SCHEMA REFINEMENT and NORMALIZATION

Eliminate or reduce redundancy

Redundancy means duplicate copies of the same data

A relation with redundancy can be redefined by decomposing it, or replacing it with smaller relations that contain the same information, but without redundancy.

Sid	Sname	Cid	Cname	Fid	Fname	Sal
S1	A	C1	C	F1	X	5K
S2	A	C1	C	F1	X	5K
S3	B	C1	C	F1	X	5K
S4	B	C2	C++	F2	Y	10K
S5	B	C1	C	F1	X	5K

Cid → Cid Cname Fid Fname Sal

Sid - Student ID, Sname - Student Name, Cid - Course ID, Fid - Faculty ID, Fname - Faculty Name, Sal- Salary

Problems Caused by Redundancy:

Storing the same information redundantly, that is, in more than one place within a database, can lead to several problems:

- **Redundant storage:** Some information is stored repeatedly.
- **Update anomalies:** If one copy of such repeated data is updated, an inconsistency is created unless all copies are similarly updated.
- **Insertion anomalies:** It may not be possible to store some information unless some other information is stored as well.
- **Deletion anomalies:** It may not be possible to delete some information without losing some other information as well.

F_1 ~~5K~~ to 7K

Sid	Sname	Cid	Cname	Fid	Fname	Sal
S1	A	C1	C	F1	X	5K
S2	A	C1	C	F1	X	5K
S3	B	C1	C	F1	X	5K
S4	B	C2	C++	F2	Y	10K
S5	B	C1	C	F1	X	5K
d1	d2					

C3 Java F3 Y 15K

C3 Java F3 Y 15K

Decomposition

A relation with redundancy can be redefined by decomposing it, or replacing it with smaller relations that contain the same information, but without redundancy.

Splitting Relation R into two or more sub relations

A decomposition of a relation schema R consists of replacing the relation schema by two (or more) relation schemas that each contain a subset of the attributes of R and together include all attributes in R.

R ✓

Sid	Sname	Cid	Cname	Fid	Fname	Sal
S1	A	C1	C	F1	X	5K
S2	A	C1	C	F1	X	5K
S3	B	C1	C	F1	X	5K
S4	B	C2	C++	F2	Y	10K
S5	B	C1	C	F1	X	5K

R1

FK

Sid	Sname	Cid
S1	A	C1
S2	A	C1
S3	B	C1
S4	B	C2
S5	B	C1

R2

Cid	Cname	Fid	Fname	Sal
C1	C	F1	X	5K
C2	C++	F2	Y	10K

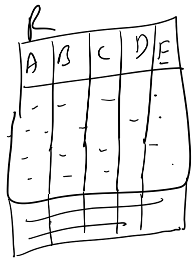
Retrieve student details that are under Faculty F1

Properties of Decomposition

R → R_1, R_2

1. Lossless-Join Decomposition

$$\boxed{R_1 \bowtie R_2 = R}$$



R

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

→ $R_1(AB)$

R_1

A	B
1	2
2	2
3	3

→ $R_2(BC)$

R_2

B	C
2	1
2	3
3	3

$(R_1 \times R_2)$

R_1.A	R_1.B	R_2.B	R_2.C
1	2	2	1
1	2	2	3
1	2	3	3
2	2	2	1
2	2	2	3
2	2	3	3
3	3	2	1
3	3	2	3
3	3	3	3

$\bowtie$ = natural Join

$$\boxed{R_1 \bowtie R_2 = \pi_{ABC}\left(\sigma_{R_1.B = R_2.B}(R_1 \times R_2)\right)}$$

$\sigma_{R_1.B = R_2.B}$

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

lossy Decomposition

$\boxed{R_1 \bowtie R_2 \supseteq R}$

lossy $\quad R_1 \bowtie R_2 \supset R$

lossless $\quad R_1 \bowtie R_2 = R$

$R_1 \bowtie R_2 \subset R$ not possible

2x2=4

1x5=5
1x7=7
5x1=5

R

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

F

R_1

A	B
1	2
2	2
3	3

$F_1 \{A \to B\}$

R_2

A	C
1	1
2	3
3	3

$F_2 : \{A \to C\}$

$$\boxed{R_1 \bowtie R_2 = \pi_{ABC}(\sigma_{R_1.A = R_2.A}(R_1 \times R_2))}$$

$F : \{A \to B, A \to C, BC \to A\}$

R

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

$$\boxed{\begin{array}{c} R_1 \cap R_2 \to R_1 \\ \text{or} \\ R_1 \cap R_2 \to R_2 \end{array}}$$

$R_1 \cap R_2$ S.K. for R_1

$R_1 \cap R_2$ S.K. for R_2

$R_1 \cap R_2 = A$

$$\boxed{\begin{array}{c} A \to R_1 \checkmark \\ A \to R_2 \checkmark \end{array}}$$

$R_1 \times R_2$

$\sigma_{R_1.A = R_2.A}$

π_{ABC}

A	B	A	C	
1	2	1	1	✓
		2	3	×
1	2	3	3	×
1	2	1	1	×
2	2	2	3	✓
2	2	3	3	×
2	2			
		1	1	×
3	3			
3	3	2	3	×
3	3	3	3	✓

Lossless join decomposition

2. **Dependency-Preserving Decomposition**

$R \quad F\{A \to B, A \to C, BC \to A\}$

$R_1 \quad R_2$

$F_1 \quad F_2$

$\{A \to B\} \quad \{A \to C\}$

$$\boxed{(F_1 \cup F_2)^+ = F^+}$$

$F_1 \cup F_2 = \{A \to B, A \to C\}$

$(BC)^+ = BC$

$\boxed{BC \to A}$ × lost

Not Dependency preserving

$$\boxed{(F_1 \cup F_2)^+ \subset F^+}$$

$(F_1 \cup F_2)^+ \quad F^+$

$\boxed{=}$ ✓

$\boxed{\subset}$ ✓

$\boxed{\supset}$ × — not possible

$$\boxed{(F_1 \cup F_2)^+ \subseteq F^+}$$

$$\boxed{(F_1 \cup F_2)^+ = F^+}$$

Dependency preserving decomposition

$$\boxed{(F_1 \cup F_2)^+ \subset F^+}$$

Not Dependency preserving