# What is Data Normalization?

* Data normalization is the process of reorganizing data within a database so that users can utilize it for further queries and analysis. Simply put, it is the process of developing clean data. This includes eliminating redundant and unstructured data and making the data appear similar across all records and fields.
* Redundant data wastes disk space and creates maintenance problems. If data that exists in more than one place must be changed, the data must be changed in exactly the same way in all locations. Ex: A customer address change is easier to implement if that data is stored only in the Customers table and nowhere else in the database.
* There are a few rules for database normalization. Each rule is called a "normal form." If the first rule is observed, the database is said to be in "first normal form." If the first three rules are observed, the database is considered to be in "third normal form." Although other levels of normalization are possible, third normal form is considered the highest level necessary for most applications.

## Data Normalization Forms

* **First normal form:** The most basic form of data normalization is 1NF which ensures there are no two same entries in a group. Eliminate repeating groups in individual tables.Create a separate table for each set of related data.Identify each set of related data with a primary key.
* **Second normal form:** In a 2NF table, all the subsets of data that can be placed in multiple rows are placed in separate tables. Create separate tables for sets of values that apply to multiple records.Relate these tables with a foreign key.
* **Third normal form:** This stage takes care of transitive functional dependencies. Remove transitive dependencies, where non-key attributes depend on other non-key attributes.
* **Other normalization forms:** Fourth normal form, also called Boyce-Codd Normal Form (BCNF), and fifth normal form do exist, but are rarely considered in practical design. Disregarding these rules may result in less than perfect database design, but shouldn't affect functionality.

## The Data Normalization Process

* A database's columns, attributes, and tables, or relationships, are organized in accordance with a set of normal form rules during normalization. These normal forms act as a kind of check and balance system to preserve the integrity of relationships between the characteristics and relations and are what direct the normalization process.
* Through a set of normal forms, the normalization process seeks to guarantee that the consistency of the database is maintained regardless of whether any data is changed, added, or destroyed.

## Advantages of Data Normalization

* Utilizing database or data redundancy through normalization
* Duplication may be eliminated.
* By normalizing, we may reduce null values.
* Minimize issues with data modification.
* It makes the queries easier.
* The database structure is more comprehensible and straightforward.
* Existing data can be added to the database without having an impact.
* Because the table is compact and more rows can fit on the data page, finding, sorting, as well as indexing may be quicker.

## Need for Data Normalization

* Every business that wants to grow and flourish must regularly implement data normalization. Eliminating errors is one of the most important things you can do to make Data Analysis less complicated and time-consuming.
* These problems frequently occur when changing, adding, or removing system information. When data input error is removed, an organization will have a well-running system that is full of beneficial, useful data.
* With normalization, a company may maximize its data and engage in data collecting at a higher, more effective level. It becomes simpler to examine data to enhance a company's operations, particularly when cross-examining.