**Normalization**

* **Definition of Normalization:**

In DBMS, Normal forms is combination of guidelines where we have to ensure that database is efficient, designed properly, organized and free from data anomalies. Normalization is used to reduce redundancy from the database table. It is the process of organizing data in the database. It is also used to remove characteristics like insertion, deletion, update anomalies. Normalization divides large data in smaller data and links them.

* **Types Of Normal Forms**

1. **First Normal Forms (1NF)**

In first normal form, if multiple values present in the cell so that it will arrange in column. If it does not contain multi-valued so that it is in 1NF.

|  |  |  |
| --- | --- | --- |
| Stud\_Id | Stud\_name | Phone no. |
| 1 | Ram | 9845864782 |
| 2 | Vishal | 9845864782  9686544524 |
| 3 | Rohan | 9764526512 |

Converting to 1NF

|  |  |  |
| --- | --- | --- |
| Stud\_id | Stud\_name | Phone no. |
| 1 | Ram | 9845864782 |
| 2 | Vishal | 9845864782 |
| 3 | Vishal | 9686544524 |
| 4 | Rohan | 9764526512 |

1. Second Normal Forms (2NF)

In second normal forms, if it is in 1NF and any non prime attribute is not dependent on any subset of the table. Every non-prime attribute must be dependent on any candidate key of the table.

|  |  |  |
| --- | --- | --- |
| Roll no | Course\_name | Course\_fee |
| 1 | Artificial Intelligence | 100000 |
| 2 | Computer Technology | 50000 |
| 3 | Electrical | 30000 |
| 4 | Data Science | 70000 |

Converting to 2NF

|  |  |
| --- | --- |
| Course\_name | Course\_fee |
| Artificial Intelligence | 100000 |
| Computer Technology | 50000 |
| Electrical | 30000 |
| Data Science | 70000 |

|  |  |
| --- | --- |
| Roll no | Course\_name |
| 1 | Artificial Intelligence |
| 2 | Computer Technology |
| 3 | Electrical |
| 4 | Data Science |

1. Third Normal Forms (3NF)

In Third Normal Forms, we need 2NF for that all non-key attributes are independent of each other. So that all columns should be directly related to the primary key and not any other column of the table. There is no transitive dependency for non-prime attributes.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Stud\_id | name | country | state | age |
| 1 | Sakshi | India | Haryana | 21 |
| 2 | Bhakti | India | Punjab | 22 |
| 3 | Avani | India | Maharashtra | 23 |

Table : Student

Student(Stud\_id, name, state, age)

State\_country (state, country)

1. Boyce-Codd Normal Forms (BCNF)

BCNF ensures that each non-key attribute is dependent on candidate key. It is the strict form of 3NF that is determinant in a table is a candidate key. It must be present super key in that table.

1. Fourth Normal Forms (4NF)

4NF table does not contain any multi valued dependencies. It is built on 1NF, 2NF, 3Nf and BCNF. Is must not containing any multi value dependencies.

|  |  |
| --- | --- |
| cid | cname |
| C1 | Math |
| C2 | Science |

|  |  |
| --- | --- |
| Sr. | name |
| S1 | Kirti |
| S2 | Nayan |

|  |  |  |  |
| --- | --- | --- | --- |
| Sr. | name | cid | cname |
| S1 | Kirti | C1 | Math |
| S2 | Nayan | C2 | Science |
| S1 | Kirti | C1 | Math |
| S2 | Nayan | C2 | Science |

1. Fifth Normal Forms (5NF)

5NF is the highest level of normalization and involves decomposing a table into smaller tables to remove data redundancy and improve data integrity. R should be present in the 4NF and it cannot be non-loss decomposed.

|  |  |  |
| --- | --- | --- |
| Agent | Company | Product |
| A1 | Tata | Cement |
| A2 | Orient | Ceiling fan |
| A3 | Samsung | Mobile |
| A4 | Asus | Laptop |
| A5 | Dell | Laptop |

|  |  |
| --- | --- |
| Agent | Company |
| A1 | Tata |
| A2 | Orient |
| A3 | Samsung |
| A4 | Asus |
| A5 | Dell |

|  |  |
| --- | --- |
| Agent | Product |
| A1 | Cement |
| A2 | Ceiling fan |
| A3 | Mobile |
| A4 | Laptop |
| A5 | Laptop |

|  |  |
| --- | --- |
| Company | Product |
| Tata | Cement |
| Orient | Ceiling fan |
| Samsung | Mobile |
| Asus | Laptop |
| Dell | Laptop |

* **Advantages of Normalization**

1. Normalization helps to reduce data redundancy.
2. Greater overall database organization.
3. Data consistency within the database.
4. Much more flexible database design.
5. Enforces the concept of relational integrity.

* **Disadvantages of Normalization**

1. You cannot start building the database before knowing what the user needs.
2. The performance degrades when normalizing the relations to higher normal forms, i.e., 4NF, 5NF.
3. It is very time-consuming and difficult to normalize relations of a higher degree.
4. Careless decomposition may lead to a bad database design, leading to serious problems.