NORMALIZATION & TYPES OF NORMALIZATION

1. The property of normalization of relations which guarantees that functional dependencies are represented in separate relations after decomposition is classified as

A. nonadditive join property
B. independency reservation property
C. dependency preservation property
D. additive join property

2. Third normal form is based on the concept of _____.

A. Closure Dependency
B. Transitive Dependency
C. Normal Dependency
D. Functional Dependency

1. What are the anomalies which can be present due to un-normalised relations?

A. Selection

B. Insertion

C. Addition

D. None of these

2. Which of the following statements is correct regarding normalization?

A. It reduces update anomalies

B. It increases insertion anomalies

C. It maximizes redundancy

D. There is no zero normal form relations

1. For a database relation R(a, b, c, d) where the domains of a, b, c, d include only the atomic values. The functional dependency a → c, b → d holds in the following relation

A. In 1NF not in 2NF

B. In 2NF not in 3NF

C. In 3NF

D. In 1NF

1) DEFINE NORMALIZATION

Normalization can be defined as :-

- A process of organizing the data in database to avoid data redundancy, insertion anomaly, update anomaly & deletion anomaly.
- A process of organizing data into tables in such a way that the results of using the database are always unambiguous and as intended. Such normalization is intrinsic to relational database theory. It may have the effect of duplicating data within the database and often results in the creation of additional tables.

Types of normalization

- 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)

First Normal Form (1NF)

First normal form enforces these criteria:

- Eliminate repeating groups in individual tables.
- Create a separate table for each set of related data.
- Identify each set of related data with a primary key

First Normal Form

Table_Product		
Product Id	Colour	Price
1	Black, red	Rs.210
2	Green	Rs.150
3	Red	Rs. 110
4	Green, blue	Rs.260
5	Black	Rs.100

This table is not in first normal form because the “Colour” column contains multiple Values.

After decomposing it into first normal form it looks like:

Product_id	Price
1	Rs.210
2	Rs.150
3	Rs. 110
4	Rs.260
5	Rs.100

Product_id	Colour
1	Black
1	Red
2	Green
3	Red
4	Green
4	Blue
5	Black

Second Normal Form (2NF)

A table is said to be in 2NF if both the following conditions hold:

- Table is in 1NF (First normal form)
- No non-prime attribute is dependent on the proper subset of any candidate key of table.

An attribute that is not part of any candidate key is known as non-prime attribute.

No partial dependency should be there in the table.

SECOND NORMAL FORM

Table purchase detail		
Customer_id	Store_id	Location
1	1	Patna
1	3	Noida
2	1	Patna
3	2	Delhi
4	3	Noida

- This table has a composite primary key i.e. customer id, store id.
- The non key attribute is location.
- In this case location depends on store id, which is part of the primary key.
- Partial dependency is there.

After decomposing it into second normal form it looks like:

Table Purchase	
Customer_id	Store_id
1	1
1	3
2	1
3	2
4	3

Table Store	
Store_id	Location
1	Patna
2	Delhi
3	Noida

Third Normal Form (3NF)

A table design is said to be in 3NF if both the following conditions hold:

- Table must be in 2NF
- Transitive functional dependency of non-prime attribute on any super key should be removed.

An attribute that is not part of any candidate key is known as non-prime attribute.

In other words 3NF can be explained like this: A table is in 3NF if it is in 2NF and for each functional dependency X-> Y at least one of the following conditions hold:

- X is a super key of table
- Y is a prime attribute of table

An attribute that is a part of one of the candidate keys is known as prime attribute.

THIRD NORMAL FORM

Table Student Details			
Book_id	Genre_id	Genre type	Price
1	1	Fiction	100
2	2	Sports	110
3	1	Fiction	120
4	3	Travel	130
5	2	sports	140

- Book_id is the Candidate key of the table
- Book_id --→ Genre_id
- Genre_id --→ Genre_type

After decomposing it into third normal form it looks like:

TABLE BOOK		
Book_id	Genre_id	Price
1	1	100
2	2	110
3	1	120
4	3	130
5	2	140

TABLE GENRE	
Genre_id	Genre type
1	Fiction
2	Sports
3	Travel

Boyce-Codd Normal Form (BCNF)

- It is an advanced version of 3NF that's why it is also referred as 3.5NF.
- BCNF is stricter than 3NF.
- A table complies with BCNF if it is in 3NF and for every functional dependency X->Y, X should be the super key or candidate key of the table.
- LHS of each Functional dependency should be candidate key or super key of the table.
- Non-prime attribute should be determined by a prime attribute only.
- No non-prime attribute should be determined by another non-prime attribute.

Boyce-Codd Normal Form

Student	Course	Teacher
Aman	DBMS	AYUSH
Aditya	DBMS	RAJ
Abhinav	E-COMM	RAHUL
Aman	E-COMM	RAHUL
abhinav	DBMS	RAJ

- KEY: {Student, Course}
- Functional dependency

{student, course} -> Teacher

Teacher-> Course

- Problem: teacher is not superkey but determines course.

After decomposing it into Boyce-Codd normal form it looks like:

Student	Course
Aman	DBMS
Aditya	DBMS
Abhinav	E-COMM
Aman	E-COMM
Abhinav	DBMS

Course	Teacher
DBMS	AYUSH
DBMS	RAJ
E-COMM	RAHUL

Fourth Normal Form (4NF)

- Fourth normal form (4NF) is a level of database normalization where there are no non-trivial multivalued dependencies other than a candidate key.
- It builds on the first three normal forms (1NF, 2NF and 3NF) and the Boyce-Codd Normal Form (BCNF).
- It states that, in addition to a database meeting the requirements of BCNF, it must not contain more than one multivalued dependency.

FOURTH NORMAL FORM

Student	Major	Hobby
Aman	Management	Football
Aman	Management	Cricket
Raj	Management	Football
Raj	Medical	Football
Ram	Management	Cricket
Aditya	Btech	Football
Abhinav	Btech	Cricket

- Key: {students, major, hobby}
- MVD: ->-> Major, hobby

After decomposing it into fourth normal form it looks like:

Student	Major
Aman	Management
Raj	Management
Raj	Medical
Ram	Management
Aditya	Btech
Abhinav	Btech

Student	Hobby
Aman	Football
Aman	Cricket
Raj	Football
Ram	Cricket
Aditya	Football
Abhinav	Cricket

Fifth Normal Form (5NF)

A database is said to be in 5NF, if and only if,

- It's in 4NF.
- If we can decompose table further to eliminate redundancy and anomaly, and when we re-join the decomposed tables by means of candidate keys, we should not be losing the original data or any new record set should not arise. In simple words, joining two or more decomposed table **should not lose records nor create new records**.

Conditions for joining the tables should follow below rules:

1. R1 U R2 must be equal to R.
2. Intersection of R1 and R2 must have atlest one common attribute
3. The common attribute must be a candidate key for R1 or R2 or both.

FIFTH NORMAL FORM

Seller	Company	Product
Aman	Coca cola company	Thumps Up
Aditya	Unilever	Ponds
Aditya	Unilever	Axe
Aditya	Uniliver	Lakme
Abhinav	P&G	Vicks
Abhinav	Pepsico	Pepsi

- Key: {seller, company, product}
- MVD: Seller ->-> Company, product Product is related to company.

After decomposing it into fifth normal form it looks like:

Seller	Product
Aman	Thumps Up
Aditya	Ponds
Aditya	Axe
Aditya	Lakme
Abhinav	Vicks
Abhinav	Pepsi

Seller	Company
Aman	Coca cola company
Aditya	Unilever
Abhinav	P&G
Abhinav	Pepsico

Continued in next slide…

Company	Product
Coca cola company	Thumps Up
Unilever	Ponds
Unilever	Axe
Unilever	Lakme
Pepsico	Pepsi
P&G	Vicks

Polling

1. A table has fields F1, F2, F3, F4, and F5, with the following functional dependencies:
   F1->F3
   F2->F4
   (F1,F2)->F5
   in terms of normalization, this table is in
   A. NF
   B. 2NF
   C. 3NF
   D. None of the above

Polling

- From the following instance of relation schema R(A, B, C) we can conclude that:

A	B	C
1	1	1
1	1	0
2	3	2
2	3	2

a) A functionally determines B and B functionally determines C.

b) A functionally determines B and B does not functionally determines C.

c) A does not functionally determines B and B does not functionally determine C.

d) None of the above.

Polling

1. A schema describes

(A) Record Relationship

(B) Data Elements

(C) Record and files

(D) All of the above

2. An abstraction concept for building composite objects from their component object is called

(A) Specialization

(B) Normalization

(C) Generalization

(D) Aggregation

Polling

1. The operation which is not considered a basic operation of relational algebra is
   (A) Join
   (B) Selection
   (C) Union
   (D) Cross product

2. In an E-R diagram double lines indicate
   (A) Total participation
   (B) Multiple participation
   (C) Cardinality N
   (D) None of the above

Polling

1. Fifth Normal form is concerned with
   (A) Functional dependency
   (B) Multivalued dependency
   (C) Join dependency
   (D) Domain-key

2. If two relations R and S are joined, then the non matching tuples of both R and S are ignored in
   (A) left outer join
   (B) right outer join
   (C) full outer join
   (D) inner join

Polling

1. Relations produced from an E-R model will always be
   (A) First normal form.
   (B) Second normal form
   (C) Third normal form
   (D) Fourth normal form

2. If both the functional dependencies : X→Y and Y→X hold for two attributes X and Y then the relationship between X and Y is
   (A) M:N
   (B) M:1
   (C) 1:1
   (D) 1:M