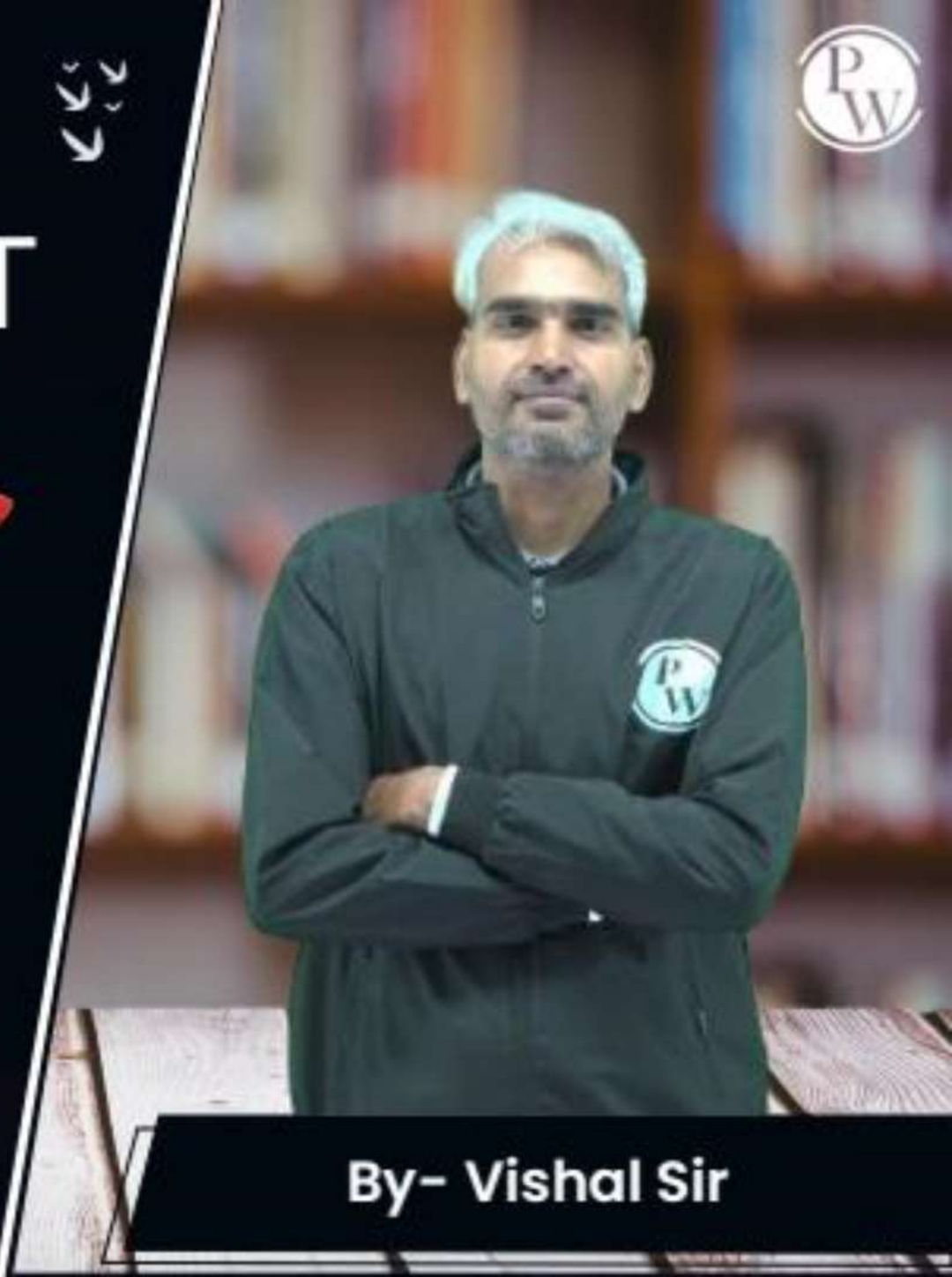
Computer Science & IT

Database Management System

Relational Model & Normal Forms

Lecture No. 09



By- Vishal Sir

Recap of Previous Lecture

Topic FD set of a sub-relation

Topic Minimal cover or Canonical cover

Slide

Topics to be Covered

- **Topic** Schema refinement (Normalization)
- **Topic** Problems because of redundancy in a relation
- **Topic** Properties of decomposition

Slide

#e.g., Consider the following FD set

$F = \{A \to BC$
$CD \to E$
$E \to C$
$D \to AEH$
$ABH \to BD$
$DH \to BC$
$\}$

Find minimal cover of F.

F' = $\{A \to BC, CD \to E, E \to C, D \to AEH, ABH \to BD, DH \to BC\}$

$AB \to C$ — trivial FD

Step ①

$A \to B$
$A \to C$
$CD \to E$
$E \to C$
$D \to A$
$D \to E$
$D \to H$
$ABH \to B$
$ABH \to D$
$DH \to B$
$DH \to C$

Step-2

$A \to B$
$A \to C$
$(C)^+ = \{C\}$ $(D)^+ = \{D, C, \ldots\}$ — C is extraneous
$D \to E$
$E \to C$
$D \to A$
$D \to E$
$D \to H$
$(A)^+ = \{A, B, C\}$
$AH \to D$
$(D)^+ = \{D, H, \ldots\}$
$D \to B$ (Redundant)
$D \to C$ (Redundant)

Step-3

$A \to B$
$A \to C$
$E \to C$
$D \to A$
$D \to E$
$D \to H$
$AH \to D$

Step-4

$A \to BC$
$E \to C$
$D \to AEH$
$AH \to D$

Minimal Cover of F.

Topic : NOTE

- Minimal cover of FD set F need not be unique, but all minimal cover are logically equivalent.

#e.g., Consider the FD set

$F = \{AB \rightarrow C, B \rightarrow A, A \rightarrow B\}$

Find the number of Minimal Covers of F.

$$F = \begin{cases} AB \rightarrow C \\ B \rightarrow A \\ A \rightarrow B \end{cases}$$

None of this two FDs can be removed

① $AB \rightarrow C$

∵ we have $A \rightarrow B$

∴ B is extraneous if A is present

Hence

Minimal Cover $Fm_1 = \{A \rightarrow C, B \rightarrow A, A \rightarrow B\}$

② $AB \rightarrow C$

∵ we have $B \rightarrow A$

∴ A is extraneous if B is present.

Hence,

Minimal Cover $Fm_2 = \{B \rightarrow C, B \rightarrow A, A \rightarrow B\}$

Topic : Schema refinement (Normalization)

duplicated information

Normalization is a process of reducing / eliminating the redundancy present in the relational table

Topic : Redundancy in a relation

If independent informations are stored in a single relational table then redundancy is Possible.

information related to students in a College

↓

w.r.t. Student information

Sid → Sname will hold.

it is not redundancy, it is important information.

Enrollment

Sid	Sname	Cid	Cname
S_1	A	C_1	OS
S_1	A	C_2	DBMS
S_2	A	C_2	DBMS
S_3	B	C_2	DBMS
S_3	B	C_3	OS

It is redundancy

information abbout Courses in a College

↓

Let Cid → Cname holds w.r.t. Course info

Consider, the FD set F w.r.t. relation Enrollment is, $F = \{ Sid \rightarrow Sname,\ Cid \rightarrow Cname \}$

∴ Candidate key of Enrollment table = (Sid, Cid)

Topic : Problems because of redundancy

- If redundancy is present in the database, then various problems are possible.

① It requires more storage space { Not a big problem }

Serious Problem

② Insertion Anomaly

③ Deletion Anomaly

④ Updation Anomaly

Insertion Anomaly :-

i.e. (Sid, Cid) is Primary Key.

Enrollment

Sid	Sname	Cid	Cname
S_1	A	C_1	OS
S_1	A	C_2	DBMS
S_2	A	C_2	DBMS
S_3	B	C_2	DBMS
S_3	B	C_3	OS
NULL	NULL	C_5	AI

Can not be NULL because it is a Part of Primary key.

{ ∴ (C_5, AI) can not be inserted }

Consider, the FD set F w.r.t. relation Enrollment is, $F = \{ Sid \rightarrow Sname,\ Cid \rightarrow Cname \}$

∴ Candidate key of Enrollment table = (Sid, Cid)

Assume (Sid Cid) is also the Primary Key of Relation

* **Insertion Anomaly** :- If independent informations are stored in the same table, then some times it may not be possible to insert one information without inserting other independent informations.

eg: In the above table we can not insert the information of a new course (C5, AI) until some students enroll for that course, because 'Sid' is a part of Primary key and it can not be set to NULL.

Deletion Anomaly

i.e. (Sid, Cid) is primary key.

Enrollment

Sid	Sname	Cid	Cname
S_1 NULL	A	C_1	OS
S_1 NULL	A	C_2	DBMS
S_2	A	C_2	DBMS
S_3	B	C_2	DBMS
S_3	B	C_3	OS

Not allowed

∴ delete complete tuple with Sid = S_1

Consider, the FD set F w.r.t. relation Enrollment is, $F = \{ Sid \rightarrow Sname,\ Cid \rightarrow Cname \}$

∴ Candidate key of Enrollment table = (Sid, Cid)

Assume (Sid Cid) is also the Primary Key of Relation

Deletion Anomaly :. Some times it may not be possible to delete one independent information without deleting other associated information.

"It may result in loss of data".

From the above table if we try to delete the information of student with "Sid = S_1", then we will have to delete the complete tuples associated with 'Sid = S_1' (because Sid can not be NULL)

As a result we will loose the information of Course "C_1" as well.

Information about Course "C_2" is still available w.r.t. Sids S_2 & S_3

Updation Anomaly :- If redundancy is present, then updation is required in all the duplicate copies, it may be a time consuming operation.

Topic : Schema refinement (Normalization)

Normalization is a process of decomposing (splitting) a relational tables into smaller tables (sub-relations) such that it eliminates/reduces the data redundancy, and it can overcome undesirable characteristics like Insertion, Updation and Deletion Anomalies

Eg: in the next slid.

i.e. (Sid, Cid) is Primary Key.

Enrollment

Sid	Sname	Cid	Cname
S_1	A	C_1	OS
S_1	A	C_2	DBMS
S_2	A	C_2	DBMS
S_3	B	C_2	DBMS
S_3	B	C_3	OS

Functional dependency set $F = \{ Sid \rightarrow Sname, Cid \rightarrow Cname \}$

Assume that the above relation is decomposed into following sub-relations.

Student

Sid	Sname
S_1	A
S_2	A
S_3	B

$F_1 = \{Sid \rightarrow Sname\}$

Course

Cid	Cname
C_1	OS
C_2	DBMS
C_3	OS

$F_2 = \{Cid \rightarrow Cname\}$

Enrollment

Sid	Cid
S_1	C_1
S_1	C_2
S_2	C_2
S_3	C_2
S_3	C_3

$F_3 = \{$ Empty $\}$

After decomposition all anomalies are overcome.

2 mins Summary

Topic Schema refinement (Normalization)

Topic Problems because of redundancy in a relation

Topic Properties of decomposition

THANK - YOU