SRI AKILANDESWARI WOMEN'S COLLEGE, WANDIWASH

PROGRAMMING IN JAVA

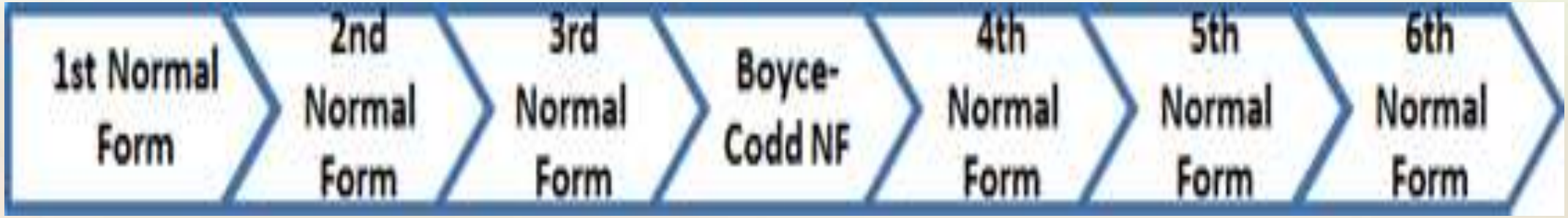
Class : UG Computer Science

Mrs.C.Thenmozhi
Assistant Professor
Department of Computer Science

SWAMY ABEDHANADHA EDUCATIONAL TRUST, WANDIWASH

Normalization

- Introduction
- Purposes of normalization
- Characteristics of normalized database
- Problems without normalization



Introduction

The process of producing a simpler and more reliable database structure is called normalization.

Purposes:

- It makes the database design efficient in performance.
- It reduce the amount of data if possible.
- It makes the database design free update, insertion and deletion anomalies.
- It makes the design according to the rules of relational databases.
- It identified relationship between entities.
- It makes a design that allows simple retrieval of data.
- It simplifies data maintenance and reduces the need to restructure data.

Characteristics of Normalized Database

A normalized database should have the following characteristics:

- Each relation must have a key field.
- All fields must contain atomic data.
- There must be no repeating fields.
- Each table must contain information about a single entity.
- Each field in a relation must depend on key fields.
- All non-key fields must be mutually independent.

Problems Without Normalization

Student table:

Rollno	Name	Branch	HOD
101	Kavitha	CS	Mr. X
102	Pradeepa	CS	Mr. X
103	Srilekha	CS	Mr. X
104	Punitha	CS	Mr. X

- Repetition of data increases the size of database.
- Other issues like: insertion ,deletion , updation.

Now how can we solved these problems by normalization.

1st Normal form

The First Normal Form, it should follow the following 4 rules:

- It should only have single(atomic) valued attributes/columns.
- Values stored in a column should be of the same domain
- All the columns in a table should have unique names.
- And the order in which data is stored, does not matter.

Example: The above 2 students have more than 1 subjects. And we have stored sub name in single column because all the column name are unique.

roll_no	name	subject
101	Akon	OS, CN
103	Ckon	Java
102	Bkon	C, C++

How to solve this Problem?

Roll_no	Name	Subject
101	Kavitha	OS
101	Pradeepa	CN
103	Srilekha	Java
102	Punitha	C
102	Sivapriya	C++

2nd Normal Form

It should 2 rules:

- It should be in the First Normal form.
- And, it should not have Partial Dependency.

What is partial dependency?

A Composite Primary Key (a primary key that is made up of multiple columns), and one of the non-key columns is functionally dependent on one, but not all of the columns that make up the Composite Primary Key.

Example:

Subject_id	Subject_name
1	Java
2	C++

- Let's create another table Score, to store the marks obtained by students in the respective subjects. We will also be saving name of the teacher who teaches that subject along with marks.

Score_id	Student_id	Subject_id	Marks	Teacher
1	10	1	70	Java Teacher
2	10	2	75	C++ Teacher
3	11	1	80	Java Teacher

- This is **Partial Dependency**, where an attribute in a table depends on only a part of the primary key and not on the whole key.

How to remove Partial Dependency?

•There can be many different solutions for this, but out objective is to remove teacher's name from Score table.

•The simplest solution is to remove columns teacher from Score table and add it to the Subject table.

Hence, the **Subject table** will become:

Subject_id	Subject_name	Teacher
1	Java	Java Teacher
2	C++	C++ Teacher
3	bio	bio Teacher

Our **Score table** is now in the second normal form, with no partial dependency.

Score_id	Student_id	Subject_id	Marks
1	10	1	70
2	10	2	75
3	11	1	80

3rd Normal Form

In the third normal form,

- It should be in the Second Normal form.
- And it should not have Transitive Dependency.

What is Transitive Dependency?

It is a condition in which an attributes is dependent on an attribute that is not a part of primary key.

for example: when column C is functionally dependent on column B, and column B is functionally dependent on column A, and column A is the primary key.

score_id	student_id	subject_id	marks	exam_name	total_marks

Primary key

- Total marks is no depend on the primary key it depend on exam name and the exam name is not a primary key.

How to remove Transitive Dependency?

The new Exam table:

Exam_id	Exam_Name	Total_Marks
1	Workshop	200
2	Mains	70
3	Practical	30

Advantage of removing Transitive Dependency

The advantage of removing transitive dependency is,

- Amount of data duplication is reduced.
- Data integrity achieved.

Boyce-Codd Normal Form (BCNF)

- It is also a 3.5 normal form because it is the latest version of 3rd normal form.

Rules for BCNF:

- It should be in the **Third Normal Form**.
- And, for any dependency A → B, A should be a **super key**.

Student_id	Subject	Professor
101	Java	C.Thenmozhi
101	C++	R.Padmashree
102	Java	N.Ambiga
103	C#	R.Saraswathi
104	Java	C.Thenmozhi

- This table satisfies the **1st Normal form** because all the values are atomic, column names are unique and all the values stored in a particular column are of same domain.
- This table also satisfies the **2nd Normal Form** as there is no **Partial Dependency** and, there is no **Transitive Dependency**, hence the table also satisfies the **3rd Normal Form** .
- But this table is not in **Boyce-Codd Normal Form**.

4th Normal form

Rules:

- It should be in the **Boyce-Codd Normal Form**.
- The table should not have any **Multi-valued Dependency**.

What is Multi-valued Dependency?

- For a dependency A → B, if for a single value of A, multiple value of B exists, then the table may have multi-valued dependency.
- Also, a table should have at-least 3 columns for it to have a multi-valued dependency.
- For relation if there is multi-valued dependency between A and B, then B and C should be independent of each other.

If all these condition are true for any relation(table), it is said to be multi- valued dependency.

S_id	Course	Hobby
1	Science	Cricket
1	Maths	Hockey
1	Science	Hockey
1	Maths	Cricket

How to satisfy 4th Normal Form?

To make the above relation satisfy the 4th normal form, we can decompose the table into 2 tables.

Course table

Stud_id	Course
1	Science
1	Maths
2	C#
2	Php

Hobbies table

Stud_id	Hobby
1	Cricket
1	Hockey
2	Cricket
2	Hockey

- A table can also have functional dependency along with multi-valued dependency.
- In that case, the functionally dependent columns are moved in a separate table and the multi-valued dependent columns are moved to separate tables.

5th Normal form

- Fifth normal form (5NF), also known as project-join normal form (PJ/NF), is a level of database normalization designed to reduce redundancy in relational databases recording multi-valued facts by isolating semantically related multiple relationships.

6th Normal form

- Sixth Normal Form is the irreducible Normal Form, there will be no further NFs after this, because the data cannot be Normalized further. ... The definition of 6NF is: a table is in 6NF when the row contains the Primary Key, and at most one, attribute.

Thank you