**What is normalization?**

**Definition :** Normalization is the process of efficiently organizing data in a database. There are two goals of the normalization process: eliminating redundant data (for example, storing the same data in more than one table) and ensuring data dependencies make sense (only storing related data in a table). Both of these are worthy goals as they reduce the amount of space a database consumes and ensure that data is logically stored. There are several benefits for using Normalization in Database.

**Benefits:**

1. Eliminate data redundancy
2. Improve performance
3. Query optimization
4. Faster update due to less number of columns in one table
5. Index improvement

Different types of Normalizations form available in the Database. Let’s see one by one.

1. First Normal Form (1NF)

 First normal form (1NF) sets the very basic rules for an organized database:

* Eliminate duplicative columns from the same table.
* Create separate tables for each group of related data and identify each row with a unique column or set of columns (the primary key).
  1. Remove repetitive groups
  2. Create Primary Key

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
| **Name** | **State** | **Country** | **Phone1** | **Phone2** | **Phone3** |
| John | 101 | 1 | 488-511-3258 | 781-896-9897 | 425-983-9812 |
| Bob | 102 | 1 | 861-856-6987 |  |  |
| Rob | 201 | 2 | 587-963-8425 | 425-698-9684 |  |
| **PK** |  |  | **[ Phone Nos ]** | | |
| ? |  |  |  | ? |  |
| **ID** | **Name** | **State** | **Country** | **Phone** |  |
| 1 | John | 101 | 1 | 488-511-3258 |  |
| 2 | John | 101 | 1 | 781-896-9897 |  |
| 3 | John | 101 | 1 | 425-983-9812 |  |
| 4 | Bob | 102 | 1 | 861-856-6987 |  |
| 5 | Rob | 201 | 2 | 587-963-8425 |  |
| 6 | Rob | 201 | 2 | 425-698-9684 |  |
|  |  |  |  |  |  |

2. **Second Normal Form (2NF)** Second normal form (2NF) further addresses the concept of removing duplicative data:

         Meet all the requirements of the first normal form.

         Remove subsets of data that apply to multiple rows of a table and place them in separate tables.

·         Create relationships between these new tables and their predecessors through the use of [foreign keys](http://databases.about.com/library/glossary/bldef-foreignkey.htm).

Remove columns which create duplicate data in a table and related a new table with Primary Key – Foreign Key relationship

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ID | Name | State | Country | Phone |  |  |  |
| 1 | John | 101 | 1 | 488-511-3258 |  |  |  |
| 2 | John | 101 | 1 | 781-896-9897 |  |  |  |
| 3 | John | 101 | 1 | 425-983-9812 |  |  |  |
| 4 | Bob | 102 | 1 | 861-856-6987 |  |  |  |
| 5 | Rob | 201 | 2 | 587-963-8425 |  |  |  |
| 6 | Rob | 201 | 2 | 425-698-9684 |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| ID | Name | State | Country |  | PhoneID | ID | Phone |
| 1 | John | 101 |  |  | 1 | 1 | 488-511-3258 |
| 2 | Bob | 102 |  | 2 | 1 | 781-896-9897 |
| 3 | Rob | 201 |  | 3 | 1 | 425-983-9812 |
|  |  |  |  |  | 4 | 2 | 587-963-8425 |
|  |  |  |  |  | 5 | 3 | 587-963-8425 |
|  |  |  |  |  | 6 | 3 | 425-698-9684 |

3. **Third Normal Form (3NF)**

Third normal form (3NF) goes one large step further:

         Meet all the requirements of the second normal form.

         Remove columns that are not dependent upon the primary key.

  Country can be derived from State also… so removing country

|  |  |  |  |
| --- | --- | --- | --- |
| ID | Name | State | Country |
| 1 | John | 101 | 1 |
| 2 | Bob | 102 | 1 |
| 3 | Rob | 201 | 2 |

**4. Fourth Normal Form (4NF)**

Finally, fourth normal form (4NF) has one additional requirement:

         Meet all the requirements of the third normal form.

         A relation is in 4NF if it has no multi-valued dependencies.

If PK is composed of multiple columns then all non-key attributes should be derived from FULL PK only. If some non-key attribute can be derived from partial PK then remove it.

The 4NF also known as BCNF NF

|  |  |  |  |
| --- | --- | --- | --- |
| TeacherID | StudentID | SubjectID | StudentName |
| 101 | 1001 | 1 | John |
| 101 | 1002 | 2 | Rob |
| 201 | 1002 | 3 | Bob |
| 201 | 1001 | 2 | Rob |
|  |  |  |  |
|  |  |  |  |
| TeacherID | StudentID | SubjectID | StudentName |
| 101 | 1001 | 1 | X |
| 101 | 1002 | 2 | X |
| 201 | 1001 | 3 | X |
| 201 | 1002 | 2 | X |

**De- Normalization.**

Normalization is the process of reducing data redundancy and maintains data integrity. This is performed by creating relationships among tables through primary and foreign keys. Normalization procedure includes 1NF, 2NF, 3NF, BCNF, and then the data is normalized.  
  
Denomalization on the contrary is the process of adding redundant data to speed up complex queries involving multiple table JOINS. One might just go to a lower form of Normalization to achieve Denormalization and better performance. Data is included in one table from another in order to eliminate the second table which reduces the number of JOINS in a query and thus achieves performance.