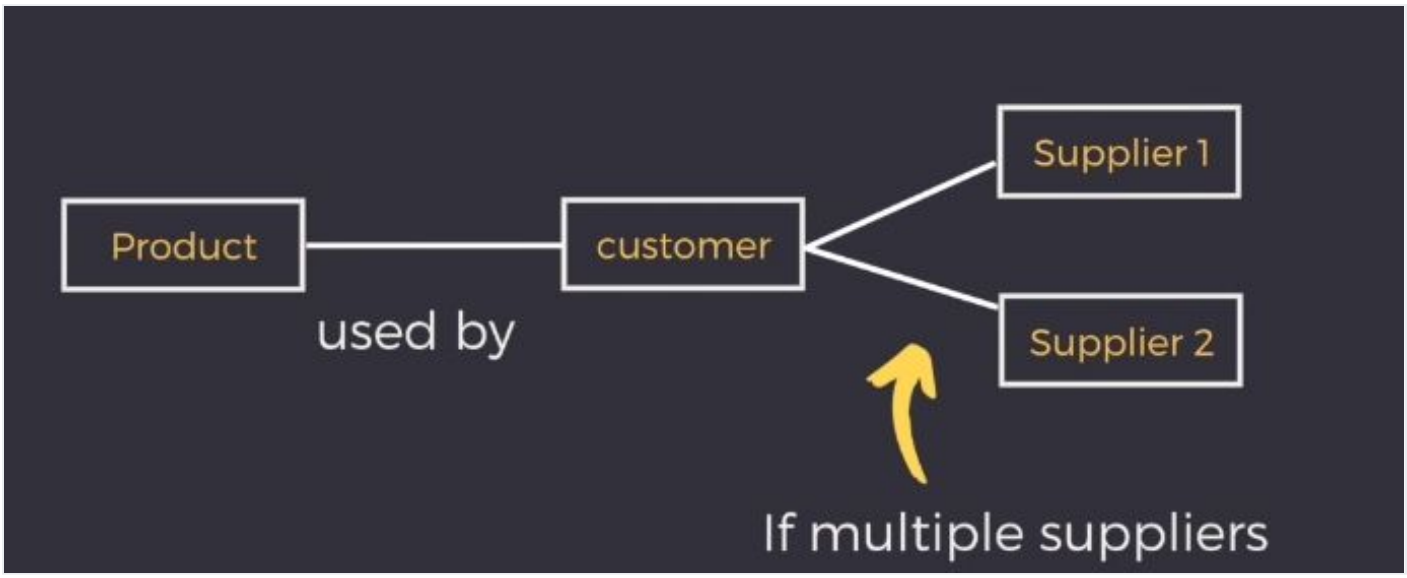

A
break down
B
C
join again
aA
If JOIN Dependency doesn't exist then
either data is lost or new entries are created.

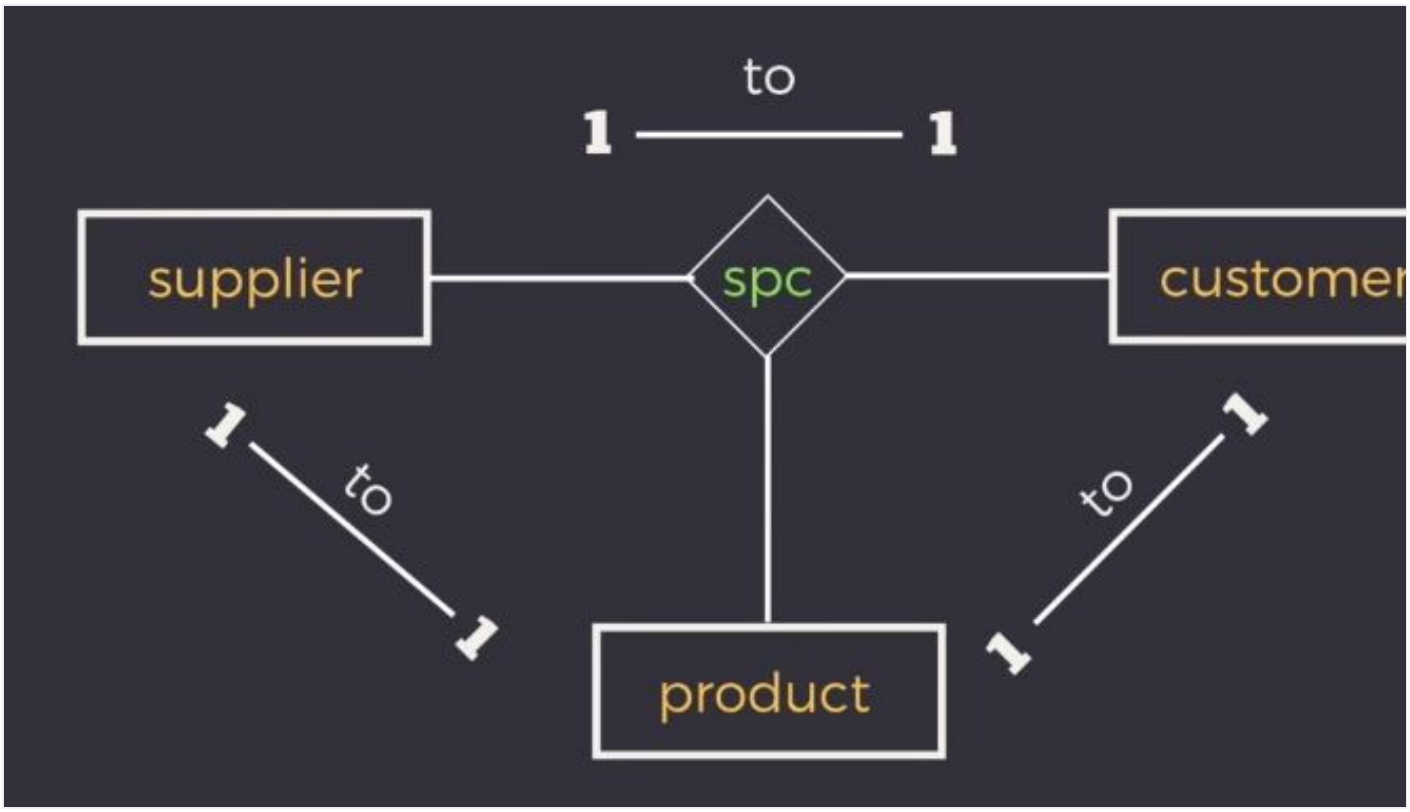

SPC Table ER Diagram
supplier
spc
customer
product
Ternary relationship as it involves 3 entities.

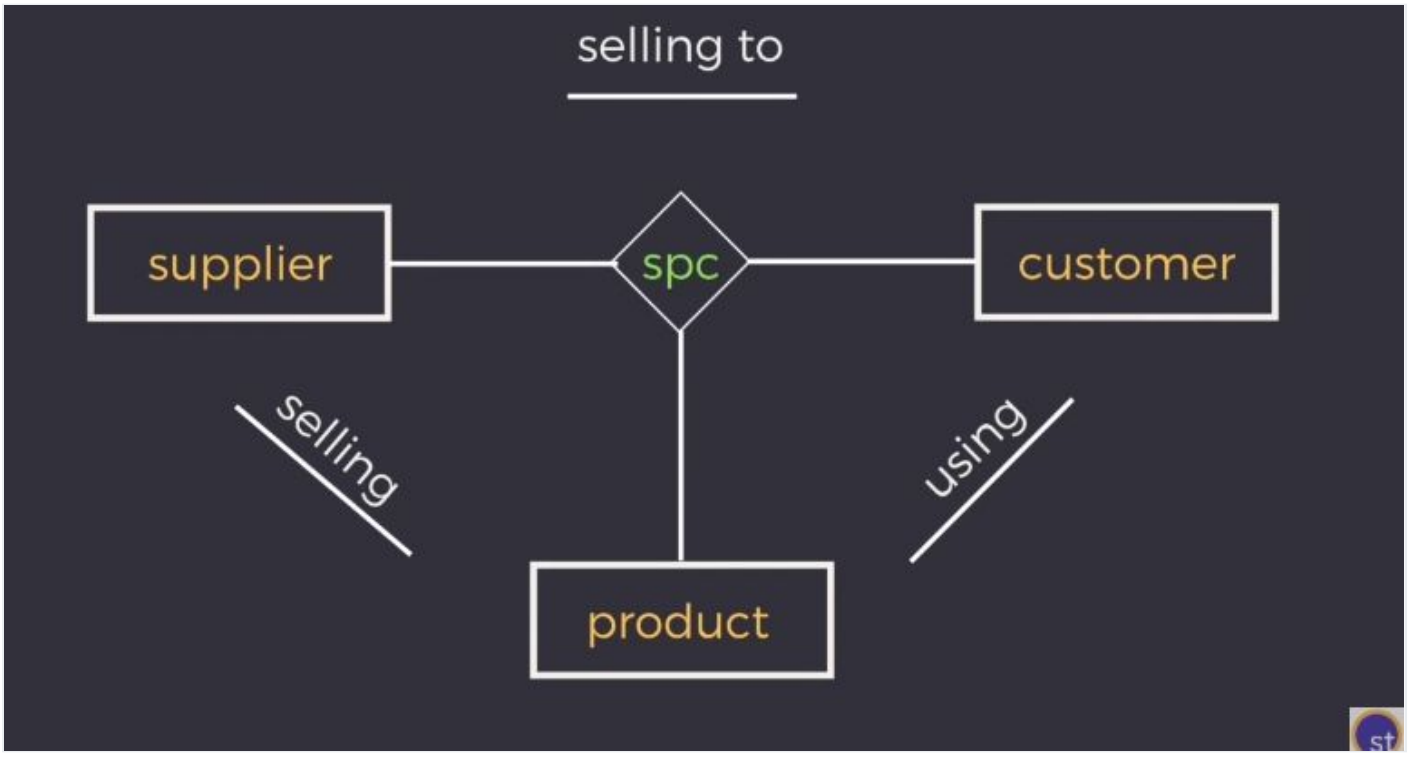

supplier
product
Customer 1
Customer 2
can supply
1 product being supplied to multiple customers.

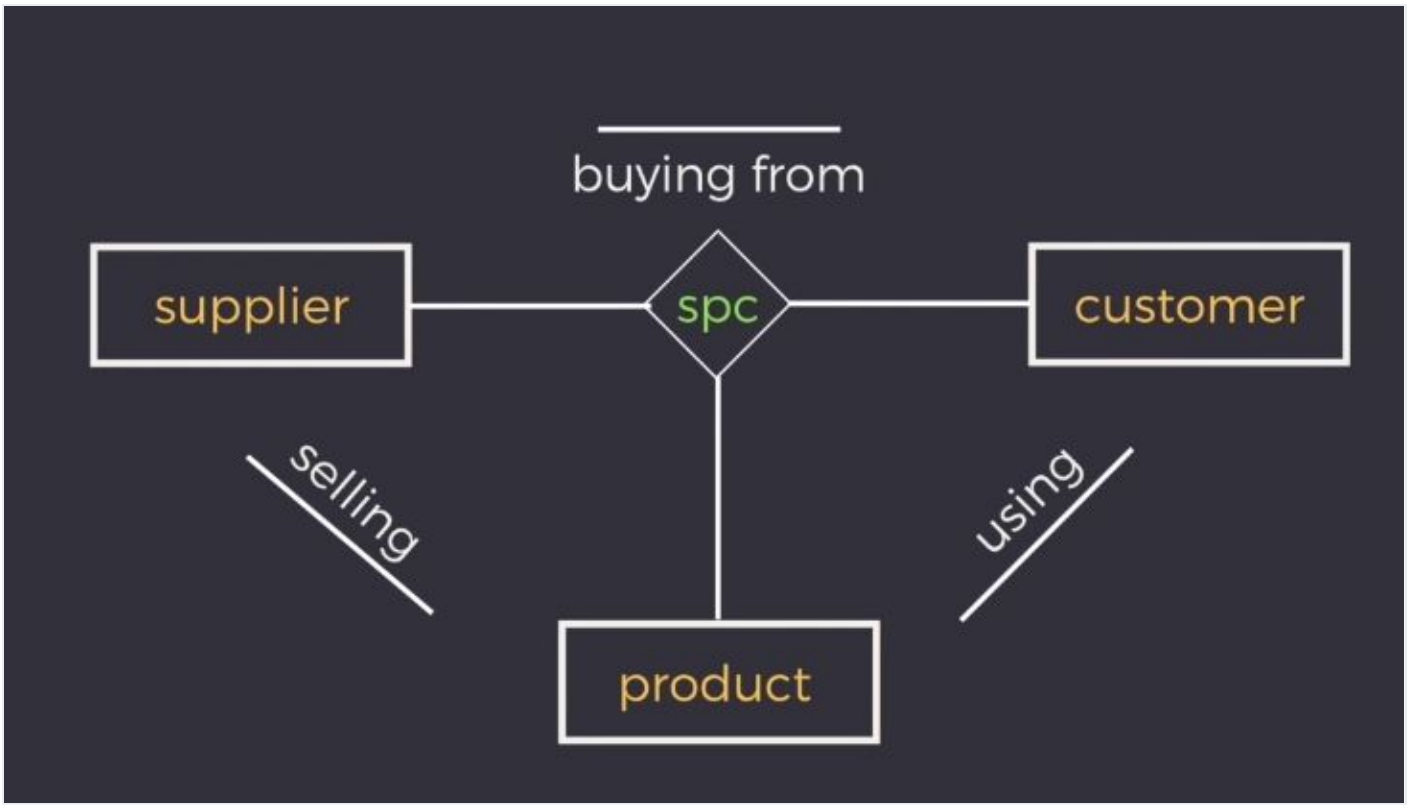

supplier
spc
customer
product

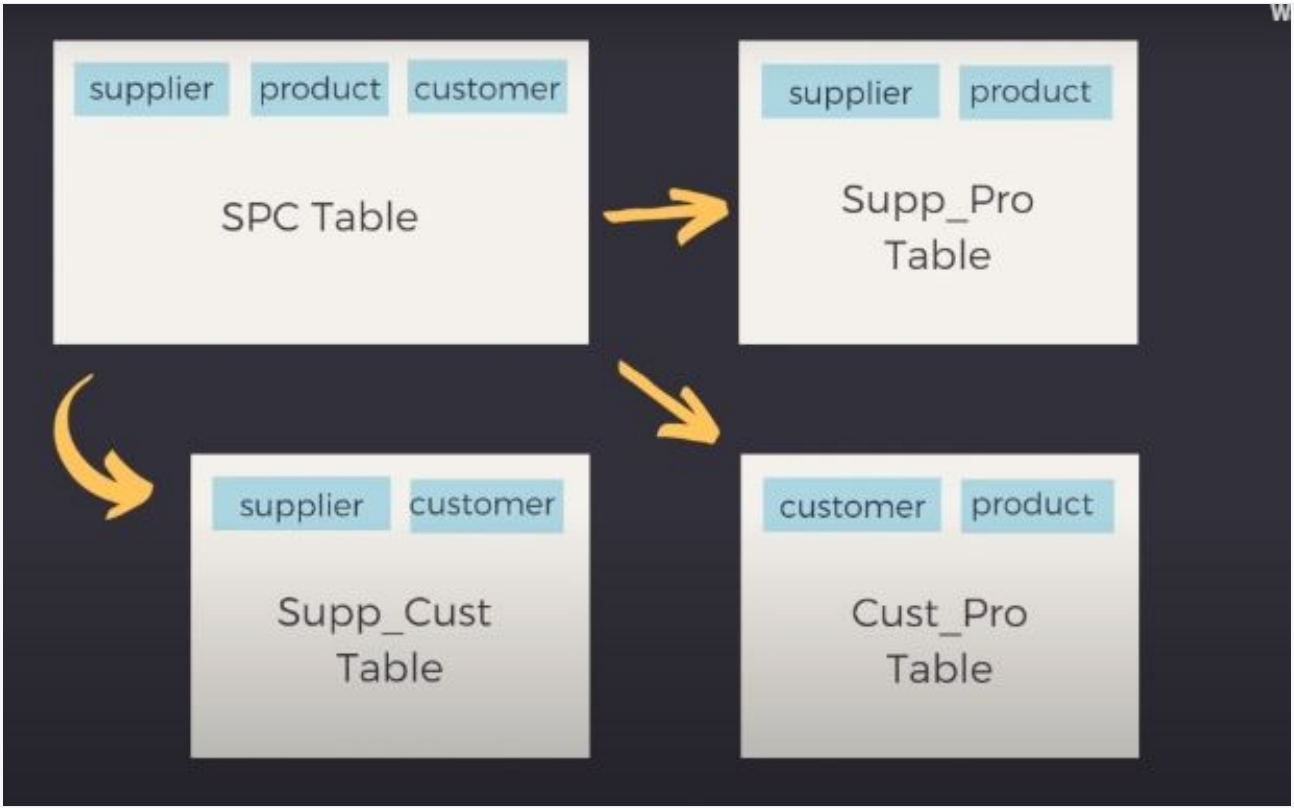

customer
supplier
Product 1
Product 2
buying from
out of the products that


Product
customer
Supplier 1
Supplier 2
used by
If multiple suppliers


to
1
1
supplier
spc
customer
1
to
1
1
to
1
product


selling to
supplier
spc
customer
selling
using
product


buying from
supplier
spc
customer
selling
using
product


supplier
product
customer
SPC Table
supplier
product
Supp_Pro Table
supplier
customer
Supp_Cust Table
customer
product
Cust_Pro Table

SPC Table

supplier	product	customer
ACME	72X SW	FORD
ACME	GEAR L	GM
ROBUSTO	E SWITCH	FORD
ROBUSTO	OBD II	MERCEDES
ALWAT	72X SW	GM
ALWAT	OBD II	MERCEDES
ALWAT	GEAR L	MERCEDES

ACME
sells 72X SW
to FORD

SUPP_PRO TABLE

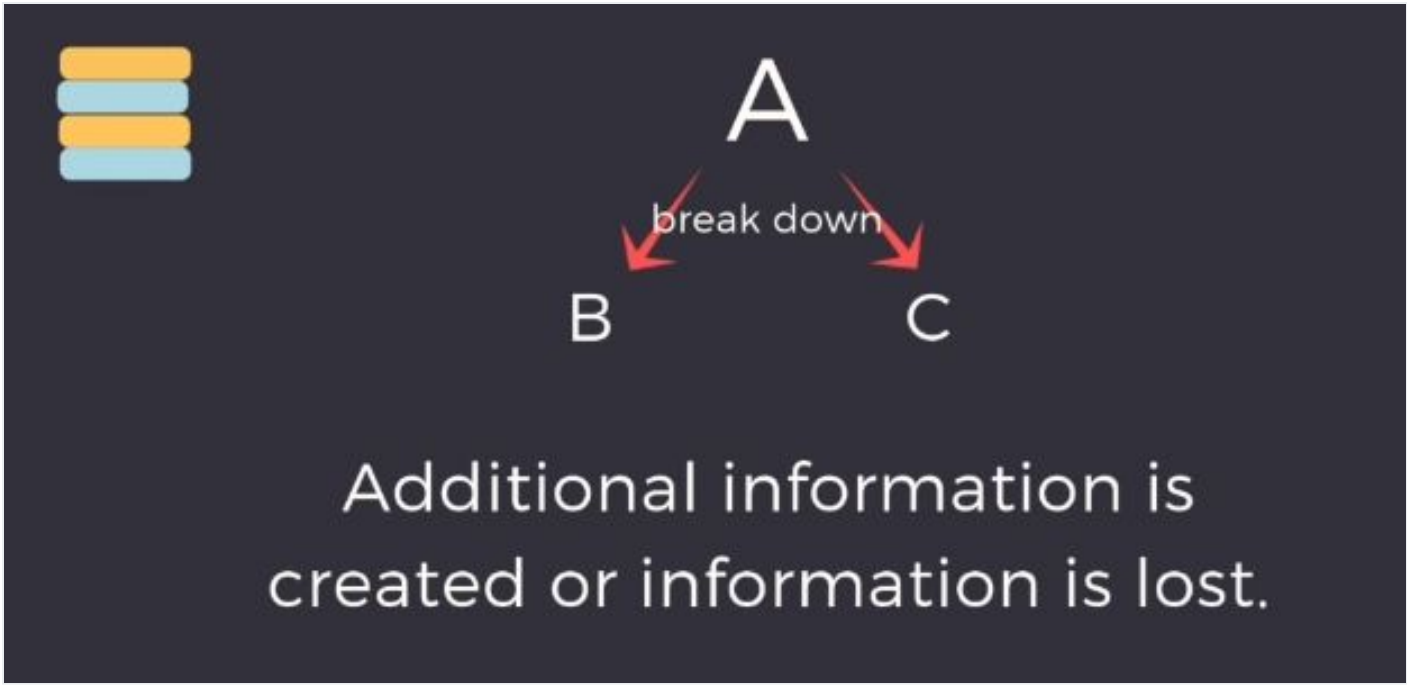
supplier	product
ACME	72X SW

SUPP_CUST TABLE

supplier	customer
ACME	FORD

CUST_PRO TABLE

customer	product
FORD	72X SW



then we should not do this breaking down of the table as it does not give surety that it is correct data,