

UNIT-IV

SCHEMA REFINEMENT

- It is a technique of organizing the data in the Database.
- It refers to refine the schema by using some techniques and one of the best techniques is Decomposition.
- It is a systematic approach of decomposing tables to eliminate data redundancy and undesirable characteristics like Insertion, updation and Deletion Anomalies.
- Insertion Anomalies (problem in insertion)
- Deletion Anomalies (problem in deletion)
- Updation Anomalies (problem in updation)

Anomalies

- Anomalies refer to inconsistencies (or) errors that occur when manipulating (or) querying data in Database.
- These anomalies can lead to inconsistent or incorrect results and can negatively impact the overall functionality and accuracy of Database.

INSERTION ANOMALIES:

It may or Not to be possible to store some information unless some other information is stored as well.

Example

SID	SName	CID	CNAME	FEE
			C	5K
S1	A	C1		5K
S2	A	C1		10K
		C2		10K
S1	A		C	10K
S3	B	C2		15K
S3	B	C3	JAVA	
			D.B	12K
<u>S4</u>	<u>ABC</u>	C4		

- If New course 'C4' is introduced. but NO student enrolled in that particular course. then we have to enter some dummy data as if "SID" is considered as primary key. It could not be Null. So, the only option is to enter dummy data which may cause problem in future.

- Because of Insertion of some data, It is forced to Insert some other Dummy data.

DELETION ANOMALIES:

- It may not be possible to delete some information without losing some other information as well.
- Deletion of "S3" student cause the deletion of course "JAVA"
- Because of deletion of some data forced to delete some other useful data

SID	SName	CID	CNAME	FEE	
S1	A	C1	C	5K	7K updation
S2	A	C1	C	5K	"
S1	A	C2	C	10K	"
S1	A	C2	C	10K	"
S3	B	C2	C	10K	"
S3	B	C3	JAVA	15K	- Deletion
S3	B	C3	JAVA	15K	- Deletion

UPDATION Anomalies:

- If one copy of repeated data is updated, an inconsistency is created unless all copies are similarly updated.
- If there is updation in fee from 5K to 7K then we have to update fee column in all the rows, else data will be inconsistent.
- To update all the rows which reflects to more I/O cost.

DECOMPOSITIONS

- It is the solution to the problem caused by Data Redundancy.
- It Means Breaking up the Large Schema into smaller Multiple Schemas.
- It Removes all the Anomalies and also helps to maintain Data Integrity.

SID	SNAME	CID	CNAME	FEE
S1	A	C1	C	5K
S2	A	C1	C	10K
S1	A	C2	C	10K
S3	B	C2	C	15K
S3	B	C3	JAVA	

(primary key ^{CID} ~~CID~~)

Student (primary key SID+CID)

SID	SNAME	CID
S1	A	C1
S2	A	C1
S1	A	C2
S3	B	C2
S3	B	C3

COURSE

CID	CNAME	FEE
C1	C	5K
C1	C	5K
C2	C	10K
C2	C	10K
C3	JAVA	15K

Problem Related to Decompositions

- It produces Number of Tables with a relatively small no. of columns.
- These columns then have to be joined using their primary Foreign key Relationship.
- It has two major Disadvantages
- Performance: All the Joins Required to merge data slow processing and place additional stress on your Hardware.
- Complex query queries: Developer have to code complex queries in order to Merge data from different tables.

Reasoning About Functional Dependency in DBM

Functional Dependency is denoted as " $\rightarrow$ "

Eg: $A \rightarrow B$

A Determines B

(OR)

B Dependent on "A"

if $t_1.A = t_2.A$ then

$$t_1.B = t_2.B$$

is called $A \rightarrow B$

"B" is functionally dependent on "A"

A	B
10	1
20	2
30	3
10	1

Functional Dependency

A functional dependency exists when the value of one (or) more attributes (the determinant) in a Relation (table) uniquely determines the values of another attribute (or) set of attributes (the dependent)

Types of Functional Dependencies:

Trivial Functional Dependency:

→ A Trivial dependency occurs when the dependent attribute is subset of Determinant attribute

$$\text{Eg: } A \rightarrow A$$

$$\text{Eg: } S.No, S.Name \rightarrow S.Name$$

Non-Trivial Functional Dependency:

A non-Trivial dependency occurs when the dependent attribute is not a subset of Determinant attribute

$$\text{Eg: } A \rightarrow B \text{ where "B" is not a part of "A"}$$

$$S.No \rightarrow S.Name$$

Partial Functional Dependency:

A Partial Dependency exists when a non-key attribute is dependant on only a part of the Primary Key (in a Composite key).

$$S.No \rightarrow \text{course}$$

Transitive Functional Dependency

A Transitive Dependency occurs when Non-Key attributes is indirectly dependent on primary key through another Non-key Attribute

$$\text{Eg: if } A \rightarrow B, B \rightarrow C, \text{ then}$$

$$A \rightarrow C$$

$$\text{If } R.No \rightarrow \text{course, course} \rightarrow \text{Fee}$$

$$\text{Then } R.No \rightarrow \text{Fee}$$

Data Redundancy :-

(7)

→ Redundancy Means having multiple copies of same data in the Database.

→ When Database is Normalized then we erase Redundancy

S.ID	S.Name	course	College	ph.No	Rank

Type's of Redundancy :-

1. Insertion Anomaly
2. Deletion Anomaly
3. Updation Anomaly

1. Insertion Anomaly :-

Inserting the values without following a specific

Rule College		course	college	Ph.No.	Rank
S.ID	S.Name				
108	Sathish	—	B SC ITS	98xxxx	10000

Entering the Table values without Entering the Course Details.

2. Deletion Anomaly :-

When ever we Delete Data from one table, it is also affecting to other table (another table data also Delete)
It is called Deletion Anomaly. By Applying

3. Updation Anomaly :-

When ever we update Details in one table, It is also get effect to other table is called Updation Anomaly.

college			Rank	
S.ID	S.Name	Course	College	Ph.No
108	Sathish	CSE	SCITS	98xxx
				10000

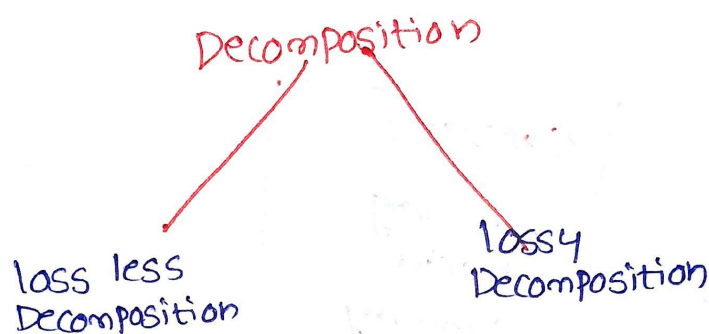
whenever we update Rank Details in college table, That will effect another tables having college as field

→ Data Inconsistency :- Condition that occurs between tables when similar data is kept in different formats in 2 different tables.

→ By Applying The Constraints we can chage Data Inconsistency to Data Consistency.

Decomposition

- It Breaks One Table into multiple Tables.
 we are using Decomposition in Relational RDBMS
 To Eliminate some of problems anomalies, inconsistency
 Redundancy
 In RDBMS is not in appropriate normal forms
 then the Decomposition is Required.



1. loss less Decomposition

In this Decomposition Guarantee that the Join of Relation will Result in the Same Relation it was Decomposed.

Eg+ Let R is a Relational Schema

$R \rightarrow R_1, R_2, R_3, \dots, R_n$

R				
stdId	std.Name	std.Age	courseId	course Name
101	X	16	100	EEE
102	Y	16	100	EEE
103	Z	17	100	EEE

R ₁			R ₂		
stdId	std.Name	std.Age	stdId	courseId	course Name

$P \cap Q \rightarrow P$ (Functional dependent on "P")
 $P \cap Q \rightarrow Q$ (Functional dependent on "Q")

Lossy Decomposition :-

⑩

- Decomposition is Dividing one table to multiple tables
- When Combining two tables we can't form the original table
- $R \rightarrow$ Relational Set
 $R_1, R_2, \dots, R_n$
 $R_1 \cup R_2 \neq R$

R

Mobile	price	category
Y12	5000	Samsung
X3	4000	Nokia
Y21	8000	Samsung

R_1

mobiles	category
Y12	Samsung
X3	Nokia
Y21	Samsung

R_2

price	category
5000	Samsung
4000	Nokia
8000	Samsung

$R_1 \cup R_2$

Mobiles	price	category
Y12	5000	Samsung
Y212	8000	Samsung
X3	400	Nokia
Y21	5000	Samsung
Y21	8000	Samsung

$$R_1 \cup R_2 \neq R$$

There will be some
Data Change (or)
Data loss

Applying 2NF to Table :-

"Normalization" :-

- Normalization is a process of evaluating and correcting table structure to minimize data redundancy, thereby reducing data anomalies.
- Normalization works through the series of stages called Normal Forms.

Normal Form :-

Normal form is defined as a set of rules that are formed in order to convert a relation into standard form.

- It involves decomposition of table with anomalies into smaller well structural relation.

- 1NF
- 2NF
- 3NF
- BCNF
- 4NF
- 5NF

1NF :-

Example Table

R.NO	Name	Lang. known	Group	Fee
1	Ramu	Telugu English	E.E.F	20k
2	Sitha	Telugu Hindi	E.C.E	30k
3	Laxman	English	C.S.E	50k

1NF:

- A Relation is 1NF if it contains no Repeated Groups
- It must have a single value.

Characteristics of 1NF:

- It Removes Repeated groups
- It identifies primary key

Applying 1NF to table

std

R.NO	Name	Language	Group	fee
1	Ramu	Telugu	E.E.E	20k
1	Ramu	English	E.E.E	20k
2	Sitha	Telugu	E.C.E	30k
2	Sitha	Hindi	E.C.E	30k
3	Laxman	English	C.S.E	50k

↓
primary key
R.NO.

2. NF:

A Relation is 2NF if and only if it contains

1NF
NO Partial Dependency.

Characteristics:

1. It satisfies 1NF
2. NO Partial Dependency (Partial Dependency is non-key attributes)

Dependency Between key and

Applying 2NF to Table :-

(13)

RNO	Name	Group	std	fee
1	Ramu	EEE		20k
2	Sitha	ECE		30k
3	Laxman	CSE		50k

R.NO	Lang. known
1	Telugu
1	English
2	Telugu
2	Hindi
3	English

3NF :-

Characteristics

- It Satisfies 2NF
- NO Transitive dependency
(Transitive Dependency Means

Dependency Between Non Key Attributes)

Applying 3NF to Table :-

R.NO	Group Name	Group
1	Ramu	EEE
2	Sitha	ECE
3	Laxman	CSE

Group	Fee
EEE	20k
ECE	30k
CSE	50k

R.NO	Group	Language known
1		Telugu
1		English
2		Telugu
2		Hindi
3		English

BCNF (Boyce-Codd Normal Form)

→ It is a special case of 3NF
 → There is no transitive dependency, Each attribute must be candidate key

Section	Subject	Faculty	Time
A	Java	Shiva	8:00
A	DBMS	Sraavan	9:00
C	Python	Ramesh	10:00
D	.Net	Manesh	11:00

Section + subject ← P. key
 faculty

From above table faculty is depending on subject, so we remove functional dependency on table.

Applying BCNF to the table

Section	Faculty Subject	Time
A	Java	8:00
A	D.B.M.S	9:00
C	Python	10:00
D	.Net	11:00

Section	Subject	Subject Faculty
	Java	Shiva
	D.B.M.S	Sraavan
	Python	Ramesh
	.Net	Manesh

4NF:-

Q15

→ It satisfies BCNF

→ No multivalued dependency. → [one Attribute is depending upon more than multiple Attributes]

Course	Faculty	Textbook
C	Ramu	Sure
	Raghu	Rahul
C++	Ravi	Udaya
	Kishore	Shakshi

Course	Faculty	TextBook
C	Ramu	SURE
	Raghu	Rahul
C	Ramu	Rahul
	Raghu	SURE
C++	Ravi	Udaya
	Kishore	Shakshi
C++	Ravi	Shakshi
	Kishore	Udaya

Applying 4NF to Faculty Table ÷

We need to Remove multivalued Dependency, we split the table into single value Dependency.

Course	Faculty
C	Ramu
	Raghu
C++	Ravi
	Kishore

Course Faculty	Text book
Ramu C	Sure
Raghu C	Rahul
Ravi C++	Shakshi
Kishore C++	Udaya

De.
help in

5NF

- It should already in 4NF
- It is also known as PJNF [Project Join Normal form]
- It should not have Join Dependency [i.e, It cannot be further non-loss decomposed]

Agent	Company	product
A1	H.P.C.L	Fuel
A1	B.P.C.L	Fuel
A1	Costrol	Engmedil
A2	RIDE	E.oil
A2	C.N.G	Fuel

Company Agent

Agent	Company
A1	H.P.C.L
A1	B.P.C.L
A1	H.P.C.L
A2	RIDE
A2	C.N.G

product

AGENT	product
A1	Fuel
A1	E.oil
A2	E.oil
A2	Fuel

Company

product	Company
Fuel	A H.P.C.L
Fuel	B.P.C.L
A2	C.N.G
Fuel	Costrol
E.oil	RIDE
E.oil	

Dependency preserving Decomposition

In DBMS, Dependency preserving Decomposition refers to the process of Breaking Down a complex database schema into simpler, smaller tables, such that all the Functional Dependencies of original schema are still enforceable without needing to perform additional Joins.

Key Concepts of Dependency preserving Decomposition

The key concept of dependency-preserving include

- Functional Dependency Preservations

It means that after De-composition the functional dependencies in the original schema must still hold true in the decomposed schema

Loss Less Join Property:

The Decomposition must allow for the original Relation to be re-constructed from the decomposed Relations without any Data loss, Ensuring no information is declared

Normalization:

The Decomposition often aims to Normalize the schema to higher Normal Forms (Like 3NF or BCNF), which further eliminates Redundancy and Dependency anomalies

Minimal Redundancy:

By Ensuring the Decomposition preserves functional dependencies, it minimizes data Redundancy and help in avoiding Data Anomalies.

Multi Valued Dependency

(18)

→ This Dependency occurs when there are more than one independent multivalued attributes in a table.

→ If a table has attributes $P, Q, \& R$ then $Q \& R$ multivalued facts of P . It is represented

→ →

Ex:-

A bike manuf company, which produce two colour (Black & RED) in each Model

Black, RED → ~~Model~~

bike-Model	Manf. Year	Colour
R1001	2008	Black
R1001	2008	RED
R2012	2009	Black
R2012	2009	RED

bike-Model → → Manf. Year

bike Model → → Colour