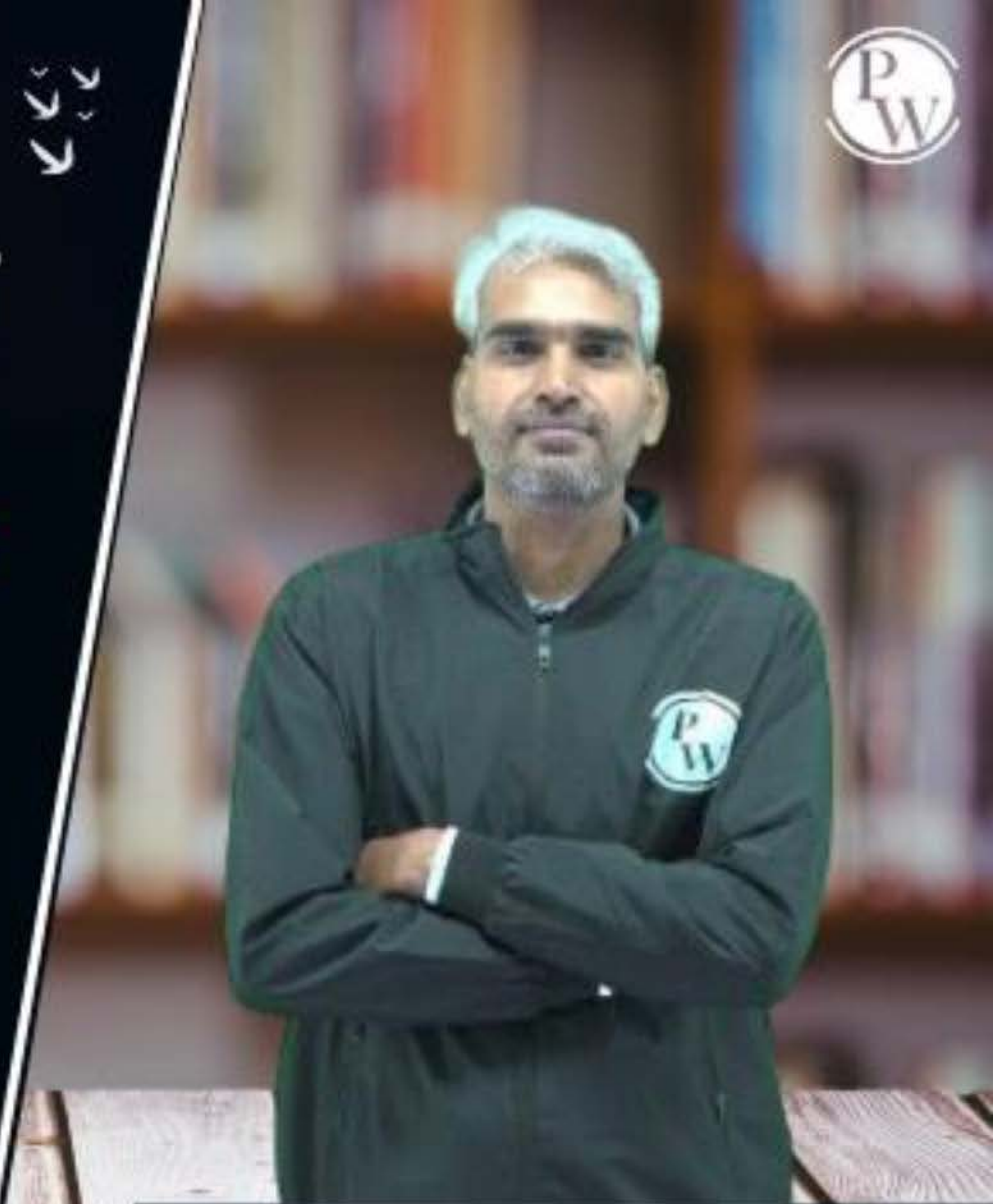
Computer Science & IT

Database Management System

Relational Model & Normal Forms

Lecture No. 07



By- Vishal Sir

Recap of Previous Lecture

Topic FD set of a subrelation

Topic Minimal cover (Canonical cover)

Topic Number of superkeys in a relation

Topics to be Covered

Topic Number of superkeys in a relation

Topic Normalization (Schema refinement)

How. ~~#e.g.,~~ Consider the following FD set

$F = \{A \rightarrow BC$

$CD \rightarrow E$

$E \rightarrow C$

$D \rightarrow AEH$

$ABH \rightarrow BD$

$DH \rightarrow BC$

$\}$

Find minimal cover of F.

	Simplify RHS.	Eliminate Extra attribute from LHS.	Eliminate Redundant FD	Union
$A \rightarrow BC$				$A \rightarrow BC$
$CD \rightarrow E$				$D \rightarrow AEH$
$E \rightarrow C$				$E \rightarrow C$
$D \rightarrow AEH$				$AH \rightarrow D$
$ABH \rightarrow BD$				
$DH \rightarrow BC$				

(For Complete process Please go through video)

Number of Super keys

Q: Let R (A, B, C, D, E) is a relation with no non-trivial functional dependency {i.e. F = { } }
Empty

then what will be the Candidate Key of relⁿ R.

Solu'n No FD in the FD set,

Hence all attributes are essential attributes

∴ C.K will be formed by combining all attributes of the relation.

* In this case, we have only one Candidate Key and Only One Super Key (it C.K itself)

H.W: Q:- Consider the relational schema $R(A_1, A_2, A_3, \ldots, A_n)$

Find the number of superkeys possible in relation R.

(i) When "A_1" is the only candidate key of relation R.

Any superset of A_1 is a super key.

Attributes: $A_1 \quad A_2 \quad A_3 \quad A_4 \quad A_5 \ldots \ldots A_n$

$$\text{Number of Supersets of } A_1 = 1 \times \underbrace{2 \times 2 \times 2 \times 2 \times \cdots \times 2}_{(n-1) \text{ times } '2'} = 2^{n-1}$$

Only '1' choice it must be present

Super Keys

H.W.

Q:- Consider the relational schema $R(A_1, A_2, A_3, \ldots, A_n)$

Find the number of superkeys possible in relation R.

(ii) When (A_1A_2) is the only candidate key of relation R.

Any super set of C.K $\{A_1, A_2\}$ is a Super key

Attributes:- $A_1 \quad A_2 \quad A_3 \quad A_4 \quad A_5 \ldots\ldots A_n$

Number of Super sets of $\{A_1, A_2\}$

$$\underbrace{1 \times 1}_{\text{Both must be present}} \times \underbrace{2 \times 2 \times 2 \times \cdots \times 2}_{(n-2) \text{ times } '2'} = 2^{n-2}$$

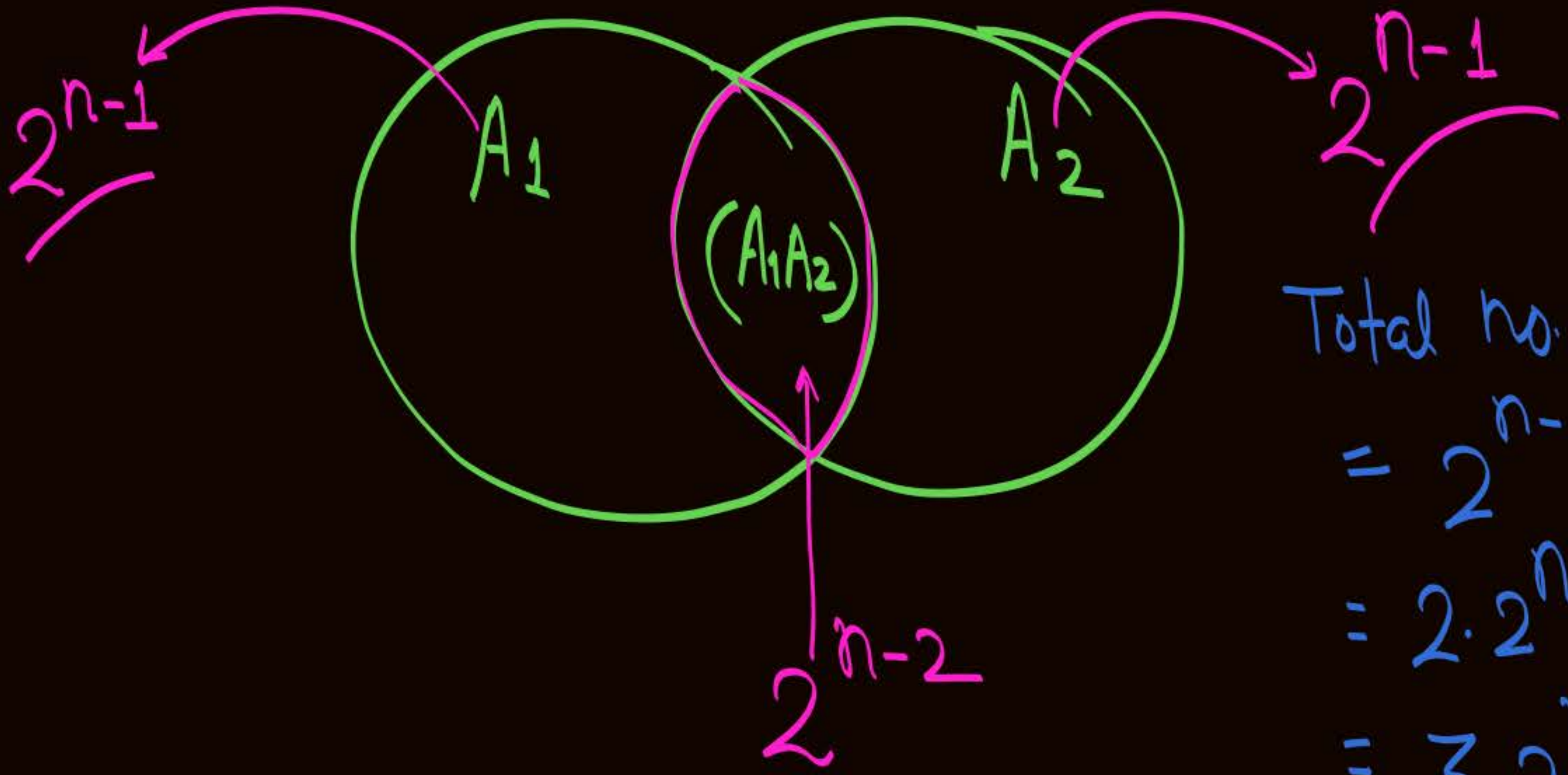
Super keys

Q:- Consider the relational schema $R(A_1, A_2, A_3, \ldots\ldots, A_n)$

H.W

Find the number of superkeys possible in relation R.

(iii) When "A_1" & "A_2" are the only two candidate keys of relation R.



Total no. of Super keys

$$= 2^{n-1} + 2^{n-1} - 2^{n-2}$$

$$= 2 \cdot 2^{n-2} + 2 \cdot 2^{n-2} - 1 \cdot 2^{n-2}$$

$$= 3 \cdot 2^{n-2}$$

Q:- Consider the relational schema $R(A_1, A_2, A_3, \ldots\ldots, A_n)$

H.W

Find the number of superkeys possible in relation R.

(iii) When "A_1" & "A_2" are the only two candidate keys of relation R.

Attributes: $A_1 \quad A_2 \quad A_3 \quad A_4 \quad A_5 \ldots\ldots A_n$

$$1 \times 1 \times 2 \times 2 \times 2 \times \ldots\ldots \times 2 = 2^{n-2}$$

(or)

$$1 \times 1 \times 2 \times 2 \times 2 \times \ldots\ldots \times 2 = + 2^{n-2}$$

(or)

$$1 \times 1 \times 2 \times 2 \times 2 \times \ldots\ldots \times 2 = + 2^{n-2}$$

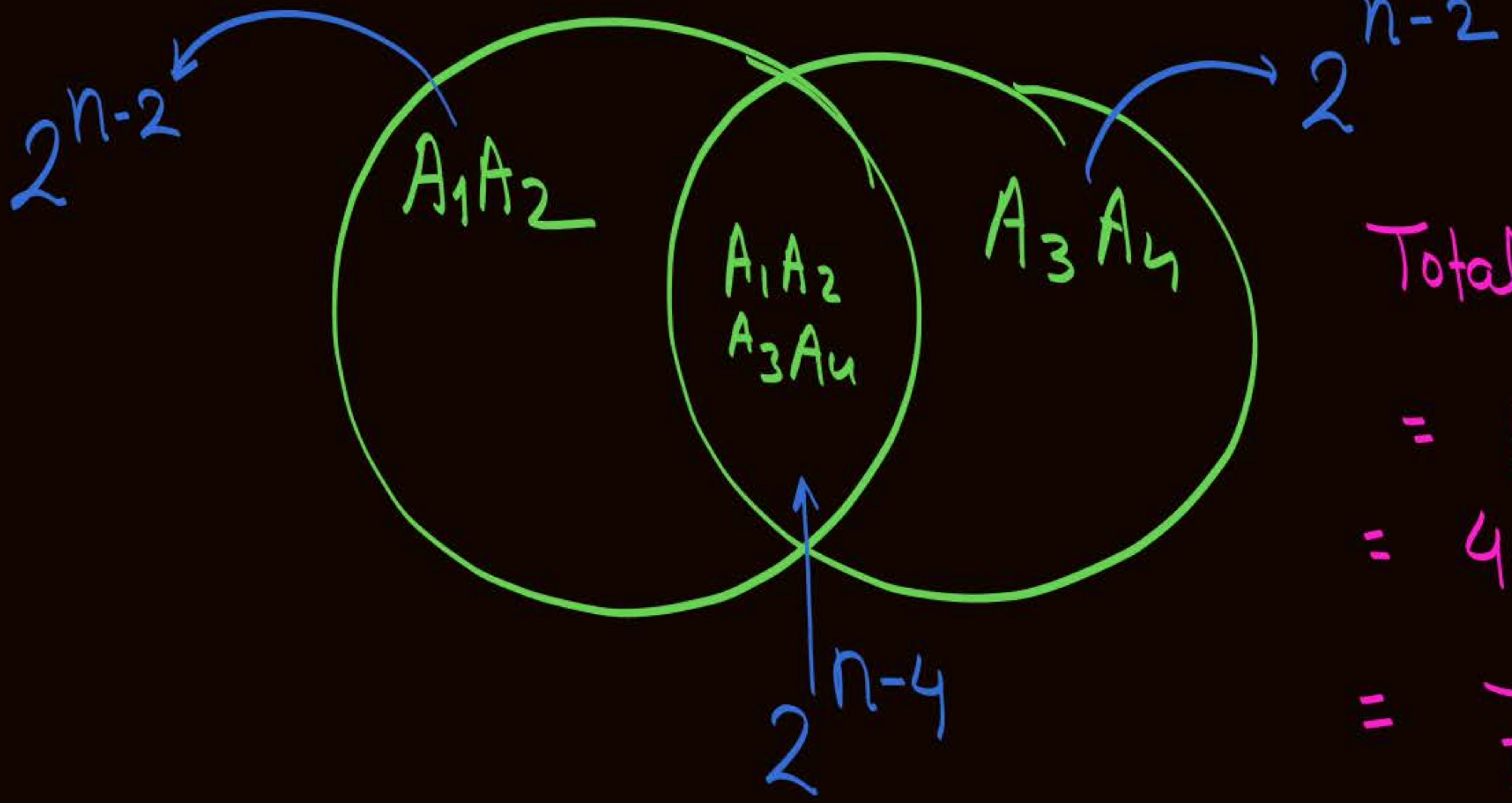
S.K = (All attributes of at least one C.K) + (0 or more attributes out of remaining)

Total no. of S.Ks $= 3 \cdot 2^{n-2}$

Q:- Consider the relational schema $R(A_1, A_2, A_3, \ldots\ldots, A_n)$

H.W. Find the number of superkeys possible in relation R.

(iv) When (A_1A_2) and (A_3A_4) are two candidate keys of relation R.



Total no. of S.K =

$$= 2^{n-2} + 2^{n-2} - 2^{n-4}$$

$$= 4\cdot 2^{n-4} + 4\cdot 2^{n-4} - 1\cdot 2^{n-4}$$

$$= 7\cdot 2^{n-4}$$

Q:- Consider the relational schema $R(A_1, A_2, A_3, \ldots\ldots, A_n)$

HW. Find the number of superkeys possible in relation R.

(iv) When (A_1A_2) and (A_3A_4) are two candidate keys of relation R.

Attributes: $A_1 \quad A_2 \quad A_3 \quad A_4 \quad (A_5 \quad A_6 \ldots\ldots A_n)$

$\checkmark 1 \times \checkmark 1 \times \times 1 \times \times 1 \times 2 \times 2 \times \ldots\ldots 2 = 2^{n-4}$

$\checkmark \quad \checkmark \quad \checkmark \quad \times$ ——— $= 2^{n-4}$

$\checkmark \quad \checkmark \quad \times \quad \checkmark$ ——— $= 2^{n-4}$

$\checkmark \quad \checkmark \quad \checkmark \quad \checkmark$ ——— $= 2^{n-4}$

$\times \quad \times \quad \checkmark \quad \checkmark$ ——— $= 2^{n-4}$

$\checkmark \quad \times \quad \checkmark \quad \checkmark$ ——— $= 2^{n-4}$

$\times \quad \checkmark \quad \checkmark \quad \checkmark$ ——— $= 2^{n-4}$

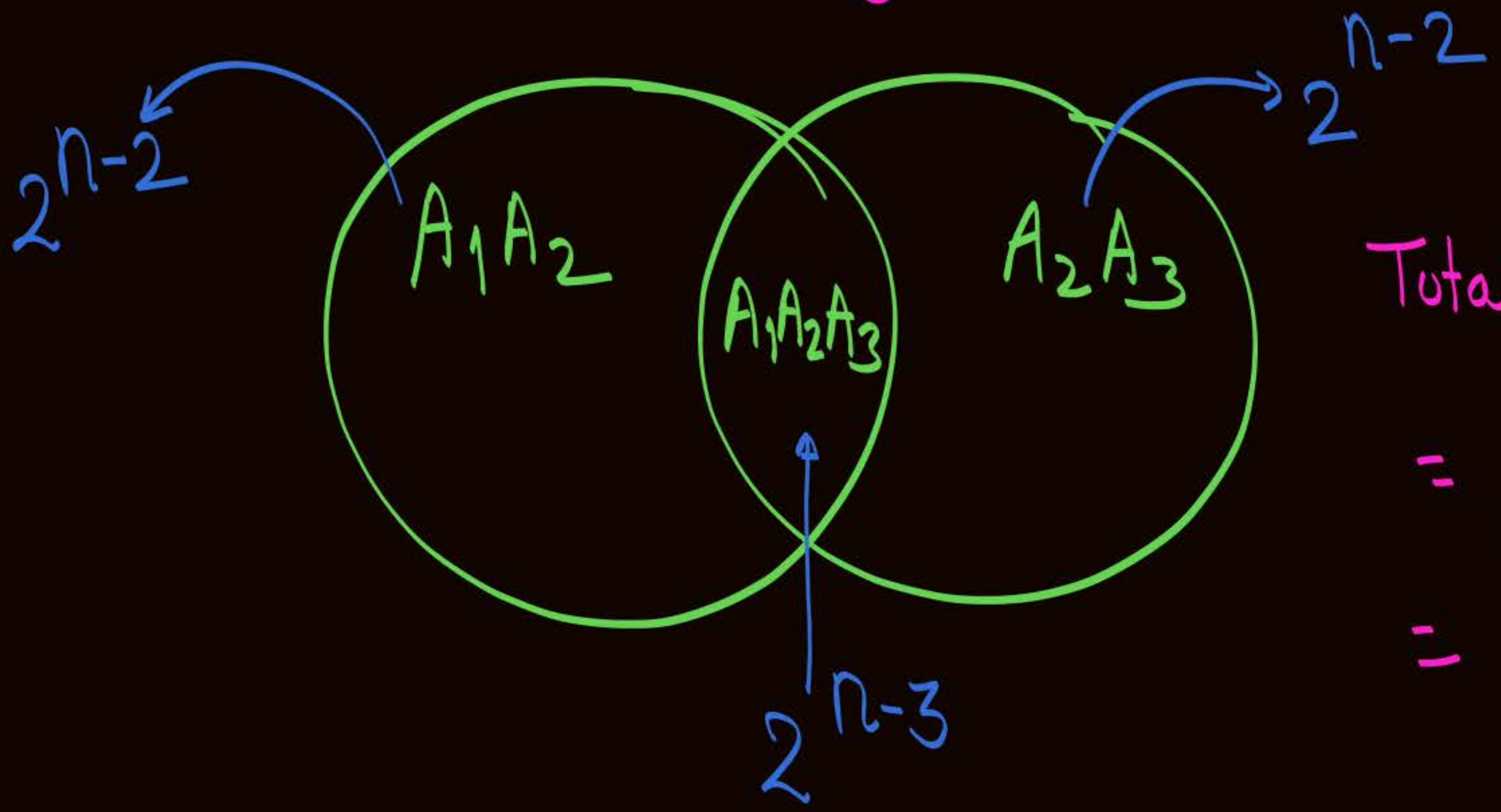
Total no. of Super Keys $= 7 \cdot 2^{n-4}$

Q:- Consider the relational schema $R(A_1, A_2, A_3, \ldots\ldots, A_n)$

H.W.

Find the number of superkeys possible in relation R.

(v) When (A_1A_2) & (A_2A_3) are only two candidate keys of relation R.



Total no. of SKs

$$= 2^{n-2} + 2^{n-2} - 2^{n-3}$$

$$= 3 \cdot 2^{n-3}$$

Q:- Consider the relational schema $R(A_1, A_2, A_3, \ldots\ldots, A_n)$

H.W.

Find the number of superkeys possible in relation R.

(V) When (A_1A_2) & (A_2A_3) are only two candidate keys of relation R.

Attributes:	A_1	A_2	A_3	A_4	A_5	--- A_n	
	✓	✓	×	ϕ	ϕ	ϕ	$= 2^{n-3}$
	×	✓	✓	ϕ	ϕ	ϕ	$= 2^{n-3}$
	✓	✓	✓	ϕ	ϕ	ϕ	$= 2^{n-3}$

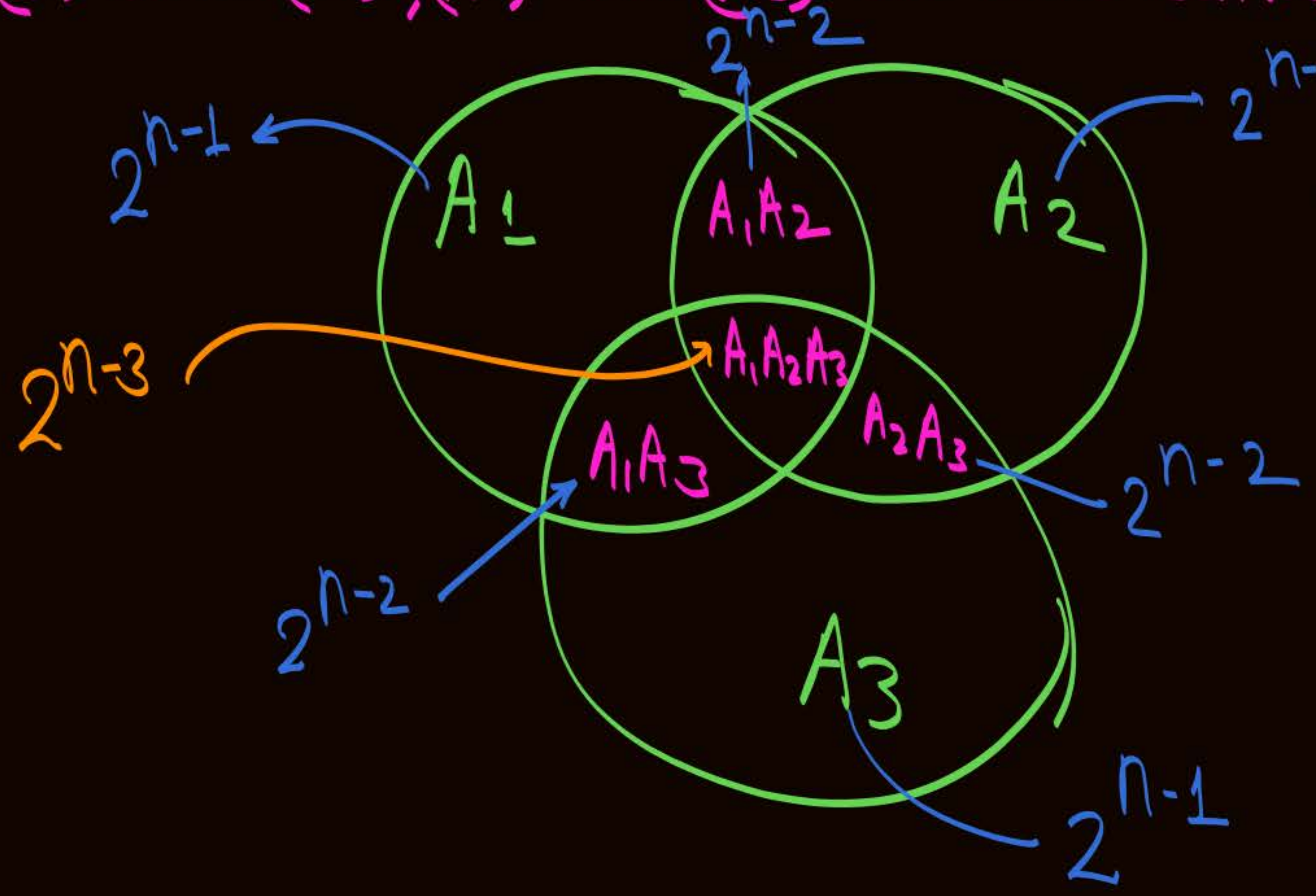
$3 \cdot 2^{n-3}$

H.W.

Q:- Consider the relational schema $R(A_1, A_2, A_3, \ldots, A_n)$

Find the number of superkeys possible in relation R.

(vi) When (A_1), (A_2) and (A_3) are three candidate keys of relation R.



Total no. of Super keys

$$= \left(2^{n-1} + 2^{n-1} + 2^{n-1} - 2^{n-2} - 2^{n-2} - 2^{n-2} + 2^{n-3}\right) = 7 \cdot 2^{n-3}$$

Total no. of subsets

No. of subsets in which none of A_1, A_2 or A_3 are present

$$= 2^n - 2^{n-3}$$

$$= 8 \cdot 2^{n-3} - 2^{n-3}$$

$$= 7 \cdot 2^{n-3}$$

$$n(A \cup B \cup C) = n(A) + n(B) + n(C) - n(A \cap B) - n(A \cap C) - n(B \cap C) + n(A \cap B \cap C)$$

H.W.

Q:- Consider the relational schema $R(A_1, A_2, A_3, \ldots, A_n)$

Find the number of superkeys possible in relation R.

When each attribute of relation R itself is a Candidate Key

$$\underline{Ans} = 2^n - 1$$

Empty Subset

Normalization

Topic : Schema refinement (Normalization)

duplicated data

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

Student information

Student_Enrollment_Info

Course information

$Sid \rightarrow Sname$

$Cid \rightarrow Cname$

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

Redundancy

it is information
(not redundancy)

Topic : Redundancy in a relation

If independent informations are stored in the same table, then redundancy is possible

Student_Enrollment_Info

Student information

Course information

$Sid \rightarrow Sname$

$Cid \rightarrow Cname$

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

it is not redundancy {it is information}

it is redundancy

Topic : Problems because of redundancy

* If redundancy is present in the relation, then various problems are possible

* (1) It requires more storage space { it is not a very big Problem }

Imp
(2) Insertion Anomaly
(3) Deletion Anomaly
(4) Updation Anomaly

Student_Enrollment_Info

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

Let,

Sid → Sname & Cid → Cname are only two FDs that holds in the relation.

∴ Candidate Key of the relation is {Sid, Cid}

If we define a primary Key for the relation then {Sid, Cid} will be Primary Key

If {Sid, Cid} is the Primary Key then Sid & Cid are not allowed to take "NULL" Values

Insertion Anomaly:-

Student_Enrollment_Info

Attributes of primary key

Sid	Sname	Cid	Cname
S_1	A	C_1	DM
S_1	A	C_2	DBMS
S_2	A	C_2	DBMS
S_3	B	C_2	DBMS
S_3	B	C_3	DM
NULL	NULL	C_4	AI

$Sid \rightarrow Sname$

$Cid \rightarrow Cname$

Try to insert the information of a new course (C_4, AI)

But it is not allowed because Sid is the attribute of Primary Key.

If no student has Enrolled for this course, then we must set Sid as NULL

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

eg: We can not insert the information of a new Course (C_4, AI) until some students enroll for that Course, because "Sid" Can not be NULL.

Deletion Anomaly :-

Suppose we want to delete the information of student with "Sid = S_1"

Student_Enrollment_Info

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

Sid → Sname

Cid → Cname

But it is not allowed because Sid can not be NULL

∴ We will have to delete the complete tuple w.r.t. Sid = S_1

Deletion Anomaly :-

Suppose we want to delete the information of student with "Sid = S_1"

→ delete tuples with Sid = S_1

Student_Enrollment_Info

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

Sid → Sname

Cid → Cname

If we delete all the tuples w.r.t. Sid = S_1, then we lose the information of course with "Cid = C_1"

Deletion Anomaly :- Some times when we try to delete one independent information, we may lose some other independent information.

eg: If we try to delete the information of student "S_1" then information of Course with $Cid = C_1$ is lost.

Updation Anomaly:-

Consider a situation where we want to update the name of Course with "Cid = C_1" to "DB&WH" from "DBMS"

Student_Enrollment_Info

$Sid \rightarrow Sname$

$Cid \rightarrow Cname$

Sid	Sname	Cid	Cname
S_1	A	C_1	DM
S_1	A	C_2	~~DBMS~~ DB&WH
S_2	A	C_2	~~DBMS~~ DB&WH
S_3	B	C_2	~~DBMS~~ DB&WH
S_3	B	C_3	DM

Updation will be required in all duplicate Copies $\Longrightarrow$ ∵ it will be a time Consuming operation

Updation Anomaly :-

Updation is required in all duplicate Copies, ∴ it will be 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

Student_Enrollment_Info

$F = \left\{ \begin{array}{l} Sid \rightarrow Sname \\ Cid \rightarrow Cname \end{array} \right\}$

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

Assume that the relation is decomposed into following three sub-relations

Student

Sid	Sname
S_1	A
S_2	A
S_3	B

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

$Ck = Sid$

Course

Cid	Cname
C_1	DM
C_2	DBMS
C_3	DM
C_4	AI

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

$Ck = Cid$

Enroll

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

$F_3 = \{ \quad \}$

$Ck = (Sid, Cid)$

After decomposition there is no redundancy

&

all the anomalies are overcome

2 mins Summary

Topic Number of superkeys in a relation

Topic Normalization (Schema refinement)

Slide

THANK - YOU

Slide