# Normalization

### Definition :-

* Normalization is a step by step process which we reduce the data redundancy and eliminates issues caused by anomalies while performing Insertions, Updates and Deletions. In this we remove unnecessary data that has become outdated or no longer useful.
* Normalization rules divides larger tables into smaller tables and links them using relationships.
* The purpose of Normalization in SQL is to eliminate redundant (repetitive) data and ensure data is stored logically.

Advantages of Normalization :

* 1. Data Consistency
  2. Integrity
  3. Minimize data redundancy

Advantages of Normalization :

1. 1NF disallow the multivalued attributes
2. Performance decreases when we normalizing the table

To higher normal forms

Reference From [: https://www.datacamp.com/tutorial/normalization-in-sql](:%20https:/www.datacamp.com/tutorial/normalization-in-sql)

## There are 6 types of Normalization Form :-

1. 1NF (First Normal Form)
2. 2NF (Second Normal Form)
3. 3NF (Third Normal Form)
4. BCNF (Boyce-Codd Normal Form)
5. 4NF (Fourth Normal Form)
6. 5NF (Fifth Normal Form)
7. 1NF (First Normal Form) :

* Each table cell should contain a single value(atomic)
* Each record needs to be unique
* Disallow the multi valued attributes

Example:-

Relation “name” is not in 1NF because of multivalued attributes in “Phone\_no.” column

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Roll no | City | Phone\_no. |
| Yash | 28 | Nagpur | 8549632157 |
| Tanmay | 12 | Pune | 9652312547,8965745030 |
| Abhanave | 25 | Mumbai | 7894578965 |
| Tejas | 19 | Bangalore | 8596321020 |

After 1 Normal Form :

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Roll no | City | Phone\_no. |
| Yash | 28 | Nagpur | 8549632157 |
| Tanmay | 12 | Pune | 9652312547 |
| Tanmay | 12 | Pune | 8965745030 |
| Abhanave | 25 | Mumbai | 7894578965 |
| Tejas | 19 | Bangalore | 8596321020 |

1. 2NF (Second Normal Form) :-

* A relation will be in 2NF if it is in 1NF
* Not contain any partial dependency(primary key should not contain duplicates values)

Example:-

In this “Roll no” is a primary Key but they are repeated more than one time so they are partial dependency ,

We remove to this duplicate values to form 2NF.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Student\_name | Roll no | City | Course | Phone\_no. |
| Yash | 28 | Nagpur | DA | 8549632157 |
| Yash | 28 | Nagpur | DA | 8549632157 |
| Tanmay | 12 | Pune | AI/ML | 9652312547 |
| Tanmay | 12 | Pune | AI/ML | 8965745030 |
| Abhanave | 25 | Mumbai | DS | 7894578965 |
| Abhanave | 25 | Mumbai | DS | 7894578965 |
| Tejas | 19 | Bangalore | CSE | 8596321020 |
| Tejas | 19 | Bangalore | CSE | 8596321020 |

After 2 Normal Form :

We create two tables which include “Roll no” , “Student\_name” , “course”

“Roll no “ is a primary key

1st Table 2nd Table

|  |  |  |
| --- | --- | --- |
| Roll no | City | Phone\_no. |
| 28 | Nagpur | 8549632157 |
| 28 | Nagpur | 8549632157 |
| 12 | Pune | 9652312547 |
| 12 | Pune | 8965745030 |
| 25 | Mumbai | 7894578965 |
| 25 | Mumbai | 7894578965 |
| 19 | Bangalore | 8596321020 |
| 19 | Bangalore | 8596321020 |

|  |  |  |
| --- | --- | --- |
| Student\_name | Roll no | Course |
| Yash | 28 | DA |
| Tanmay | 12 | AI/ML |
| Abhanave | 25 | DS |
| Tejas | 19 | CSE |

1. 3NF (Third Normal Form) :-

* The third normal form states that you should eliminate fields in a table that do not depend on the key.
* A Table is already in 2 NF
* Non-Primary key columns shouldn’t depend on the other non-Primary key columns
* There is no transitive functional dependency

Example :-

“Roll no.” is a primary key and other all columns is a non primary key , So first the table is in the 2NF form then we convert them into 3NF form and there is no transitive dependency.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Student\_name | Roll no | City | Course | Phone\_no. |
| Yash | 28 | Nagpur | DA | 8549632157 |
| Tanmay | 12 | Pune | AI/ML | 8965745030 |
| Abhanave | 25 | Mumbai | DS | 7894578965 |
| Tejas | 19 | Bangalore | CSE | 8596321020 |

After 3 Normal Form :

|  |  |  |
| --- | --- | --- |
| Student\_name | Roll no | Course |
| Yash | 28 | DA |
| Tanmay | 12 | AI/ML |
| Abhanave | 25 | DS |
| Tejas | 19 | CSE |

No transitive dependency. transitive dependency

So this is in 3NF

|  |  |
| --- | --- |
| City | Phone\_no. |
| Nagpur | 8549632157 |
| Nagpur | 8549632157 |
| Pune | 9652312547 |
| Pune | 8965745030 |
| Mumbai | 7894578965 |
| Mumbai | 7894578965 |
| Bangalore | 8596321020 |
| Bangalore | 8596321020 |

1. BCNF (Boyce-Codd Normal Form) :-

* Even when a database is in 3rd Normal Form, still there would be anomalies resulted if it has more than one Candidate Key.
* Sometimes is BCNF is also referred as 3.5 Normal Form.
* For any dependency A -> B (A should be a superkey)

Example :-

|  |  |  |  |
| --- | --- | --- | --- |
| Student\_name | Roll no | City | Course |
| Yash | 28 | Nagpur | DA |
| Tanmay | 12 | Pune | AI/ML |
| Abhanave | 25 | Mumbai | DS |
| Tejas | 19 | Bangalore | CSE |

After BCNF Normal Form :

Student\_name -> Roll no City -> Course

Student\_name = Super Key City = super key

Roll no depends on Student\_name column Course depends on City Column

|  |  |
| --- | --- |
| Student\_name | Roll no |
| Yash | 28 |
| Tanmay | 12 |
| Abhanave | 25 |
| Tejas | 19 |

|  |  |
| --- | --- |
| City | Course |
| Nagpur | DA |
| Pune | AI/ML |
| Mumbai | DS |
| Bangalore | CSE |

1. 4NF (Fourth Normal Form) :-

* If no database table instance contains two or more independent and multivalued data describing the relevant entity then it is in 4th Normal Form.
* This is a normalization level that builds on BCNF by dealing with multi-valued dependencies.

Example :-

The Below Table is in BCNF form in which Roll\_no is a primary Key and city depends on Roll no and course depends on city.

|  |  |  |  |
| --- | --- | --- | --- |
| Roll no | Student\_name | City | Course |
| 28 | Yash | Nagpur | DA |
| 12 | Tanmay | Pune | AI/ML |
| 25 | Abhanave | Mumbai | DS |
| 19 | Tejas | Bangalore | CSE |

After 4 Normal Form :

Remove the multivalued dependency (only one columns depends on one primary key column)

|  |  |
| --- | --- |
| Roll no | City |
| 28 | Nagpur |
| 12 | Pune |
| 25 | Mumbai |
| 19 | Bangalore |

|  |  |
| --- | --- |
| Roll no | Student\_name |
| 28 | Yash |
| 12 | Tanmay |
| 25 | Abhanave |
| 19 | Tejas |

|  |  |
| --- | --- |
| Roll no | Course |
| 28 | DA |
| 12 | AI/ML |
| 25 | DS |
| 19 | CSE |

1. 5NF (Fifth Normal Form) :-

* A table is in 5th Normal Form only if it is in 4NF and it cannot be decomposed into any number of smaller tables without loss of data.

Example :-

In this Form we divide the tables into smaller tables

|  |  |  |
| --- | --- | --- |
| Student\_name | Roll no | Course |
| Yash | 28 | DA |
| Tanmay | 12 | AI/ML |
| Abhanave | 25 | DS |
| Tejas | 19 | CSE |

After 5 Normal Form :

|  |  |
| --- | --- |
| Student\_name | Roll no |
| Yash | 28 |
| Tanmay | 12 |
| Abhanave | 25 |
| Tejas | 19 |

|  |  |
| --- | --- |
| Roll no | Course |
| 28 | DA |
| 12 | AI/ML |
| 25 | DS |
| 19 | CSE |

|  |  |
| --- | --- |
| Student\_name | Course |
| Yash | DA |
| Tanmay | AI/ML |
| Abhanave | DS |
| Tejas | CSE |