Module - IV **Normalization**

Normalization

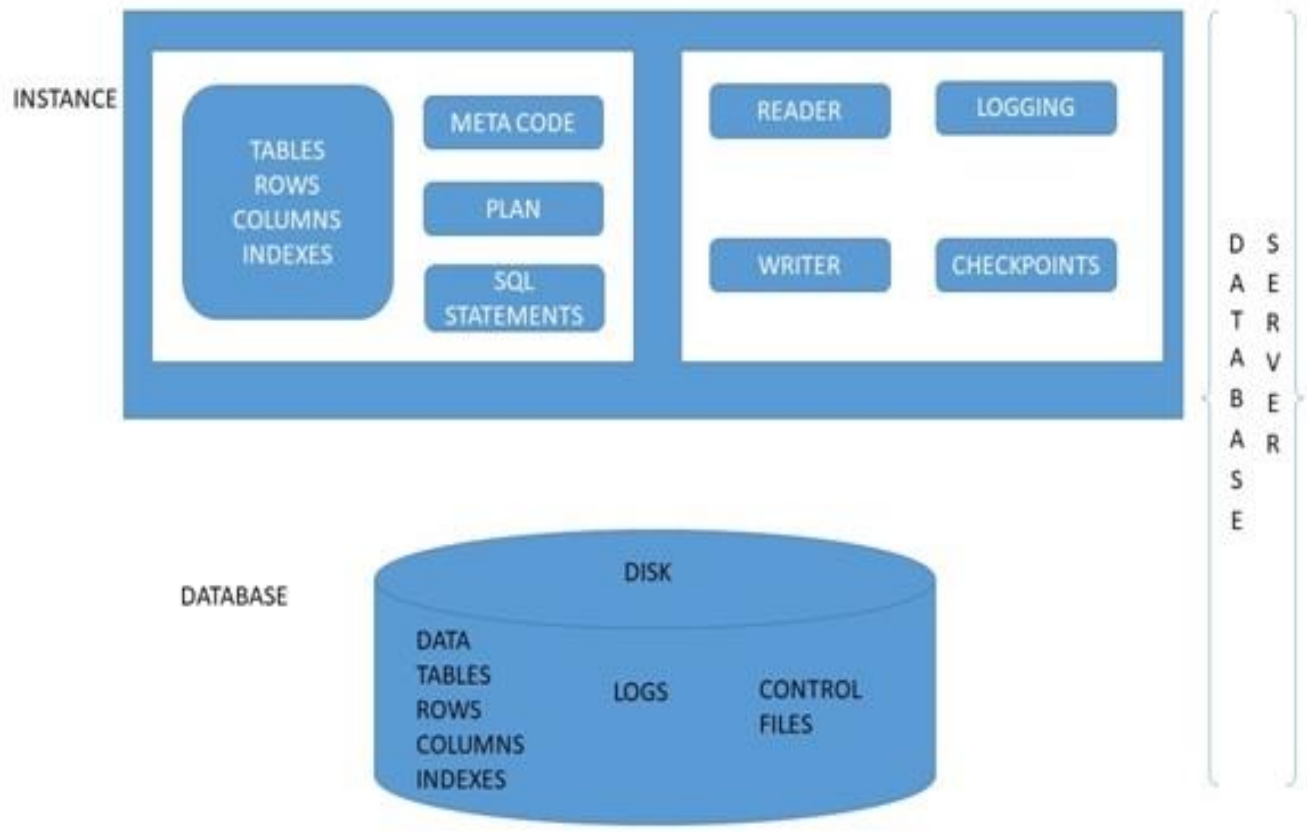
Relational database design: Introduction, Functional Dependencies (FDs), Normalization forrelational databases: 1NF, 2NF, 3NF and BCNF, Basic definitions of Multi Valued Dependencies, 4NF & 5NF.

Relational Database Management System (RDMS)

Relational database design (RDD) models' information and data into a set of tables with rows and columns. Each row of a relation/table represents a record, and each column represents an attribute of data. The Structured Query Language (SQL) is used to manipulate relational databases. The design of a relational database is composed of four stages, where the data are modeled into a set of related tables. The stages are −

- Define relations/attributes
- Define primary keys
- Define relationships
- Normalization

Relational databases differ from other databases in their approach to organizing data and performing transactions. In an RDD, the data are organized into tables and all types of data access are carried out via controlled transactions. Relational database design satisfies the ACID (atomicity, consistency, integrity, and durability) properties required from a database design. Relational database design mandates the use of a database server in applications for dealing with data management problems.



FUNCTIONAL DEPENDENCY

In a relation, an attribute depends on another attribute is called as *functional dependency*. The functional dependency (FD) is represented by the right arrow (→). For example, the relation EMPLOYEE contains the attributes Eno, Ename, Doj and Service. The attribute *Service* is functionally dependent on *Doj*.

EMPLOYEE : Doj → Service

In a functional dependency, the attribute on the left side of the right arrow is called *determinant*, the right side of the right arrow is called *dependent*. In the above example, determinant is *Doj* and dependent is *Service*.

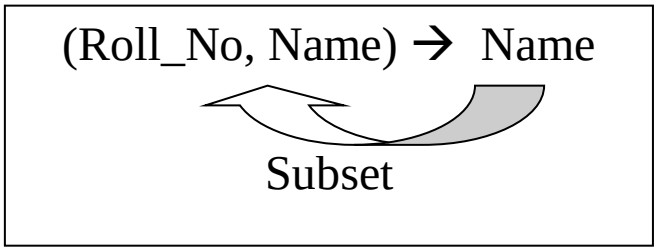
Types of Functional dependencies:

1. Trivial Functional Dependency:

In Trivial Functional Dependency, a dependent is always a subset of the determinant. i.e., if X→Y and Y is the subset of X, then it is called *trivial functional dependency*.

Eg:

Roll_No	Name	Age
1	Smith	17
2	Allen	18
3	Martin	18



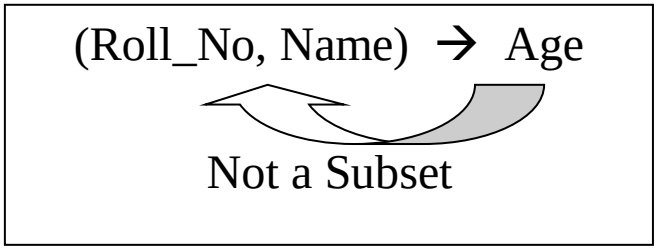
Here, ***Roll_No, Name → Name*** is a trivial functional dependency. Since the dependent ***Name*** is a subset of determinant set ***Roll_No, Name***.

2. Non-Trivial Functional Dependency:

In Non-Trivial Functional Dependency, a dependent is strictly not a subset of the determinant. i.e., if X→Y and Y is not a subset of X, then it is called *non-trivial functional dependency*.

Eg:

Roll_No	Name	Age
1	Smith	17
2	Allen	18
3	Martin	18



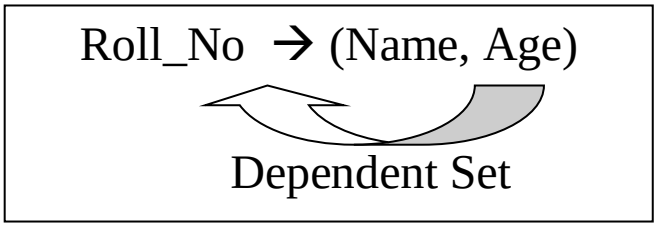
Here, ***Roll_No, Name → Age*** is a non-trivial functional dependency. Since the dependent ***Age*** is not a subset of determinant set ***Roll_No, Name***.

3. Multivalued Functional Dependency:

In Multi-Valued Functional Dependency, entities of the dependent set are not dependent on each other. i.e., if A→{B,C} and there exists no functional dependency between B and C, then it is called *multi-valued functional dependency*.

Eg:

Roll_No	Name	Age
1	Smith	17
2	Allen	18
3	Martin	18



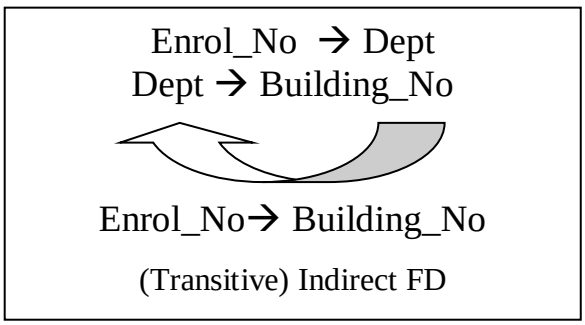
Here, ***Roll_No → Name, Age*** is a multi-trivial functional dependency. Since the dependents ***Name*** and ***Age*** are not dependent on each other (i.e ***Name → Age*** or ***Age → Name*** doesn't exist).

4. Transitive Dependency:

In Transitive Functional Dependency, dependent is indirectly dependent on determinant.(Or) A non key attribute is depends on another non key attribute called Transitive Dependency i.e., if A→B & B→C, then according to axiom of transitivity, A→C. This is a *transitive function dependency*.

Eg:

Enrol_No	Name	Dept	Building_No
1	Smith	EC	42
2	Allen	IT	43
3	Martin	CS	44



Here, ***Enrol_No→Dept*** and ***Dept→Building_No***, Hence, according to the axiom of transitivity, ***Enrol_No→Building_No*** is a valid this is an indirect functional dependency, hence called transitive functional dependency.

NORMALIZATION

Normal Forms:-

Def:- The process of decomposing the relations into smaller and well structured relations for reducing the amount of redundancy and inconsistency will be called "Normalization".

If the table contains redundancy, the following anomalies (errors) are occurred.

i) Insert Anomaly
ii) Delete Anomaly
iii) Update Anomaly

i) Insert Anomaly:- The errors that are occurred, when we insert a new record. Those errors are called as "Insert Anomalies".

ii) Delete Anomaly:- The errors that are occurred, when we delete a existing a record. Those errors are called as "Delete Anomalies"

iii) Update Anomaly:-The errors that are occurred, when we update a existing record. Those errors are called "Update Anomalies".

For achieving the normalization we have to apply set of rules. These rules are known as "Normal Forms". They are

1) Basic Normal Forms

2) Advanced Normal Forms **(OR)** Higher Normal Forms

Functional Dependency:- A functional dependency is a constraint between two or more attributes. For example consider two attributes A and B such that "the attribute" "B" is functionally depends on the attribute "A". It can be represented in the following way

i) Fully Functional Dependency
ii) Partial Functional Dependency
iii) Transitive Dependency

i) Fully Functional Dependency:- In a relation a non-key attributes dependence on key attributes then that dependency is called "Fully Functional Dependency".

ii) Partial Functional Dependency:- In a relation a non-key attributes depends on part of key attribute that dependency is called "Partial Functional Dependency"

iii) Transitive Dependency:- In a relation a non- key attribute depends on another non-key attribute. Then that dependency is called "Transitive Dependency"

Ex:- R(A,B,C,D)

Here A,B is a composite key i.e., key attribute

A,B➔ C Fully functional dependency

B➔C
A➔D } Partial Dependency

C➔D Transitive dependency.

Determinant:- The attribute on the left hand side of the arrow ➔ in the functional dependency will be called as "Determinant".

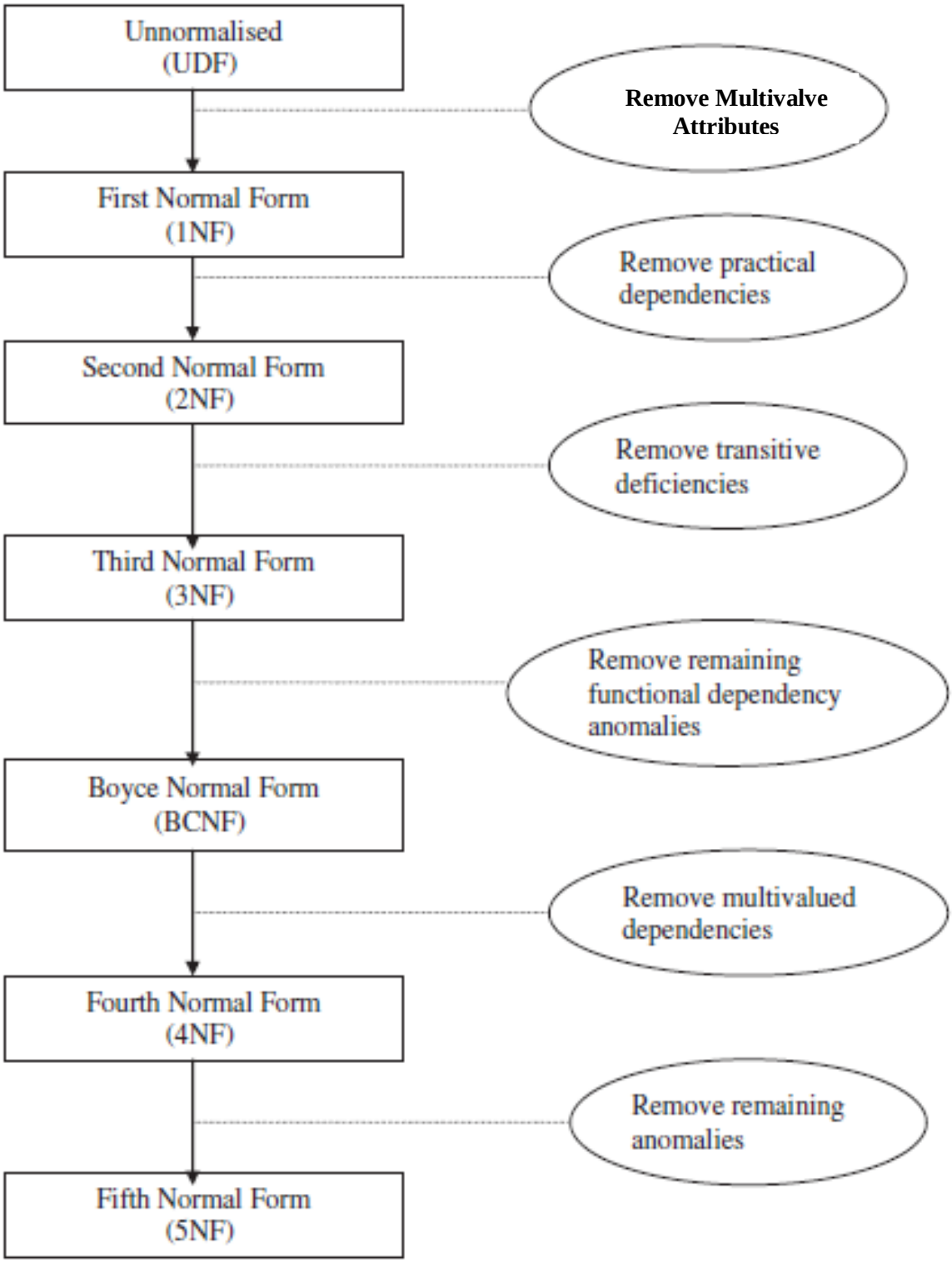
Ex:- R(A,B,C,D)

A➔B C➔D Here A,C are called Determinants and B,D are called dependents

For achieving the normalization we have to apply set of rules. These rules are known as "Normal Forms". They are

1) Basic Normal Forms

2) Advanced Normal Forms **(OR)** Higher Normal Forms



1. Basic Normal Forms:- There are 3 types of Basic Normal Forms. They are

i) First Normal Form(1NF)

ii) Second Normal Form (2NF)

iii) Third Normal Form (3NF)

i) First Normal Form:- A relation is said to be in 1NF, if it does not contain "Multivalued Attributes". i.e all the values of columns are "**Atomic".**

Ex:-

EMPNO	ENAME	SAL	COURSE_NAME	DATE_COMPLETED
100	Smith	5000	C	05-Oct-2010
			C++	10-Nov-2010
101	Clark	6000	Java	11-Jan-2011
			Oracle	15-Feb-2011

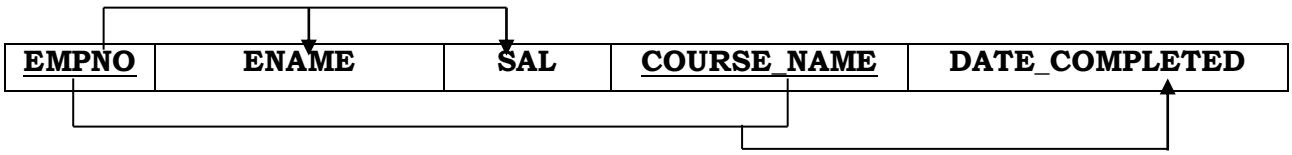
In the above relation is not in First Normal Form. Because the relation having Multiple values of a column. To convert the above relation into 1NF follow below.

EMPNO	ENAME	SAL	COURSE_NAME	DATE_COMPLETED
100	Smith	5000	C	05-Oct-2010
100	Smith	5000	C++	10-Nov-2010
101	Clark	6000	Java	11-Jan-2011
101	Clark	6000	Oracle	15-Feb-2011

Hence the above relation is in First Normal Form

ii) Second Normal Form:- A relation is said to be in 2nd NF, if it satisfies the following properties.

- The relation must be in 1NF
- It does not contain partial functional dependency



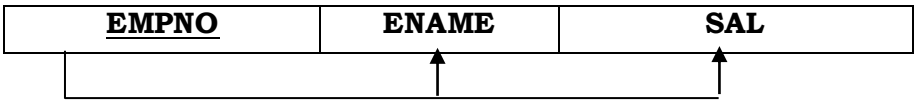
The functional dependencies are

EMPNO, COURSE_NAME➔ DATE_COMPLETED

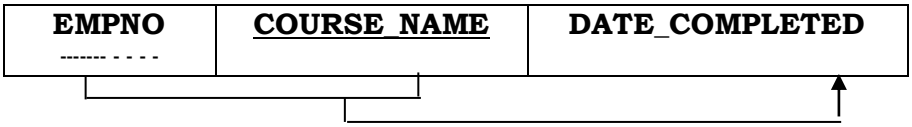
EMPNO➔ENAME, SAL

In second functional dependency there is a partial functional dependency so we have to decompose the relation the following.

EMPLOYEE



COURSE



Hence the above two relations is in 2nd NF

iii) Third Normal Form:- A relation is said to be in 3NF, if it satisfies the following properties.

- The relation must be in 2NF
- It does not contain transitive dependency

Consider the following Relation

CUSTOMER

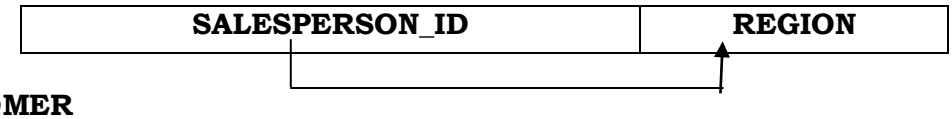
The functional dependencies are

CUST_ID→ CUST_NAME, SALESPERSON_ID, REGION

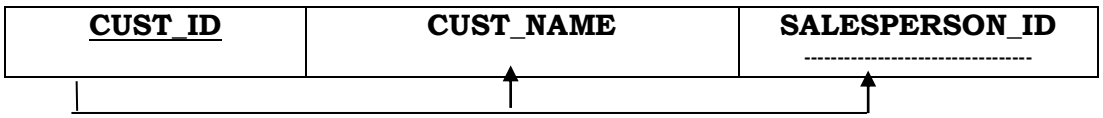
SALESPERSON_ID→ REGION

In Second functional dependency there is a transitive dependency so we have to decompose the relation.

SALESPERSON



CUSTOMER



Hence the relation is in 3 NF

Advanced or Higher Normal forms:-

There are 3 types of Higher Normal forms. They are

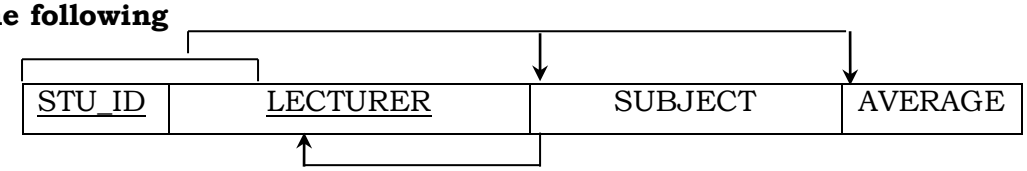
i) Boyce Codd Normal Form(BCNF)
ii) Fourth Normal Form(4NF)
iii) Fifth Normal Form (5NF)

i) Boyce Codd Normal Form(BCNF) :- A relation is said to be in BCNF, if it satisfies the following properties.

- The relation must be in 3NF
- The determinant must be a candidate key.

Candidate Key:- A field or combination of fields used for identifying a single record will be called as a "Candidate Key".

Consider the following



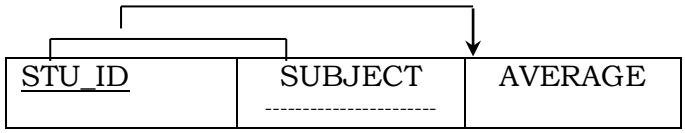
The functional dependencies are

STU_ID, LECTURER → SUBJECT, AVERAGE

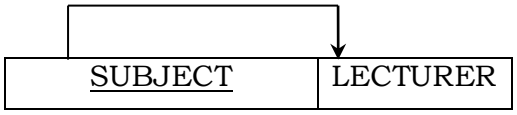
SUBJECT→LECTURER

The above relation is in 3NF but not BCNF. So we can decompose the relation. After decomposing the following

STUDENT



SUBJECT



Hence, the above two relation are in BCNF

ii) Fourth Normal Form:- A relationship said to be in 4NF, If it satisfies the following properties.

- The relation must be in BCNF
- It does not contain multivalued dependencies.

Multivalued Dependency:- A Multivalued dependency exists when there are at least 3 attributes let a relation with 3 relations A, B, and C. For a single value of the attribute 'A', there can be many values of the attributes 'B' and 'C'. But the attributes 'B' and 'C' are independent each other. Consider the following relation

SUBJECT	LECTURER	TEXT BOOK
Computers	Thanuja	T1
	Srinivas	T2
	Vasavi	
Maths	Chiranjeevi	T3
	Subbarao	T4

The above information can also be represented as below.

SUBJECT	LECTURER	TEXT BOOK
Computers	Thanuja	T1
Computers	Thanuja	T2
Computers	Srinivas	T1
Computers	Srinivas	T2
Computers	Vasavi	T1
Computers	Vasavi	T2
Maths	Chiranjeevi	T3
Maths	Chiranjeevi	T4
Maths	Subbarao	T3
Maths	Subbarao	T4

The above relation having the data redundancy so we can decompose the relation. After decomposing the relation as show below

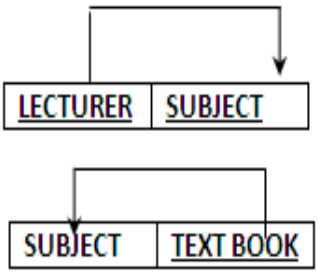
COURSE

SUBJECT	LECTURER
Computers	Thanuja
Computers	Srinivas
Computers	Vasavi
Maths	Chiranjeevi
Maths	Subbarao

BOOK

SUBJECT	TEXTBOOK
Computers	T1
Computers	T2
Maths	T3
Maths	T4

OR



Hence the above two relations are in 4NF

iii) Fifth Normal Form(5NF): A relation said to be in 5 NF, if it satisfied the following properties

- The relation must be in 4NF
- It does not contain Joined dependency

Joined dependency:- Joined dependency means the relation contains minimum 3 attributes and each attribute may functionally dependent on remaining attributes.

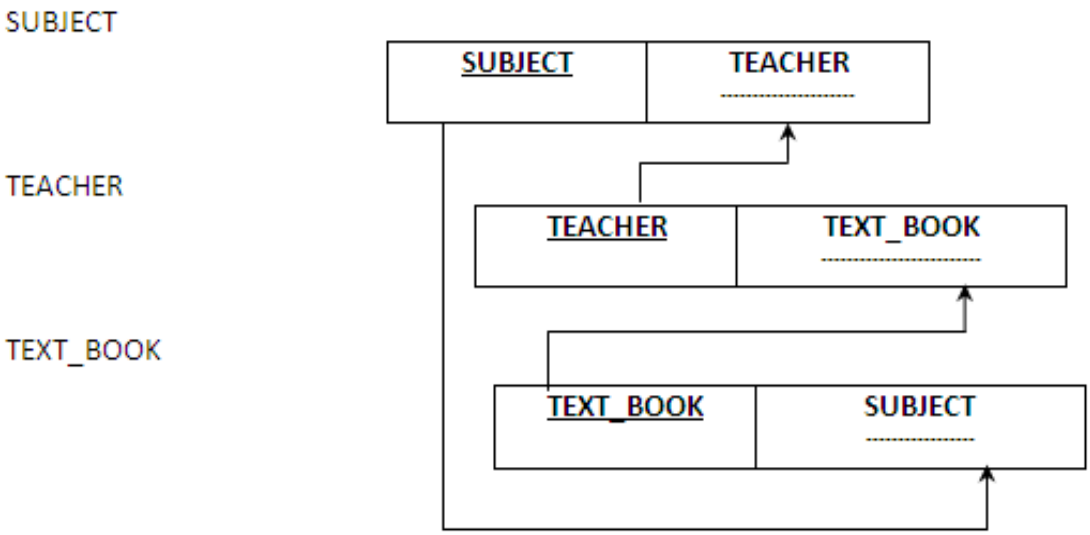
Consider the following relation "CLASS"

SUBJECT	TEACHER	TEXT_BOOK

Hear primary key is the composite key of SUBJECT, TEACHER and TEXT_BOOK. The functional dependencies are

SUBJECT, TEACHER → TEXT_BOOK
TEACHER, TEXT_BOOK → SUBJECT } Joined
TEXT_BOOK, SUBJECT → TEACHER

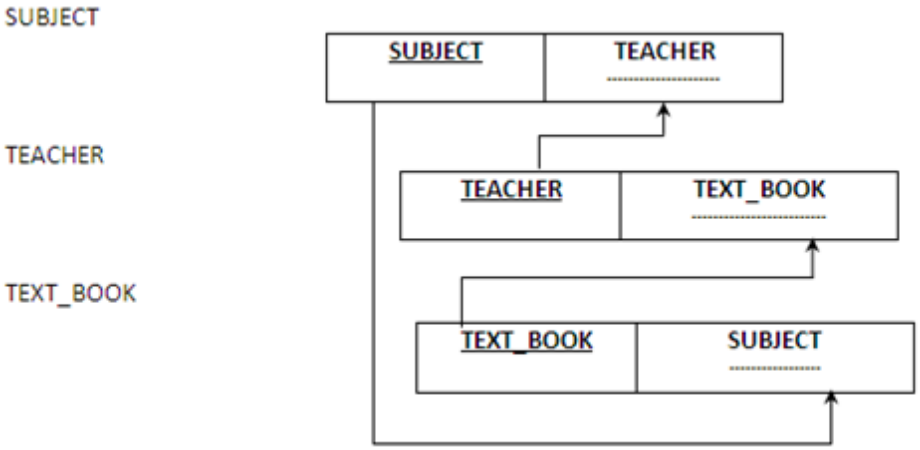
After decomposing the above relation



Hence the above 3 relations are in 5 NF

DENORMALIZATION

Denormalizaiton is the process of transforming normalized relations into unnormalized relations. The problem with normalization is that takes tables are decomposed into no of database tables. There fore retrieving information these tables take much time. So system speed will be reduced. For example the name of the an account holder has to be displayed along with the account number and cash balance. Every time the account is accessed. In our Normalization schema, this required a join of "account" with "depositor". One alternative is to create a relation containing all the attributes of "account" and "depositor". This makes displaying the account information faster.



Dear One,

If you are in right path, then no one can judge your mistakes, without your presence

By
Venukumar .DVH
M.Sc., B.Ed., M.Ed., M.Tech

ALL THE BEST