Normalization :- Database design is based on ER model but there exists some amount of inconsistency, redudancy, ambiguity. Inorder to remove the problem from the database design we need to do some refinement process, that process is called as normalization.

Normalization process involves bottom-up approach

Normalization is mainly used to eliminate data duplication and it allows insert update and delete operation works more efficiently and improve performance of database.

Normalization process

Users → users requirement specification ← Sources describing the enterprise such as data dictionary and data model

↓

Unnormalize form (UNF)

Remove repeating groups

↓

First Normal form (1NF)

Remove partial dependency

↓

Second Normal form (2NF)

Remove Transitive dependency

↓

Third Normal form (3NF)

Make sure every determinant is a candidate key

↓

Boyce-Codd Normal form (BCNF)

Remove multivalue dependencies

↓

Fourth-Normal form (4NF)

Remove join dependencies

↓

Fifth-Normal form (5NF)

Normalization is considered as a strong mathematical foundation. Before we discuss normalization lets discuss a few building blocks of normalization

They are i] Determinant

ii] Functional dependency

iii] Full functional dependency

iv] partial dependency

v] Transitive dependency (or) indirect dependency.

Determinant:- In a given relation R(X,Y) X and Y are attributes, where attribute X is said to be determinant if it is uniquely defines the value of attribute Y.

ii] The dependency of an attribute is represented as

X → Y

eg:- Marks → grade

Here, the determinant may or maynot be a key attribute

Functional dependency

⇒ In a given relation R X and Y are attributes, where Y is functionally dependent on attribute X if and only if attribute X determines exactly one value of attribute Y. Here X may be composite in nature.

eg:- consider a relation

report (cid, sid, sname, cname, DOB, marks, grade)

FD1, marks → grade

FD2, cid → cname

FD3 sid → sname

FD4 sid, cid → marks.

Partial Function dependency:-

In a given relation R x and y are attribute, where y is partially function dependent on x if and only if attribute y depends on attribute x, where x may be composite in nature

Full functional dependency:-

In a given relation R. x and y are attribute where y is fully functional dependent on attribute x if and only if attribute y is not functionally dependent on a subset of attribute x, where x may be composite in nature

Transitive Functional Dependency:-

Sid, cid → Marks
Marks → grade
Sid, cid → grade

Indirect dependency

In a given relation R x, y and z are attributes, where attribute z is fully functionally dependent on attribute y, attribute y is fully functionally dependent on x, so that attribute z is transitive dependend on attribute x, where x may be composite in nature

i] Normal form (1NF):- A relation R is said to be in first Normal form if and only if it contains atomic values i.e all the attributes of relation must contain atomic values. Atomic means which cannot be divisible further (or) the smallest level to which the data can be broken down and remain meaningful is called atomic.

Rules to maintain 1st normal form

i] All columns (or) attribute must contain a single value

ii] Attributes or Columns in a relation must have a unique name

iii] Domain of attribute must not be changed.

iv] Order of data must be in any manner.

items	Hues	cost
1.	Pink, black	15.99
2.	red	23.99
3.	black	17.50
4.	red, gray	9.99
5.	brown	29.99

=>

items	Hues	cost
1	Pink	15.99
1	black	15.99
2	red	23.99
3	black	17.50
4	red	9.99
4	grey	9.99
5	brown	29.99

ii] Second Normal form (2NF) :- Relation R is said to be in 2NF if and only if

1] it must be in 1NF

ii] no partial dependency exists b/w key and nonkey attributes

①

Candidate - Id	Course - Id	Candidate name	Course_name
C829	A09	Beverly	CSS
C736	A07	Sheldon	PHP
C546	A03	Lenoard	HTML
C952	A05	Zach	Ruby

(2)

Teacher_ID	Subject	Teacher_age
25	chemistry	30
25	Biology	30
47	English	35
83	Math	38
83	computer	38

1st table in Second Normal form

(1)

Candidate_ID	Course_id	Candidate name
C829	A09	Beverly
C736	A07	Sheldon
C546	A03	Leonard
C952	A05	Zach

Course Id	Course_name
C829	CSS
C736	PHP
C546	HTML
C952	Ruby

(2) 2nd table in Second Normal form

Teacher_id	Subject
25	Chemistry
25	Biology
47	English
83	Math
83	computer

Teacher_id	Teacher_age
25	30
25	30
47	35
83	38
83	38

Third Normal form (3NF) :- A relation R is said to be in 3NF if and only if

i] it must be in the 2NF

ii] No transitive dependency exists b/w key and non-key attributes.

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

Emp-zip	Emp-state	Empcity
201010	UP	Noida
02228	US	Boston
60007	US	Chicago
06389	UK	Norwich
462007	MP	Bhopal

Emp-id	Empname	Empzip
222	Harry	201010
333	stephan	02228
444	lan	60007
555	katharine	06389
666	john	462007

Boyce-Codd Normal form (BCNF) :- which will be defined as 3.5NF. In a given relation R is said to be BCNF if and only if

i] it must be in the 3NF.

ii] Every determinant must be a candidate key

iii] BCNF is also called as stronger 3NF but every 3NF is not a BCNF. The determinant which is a candidate key in a relation then only the relation is said to be BCNF whereas in 3NF determinant may or may not be a key attribute

Sid-ID	email-ID	cid	Marks
S101	AAA@gmail.com	M3	60
S101	AAA@gmail.com	P3	70
S102	BBB@gmail.com	C3	80
S102	BBB@gmail.com	M3	90
S103	ccc@gmail.com	P3	70
S103	ccc@gmail.com	C3	80

In the above table we have two candidate key.
1) sid, cid
2) email-id, cid.
and marks is a non key attribute
Hence the above table is in 3NF because the nonkey attribute i.e marks (or) non-transitively and fully functionally depend upon the key attribute of the given relation.

1) Sid-id which will decides email-id.
2) email-id decides student-id
3) student-id and course-id which will decides the rest of the attribute of the relation
4) course-id and email-id which will decide the rest of the attributes in the given relation.

In the above 4 determinants two determinants combination of {sid-Id, cid} and {cid, email-Id} are candidate keys but ii] Sid-id decides email-id, but sid-id alone is not a candidate key. Similarly email-id decides sid-id but email_id alone is not a candidate key

iii] To make the above relation in BCNF we need to decompose the relation into two relations
a] relation contains attribute of sid-id, email-id.
b] another relation contains attribute of sid-id, cid, marks

Result -R1

sid_id	email_id
S101	AAA@gmail.com
S102	BBB@gmail.com
S103	CCC@gmail.com

sid_id	cid	marks
S101	M3	60
S101	P3	70
S102	C3	80
S102	M3	90
S103	P3	70
S103	C3	80

i] The above relations are not only in 3NF but they are also in BCNF

ii] In table 1 sid-id decides email-id and email-id decides sid-id where both the determinants are candidate keys as well as the second table sid, cid decides marks where the determinant is the candidate key so that both the tables are in 3NF as well as BCNF

Properties of relational decomposition

Decomposition is a tool which is used to eliminate redudancy and it also allows not to occur any new Problems.

ii] Decomposition is used to check the recover the original relation from decomposed smaller relation without having loss of information which is called as loss less join decomposition property (or) non-additive decomposition property and it is also used to check whether the integrity constraints working efficiently or not. By using a property called as dependency preserving property.

The Two properties of relational decomposition

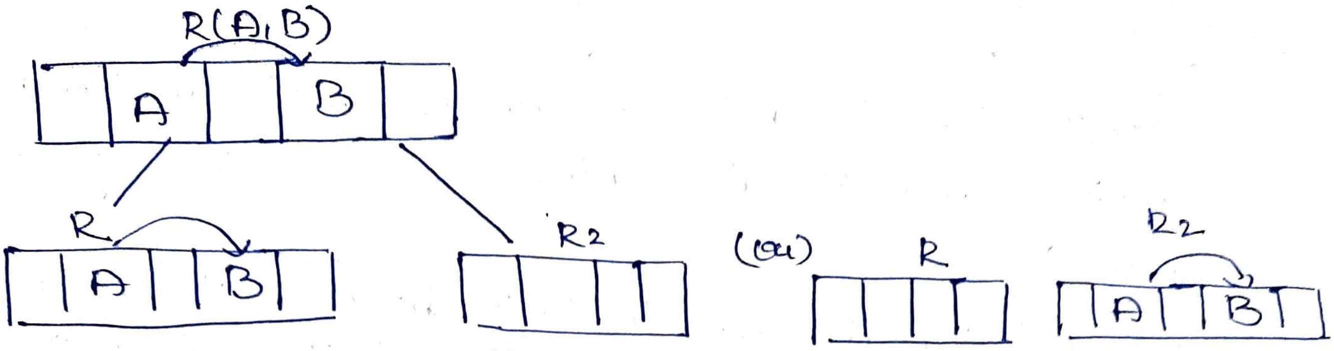
1) Lossless join decomposition property (or) non-additive decomposition property

2) Dependency preserving property.

Dependency preserving decomposition property

When there is a functional dependency of A determines B then whenever we see that attribute value of A, we must see the same attribute value for B which is called as preserving functional dependency.

ii] In a given relation R where A and B are attributes where B is functionally dependent on A (A -> B). So when we decompose the relation R into two relations R_1 and R_2, it must also hold functional dependency.



If a relation R is decomposed into multiple relation or smaller relation then the attributes which are holding functional dependencies must be present in any of the smaller relation.

Lossless decomposition property

whenever a given relation R is decomposed into smaller relations and if you join the smaller relation in order to get the original relation by using join i.e natural join without having any loss of information then that type of property is called as lossless join decomposition property (or) non-additive join decomposition property

R

R1 R2 (Decompose)

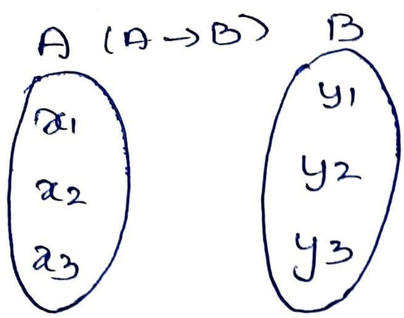
R (join)

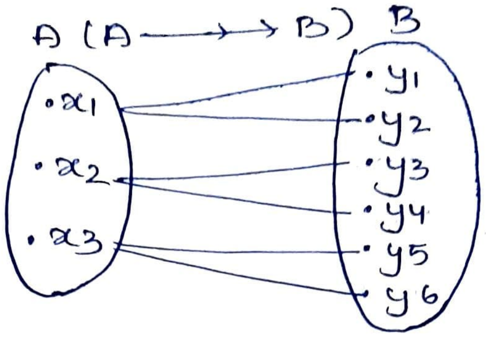
Multivalued dependencies

Multivalued dependencies is a dependency where one attribute value is potentially muti-valued fact about another attribute

$A \rightarrow B$ (FD)

$A \twoheadrightarrow B$ (MVD)





Multivalue dependencies :- Multivalue dependencies are the dependency where two or more independent attribute or multivalued fact about the same attribute within the same relation.

Fourth Normal form (4NF) :- A relation R is said to be in the 4NF if and only if it satisifies the following condition

i] Relation must be in 3NF (or) BCNF

ii] No multivalue dependencies exist in the relation

Rules to transform a relation into 4NF

Consider a relation R with the following attributes A, B, C ⇒ R(A, B, C) where B and C are multivalued facts of attribute A which exists a multivalued dependency in relation so that we need to decompose the relation into two relations, one relation contain (A, B) as attribute which comes another relation contain (A, C) as attribute which comes under 4NF

R(A, B, C)

R(A, B) R2(A, C)

Course-name	Sname	textbook
C	Ravi	X1
C	Ravi	X2
C	Arjun	X3
Java	Shiva	core-java
Java	Shiva	adv-java
SQL	Raj	B1
SQL	Raj	B1

cname	Sname
C	Ravi
C	Arjun
java	Shiva
SQL	Raj

cname	sname
C	X1
C	X2
C	X3
java	core-java
java	core-java
SQL	B1
SQL	B2

Fifth Normal form (5NF):- Join dependency refer lossless join decomposition property. This normal form is rarely used for practicals, but is mainly useful for theoritical purpose

ii] join dependency refers decomposition a relation into multiple relation without any loss of information and it should maintain the dependencies of original relation.

iii] A relation R is said to be in 5NF if and only if the following conditions are satisified.

a] It must be in 4NF

b] If the relation exist join dependency the decomposition The relation and the decomposition should be done in a way that further it should not be non-loss decomposed.

Factory	Component	Project
GM	Engine	MPC
GM	Gearbox	125A
Honda	Engine	125A
GM	Engine	125A

A factory produces a component and the component is delivery to a project.

R1 : Factory, component $\quad R_1 \bowtie R_2 \bowtie R_3 = R.$

R2 : Component, project

R3 : Factory, Project