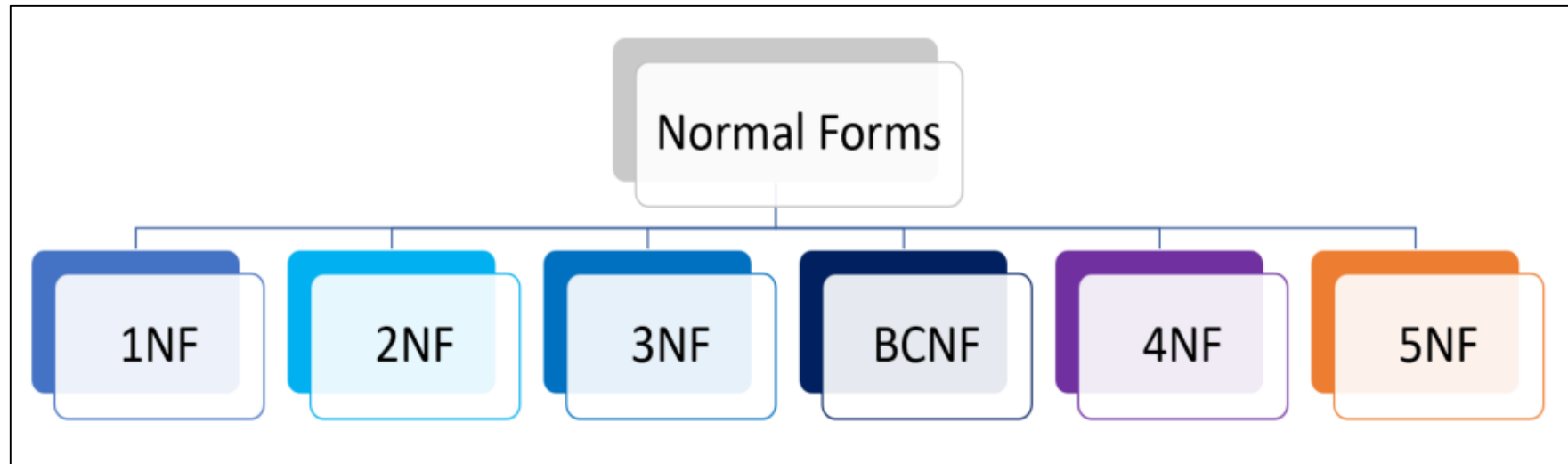


Database Management Systems (BCSC-1003)

Topic: Normalization



Dr. Nikhil Govil

Assistant Professor, Dept. of CEA, GLA University, Mathura.

Normalization

- Normalization is a database design technique that reduces data redundancy and eliminates undesirable characteristics like Insertion, Update and Deletion Anomalies.
- Normalization rules divides larger tables into smaller tables and links them using relationships.
- The purpose of Normalization in SQL is to eliminate redundant (repetitive) data and ensure data is stored logically.

Types of Normal Form

Here is a list of Normal Forms

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

First Normal Form (1NF)

- A relation will be in 1NF if it contains an atomic value.
- It states that an attribute of a table cannot hold multiple values. It must hold only single-valued attribute.
- First normal form disallows the multi-valued attribute, composite attribute, and their combinations.

First Normal Form (1NF)

Example:

- Relation EMPLOYEE is not in 1NF because of multi-valued attribute EMP_PHONE.

EMPLOYEE table:

EMP_ID	EMP_NAME	EMP_PHONE	STATE
2925	RAJESH	9500407832, 9088170884	TAMILNADU
6796	KISHORE	9510508668	KARNATAKA
4592	ALOK	8455661732	PUNJAB
3565	HEMANT	9759105522, 9414525254	GUJRAT

First Normal Form (1NF)

The decomposition of the EMPLOYEE table into 1NF has been shown below:

EMP_ID	EMP_NAME	EMP_PHONE	STATE
2925	RAJESH	9500407832	TAMILNADU
2925	RAJESH	9088170884	TAMILNADU
6796	KISHORE	9510508668	KARNATAKA
4592	ALOK	8455661732	PUNJAB
3565	HEMANT	9759105522	GUJRAT
3565	HEMANT	9414525254	GUJRAT

Second Normal Form (2NF)

- In the 2NF, relational must be in 1NF.
- In the second normal form, all non-key attributes are fully functional dependent on the primary key.

Example:

Let's assume, a college can store the data of teachers and the subjects they teach. In a college, a teacher can teach more than one subject.

TEACHER table:

TEACHER_ID	SUBJECT	TEACHER_AGE
2529	Information Technology	30
2529	Soft Computing	30
6535	Data Communication	45
3578	Soft Computing	35
3578	Digital Image Processing	35

Second Normal Form (2NF)

- In the given table, non-prime attribute TEACHER_AGE is dependent on TEACHER_ID. That's why it violates the rule for 2NF.
- To convert the given table into 2NF, we decompose it into two tables:

TEACHER_DETAIL Table and

TEACHER_ID	TEACHER_AGE
2529	30
6535	45
3578	35

TEACHER_SUBJECT Table:

TEACHER_ID	SUBJECT
2529	Information Technology
2529	Soft Computing
6535	Data Communication
3578	Soft Computing
3578	Digital Image Processing

Third Normal Form (3NF)

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

A relation is in third normal form if it holds at least one of the following conditions for every non-trivial function 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.

Third Normal Form (3NF)

EMPLOYEE_DETAIL table:

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

Super key in the table above:

{EMP_ID}, {EMP_ID, EMP_NAME}, {EMP_ID, EMP_NAME, EMP_ZIP}.... so on.

Candidate key: {EMP_ID}

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

Here, EMP_STATE & EMP_CITY dependent on EMP_ZIP and EMP_ZIP dependent on EMP_ID. The non-prime attributes (EMP_STATE, EMP_CITY) transitively dependent on super key (EMP_ID). It violates the rule of third normal form.

Third Normal Form (3NF)

- That's why we need to move the EMP_CITY and EMP_STATE to the new <EMPLOYEE_ZIP> table, with EMP_ZIP as a Primary key.

EMPLOYEE table:

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

EMPLOYEE_ZIP Table:

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

Boyce-Codd Normal Form (BCNF)

- BCNF is the advance version of 3NF. It is stricter than 3NF.
- A table is in BCNF if every functional dependency $X \rightarrow Y$, X is the super key of the table.
- For BCNF, the table should be in 3NF, and for every FD, LHS is super key.

Employee Table

Emp_ID	Emp_Country	Emp_Dept	Dept_Type	Dept_No
264	India	Designing	D394	283
264	India	Testing	D394	300
364	UK	Stores	D283	232
364	UK	Development	D283	549

Boyce-Codd Normal Form (BCNF)

In the above table Functional dependencies are as follows:

- $EMP_ID \rightarrow EMP_COUNTRY$
- $EMP_DEPT \rightarrow \{DEPT_TYPE, EMP_DEPT_NO\}$

EMP_COUNTRY Table

Emp_ID	Emp_Country
264	India
364	UK

EMP_DEPT Table

Emp_Dept	Dept_Type	Dept_No
Designing	D394	283
Testing	D394	300
Stores	D283	232
Development	D283	549

EMP_DEPT_MAPPING Table

Emp_ID	Emp_Dept
264	Designing
264	Testing
364	Stores
364	Development

References



- Korth, Silbertz and Sudarshan (1998), “Database Concepts”, 4th Edition, TMH.
- Elmasri and Navathe (2010), “Fundamentals of Database Systems”, 5th Edition, Addison Wesley.
- Date C J,” An Introduction to Database Systems”, 8th Edition, Addison Wesley.
- M. Tamer Oezsu, Patrick Valduriez (2011). “Principles of Distributed Database Systems”, 2nd Edition, Prentice Hall.

*Thank
you*

