

Database Normalization

Normalization is a process used to organize a database into tables and columns to reduce data redundancy and improve data integrity. It involves restructuring tables to ensure data is stored efficiently while maintaining consistency. This process follows a series of rules, known as normal forms, that guide how tables should be structured.

The goals of normalization include:

- **Minimizing redundancy:** Ensuring that data is not repeated unnecessarily across tables.
- **Eliminating update anomalies:** Preventing problems that arise when modifying, deleting, or inserting data.
- **Improving data integrity:** Ensuring that the database reflects accurate and consistent information.

The most commonly applied normal forms are the First Normal Form (1NF), Second Normal Form (2NF), and Third Normal Form (3NF).

First Normal Form (1NF)

A table is in First Normal Form (1NF) if:

- All columns contain atomic (indivisible) values.
- All columns do not contain multi-valued attributes.
- Each row in the table is unique, identified by a primary key.
- There are no repeating groups or arrays within a single row.

Second Normal Form (2NF)

A table is in Second Normal Form (2NF) if:

- It is already in 1NF.
- All non-key attributes are fully dependent on the entire primary key, not just part of it.

This normal form applies to tables that have a composite primary key (a key made of multiple columns). It eliminates partial dependencies, where an attribute is dependent on only part of the primary key.

Third Normal Form (3NF)

A table is in Third Normal Form (3NF) if:

- It is already in 2NF.
- There are no transitive dependencies, meaning that non-key attributes should not depend on other non-key attributes.

Transitive dependencies occur when a non-key attribute depends on another non-key attribute rather than directly on the primary key.