DATABASE NORMALIZATION

Radhika Rani Chintala

Start with Understanding

- Problems without Normalization
- How Normalization solves these problems

What is Normalization ?

- A technique of organizing the data into multiple related tables, to minimize **DATA REDUNDANCY.**

What is Data Redundancy ?

Why should we reduce it?

- Repetition of similar data at multiple places.

- Repetition of data increases the size of database.
- Other issues:
  - Insertion problems
  - Deletion problems
  - Updation problems

Example

Students Table

Sid	Sname	Branch	HOD	Office_Tel
1	Anandi	CSE	Mr. X	3031
2	Jaya	CSE	Mr. X	3031
3	Radhi	CSE	Mr. X	3031
4	Satya	CSE	Mr. X	3031

Branch name, HOD name and Office telephone number are same for all the rows and hence they are **redundant data**.

Issues due to Data Redundancy

- Increases size of database
- Insertion Anomaly
- Deletion Anomaly
- Updation Anomaly

Insertion Anomaly

Sid	Sname	Branch	HOD	Office_Tel
1	Anandi	CSE	Mr. X	3031
2	Jaya	CSE	Mr. X	3031
3	Radhi	CSE	Mr. X	3031
4	Satya	CSE	Mr. X	3031
5				
.				
.				
100				

To add a new student, we have to repeat the same data that holds the branch information(Branch, HOD, Office_Tel)

Insertion Anomaly

- Inserting redundant data for new row is called Data Insertion problem or anomaly

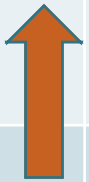
- **Reason for data repetition:**
  - Two different but related data is stored in the same table (Student data and Branch data)

Deletion Anomaly

- When we try to delete student information(row), the branch information is also deleted

Deletion Anomaly

Sid	Sname	Branch	HOD	Office_Tel
1	Anandi	CSE	Mr. X	3031



We have not stored the branch information anywhere else.

When we delete the last row, We lose the branch information completely along with student information.

Radhika Rani Chintala

Deletion Anomaly

- Loss of related dataset when some other dataset is deleted is known as Deletion anomaly.

Updation Anomaly

Sid	Sname	Branch	HOD	Office_Tel
1	Anandi	CSE	Mr. X	3031
2	Jaya	CSE	Mr. X	3031
3	Radhi	CSE	Mr. X	3031
4	Satya	CSE	Mr. X	3031
5	Rani	CSE	Mr. X	3031

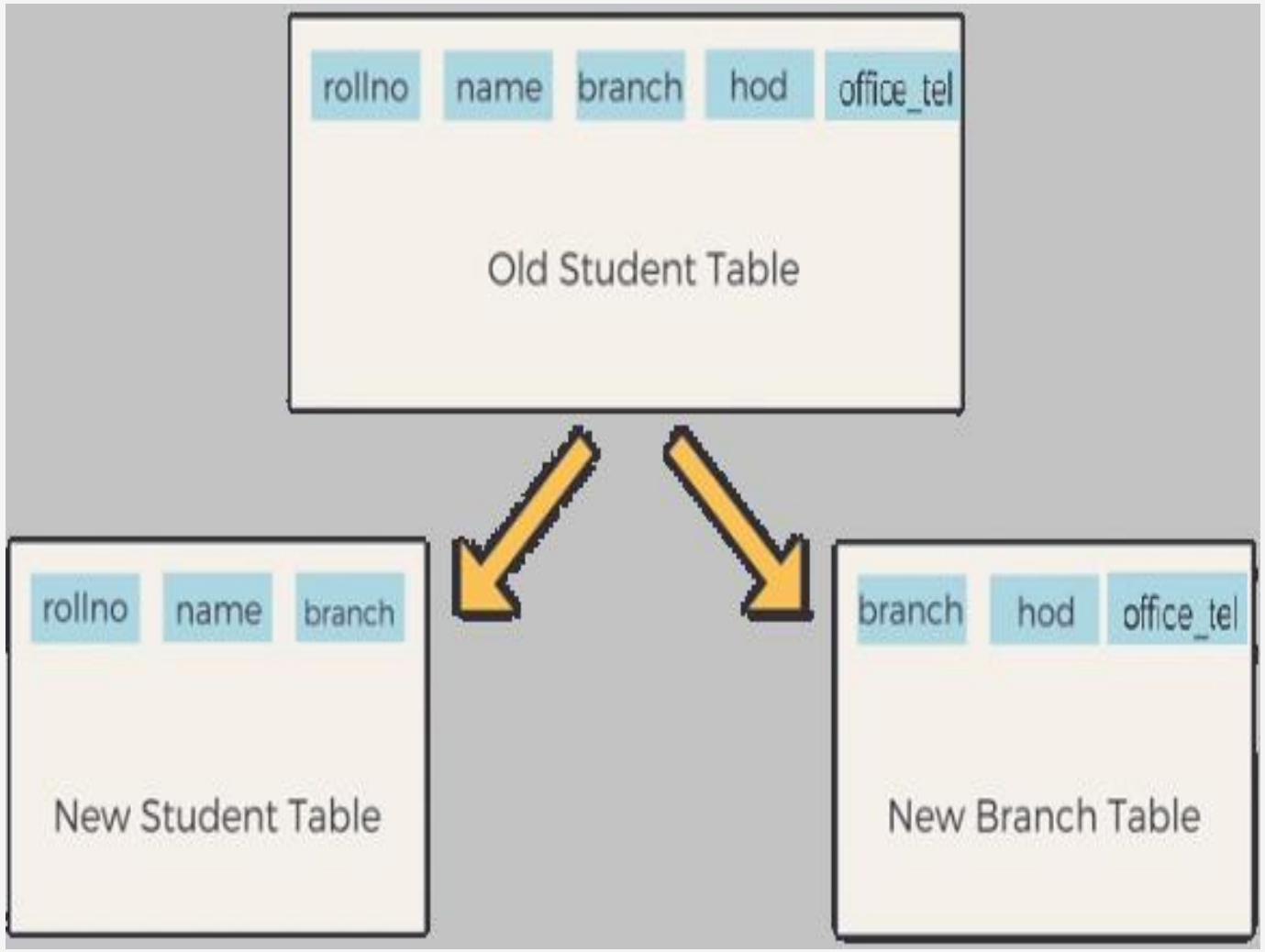
Mr. X leaves and **Mr. Y** joins as the new HOD for CSE, then you need to modify each row with new HOD

Data Redundancy

- Repetition of data hence needs extra space
- Leads to insertion, deletion and updation issues

How normalization will solve all these problems

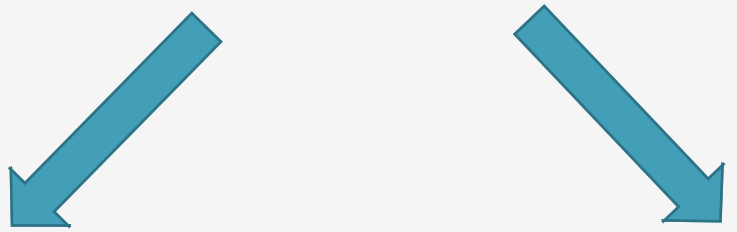
???



Radhika Rani Chintala

Redundancy is reduced and thus reducing the problems of *insertion,deletion and updation*

Sid	Sname	Branch	HOD	Office_Tel
1	Anandi	CSE	Mr. X	3031
2	Jaya	CSE	Mr. X	3031
3	Radhi	CSE	Mr. X	3031



STUDENT TABLE

Sid	Sname	Branch
1	Anandi	CSE
2	Jaya	CSE
3	Radhi	CSE

BRANCH TABLE

Branch	HOD	Office_Tel
CSE	Mr. X	3031

Normalization is Good

- It follows Divide and Rule
- Divides the data into separate independent logical entities and relating them using some common entity(Branch)

Types of Normalization

- 1st Normal Form(1NF)
- 2nd Normal Form(2NF)
- 3rd Normal Form(3NF)
- Boyce-Codd Normal Form (BCNF)
- 4th Normal Form(4NF)
- 5th Normal Form(5NF)

Radhika Rani Chintala

1st Normal Form (1NF)

For a table to be in the First Normal Form, it should follow the following 4 rules:

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

Radhika Rani Chintala

1NF Example

Roll_no	Name	Subject
101	Akon	OS, CN
103	Ckon	Java
102	Bkon	C, C++

Three of the rules were satisfied

Roll_no	Name	Subject
101	Akon	OS
101	Akon	CN
103	Ckon	Java
102	Bkon	C
102	Bkon	C++

2nd Normal Form (2NF)

For a table to be in the Second Normal Form:

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

Radhika Rani Chintala

Partial Dependency ?

Dependency / Functional Dependency:

If the information stored in a table can uniquely determine another information in the same table, then it is called Functional Dependency.

P -> Q : *P functionally determines Q*

/ Q is functionally dependent on P

Partial Dependency:

Partial Dependency occurs when a nonprime attribute is functionally dependent on part of a candidate key.

Radhika Rani Chintala

2NF Example

Jump to Excel

Radhika Rani Chintala

2NF

- For a table to be in the Second Normal form, it should be in the First Normal form and it should not have Partial Dependency.
- Partial Dependency exists, when for a composite primary key, any attribute in the table depends only on a part of the primary key and not on the complete primary key.
- To remove Partial dependency, we can divide the table, remove the attribute which is causing partial dependency, and move it to some other table where it fits in well.

3rd Normal Form (3NF)

A table is said to be in the Third Normal form when,

1. *It is in the Second Normal form.*
2. *And, it doesn't have Transitive Dependency.*

Radhika Rani Chintala

Transitive Dependency ?

When a non-prime attribute depends on other non-prime attributes rather than depending upon the prime attributes or primary key, then we say it is ***Transitive Dependency***

How to remove Transitive Dependency?

- Again the solution is very simple.
- Take out the columns that caused transitive dependency from the existing table and move them to some other table where it fits in well.

Radhika Rani Chintala

3NF Example

Let's go to Excel

Radhika Rani Chintala

Boyce-Codd Normal Form (BCNF)

For a table to be in BCNF, following conditions must be satisfied:

- *R must be in 3rd Normal Form*
- *And, for each functional dependency (X → Y), X should be a super Key.*

Radhika Rani Chintala

Super Key

Candidate Key: The minimal set of attribute which can uniquely identify a tuple is known as candidate key.

Ex : Student table: {**Sno**} or {**Mobile**} ; Course table: {**Sno, Course_No**}

Super Key: The set of attributes which can uniquely identify a tuple is known as Super Key. A candidate key is a super key but vice versa is not true.

Ex : Student table: **{Sno}, {Sno, Sname}**

Primary Key: There can be more than one candidate key in relation out of which one can be chosen as the primary key.

Ex : Student table: **{Sno}**

Sno	SName	Mobile	Age	Address

Sno	Course_No	Course_Name

BCNF Example

Let's move to Excel

Radhika Rani Chintala

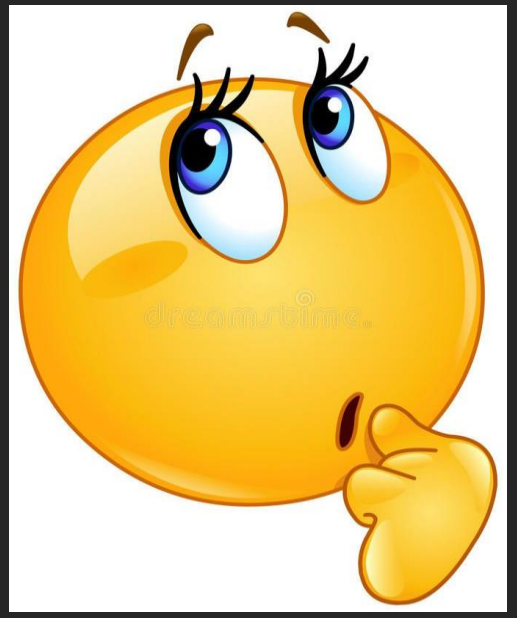
4th Normal Form (4NF)

A table is said to be in the Fourth Normal Form when,

1. *It is in the Boyce-Codd Normal Form.*
2. *And, it doesn't have Multi-Valued Dependency.*

Radhika Rani Chintala

Multi-Valued Dependency



?

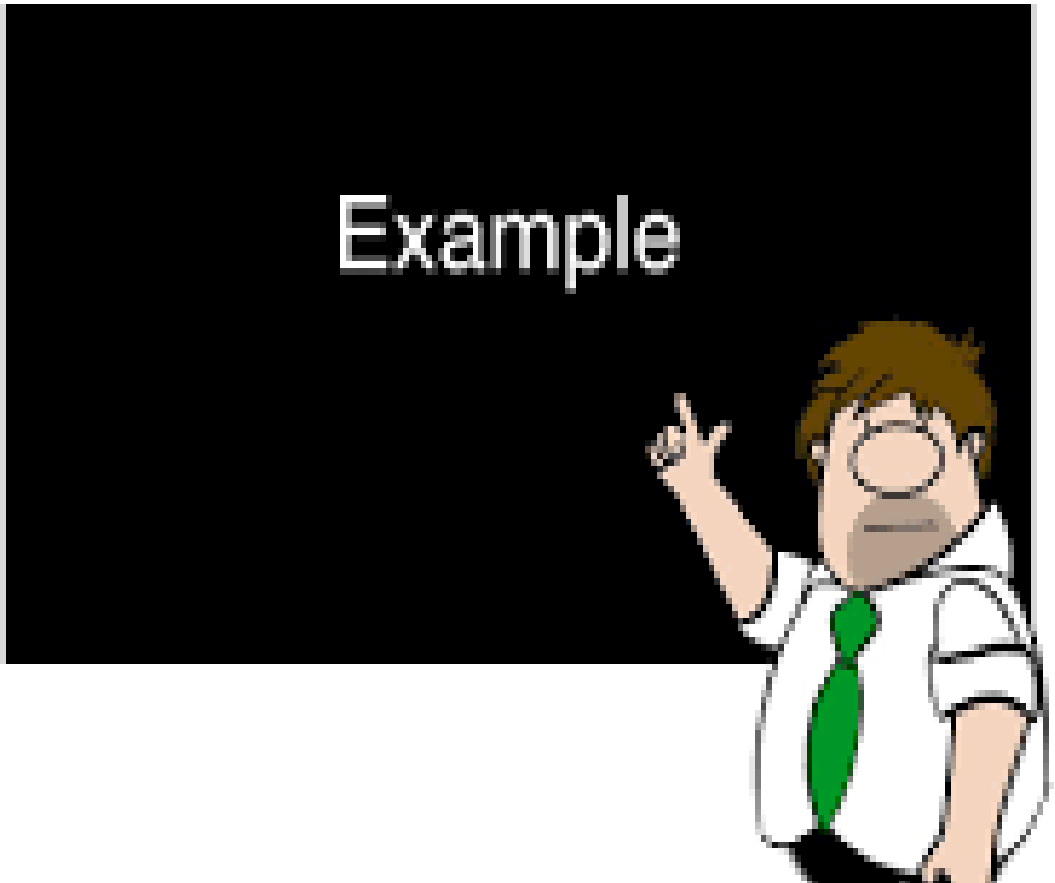
A table is said to have multi-valued dependency, if the following conditions are true.

1. 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.

2. Also, a table should have at-least 3 columns for it to have a multi-valued dependency.

3. And, for a relation `R(A,B,C)`, if there is a multi-valued dependency between, A and B, then B and C should be independent of each other.

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



4NF Example

Move to Excel

Radhika Rani Chintala

5th Normal Form (5NF)

A table is said to be in the Fifth Normal Form when,

1. *It is in 4NF*
2. *And, it doesn't have any Join Dependency.*

- **5NF is satisfied when all the tables are broken into as many tables as possible in order to avoid redundancy.**
- **Once it is in fifth normal form it cannot be broken into smaller relations without changing the facts or the meaning.**

Radhika Rani Chintala

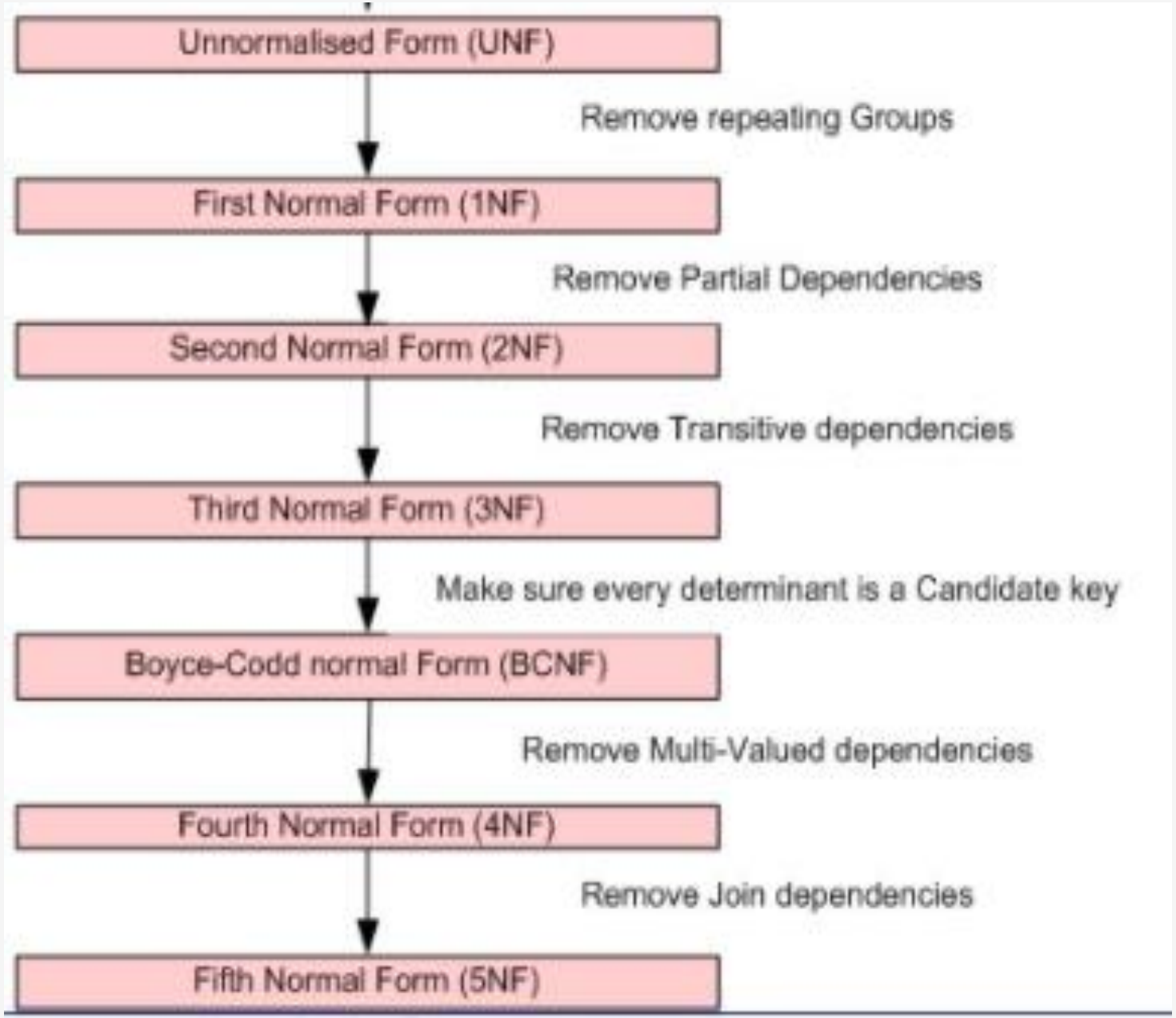
Join Dependency

?

- If a table can be recreated by joining multiple tables and each of this table have a subset of the attributes of the table, then the table is in **Join Dependency**.
- 5NF is also known as Project-join normal form (PJNF).
- Fifth Normal Form in Database Normalization is generally not implemented in real life database design.

Summary


I just need
the main ideas


Unnormalised Form (UNF)
Remove repeating Groups
First Normal Form (1NF)
Remove Partial Dependencies
Second Normal Form (2NF)
Remove Transitive dependencies
Third Normal Form (3NF)
Make sure every determinant is a Candidate key
Boyce-Codd normal Form (BCNF)
Remove Multi-Valued dependencies
Fourth Normal Form (4NF)
Remove Join dependencies
Fifth Normal Form (5NF)

THE END

Radhika Rani Chintala