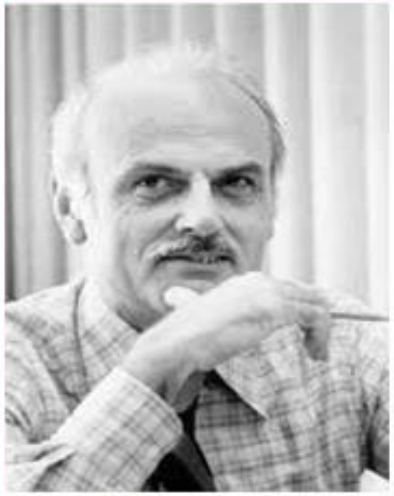
NORMALIZATION

- Normalization is developed by IBM researcher E.F Codd in 1970's.
- Normalization is a process of organizing the data in database to avoid redundancy, insert anomaly, update anomaly and delete anomaly.

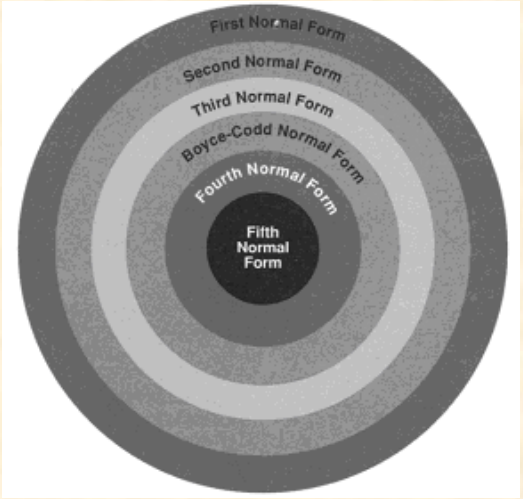


E.F. Codd

- Normalization split a large table into smaller tables and define relationships between to increase the clarity in organizing data.
- The words normalization and normal form refers to the structure of database.

NORMALIZATION RULES

Database normalization process are divided into following normal forms:



1. First Normal Form (1NF)
2. Second Normal Form (2NF)
3. Third Normal Form (3NF)
4. Boyce-Codd Normal Form (BCNF)
5. Fourth Normal Form (4NF)
6. Fifth Normal Form (5NF)

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 data does not contain any repeating column groups.

Employee

DeptName	Employee
IT	Raju, Rama, Ravan
HR	Rani, Sita

First Normal Form (1NF)

Repeating Column Groups:

DeptName	Employee1	Employee2	Employee3
IT	Raju	Rama	Ravan
HR	Rani	Sita	

Problems of Repeating Column Groups:

More than 3 Employees – Require to change table structure

Less than 3 Employees – Wasted disk space

First Normal Form (1NF)

Table Design in 1NF:

DeptName	Employee
IT	Raju
IT	Rama
IT	Ravan
HR	Rani
HR	Sita

Second Normal Form (2NF)

A table is said to be 2 NF, if

1. A table meets all the conditions of 1NF.
2. No partial dependency exists between non-key attributes and key attributes
3. Move redundant data to a separate table.
4. Create relationship between these tables using foreign keys.

Second Normal Form (2NF)

EmpId	EmpName	Gender	Salary	DeptId	DeptName	DepLocation
1	Raju	Male	20000	10	IT	Hyderabad
2	Rani	Female	15000	20	HR	Chennai
3	Rama	Male	30000	10	IT	Hyderabad
4	Sita	Female	25000	20	HR	Chennai
5	Ravan	Male	20000	10	IT	Hyderabad

Problems of Data Redundancy:

1. Disk space wastage
2. Data Inconsistency
3. DML Queries can become slow.

Second Normal Form (2NF)

Table Design in 2NF:

DeptId	DeptName	DepLocation
10	IT	Hyderabad
20	HR	Chennai

EmpId	EmpName	Gender	Salary	DeptId
1	Raju	Male	20000	10
2	Rani	Female	15000	20
3	Rama	Male	30000	10
4	Sita	Female	25000	20
5	Ravan	Male	20000	10

Third Normal Form (3NF)

A table is said to be 3NF, if

1. Meets all the conditions of 1NF & 2NF
2. No transitive dependency exists between non-key attributes and key attributes.

Third Normal Form (3NF)

EmpId	EmpName	Gender	Salary	AnnualSalary	DeptId	DeptName	DepLocation
1	Raju	Male	20000	240000	10	IT	Hyderabad
2	Rani	Female	15000	180000	20	HR	Chennai
3	Rama	Male	30000	360000	10	IT	Hyderabad
4	Sita	Female	25000	300000	20	HR	Chennai
5	Ravan	Male	20000	240000	10	IT	Hyderabad

EmpId→ EmpName, Gender, Salary, AnnualSalary, DeptId, DeptName, DepLocation

Salary→AnnualSalary

DeptId→DeptName, DeptLocation

Third Normal Form (3NF)

DeptId	DeptName	DepLocation
10	IT	Hyderabad
20	HR	Chennai

Salary	AnnualSalary
20000	240000
15000	180000
30000	360000
25000	300000

EmpId	EmpName	Gender	Salary	DeptId
1	Raju	Male	20000	10
2	Rani	Female	15000	20
3	Rama	Male	30000	10
4	Sita	Female	25000	20
5	Ravan	Male	20000	10

Note: In real-time most database's are in 3NF.

DATA Hills