Outline

- Normalization basics
- Normal Forms

1. 1NF (First Normal Form)
2. 2NF
3. 3NF
4. BCNF (Boyce Codd Normal Form)

Normalization

Process of analyzing the given relation schemas based on the FD's and candidate keys to achieve the desirable properties of

1) Minimizing the redundancy.

2) Minimizing insertion, deletion and update anomalies.

Insertion anomalies

It is difficult to insert a new department that has no employees as yet in the EMP_DEPT relation. The only way to do is to place NULL values in the atttributes for employee. This causes a problem because SSN is primary key of EMP_DEPT and each tuple is supposed to represent an employee entity - not department entity.

Deletion anomalies

If we delete from EMP_DEPT an employee tuple that happens to represent the last employee working for a particular department, the information concerning the department is lost from the database.

Update anomalies

In EMP_DEPT, if we change the value of one of the attributes of a particular department- say the manager of department 5- we must update all employee who work in that department; otherwise the database will be inconsistent.

Normal Forms

- 1NF (Atomic)
- 2NF (No partial dependencies)
- 3NF (No transitive dependencies)
- BCNF (Stronger than 3NF)

First Normal Form (1NF)

The relation is in 1NF if we say domains of all attributes of R are atomic.

Example of 1NF

Roll No	Name	Phone No.
1	Harsh	9890385807
		9920297878
2	Kumar	9561922284
3	Prashant	9819740847
		9920625758

Conversion in 1NF

Roll No	Name	Phone No.
1	Harsh	9890385807
1	Harsh	9920297878
2	Kumar	9561922284
3	Prashant	9819740847
3	Prashant	9920625758

Second Normal Form

A relation schema R is in 2NF if it is in 1NF and every nonprime attirbute A in R is fully functionally dependent on Candidate key of R.

Example 1 on 2NF

Student (sid, sname, cname)

F.D. = { sid, cname -> sname, sid -> sname}

Then here **Candidate key is (sid ,cname)**

But sname is dependent on part of the candidate key sid for second F.D. Hence it is not in 2NF.

- 2NF decomposition =

  Table 1 (sid, sname) and

  Table 2 (sid, cname)

Example 2 on 2NF

- Guest (phone, name, address, room, floor, stay)

  F.D. = { phone , name --> address,
  phone --> room,
  name --> floor, stay}
- Not in 2NF
- Decomposition in 2NF is:

  Table1 (phone,name, address)

  Table 2 (phone, room)

  Table 3 (name, floor, stay)

Example 3 on 2NF

- Report (rid, title, author,subject)

  F.D. = { rid --> title, rid --> author,

  rid -->subject, author --> subject }

- In 2NF

Third Normal Form

A relation schema R is in 3NF if it is in 2NF and there should not be any transitive dependencies present inside the relation.

Conditions of 3NF

A relational schema R is in 3NF if for every F.D. X --> A associated with R either

- A should be a subset of X (i.e. trivial F.D.) or
- X is superkey of R or
- A is a part of some key (not just superkey)

Example 1 on 3NF

Sup_city (sid, status, city) and F.D. are

{ sid --> status, sid --> city and city --> status}

- It is already in 2NF
- After applying 3 conditions, it is not in 3NF
- Hence decomposition in 3NF is

  Table 1 (sid,city) and Table 2 (city, status)

Example 2 on 3NF

dept_advisor (sid, iid, dname)

F.D. = { sid, dname --> iid, iid --> dname}

- C.K. are (sid,dname) and (sid,iid)
- Hence super keys are : (sid,dname) and (sid,iid)
- Hence after applying conditions the relation is in 3NF.

Boyce Codd Normal Form (BCNF)

A relation schema is in BCNF if whenever a nontrivial functional dependency X -> A holds in R, then X is a super key of R.

Example 1 of BCNF

- R(A , B , C) and F.D. = { A --> B , B --> C}
- Not in BCNF
- Decomposition is

  Table 1 (A,B)

  Table 2 (B, C)

Example 2 of BCNF

- R(A , B , C) and

  F.D. = { A --> B , B --> C , C --> A}
- In BCNF