**SQL Normalization**

# Normalization

Normalization is the process of designing a database effectively such that we can avoid

data redundancy [data duplication], in turn which will help us to avoid anomalies such

insertion, updation and deletion anomalies.

# Different Levels

1. 1NF – First Normal Form
2. 2NF – Second Normal Form
3. 3NF – Third Normal Form
4. 4NF – Fourth Normal Form
5. BCNF – Boyce-Codd Normal Form
6. 5NF – Fifth Normal Form
7. 6NF – Sixth Normal Form

Though we have 6 Normal Form, the golden standard of normalization is 3rd Normal Form.

Most of the companies try to normalize their data till 3rd Normal Form.

Based on the need, the company / project can go beyond the third Normal Form.

# 1NF Rule

1. Every column / attribute needs to have a single value.
2. Each row should be unique. Either through a single or multiple columns. Not mandatory to have primary key.

# 2NF Rule

1. Must be in 1NF
2. All Non-Key attributes must be fully dependent on candidate key. i.e., If a non-key column is partially dependent on candidate key (subset of columns forming candidate key) then split them into separate tables.
3. Every table should have primary key and relationship between the tables should be formed using foreign key.

# Candidate Key

1. Set of columns which uniquely identify a record.
2. A table can have multiple candidate keys because there can be multiple set of columns which uniquely identify a record / row in a table.

# Non-Key Columns

Columns which are not part of the candidate key or primary key.

# Partial Dependency

If your candidate key is a combination of 2 columns (or multiple columns) then every non-key column (columns which are not part of the candidate key) should be fully dependent on all the columns. If there is any non-key column which depends only on one of the candidate key columns then this results in partial dependency.