Relational Database Design

Normalization

Normalization

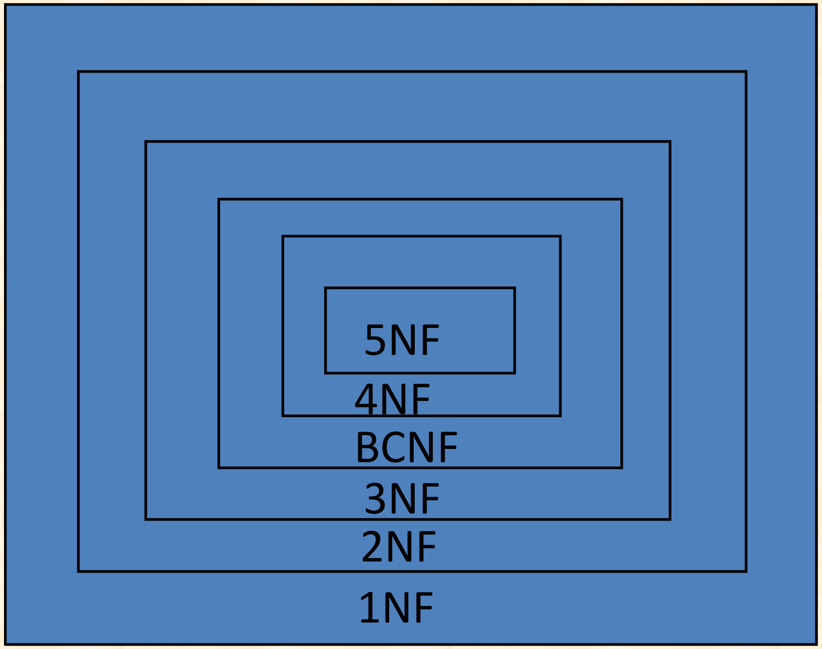
- A design technique used to design Relational Model.
- Can be defined as a process during which redundant relation schemas are decomposed by breaking up their attributes into smaller relation schemas that possess desirable properties.
- Basic objective of normalization is to reduce redundancy.
- Storing information several times leads to –
  - Wastage of storage space
  - High cost
  - Inconsistencies

Properties of a Normalized Relation

- No data values should be duplicated in different rows unnecessarily.
- A value must be specified for every attribute in a row.
- Each relation should be self- contained. In other words, if a row is deleted, important information should not be accidentally lost.
- When a row is added to a relation, other relations in a database should not be affected.
- A value of an attribute in a tuple may be changed independent of other tuples in the relation and other relations.

Types of Normal Forms

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



Dependencies

- Functional Dependency (FD):- An association between two attributes of the same relational database.
- One attribute is called determinant and the other is called determined.
- For each value of the determinant there is associated one and only one value of the determined.
- If A is the determinant and B is determined ,then we can say that A functionally determines B and is represented as A → B.

For each value of A ,there is only one value of B

A	B
1	1
2	4
3	9
4	16

Fully Functional Dependency (FFD)

- Is defined as attribute Y is FFD on attribute X if it is FD on X and not FD on any proper subset of X.
- E.g. (A,B) → C

  X Y

- C depends on both A & B and not on A or B.

Multivalued Dependency(MVD)

- MVD is when one attribute value is potentially a multi valued fact about another.
- If a relation R having A,B,C as attributes, B and C are multivalued facts about A which is represented as A→→B and A →→C, then multivalued dependency exist only if B and C are independent of each other.

Transitive Dependency

- Assume that A,B,C are the attributes of a relation R. If in a relation

  $A \rightarrow B$

  $B \rightarrow C$

Then $A \rightarrow C$

- This is transitive dependency

First Normal Form

- A relation is in 1NF if and only if all underlying domains contain atomic values or single value only.

Course code	Course name	Teacher name	Roll no	Name	System _used	Hourly _rate	Total_ hrs
C1	Visual Basic	ABC	100 101	A1 A2	P-1 P-II	20 30	7 3
C2	Oracle	DEF	100 104	A1 A7	P-I P-III	20 37	7 3
C3	C++	KJP	106	A7	P-IV	40	3

Table -1 Unnormalized Relation

1NF

- Two approaches used to convert a relation to 1NF-
  - Flattening the table
  - Decomposition of the table
- Flattening the table:- It removes repeating groups by filling in the "missing" entries of each incomplete row of the table with copies of their non – repeating attributes.

Course code	Course name	Teacher name	Roll no	Name	System_used	Hourly_rate	Total_hrs
C1	Visual Basic	ABC	100	A1	P-1	20	7
C1	Visual Basic	ABC	101	A2	P-II	30	3
C2	Oracle	DEF	100	A1	P-I	20	7
C2	Oracle	DEF	104	A7	P-III	37	3
C3	C++	KJP	106	A7	P-IV	40	3

Table 2 Normalized relation

1NF (Contd...)

2. Decomposition of the table:- In this original table is decomposed into two new tables.

- Involves separating the attributes of the relation to create the schemas of two new relations.
- Before decomposing the original table, it is necessary to identify an attribute or set of attributes that can be used as table identifier.
- Rule of decomposition
  - One of the two tables contains the table identifier of the original table and all the non- repeating attributes,
  - The other table contains a copy of the table identifier and all the repeating attributes.

1NF (Contd…)

Course code	Course name	Teacher name
C1	Visual Basic	ABC
C2	Oracle	DEF
C3	C++	KJP

PK

Table 3 Course

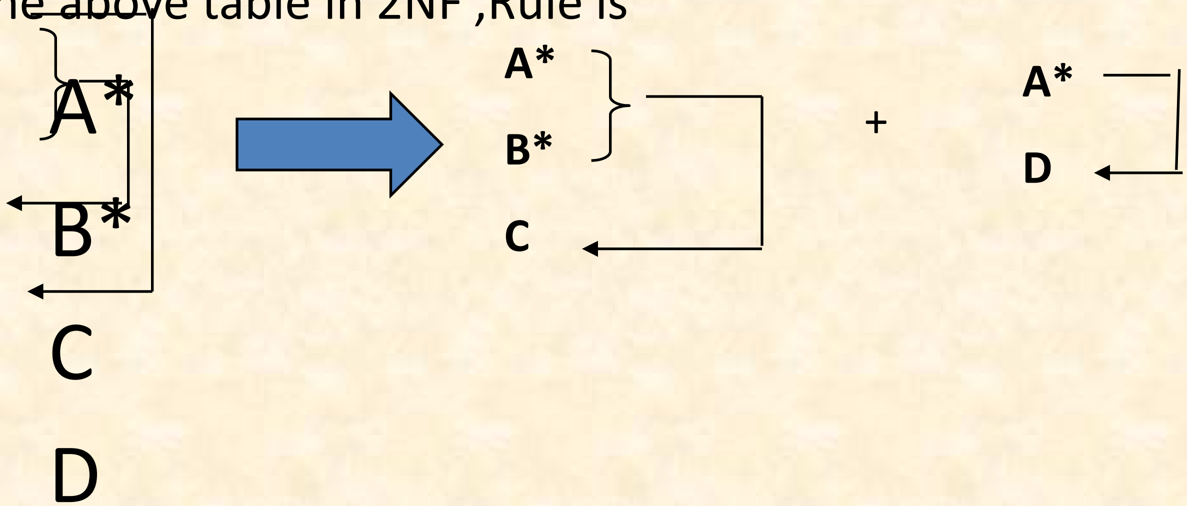
Course code	Roll no	Name	System_used	Hourly_rate	Total_hrs
C1	100	A1	P-1	20	7
C1	101	A2	P-II	30	3
C2	100	A1	P-I	20	6
C2	104	A7	P-III	37	3
C3	106	A7	P-IV	40	3

Table 4 Course student

PK

Second Normal Form

- A relation is said to be in 2NF if and only if it is in 1NF and every non - key attribute is fully dependent on the primary key.
- Table 4 Course_student is in 1NF but not in 2NF because non key attributes name, system_ used and hourly_ rate are not fully dependent on PK (Course code, Rollno).
- To convert the above table in 2NF ,Rule is



- The first relation contains the PK and the attributes that are fully dependent on the PK.
- Second relation contains the attributes that are partially dependent on the key and the key attribute on which it is dependent.

2NF(Contd...)

- Transformation of Course_student into 2NF

Course Code	Roll no	Total _hrs
C1	100	7
C1	101	3
C2	100	6
C2	104	3
c3	106	3

Table 5 Hours assigned

Contains attribute Total_hrs which is fully dependent on PK (Course code, Rollno)

Table 6 System Charge

Contains attributes that are partially dependent

Roll no	Name	System_used	Hourly_rate
100	A1	P-I	20
101	A2	P-II	30
104	A7	P-III	37
106	A7	P-IV	40

Third Normal Form

- A relation is said to be in 3NF if it satisfies the following conditions simultaneously –
  - R is already in 2NF
  - No non prime attribute is transitively dependent on the key i.e no non prime attribute functionally determines any other non –prime attribute.
- Table 6 System charge is not in 3NF because there is transitive dependency of a non prime attribute on the PK of the relation.
- In table 6 ,non-prime attribute Hourly_rate is transitively dependent on the pk rollno through FD.

Rollno → System_used

System_used → Hourly_rate

=> Rollno → Hourly_rate

This is transitive dependency

3NF(Contd…)

- Rule to convert to 3NF

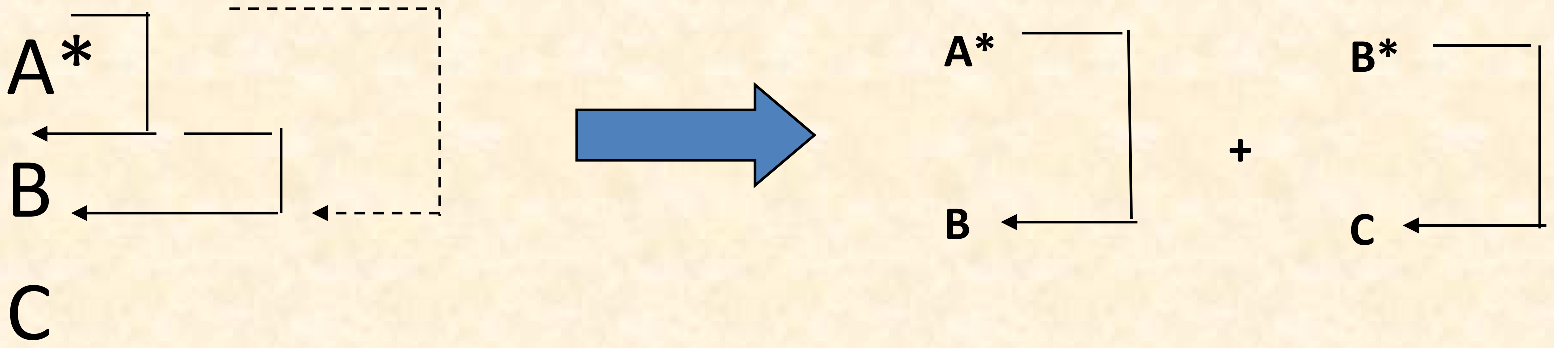


Table 6 can be represented in 3NF by dividing it into two relations.

- First relation contains the pk and the attributes that are dependent on it.
- Second relation contains the non-prime attribute and another non-prime attribute on which it is FD and making it as pk.

3NF(Contd…)

Table 7 System used

Rollno	Name	System_used
100	A1	P-I
101	A2	P-II
104	A7	P-III
106	A7	P-IV

PK

Table 8 Charges

System_used	Hourly_rate
P-I	20
P-II	30
P-III	37
P-IV	40

PK

Boyce Codd Normal Form

- BCNF states that –

  A relation R is said to be in BCNF if and only if every determinant is a candidate key.

- Consider table 6, System Charge (Course code, Rollno, Name, System_used, Hourly_rate, Total_hrs)

  (Course code, Rollno) → Total_hrs

  Rollno → (Name, System_used, Hourly_rate)

Rollno is a determinant but not a candidate key.

System_used is again a determinant but not unique.

So, Table 4 is not in BCNF.

BCNF (Contd…)

- To make it in BCNF ,divide it into three tables-
- First table Includes Course_time (Course code, rollno,Total_hrs)
- Second table includes Student_system (Rollno, system_used, Hourly_rate)
- Third includes Charge (System_used, Hourly_rate)

Difference between 3NF and BCNF

- A table in 3NF has to be in 2NF but in BCNF, a relation need not be in 2NF or 1NF.
- BCNF is stronger definition of 3NF.

Fourth Normal Form

- A relation is said to be in 4NF if and only if-
  - If R is already in 3NF or BCNF
  - If it contains no multi valued dependencies.
- Two things must be noted-
  - Firstly, in order for a table to contain MVD, it must have 3 or more attributes.
  - It is possible to have table containing two or more attributes which are independent multi valued facts about another attribute.

4NF(Contd…)

- Consider a relation

Course	Student_name	Textbook
Physics	Ankit	Mechanics
Physics	Ankit	Optics
Physics	Rahat	Mechanics
Physics	Rahat	Optics
Chemistry	Ankit	Organic
Chemistry	Ankit	Inorganic
English	Raj	Literature
English	Raj	Grammar

Table 9 Course_student_book

MVDs in the above table are-

Course →→ Student_name

Course →→ Textbook

4NF(contd…)

- Problems with MVD
  - If a new student join the physics course, then we have to make two insertions for that student in the database which is equal to the no. of physics books. And if there are hundred textbooks for a subject?
  - if a new book is introduced for a course , then again we'll have to make multiple insertions which is equal to the no. of students for that course.

So, there is high degree of redundancy which will lead to update problems.

Rule to transform a relation to 4NF-

A relation R having A,B,C as attributes can be no loss decomposed into two tables R1(A,B),R2(A,C) if and only if MVD A →→B/C hold in R.

4NF(Contd...)

- To convert table 9 into 4NF,two separate tables are formed-
  - **Course_student → (Course, Student_name)**
  - **Course_book → (Course, Textbook)**

Course	Student_ name
Physics	Ankit
Physics	Rahat
Chemistry	Ankit
English	Raj

Table 10 Course_student

Course	Textbook
Physics	Mechanics
Physics	Optics
Chemistry	Organic
Chemistry	Inorganic
English	Literature
English	Grammar

Table 11 Course_book

Fifth Normal Form

- A relation is said to be in 5NF if and only if-
  - R is already in 4NF
  - It cannot be further non- loss decomposed.
- Non loss decomposition is possible because of the availability of the join operator.
- In 5NF,consideration must be given to tables where non-loss decomposition can be achieved by decomposition into 3 or more separate tables.
- In join, a new table is formed which contains all the columns from both the joined tables whose tuples are defined by the restriction applied.

5NF(Contd…)

- E.g. consider a relation

Agent	Company	Product name
Suneet	ABC	Nut
Raj	ABC	Bolt
Raj	ABC	Nut
Suneet	CDE	Bolt
Suneet	ABC	Bolt

Table 12 It states that ABC makes both Nuts and Bolts but CDE makes only Bolts.

This table can be decomposed into 3 tables –

- **P1 (Agent, Company)**
- **P2 (Agent, Product name)**
- **P3 (Company, Product name)**

5NF(Contd…)

- P1

Agent	Company
Suneet	ABC
Suneet	CDE
Raj	ABC

P2

Agent	Product name
Suneet	Nut
Suneet	Bolt
Raj	Bolt
Raj	Nut

P3

Company	Product name
ABC	Nut
ABC	Bolt
CDE	Bolt

5NF(Contd...)

- Join P1 and P2

Agent	Company	Product_ name
Suneet	ABC	Nut
Suneet	ABC	Bolt
Suneet	CDE	Nut*
Suneet	CDE	Bolt
Raj	ABC	Bolt
Raj	ABC	Nut

Table P4

- **Join P4 with P3**

Agent	Company	Product_ name
Suneet	ABC	Nut
Suneet	ABC	Bolt
Suneet	CDE	Bolt
Raj	ABC	Bolt
Raj	ABC	Nut

Correct recomposition of the original table. So, the original table is in 5NF.

Further Normal Forms

- **Join dependencies** generalize multivalued dependencies
  - lead to **project-join normal form (**PJNF) (also called **fifth normal form**)
- A class of even more general constraints, leads to a normal form called **domain-key normal form**.
- Problem with these generalized constraints: are hard to reason with, and no set of sound and complete set of inference rules exists.
- Hence rarely used