**Data normalization** is the process of organizing and transforming data to improve its structural integrity, accuracy, and consistency. Data normalization is also an important part of database design.

Data normalization is adopted because it helps to ensure that data will be consistent. This is important because if the data is inconsistent, it will be more difficult to derive useful insights from it.

Data normalization also helps to ensure that data is reliable and accurate. By ensuring that the data is consistent and reliable so that we can have greater confidence in our analyses.

Ref.( [data was sourced from here](https://u-next.com/blogs/analytics/what-is-data-normalization-and-why-is-it-important/#:~:text=Cost%20reduction:%20Normalized%20data%20helps%20you%20reduce,applications%20because%20all%20the%20data%20is%20standardized))

**TLDR;Normalization** in database design is the process of organizing the data to minimize redundancy and improve data integrity. The main goal is to divide large tables into smaller, more manageable ones and to define relationships between them. This makes the database more efficient and reduces the chance of data anomalies.

**Steps to Normalize a Database**

**1] Understand Your Data**

Begin by understanding the data and the relationships between different data elements. This involves identifying the key pieces of information (entities) and how they interact.

**2] Define Primary Keys**

Each table should have a primary key, a unique identifier for the records in that table. This ensures that each record can be uniquely identified.

**3] Apply the Normal Forms**

First Normal Form (1NF)

* Ensure that each table has a primary key.
* Eliminate duplicate columns from the same table.
* Create separate tables for each group of related data and identify each row with a unique column (the primary key).

Example:

Before 1NF:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| OrderID | CustomerName | Product1 | Product2 | Product3 |
| 1 | Matthew Collins | Thermodynamics 101 | Chemistry for Beginners |  |
| 2 | Sean Ryker | What is Compounding? |  |  |

After 1NF:

|  |  |
| --- | --- |
| OrderID | CustomerName |
| 1 | Matthew Collins |
| 2 | Sean Ryker |
| OrderID | Product |
| 1 | Thermodynamics 101 |
| 1 | Chemistry for Beginners |
| 2 | What is Compounding? |

***Second Normal Form (2NF)***

Ensure that the table is in 1NF, move data that is only dependent on part of a composite primary key to a separate table.

Before 2NF:

|  |  |  |
| --- | --- | --- |
| OrderID | CustomerID | ProductID |
| 1 | 101 | 201 |
| 1 | 101 | 202 |
| 2 | 102 | 203 |

After 2NF:

|  |  |
| --- | --- |
| OrderID | CustomerID |
| 1 | 101 |
| 2 | 102 |
| OrderID | ProductID |
| 1 | 201 |
| 1 | 202 |
| 2 | 203 |
| CustomerID | CustomerName |
| 101 | Matthew Collins |
| 102 | Sean Ryker |

***Third Normal Form (3NF)***

Ensure that the table is in 2NF, remove columns that are not dependent on the primary key.

Before 3NF:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| OrderID | CustomerID | CustomerName | ProductID | ProductName |
| 1 | 101 | Matthew Collins | 201 | Thermodynamics 101 |
| 2 | 102 | Sean Ryker | 203 | What is Compounding? |

After 3NF:

|  |  |
| --- | --- |
| OrderID | CustomerID |
| 1 | 101 |
| 2 | 102 |
| CustomerID | CustomerName |
| 101 | Matthew Collins |
| 102 | Sean Ryker |
| ProductID | ProductName |
| 201 | Thermodynamics 101 |
| 203 | What is Compounding? |
| OrderID | ProductID |
| 1 | 201 |
| 1 | 202 |
| 2 | 203 |

**4] Test the Database**

After implementation, test the database to ensure it performs efficiently and maintains data integrity. Run queries to verify that the relationships and data are correctly set up.

**References:**

1. **(**<https://www.essentialsql.com/get-ready-to-learn-sql-database-normalization-explained-in-simple-english>)
2. (<https://www.databasestar.com/database-normalization>/)