Normalization is a database design technique which organizes tables in a manner that reduces redundancy and dependency of data.

It divides larger tables to smaller tables and link them using relationships.

Purpose of Normalization,

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

Problems without Normalization

Insertion, deletion, updation of database may face to data loss, without normalization.
For example:

S_id	S_Name	S_Address	S_Subject

101	Gourav	Burhanpur	Maths
102	Jitesh	Khandwa	Hindi
103	Ritesh	Sanawad	Hindi
104	Avinash	Burhanpur	Sciecne

Updation Anamoly

To update student address who repeated more than once we need to update S_Address column in all the rows, else data will become inconsistent.

Insertion Anamoly

Suppose we are not having ay value to insert in S_Subject, than we need to insert null there, which lead to insertion anamoly.

Deletion Anamoly

If we need to drop a subject , than full row need to be dropped, entire student record will be deleted, which lead to deletion anamoly.

Normalization rules

Normalization rule are divided into following normal form.

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

First Normal Form (1NF)

- All rows unique.
- All columns unique.
- Each row must have a primary key or group of keys act as primary key.

For example:

Table not in normal form-

Student	Age	Subject
Gourav	19	Hindi, Science
Rohit	18	Science
Jitesh	16	Science

In 1NF, no colums contain more than one value.

1NF of above table will be-

Student	Age	Subject
Gourav	19	Hindi
Gourav	19	Science
Rohit	18	Science
Jitesh	16	Science

Disadvantages (1NF)

- Data redundancy increases.

Advantages (1NF)

- Each row will be unique.

Second Normal Form (2NF)

In 1NF,

Student	Age	Subject
Gourav	19	Hindi
Gourav	19	Science
Rohit	18	Science
Jitesh	16	Science

In above table two rows are identical in column Student and Age.Age depends on Student.Subject also depends on student.

2NF will be,

Student	Age
Gourav	19
Rohit	18

Jitesh	16

Student	Subject
Gourav	Hindi
Gourav	Science
Rohit	Science
Jitesh	Science

Advantages (2NF):

Now above 2NF will not suffer update anomaly now.

Third Normal Form (3NF)

- Every non-prime attribute of table must depend on primary key.
- Non-prime attribute must notbe determined by another non-prime attribute. (called transitive functional dependency)
- Table must be in 2NF.

For example,

S_ID	S_Name	DOB	Street	City	State	Pincode

In this table S_ID is a primary key.

Street, City and State depends upon Pincode.

The dependency between Pincode and other fields is called transitive dependency.

So,
To apply 3NF, we need to move the Street, City and State to new table, with Pincode as primary key.

S_ID	S_Name	DOB	Pincode

Pincode	Street	City	State

Advantages (3NF)

- Data duplication is reduced.
- Data integrity achieved.

Boyce and Codd Normal Form (BCNF)

Boyce and Codd Normal Form is a higher version of the Third Normal form.

This form deals with certain type of anamoly that is not handled by 3NF.
A 3NF table which does not have multiple overlapping candidate keys is said to be in BCNF.
A 3NF table needed following conditions to be in BCNF

- It must be in 3NF.
- For each functional dependency (X->Y), X should be a super key.

Consider the following relationship: R(W,X,Y,Z)

And following dependencies:

W-> XYZ

XY-> WZ

Z-> X

Above table is already in 3NF, Keys are W and XY.

In FD,

W-> XYZ, W is the super key.

XY-> WZ, XY is also a key.

Z->X, D is not a key.

So, we can break relationship in to (W,X,Y,Z)