Unit –IV Normalization

Que : Normalization

- Normalization is a process in database design that organizes data in a relational database to minimize redundancy and dependency issues.

- **First Normal Form (1NF)**
- **Second Normal Form (2NF)**
- **Third Normal Form (3NF)**

First Normal Form (1NF):

- In 1NF, a table must have a primary key, and each column should contain atomic values (indivisible values).
- There should be no repeating groups or arrays within the table.

Table representing student information.

Student ID	Student Name	Subjects
1	Amit	Math, Science
2	Ram	English, History, Art
3	Shyam	Physics

The "Subjects" column contains multiple values separated by commas

Student ID	Student Name	Subject
1	Amit	Math
1	Amit	Science
2	Ram	English
2	Ram	History
2	Ram	Art
3	Shyam	Physics

In this modified table, each subject is stored in a separate row with a corresponding student ID and student name.

Second Normal Form (2NF):

- In 2NF, a table must be in 1NF, and all non-key attributes should be fully dependent on the primary key.
- If any non-key attribute depends on only a part of the primary key, it should be moved to a separate table.

Student ID	Student Name	Subject	Professor
1	Amit	Math	Mr.Patel
2	Ram	English	Mr.Pandya
2	Ram	History	Mr.Pandya
2	Ram	Art	Mr.Shah
3	Shyam	Physics	Mr.Patil

- In this table, the primary key is composed of both the "Student ID" and "Subject" columns.
- However, the "Professor" column depends only on the "Subject" column and not on the entire primary key.
- This violates the second normal form (2NF), which states that non-key attributes should be fully dependent on the entire primary key.

Table 1: Students

Student ID	Student Name
1	Amit
2	Ram
3	Shyam

Table 2: Subjects

Student ID	Subject	Professor
1	Math	Mr.Patel
2	English	Mr.Pandya
2	History	Mr.Pandya
2	Art	Mr.Shah
3	Physics	Mr.Patil

- Now, the "Students" table contains unique student information, while the "Subjects" table contains the relationship between students, subjects, and professors.
- The "Student ID" column serves as the primary key for the "Students" table, and in the "Subjects" table, the combination of "Student ID" and "Subject" forms the composite primary key.x
- By separating the data into these two tables, we ensure that non-key attributes, such as the "Professor" column, are fully dependent on the primary key. This satisfies the requirements of the second normal form (2NF).

Third Normal Form (3NF):

- In 3NF, a table must be in 2NF, and there should be no transitive dependencies.
- This means that non-key attributes should depend only on the primary key and not on other non-key attributes.

Here's a modified version of the "Subjects" table to demonstrate third normal form (3NF):

Table 1: Students

Student ID	Student Name
1	Amit
2	Ram
3	Shyam

Table 2: Subjects

Subject ID	Subject
1	Math
2	English
3	History
4	Art
5	Physics

Table 3: Professors

Professor ID	Professor
1	Mr.Patel
2	Mr.Pandya
3	Mr.Shah
4	Mr.Patil

Table4: Student_Subjects

Student ID	Subject ID
1	1
2	2
2	3
2	4
3	5

- The "Subjects" table now has a separate primary key column, "Subject ID," which uniquely identifies each subject.
- The "Professors" table also has a separate primary key column, "Professor ID," which uniquely identifies each professor.

- The "Student_Subjects" table represents the many-to-many relationship between students and subjects, using the foreign key columns "Student ID" and "Subject ID."
- This table serves as the junction table to link the "Students" and "Subjects" tables.

- By splitting the data into these separate tables and eliminating transitive dependencies, we have achieved the requirements of the third normal form (3NF).
- The tables are now more organized, and each attribute depends only on the primary key in their respective tables.