# Normalization

Normalization is a database design technique used to organize and structure relational databases efficiently. It aims to minimize data redundancy, prevent update anomalies, and maintain data integrity, ensuring that data is stored in a consistent and optimized manner. Normalization is achieved by breaking down a database into multiple related tables and applying specific rules or normal forms to eliminate data duplication and improve data integrity.

There are several normal forms, each building upon the previous one, and they are denoted by numerals (1NF, 2NF, 3NF, etc.). Let's go through each normal form in detail:

## First Normal Form (1NF):

In 1NF, each table cell can only hold a single value (atomicity), and each column has a unique name.

The table should have a primary key that uniquely identifies each row.

There should be no repeating groups or arrays within a table cell.

Data redundancy is minimized by organizing data into separate tables.

## Second Normal Form (2NF):

For a table to be in 2NF, it must already be in 1NF.

In 2NF, each non-key column must be fully functionally dependent on the entire primary key, meaning that it should not have partial dependencies on only a part of the primary key.

If there are partial dependencies, the table is split into separate tables to ensure that each non-key attribute depends on the entire primary key.

## Third Normal Form (3NF):

For a table to be in 3NF, it must already be in 2NF.

In 3NF, each non-key column must be transitively dependent on the primary key. In other words, there should be no non-key attributes that depend on other non-key attributes.

If transitive dependencies exist, they are resolved by breaking down the table into separate tables, ensuring that non-key attributes depend only on the primary key.

## Boyce-Codd Normal Form (BCNF):

BCNF is an extension of 3NF and is primarily concerned with functional dependencies and superkeys.

A table is in BCNF if for every non-trivial functional dependency (A → B), where A is a candidate key, B is a candidate key. In simpler terms, every determinant (A) should uniquely determine every non-key attribute (B).

## Fourth Normal Form (4NF):

4NF deals with multi-valued dependencies.

A table is in 4NF if it has no multi-valued dependencies. Multi-valued dependencies occur when a non-key attribute can have multiple independent values for a single value of the primary key.

If multi-valued dependencies exist, the table is broken down into separate tables, each containing a single-valued dependency.

## Fifth Normal Form (5NF) or Project-Join Normal Form (PJNF):

5NF is the highest level of normalization and deals with join dependencies.

A table is in 5NF if it has no join dependencies. Join dependencies occur when a table can be decomposed into smaller tables without losing any information.

Achieving 5NF often involves using a technique called "projection-join," where related data is combined using joins during queries.