Vinoothna Bavireddy

Normalization

Normalization refers to the structure of a database. It was first developed by an IBM researcher E.F.Codd in the 1970s. It is defined as organizing the data by following a set of rules called "forms". This process avoids redundancy that is caused by updating, inserting or deleting the data.

The different types of normalization include 1NF,2NF,3NF,4NF,5NF.

1NF: This type of normalization involves unique columns and can hold a single value.
For example: Patient encounter in the outpatient department

Patient	Encounter Date	Department
Jon Snow	08/23/1998	Orthopedics
Mercille	07/05/2000	Cardiology, Neurology
Steve	05/08/2011, 03/09/2012	Radiology

This table is not normalized as it contains multiple values in one column which can be normalized using 1NF as shown below.

Patient	Encounter Date	Department
Jon Snow	08/23/1998	Orthopedics
Mercille	07/05/2000	Cardiology
Mercille	07/05/2000	Neurology
Steve	05/08/2011	Radiology
Steve	03/09/2012	Radiology

2NF: The Two condition to consider a table 2NF are

- The data should already be normalized in the form of 1 NF
- All attributes within the entity should depend solely on the unique identifier of the entity such as a primary key.

For example: Patients and their doctor visits.

Patient_ID	Doctor	Department

1	Dr. Nancy	Orthopedics
1	Dr.Shannon	Cardiology
2	Dr. Ryan	Neurology
3	Dr. Saurabh	Radiology
4	Dr. Kevin	Oncology

In a 2NF this can be represented as:

Patient_ID	Doctor
1	Dr. Nancy
1	Dr.Shannon
2	Dr. Ryan
3	Dr. Saurabh
4	Dr. Kevin

Patient_ID	Department
1	Orthopedics
1	Cardiology
2	Neurology
3	Radiology
4	Oncology

3NF: The conditions for 3NF are that the data

- Should be in 2NF
- for each functional dependency X-> Y at least one of the following conditions hold:
    - X is a super key of table
    - Y is a prime attribute of table

For example: Patient information in a hospital database can be stored as:

Patient_id	First_nam	Last_Name	Age	Gender	Address	Department_no

	e					
1	Jon	Snow	30	Male	Michigan	3
2	Nancy	Carlier	45	Female	New York	2
3	Steve	Marlow	60	Male	Texas	1

This data can be normalized using 3NF by splitting into 3NF by making different tables

Patient Table:

Patient_id	First_name	Last_Name	Age	Gender	Address
1	Jon	Snow	30	Male	Michigan
2	Nancy	Carlier	45	Female	New York
3	Steve	Marlow	60	Male	Texas

Department Table:

Department_No	Name	Physician Name
1	Cardiology	Dr. Carl
2	Oncology	Dr. Mendlow
3	Neurology	Dr. Osborne

4NF: The table should not have any Multi-valued Dependency.
A table is said to have multi-valued dependency, if the following conditions are true,

- For a dependency X → Y, if for a single value of X, multiple value of Y exists, then the table may have multi-valued dependency.
- A table must have at-least 3 columns for it to have a multivalued dependency.
- Lastly, for a relation R(X,Y,Z), if there is a multi-valued dependency between, X and Y, then Y and Z should be independent of each other.

5NF: This is generally not implemented in real life database design. The condition here is that A composite key shouldn't have any cyclic dependencies.

Vinoothna Bavireddy

References:

1. CHAITANYA SINGH. (2015, May 2). Normalization in dbms: 1nf, 2nf, 3nf and BCNF in database. beginnersbook.com. https://beginnersbook.com/2015/05/normalization-in-dbms/
2. Database normalization. (n.d.). W3schools Online Programming Tutorials. https://www.w3schools.in/dbms/database-normalization/
3. 4th normal form (4NF) and mutli-valued dependency in database normalization. (n.d.). Studytonight - Best place to Learn Coding Online. https://www.studytonight.com/dbms/fourth-normal-form.php