NORMALISATION.

Introduction to Normalisation.

Definition

1. Normalisation is the Process of organizing data in a database to reduce redundancy and improve data integrity. It involves dividing large tables into smaller, more manageable tables and defining relationships between them.

2. Normalisation is the Process of organizing data in a database to reduce redundancy and improve data integrity.

Step: 2.

1. organizing Data: Structuring the database in a way that makes it efficient and easy to use

2. Reduce Redundancy: Minimizing duplicate data in the database
eg:- Instead of storing a Customer's Name & address in every record, Store it once in Customer table and link it to the orders.

Data integrity:
Ensuring the accuracy and consistency of data over its lifecycle.

eg:- Ensuring that the customer's Phone number is the same in all records.

Step-3: Purpose of Normalization.

1. Eliminate Redundancy:
By storing data in one Place and referencing it, we avoid unneccesary repetition.

2. Data Integrity: By organizing data Properly, we maintain its accuracy and consistency.

3. Optimize Query Performance:

· Explanation: Efficiently organized data can be retrived faster.

step:4

· Before Noralization (Denormalized Tables)

Stud_Id	Name	Course1	Course2	Course3
1	Shubham	Math	Science	History
2	Sourabh	Math	English	History

After Normalization.

Students table

Stud Id	Name
1	Shubhan
2	Sourabh

Course table:

Course_Id	Course Name
101	Math
102	science

Enrollment Table:

StudeId	Course_Id.
1	101
1	102
1	103
2	101
2	104
2	103

Step 5: Summarize the Benefits

1. Reduce Redundancy: Minimize duplicate data

2. Ensures Data integrity: Maintain accurate and Consistent Data

3. ~~Optiz~~ Optimize Performance: Speeds up data retrival.

Here. StudentId, Course together form the Primary key.

. 'Instructor' should depend on both 'StudentID' and 'Course' together not Just one of them.

. A Partial Dependency occurs when a non-key attribute is dependent on only Part of the Compos Primary key.

Converting to 2NF:

. To Eliminate Partial Dependency, seperate the table into multiple tables.

Eg.

Std_Id	Name	Course	Instructor
1	Shubham	Math	Dr. Smith
1	shubham	Science	Dr. Brown
2	Sourabh	English	Dr. John
2	Sourabh.	History	Dr. White.

Normalized table.

Students table

Stud_Id	Name
1	Shubham
2	Sourabh

Table.

Consider a table in 2NF with transitive dependencies

StudentID	Course	Instructor	Dept
1	Math	Dr.Smith	Math
1	Science	Dr.Brown	Science
2	English	Dr.John	Arts
2	History	Dr.White	Arts.

To Convert to 3NF.

Course Table.

Course	Instructor.
Math	Dr.Smith
Science	Dr.Brown
English	Dr.John
History	Dr.White

Instructors Table.

Instructor	Dept.
Dr.Smith	Math
Dr.Brown	Science
Dr. Johnson	Arts
Dr. White	Arts.

Boyce-Codd Normal Form (BCNF)

- It is in 3NF
- For every functional dependency (x->y) x Shou be a Super Key.

Eg:-
Consider a table in 3NF with Potential anomalies

Course	Instructor	Room.
Math	Dr. Smith	R101
Science	Dr. Brown	R102
English	Dr. Johnson	R103
History	Dr. White	R104

If each instructor is assigned to a Specifi room, we need to ensure no anomalies by ensurin X is a Super Key.

Rooms Table.

Instructor	Room
Dr. Smith	R101
Dr. Brown	R102
Dr. Johnson	R103
Dr. White	R104

Courses table-

Course ID	Course Name
101	Math
102	Science
103	English
104	History

Instructors table

~~tat~~ Course ID	Instructor
101	Dr. Smith
102	Dr. Brown
103	Dr. Johnson
104	Dr. White-

Enrollment Table

Stud_ID	CourseID.
1	101
1	102
2	103
2	104.

- Third Normal Form (3NF)

Conditions

- It is in 2NF
- All non key attributes are not only fully functional dependent on the Primary key but also non transitive (i.e, no transitive dependency).

To Convert to 1NF.

Stud_ID	Name	Course
1	Shubham	Math
1	Shubham	Science
2	Sourabh	English
2.	Sourabh	History

Second Normal Form (2NF)
Definition

A table is in the second normal form if.

- It is in 1NF.
- All non-key attributes are fully functional Dependent on the Primary key.
- In other words, the attribute depends on the entire Primary key, not just a subset of it.

Eg:-

Stud_ID	Course	Instructor
1	Math	Dr. Smith
1	Science	Dr. Brown
2	English	Dr. John
2	History	Dr. White

Benefits of Normalization.

1) Data Integrity:- Ensures accuracy and Consistency

2) Elimination of Redundancy:- Reduces storage space and Prevents data anomalies

3) Simplified Queries:- Makes it easier to write and maintain SQL Queries.

4) Enhanced Performance:- Improves Database Performance through optimized query execution.

-> Transitive Dependency

- A non-key attribute is dependent on another non-key attribute rather than directly on the Primary key.

- In simple terms, if 'A -> B' + 'B -> C', then 'A - C' is a transitive dependency, Where 'A' is the Primary key and 'B' + 'C' are non-key attributes

Normal Forms.

- Normalization is achieved through a series of rules applied to the database schema, Known as Normal Forms.

1) First Normal Form (1NF)
2) Second Normal Form (2NF)
3) Third Normal Form (3NF)
4) Boyce-Codd Normal Form (BCNF)

- First Normal Form. (1NF)

Definition.
A table is in the First Normal form if.

- All Column Values are atomic (indivisible)

- Each Column Contains only one type of data

- Each Column Contains unique Values.

Eg:-

Stud_ID	Student Name	Courses
1	Shubham	Math, science
2	Saurabh	English, history.