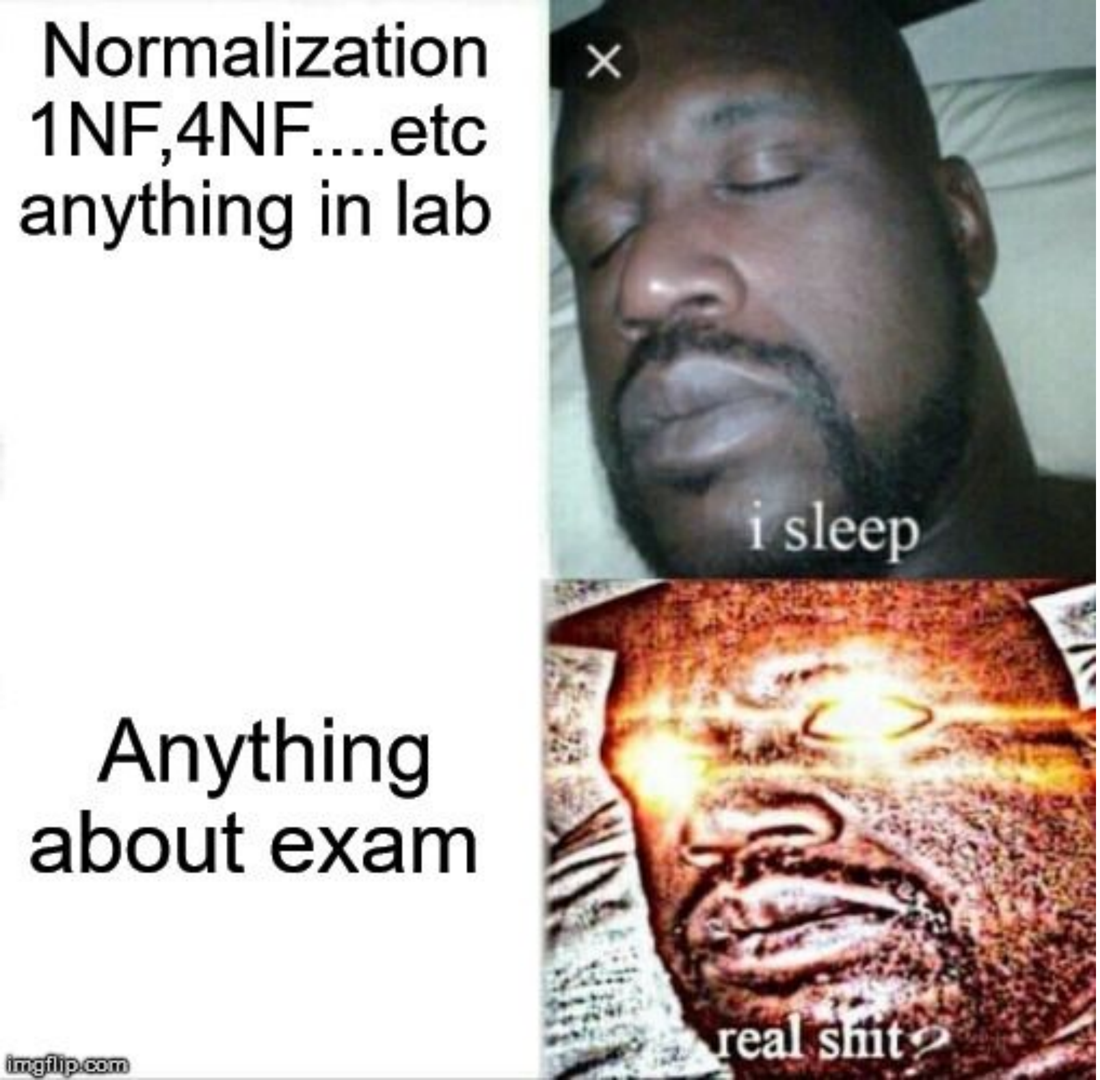

Normalization
1NF,4NF....etc
anything in lab
i sleep
Anything
about exam
real shit
imgflip.com

Databases - Tutorial 07
Normalization in SQL

Hamza Salem - Innopolis University

EMPLOYEE

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno

DEPARTMENT

Dname	Dnumber	Mgr_ssn	Mgr_start_date

DEPT_LOCATIONS

Dnumber	Dlocation

PROJECT

Pname	Pnumber	Plocation	Dnum

WORKS_ON

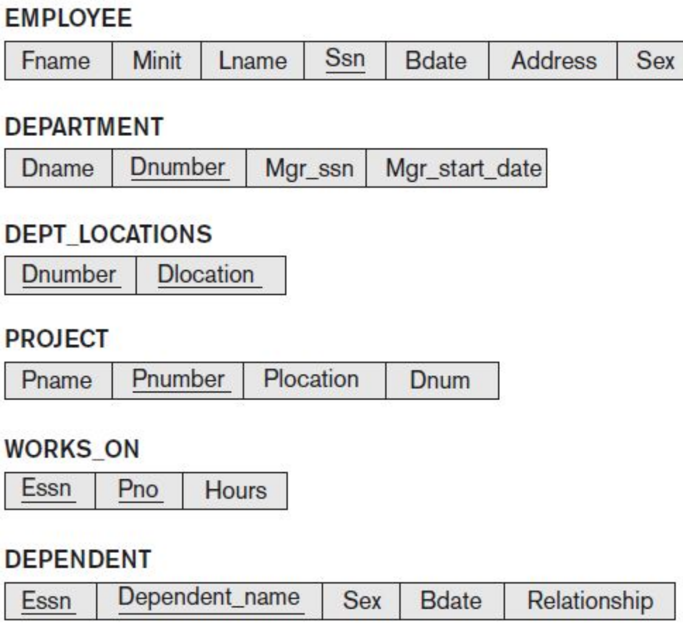
Essn	Pno	Hours

DEPENDENT

Essn	Dependent_name	Sex	Bdate	Relationship

Query 14. Retrieve all employees in department 5 whose salary is between $30,000 and $40,000.

Query 15. Retrieve a list of employees and the projects they are working on, ordered by department and, within each department, ordered alphabetically by last name, then first name.



EMPLOYEE

Fname	Minit	Lname	Ssn	Bdate	Address	Sex	Salary	Super_ssn	Dno

DEPARTMENT

Dname	Dnumber	Mgr_ssn	Mgr_start_date

DEPT_LOCATIONS

Dnumber	Dlocation

PROJECT

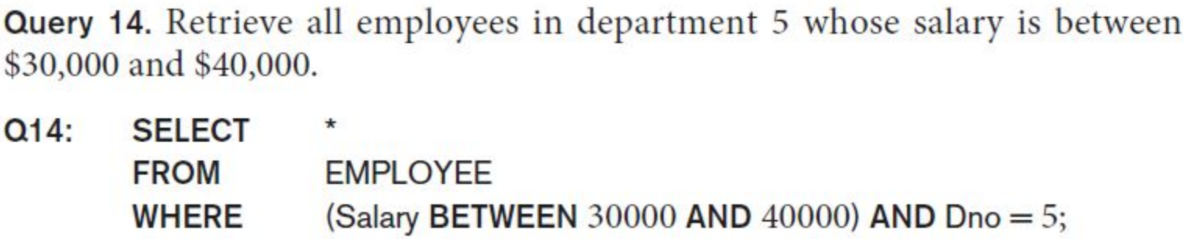
Pname	Pnumber	Plocation	Dnum

WORKS_ON

Essn	Pno	Hours

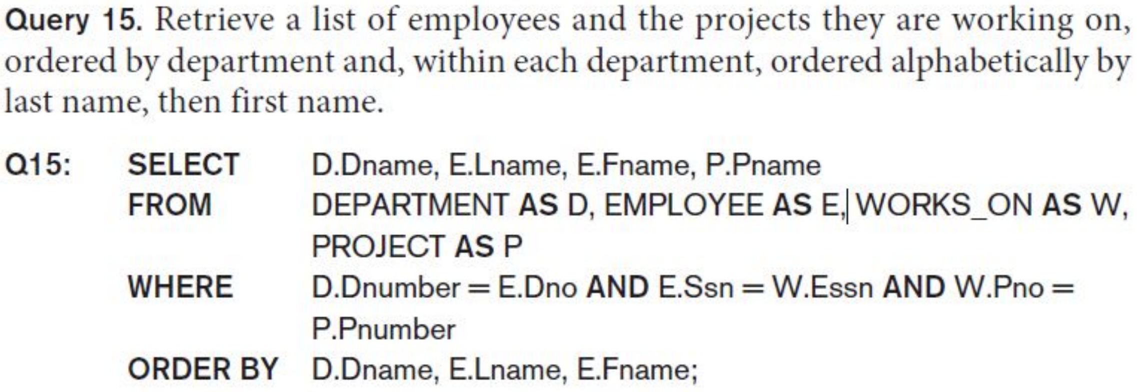
DEPENDENT

Essn	Dependent_name	Sex	Bdate	Relationship



Query 14. Retrieve all employees in department 5 whose salary is between $30,000 and $40,000.

```
Q14:    SELECT     *
        FROM       EMPLOYEE
        WHERE      (Salary BETWEEN 30000 AND 40000) AND Dno = 5;
```



Query 15. Retrieve a list of employees and the projects they are working on, ordered by department and, within each department, ordered alphabetically by last name, then first name.

```
Q15:    SELECT     D.Dname, E.Lname, E.Fname, P.Pname
        FROM       DEPARTMENT AS D, EMPLOYEE AS E, WORKS_ON AS W,
                   PROJECT AS P
        WHERE      D.Dnumber = E.Dno AND E.Ssn = W.Essn AND W.Pno =
                   P.Pnumber
        ORDER BY   D.Dname, E.Lname, E.Fname;
```

Students who missed ⅗ labs from the first 5 labs

Import DVDrental DB and Submit the following queries

DB link: https://www.postgresqltutorial.com/postgresql-sample-database/

Q1: Find customers that have rented movies priced $14?

Q2: Find customers that have rented movies priced $14 or that they spent more than $2.99 on individual rentals, but have spent a total less than $5 ?

Q3: what is the statement for this query..

```
SELECT a. customer id, a.first name, a.last name, b.total FROM customer a INNER JOIN (SELECT customer id, SUM(amount) as
total FROM payment GROUP BY customer_id ORDER BY total desc LIMIT 10) b ON a.customer_id=b.customer_id;
```

Q4: Display How Many Rentals Were Returned Late, On Time, Or Have Not Been Returned?

- Submit using github and add **@enghamzasalem** as collaborator.
- Full garde is **70%** of the first 5 labs.
- Q1,Q2,Q4 should include the output of the query as text file too.

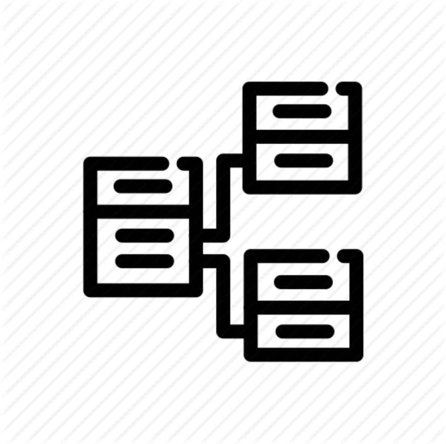
Contents

- Normalization

Organizing

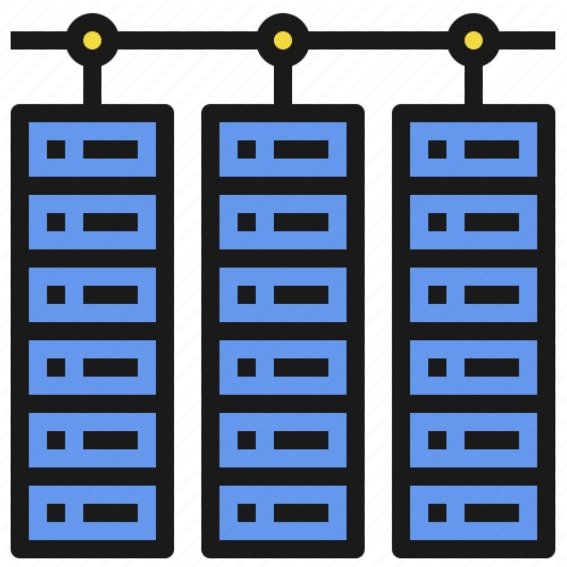
Normalization in Databases

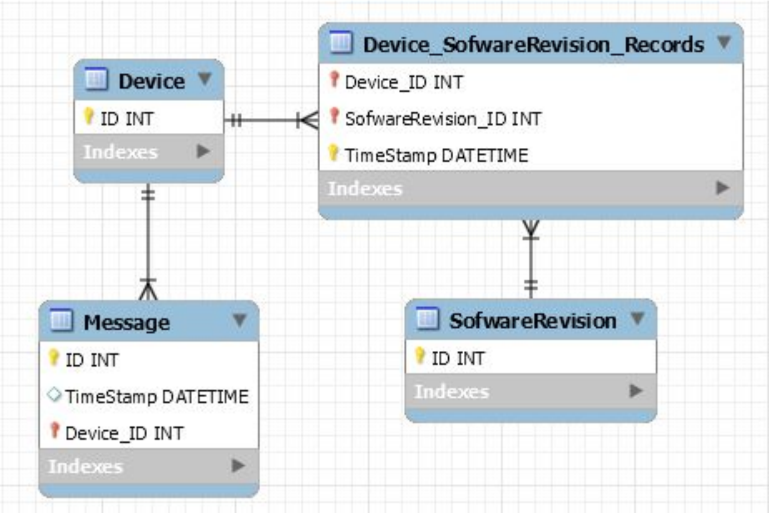
Database Normalization is a technique of organizing the data in the database. Normalization is a systematic approach of **decomposing tables** to eliminate data redundancy(repetition) and undesirable characteristics like Insertion, Update and Deletion Anomalies. It is a **multi-step process that puts data into tabular form, removing duplicated data from the relation tables**.



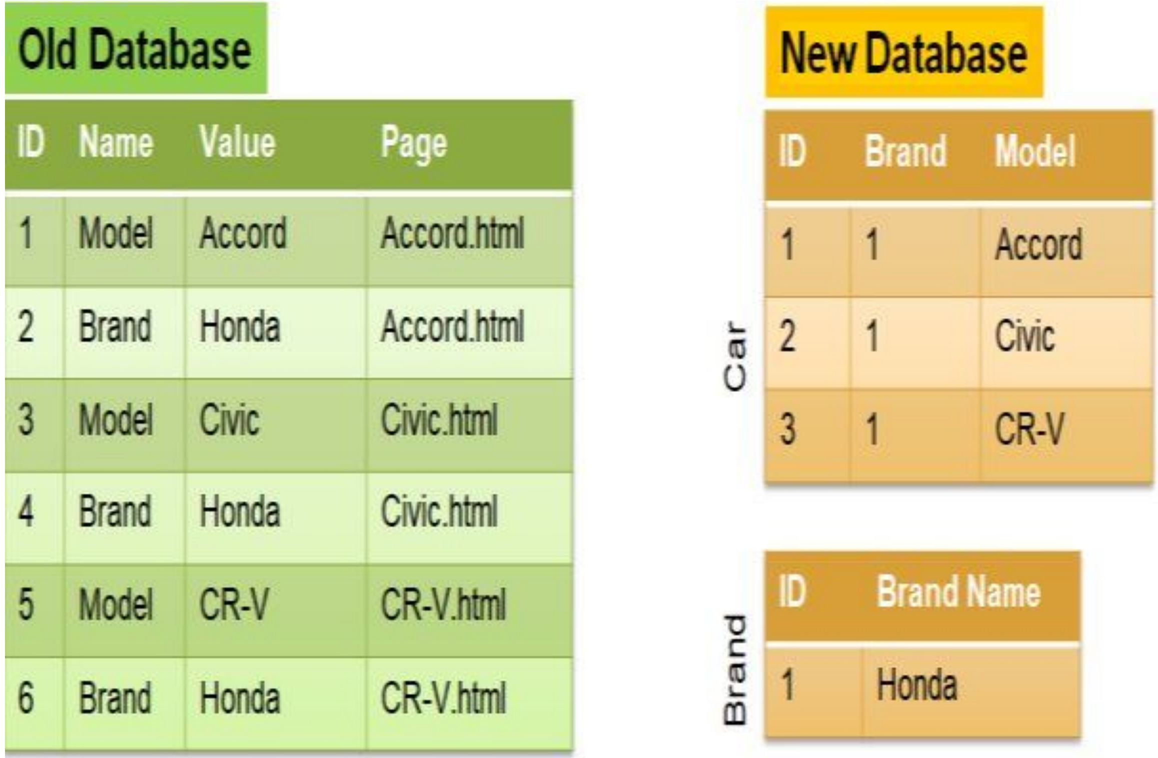
Why ?What is the purpose?

- Eliminating redundant(useless) data.
- Ensuring data dependencies make sense i.e data is logically stored.





For Example:



Old Database

ID	Name	Value	Page
1	Model	Accord	Accord.html
2	Brand	Honda	Accord.html
3	Model	Civic	Civic.html
4	Brand	Honda	Civic.html
5	Model	CR-V	CR-V.html
6	Brand	Honda	CR-V.html

New Database

Car

ID	Brand	Model
1	1	Accord
2	1	Civic
3	1	CR-V

Brand

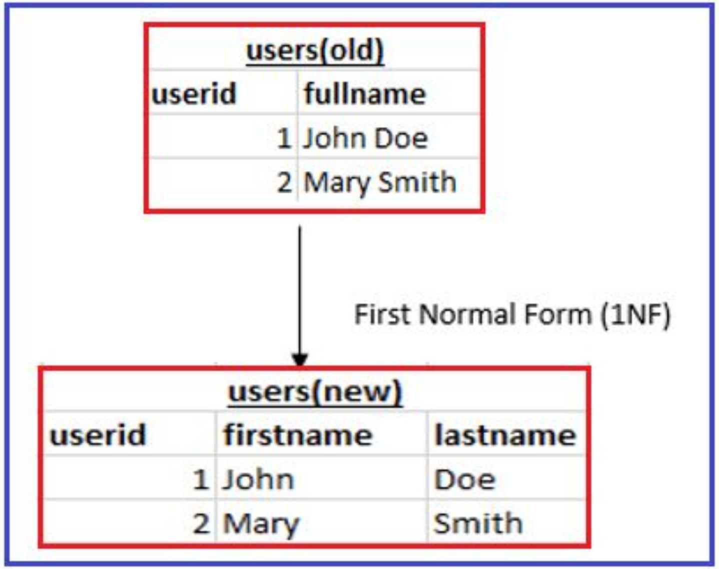
ID	Brand Name
1	Honda

Data Normalization form

1) First Normal Form
2) Second Normal Form
3) Third Normal Form
4) BCNF

First Normal Form (1NF)

- For a table to be in the First Normal Form, it should follow the following 4 rules:
  - **It should only have single(atomic) valued attributes/columns.**
  - **Values stored in a column should be of the same domain**
  - All the columns in a table should have unique names.
  - And the order in which data is stored, does not matter.



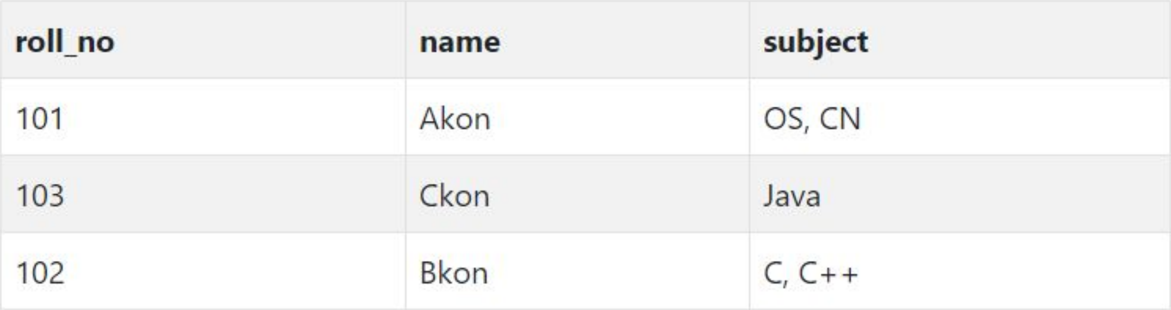
users(old)

userid	fullname
1	John Doe
2	Mary Smith

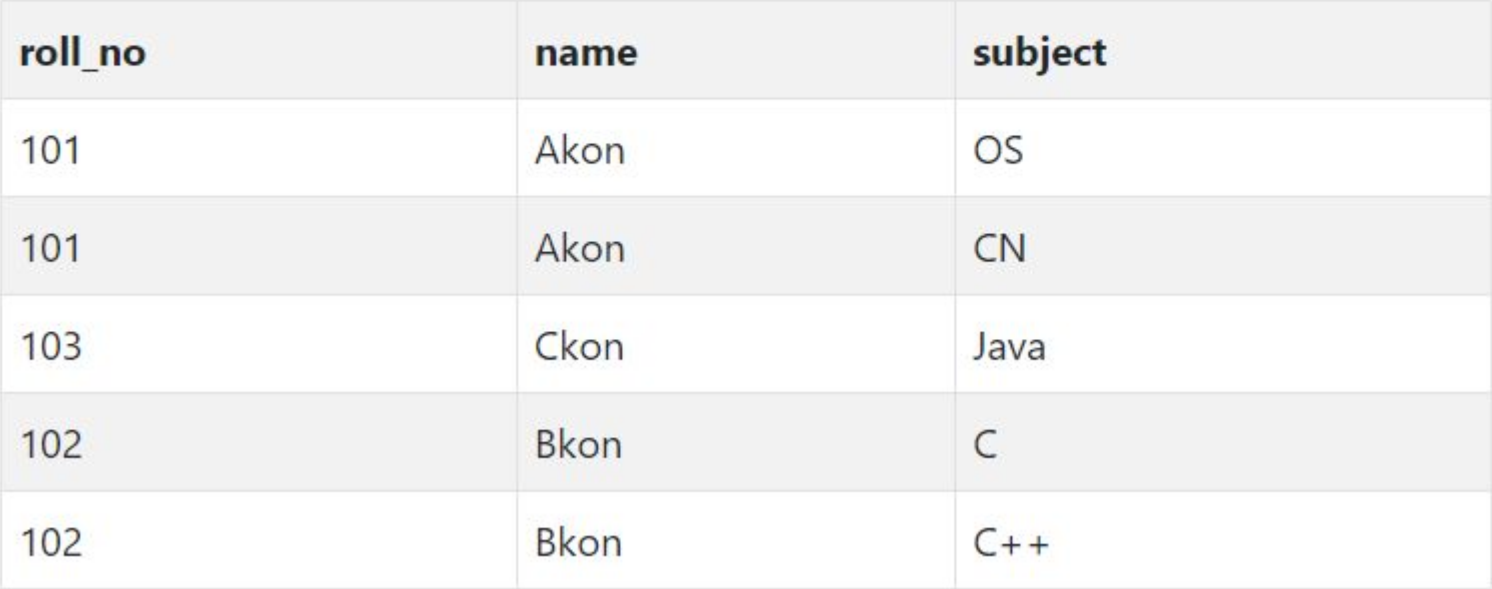
First Normal Form (1NF)

users(new)

userid	firstname	lastname
1	John	Doe
2	Mary	Smith



roll_no	name	subject
101	Akon	OS, CN
103	Ckon	Java
102	Bkon	C, C++



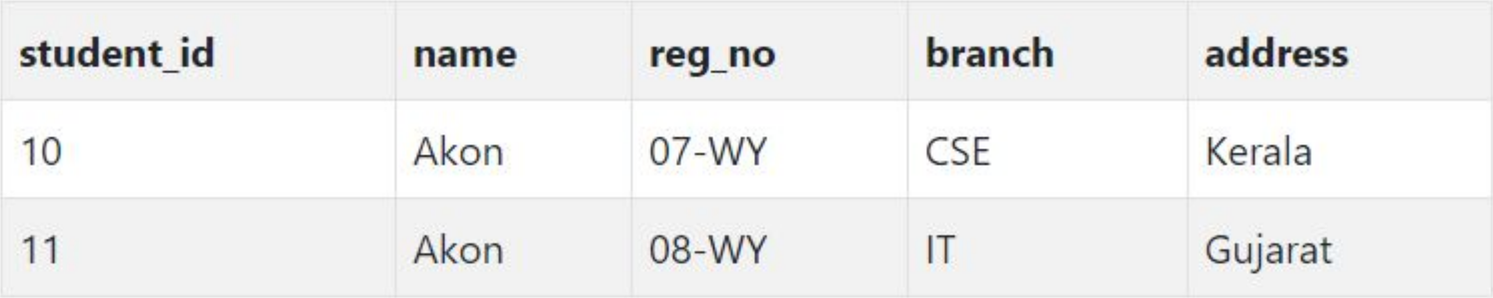
roll_no	name	subject
101	Akon	OS
101	Akon	CN
103	Ckon	Java
102	Bkon	C
102	Bkon	C++

Second Normal Form (2NF)

- For a table to be in the Second Normal Form,
  - It should be in the First Normal form.
  - And, it should not have **Partial Dependency.**

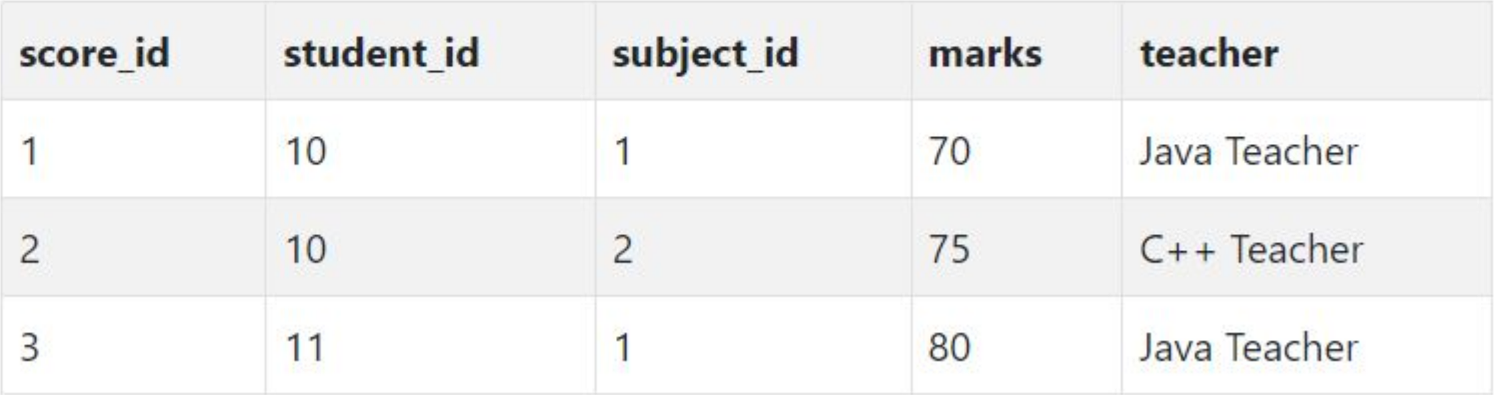
Functional dependency

But what is “Dependency” ?

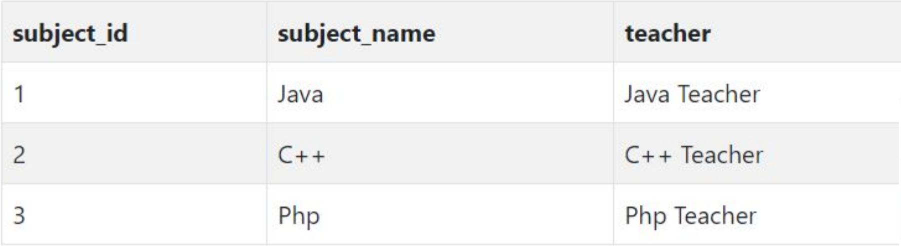


student_id	name	reg_no	branch	address
10	Akon	07-WY	CSE	Kerala
11	Akon	08-WY	IT	Gujarat

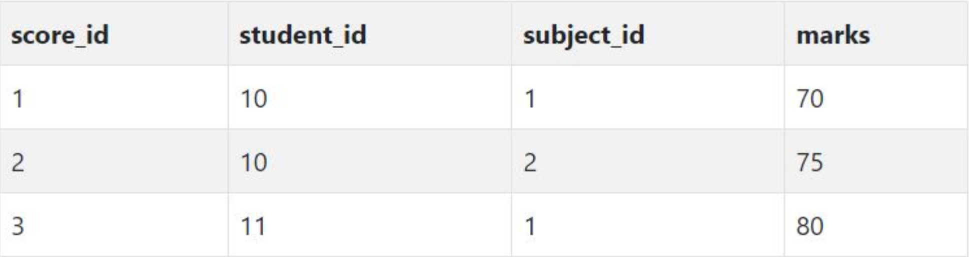
Partial dependency



score_id	student_id	subject_id	marks	teacher
1	10	1	70	Java Teacher
2	10	2	75	C++ Teacher
3	11	1	80	Java Teacher

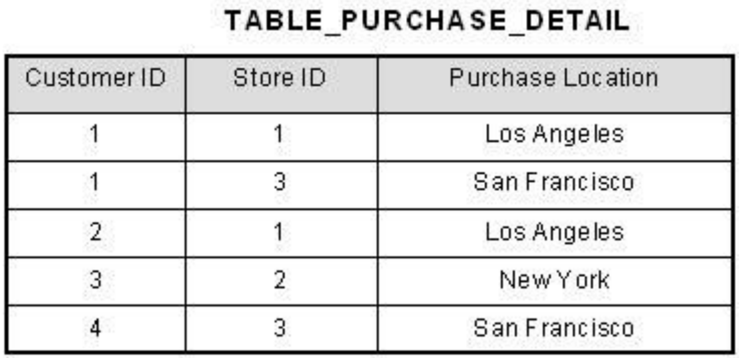


subject_id	subject_name	teacher
1	Java	Java Teacher
2	C++	C++ Teacher
3	Php	Php Teacher



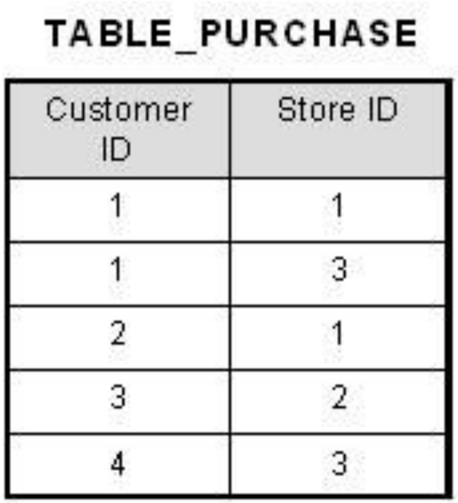
score_id	student_id	subject_id	marks
1	10	1	70
2	10	2	75
3	11	1	80

Example :



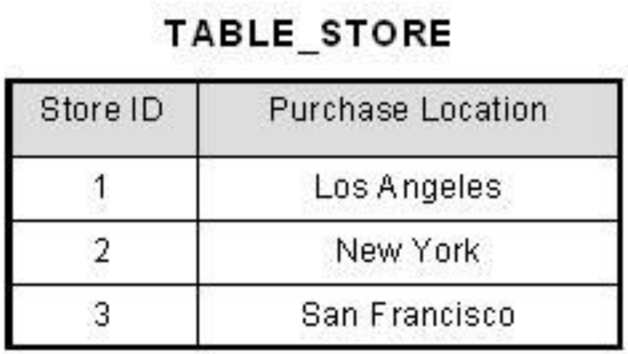
TABLE_PURCHASE_DETAIL

Customer ID	Store ID	Purchase Location
1	1	Los Angeles
1	3	San Francisco
2	1	Los Angeles
3	2	New York
4	3	San Francisco



TABLE_PURCHASE

Customer ID	Store ID
1	1
1	3
2	1
3	2
4	3



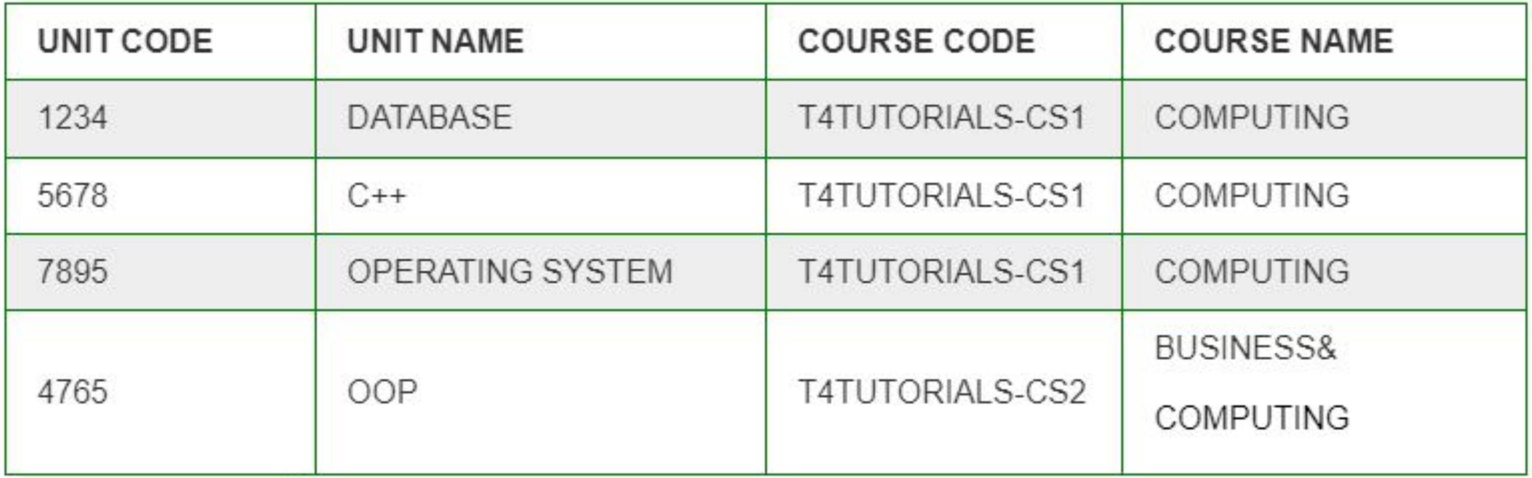
TABLE_STORE

Store ID	Purchase Location
1	Los Angeles
2	New York
3	San Francisco

Third Normal Form (3NF)

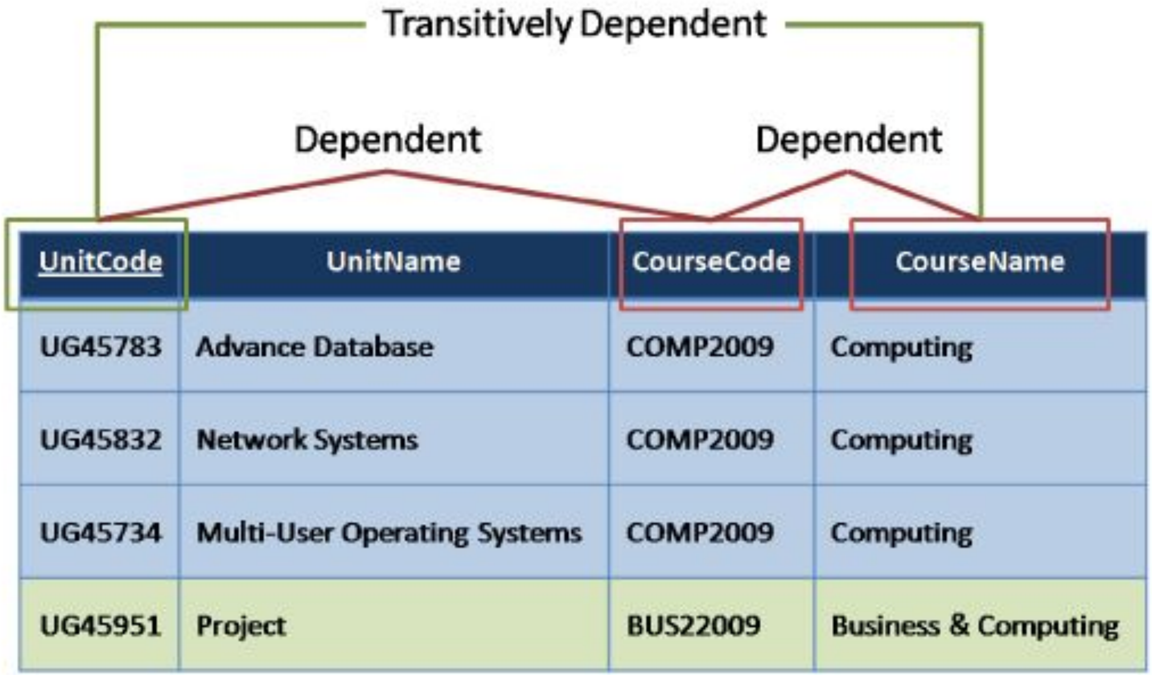
A table is said to be in the Third Normal Form when,

1. It is in the Second Normal form.
2. And, it doesn't have Transitive Dependency.



UNIT CODE	UNIT NAME	COURSE CODE	COURSE NAME
1234	DATABASE	T4TUTORIALS-CS1	COMPUTING
5678	C++	T4TUTORIALS-CS1	COMPUTING
7895	OPERATING SYSTEM	T4TUTORIALS-CS1	COMPUTING
4765	OOP	T4TUTORIALS-CS2	BUSINESS& COMPUTING

Transitive Dependency

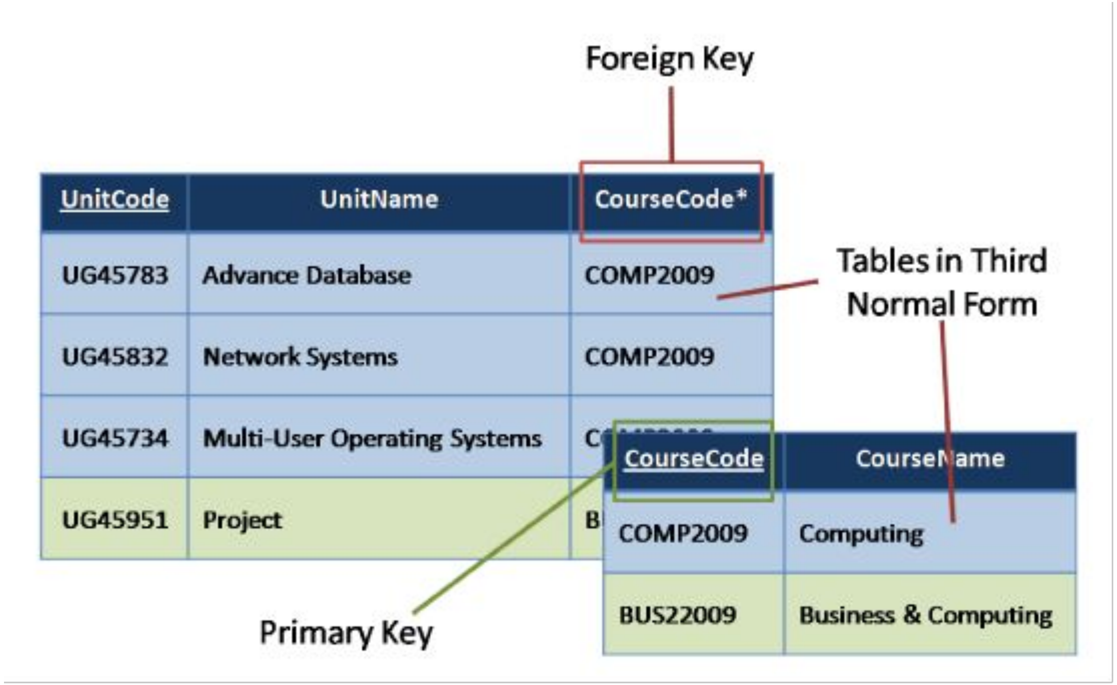

Transitively Dependent
Dependent
Dependent

UnitCode	UnitName	CourseCode	CourseName
UG45783	Advance Database	COMP2009	Computing
UG45832	Network Systems	COMP2009	Computing
UG45734	Multi-User Operating Systems	COMP2009	Computing
UG45951	Project	BUS22009	Business & Computing

Transitive Dependency

The advantage of removing transitive dependency is,

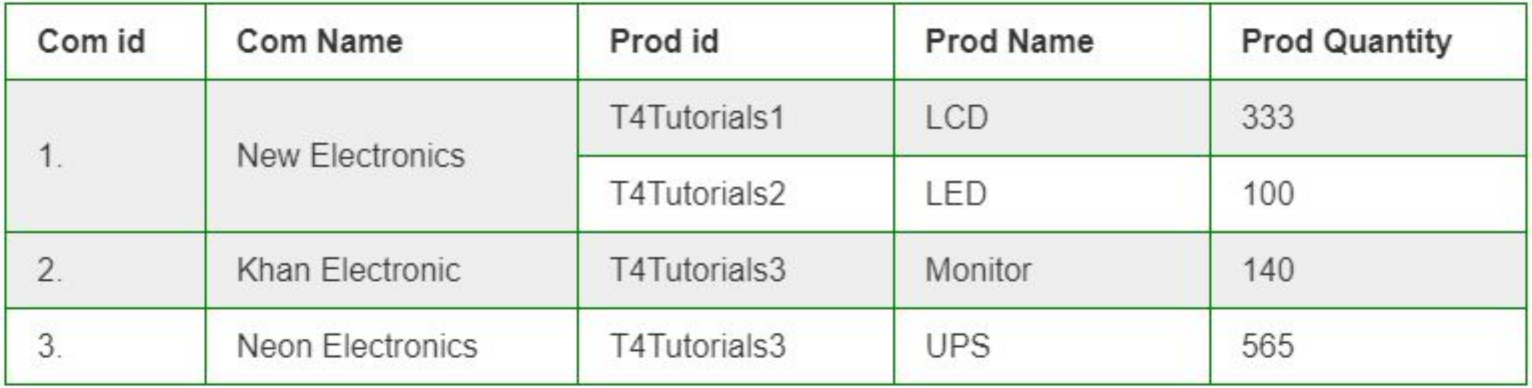
- Amount of data duplication is reduced.
- Data integrity achieved.



UnitCode	UnitName	CourseCode*
UG45783	Advance Database	COMP2009
UG45832	Network Systems	COMP2009
UG45734	Multi-User Operating Systems	C
UG45951	Project	B

CourseCode	CourseName
COMP2009	Computing
BUS22009	Business & Computing

Let us try



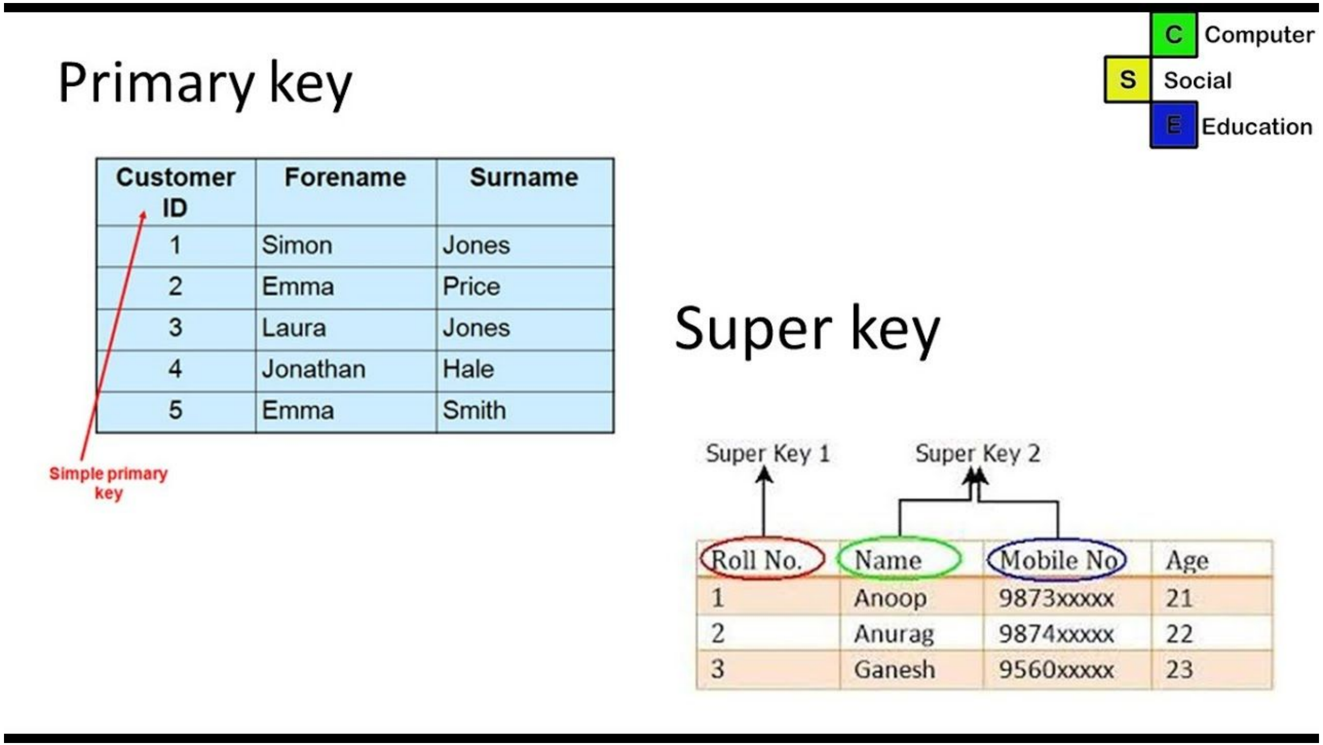
Com id	Com Name	Prod id	Prod Name	Prod Quantity
1.	New Electronics	T4Tutorials1	LCD	333
		T4Tutorials2	LED	100
2.	Khan Electronic	T4Tutorials3	Monitor	140
3.	Neon Electronics	T4Tutorials3	UPS	565

Boyce-Codd Normal Form (BCNF)

For a table to satisfy the Boyce-Codd Normal Form, it should satisfy the following two conditions:

1. It should be in the Third Normal Form.
2. And, for any dependency A → B, **A should be a super key.**

Super key vs primary key



C Computer
S Social
E Education

Primary key

Customer ID	Forename	Surname
1	Simon	Jones
2	Emma	Price
3	Laura	Jones
4	Jonathan	Hale
5	Emma	Smith

Simple primary key

Super key

Super Key 1
Super Key 2

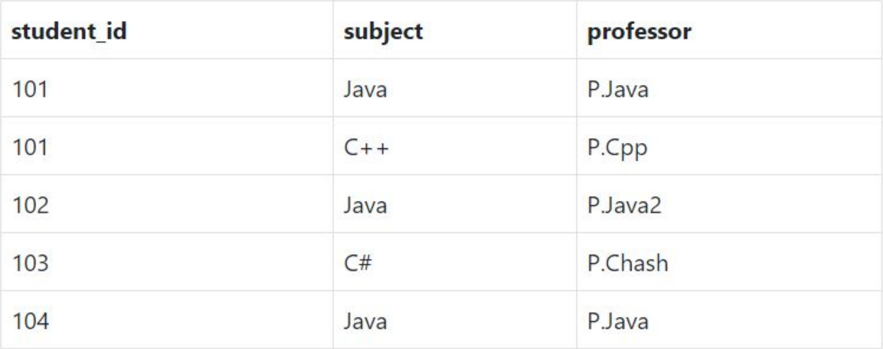
Roll No.	Name	Mobile No	Age
1	Anoop	9873xxxxx	21
2	Anurag	9874xxxxx	22
3	Ganesh	9560xxxxx	23

Dependency A → B, A should be a super key

This table satisfies the 1st Normal form because all the values are atomic, column names are unique and all the values stored in a particular column are of same domain.

This table also satisfies the 2nd Normal Form as there is no Partial Dependency.

And, there is no Transitive Dependency, hence the table also satisfies the 3rd Normal Form.



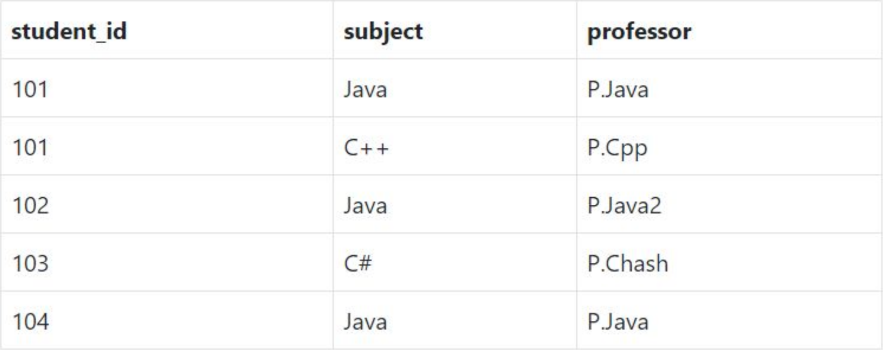
student_id	subject	professor
101	Java	P.Java
101	C++	P.Cpp
102	Java	P.Java2
103	C#	P.Chash
104	Java	P.Java

Dependency A ➜ B, A should be a super key

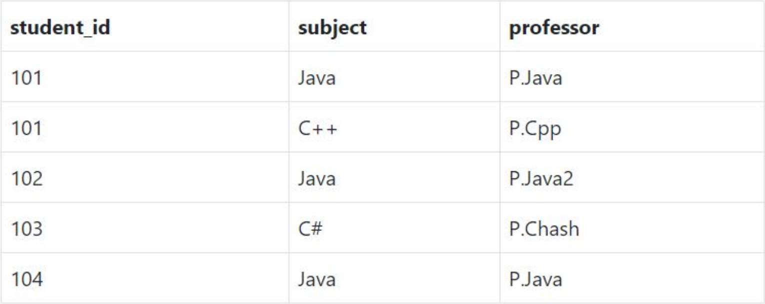
student_id, subject form primary key

But, there is one more dependency, professor → subject.

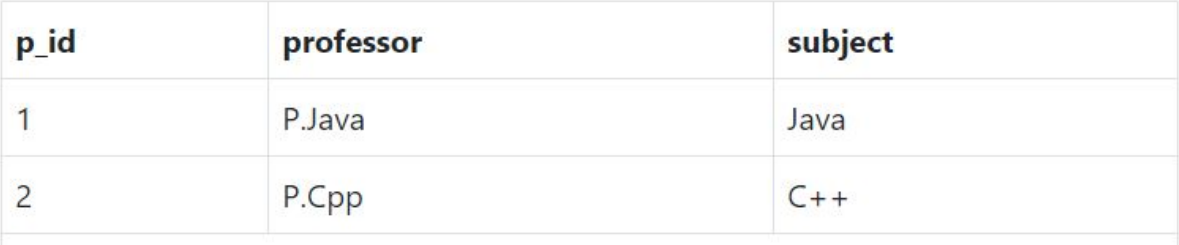
And while subject is a prime attribute, professor is a non-prime attribute, which is not allowed by BCNF



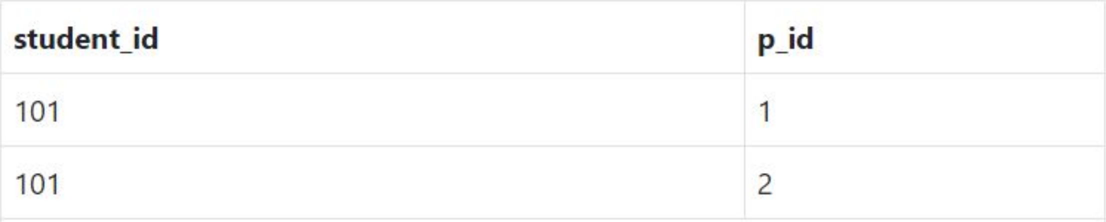
student_id	subject	professor
101	Java	P.Java
101	C++	P.Cpp
102	Java	P.Java2
103	C#	P.Chash
104	Java	P.Java



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



p_id	professor	subject
1	P.Java	Java
2	P.Cpp	C++



student_id	p_id
101	1
101	2

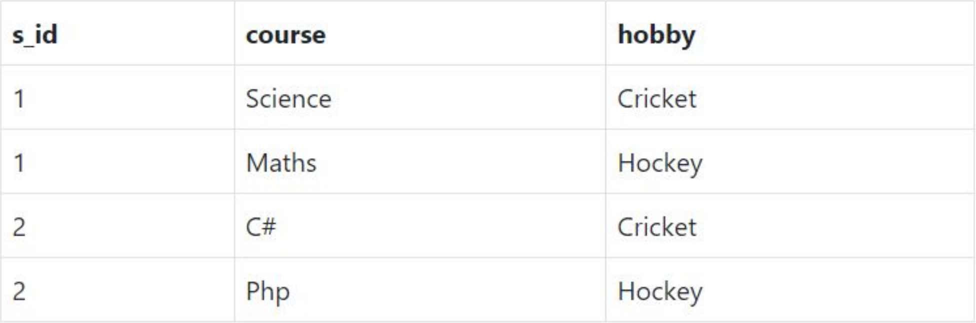
Fourth Normal Form (4NF)

For a table to satisfy the Fourth Normal Form, it should satisfy the following two conditions:

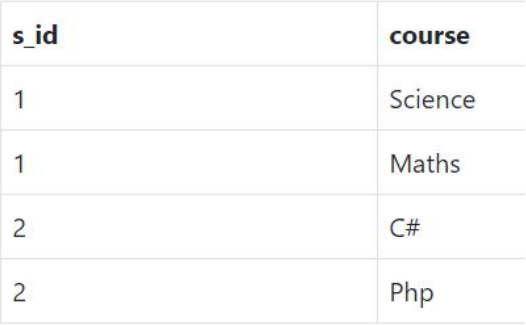
1. It should be in the Boyce-Codd Normal Form.
2. And, the table should not have any **Multi-valued Dependency**.

Multi-valued Dependency

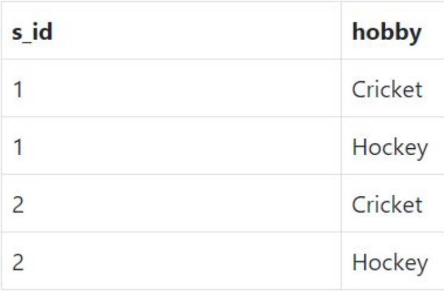
1. **For a dependency A → B, if for a single value of A, multiple value of B exists, then the table may have multi-valued dependency.**
2. **Also, a table should have at-least 3 columns for it to have a multivalued dependency.**
3. **And, for a relation R(A,B,C), if there is a multi-valued dependency between, A and B, then B and C should be independent of each other.**



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



s_id	**course**
1	Science
1	Maths
2	C#
2	Php



s_id	**hobby**
1	Cricket
1	Hockey
2	Cricket
2	Hockey

Useful links

- https://t4tutorials.com/third-normal-form-examples-3nf/
- https://www.geeksforgeeks.org/difference-between-primary-key-and-super-key/#:~:text=1.-,Super%20Key%20is%20an%20attribute%20(or%20set%20of%20attributes)%20that,all%20attributes%20in%20a%20relation.
-