Normalization

Normalization is the process of minimizing redundancy from a relation or set of relations

Types of Normalization
1. First Normal Form (1NF)
2. Second Normal Form (2NF)
3. Third Normal Form (3NF)
4. Boyce Codd Normal Form (BCNF)

1. First Normal Form (1NF)
- A relation will be in 1NF if it contains atomic value.
- It states that an attribute of the table cannot hold multiple values. It must hold only a single valued attribute.

EMP_ID	EMP_NAME	EMP_PHONE	EMP_STATE
14	John	7272826385, 9064738238	UP
20	Harry	8574783832	Bihar
12	Sam	7390372389, 8589830302	Punjab

The decomposition of the Employee table into 1NF has been shown below.

EMP_ID	EMP_NAME	EMP_PHONE	EMP_STATE
14	John	7272826385	UP
14	John	9064738238	UP
20	Harry	8574783832	Bihar
12	Sam	7390372389	Punjab
12	Sam	8589830302	Punjab

2. Second Normal Form (2NF)

- In the 2NF, relation must be in 1NF.

- In Second Normal Form, all non-key attributes are fully functional dependency on the primary key.

TEACHER_ID	SUBJECT	TEACHER_AGE
25	Chemistry	30
25	Biology	30
47	English	35
83	Math	38
83	Computer	38

In the given table, non-prime attribute TEACHER_AGE is dependent on TEACHER_ID which is a proper subset of a candidate key. That's why it violates the rule for 2NF.

TEACHER_ID	TEACHER_AGE
25	30
47	35
83	38

TEACHER_ID	SUBJECT
25	Chemistry
25	Biology
47	English
83	Math
83	Computer

3. Third Normal Form (3NF)

- A relation will be in 3NF if it is in 2NF and does not contain any transitive partial dependencies.
- 3NF is used to reduce data duplication. It is also used to achieve data integrity.
- If there is no transitive dependency for non-prime attributes, then the relation must be in the third normal form.

A relation is in third normal form if it holds at least one of the following conditions for every non-trivial functional dependency $X \rightarrow Y$.

1. X is a super key.

2. Y is a prime attribute, i.e., each element of Y is part of some candidate key.

EMP_ID	EMP_NAME	EMP_ZIP	EMP_STATE	EMP_CITY
222	Harry	201010	UP	Noida
333	Stephan	02228	US	Boston
444	Lan	60007	US	Chicago
555	Katharine	06389	UK	Norwich
666	John	462007	MP	Bhopal

In the given table, all attributes except EMP_ID are non-prime.

EMP_ID	EMP_NAME	EMP_ZIP
222	Harry	201010
333	Stephan	02228
444	Lan	60007
555	Katharine	06389
666	John	462007

EMP_ZIP	EMP_STATE	EMP_CITY
201010	UP	Noida
02228	US	Boston
60007	US	Chicago
06389	UK	Norwich
462007	MP	Bhopal

4. Boyce Codd Normal Form (BCNF)

- BCNF is the advanced version of 3NF. It is stricter than 3NF.

- A table is in BCNF if every functional dependency X → Y, X is the super key of the table.

- For BCNF, the table should be in 3NF, and for every FD, LHS is a super key.

EMP_ID	EMP_COUNTRY	EMP_DEPT	DEPT_TYPE	EMP_DEPT_NO
264	India	Designing	D394	283
264	India	Testing	D394	300
364	UK	Stores	D283	232
364	UK	Developing	D283	549

EMP_ID → EMP_COUNTRY

EMP_DEPT → {DEPT_TYPE, EMP_DEPT_NO}

EMP_ID	EMP_COUNTRY
264	India
364	UK

EMP_DEPT	DEPT_TYPE	EMP_DEPT_NO
Designing	D394	283
Testing	D394	300
Stores	D283	232
Developing	D283	549

EMP_ID	EMP_DEPT_NO
264	283
264	300
364	232
364	549

Candidate keys:

For the first table: EMP_ID
For the second table: EMP_DEPT
For the third table: {EMP_ID, EMP_DEPT}

Now, this is in BCNF because the left side part of both the functional dependencies is a key.