

What is Normalization

Database normalization is the process of organizing data to minimize data redundancy (data duplication), which in turn ensures data consistency.

EmployeeName	Gender	Salary	DeptName	DeptHead	DeptLocation
Sam	Male	4500	IT	John	London
Pam	Female	2300	HR	Mike	Sydney
Simon	Male	1345	IT	John	London
Mary	Female	2567	HR	Mike	Sydney
Todd	Male	6890	IT	John	London

Problems of Data Redundancy

1. Disk Space Wastage
2. Data Inconsistency
3. DML queries can become slow

Normalized Table Design

DeptId	DeptName	DeptHead	DeptLocation
1	IT	John	London
2	HR	Mike	Sydney

EmployeeId	EmployeeName	Gender	Salary	DeptId
1	Sam	Male	4500	1
2	Pam	Female	2300	2
3	Simon	Male	1345	1
4	Mary	Female	2567	2
5	Todd	Male	6890	1

Database normalization is a step by step process. There are 6 normal forms, First Normal form (1NF) thru Sixth Normal Form (6NF). Most databases are in **third normal form (3NF)**. There are certain rules, that each normal form should follow.

First Normal Form (1NF)

A table is said to be in 1NF, if

1. The data in each column should be atomic. No multiple values, separated by comma.
2. The table does not contain any repeating column groups
3. Identify each record uniquely using primary key.

Non Atomic Employee Column

DeptName	Employee
IT	Sam, Mike, Shan
HR	Pam

Problems of Non Atomic Columns

It is not possible to SELECT, INSERT, UPDATE and DELETE just one employee

No Repeating Column Groups

DeptName	Employee1	Employee2	Employee3
IT	Sam	Mike	Shan
HR	Pam		

Problems of Repeating Column Groups

More than 3 employees - Table structure change required
Less than 3 employees - Wasted disk space

First Normal Form
1 NF

Primary Key

DeptId	DeptName
1	IT
2	HR

Foreign Key

DeptId	Employee
1	Sam
1	Mike
1	Shan
2	Pam

First Normal Form
1 NF

Second Normal Form (2NF)

A table is said to be in 2NF, if

1. The table meets **all the conditions of 1NF**
2. Move **redundant data** to a separate table
3. Create **relationship** between these tables using **foreign keys**.

EmpId	EmployeeName	Gender	Salary	DeptName	DeptHead	DeptLocation
1	Sam	Male	4500	IT	John	London
2	Pam	Female	2300	HR	Mike	Sydney
3	Simon	Male	1345	IT	John	London
4	Mary	Female	2567	HR	Mike	Sydney
5	Todd	Male	6890	IT	John	London

Problems of Data Redundancy

1. Disk Space Wastage
2. Data Inconsistency
3. DML queries can become slow

Table Design in Second Normal Form

DeptId	DeptName	DeptHead	DeptLocation
1	IT	John	London
2	HR	Mike	Sydney

EmpId	EmployeeName	Gender	Salary	DeptId
1	Sam	Male	4500	1
2	Pam	Female	2300	2
3	Simon	Male	1345	1
4	Mary	Female	2567	2
5	Todd	Male	6890	1

Third Normal Form (3NF)

A table is said to be in 3NF, if the table

1. Meets all the conditions of 1NF and 2NF
2. Does not contain columns (attributes) that are **not fully dependent** upon the **primary key**

EmpId	EmployeeName	Gender	Salary	AnnualSalary	DeptId
1	Sam	Male	4500	54000	1
2	Pam	Female	2300	27600	2
3	Simon	Male	1345	16140	1
4	Mary	Female	2567	30804	2
5	Todd	Male	6890	82680	1

EmpId	EmployeeName	Gender	Salary	DeptName	DeptHead
1	Sam	Male	4500	IT	John
2	Pam	Female	2300	HR	Mike
3	Simon	Male	1345	IT	John
4	Mary	Female	2567	HR	Mike
5	Todd	Male	6890	IT	John

EmpId	EmployeeName	Gender	Salary	DeptId
1	Sam	Male	4500	1
2	Pam	Female	2300	2
3	Simon	Male	1345	1
4	Mary	Female	2567	2
5	Todd	Male	6890	1

DeptId	DeptName	DeptHead
1	IT	John
2	HR	Mike