Decomposition in DBMS

Decomposition of a Relation-

Definition: The process of breaking up or dividing a single relation into two or more sub relations is called as decomposition of a relation.

Properties of Decomposition-

The following two properties must be followed when decomposing a given relation-

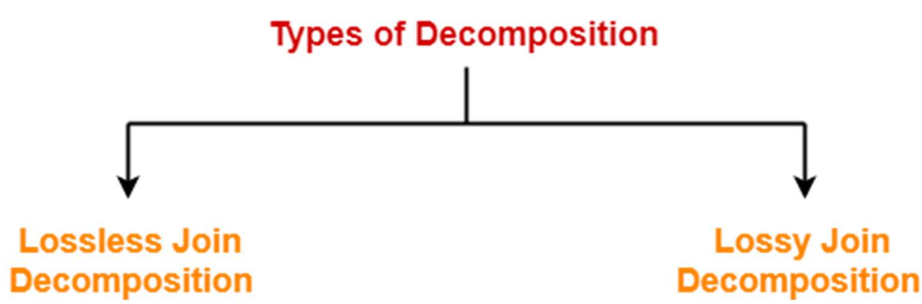
1. Lossless decomposition-

Lossless decomposition ensures-

- No information is lost from the original relation during decomposition.
- When the sub relations are joined back, the same relation is obtained that was decomposed. Every decomposition must always be lossless.

Types of Decomposition-

Decomposition of a relation can be completed in the following two ways-



1. Lossless Join Decomposition-

- Consider there is a relation R which is decomposed into sub relations R_1 , R_2 , …. , R_n.
- This decomposition is called lossless join decomposition when the join of the sub relations results in the same relation R that was decomposed.
- For lossless join decomposition, we always have-

$$\boxed{\mathbf{R_1 \bowtie R_2 \bowtie R_3 \ldots\ldots \bowtie R_n = R}}$$

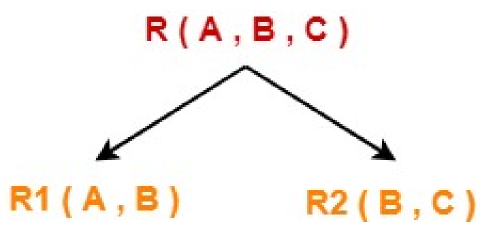
where ⋈ is a natural join operator

Example-

Consider the following relation R(A , B , C)-

A	B	C
1	2	1
2	5	3
3	3	3

Consider this relation is decomposed into two sub relations R_1(A , B) and R_2(B , C)-



The two sub relations are-

A	B
1	2
2	5
3	3

R_1(A , B)

B	C
2	1
5	3
3	3

R_2(B , C)

Now, let us check whether this decomposition is lossless or not. For lossless decomposition, we must have- $\mathbf{R_1 \bowtie R_2 = R}$

Now, if we perform the natural join ($\bowtie$) of the sub relations R_1 and R_2 , we get-

A	B	C
1	2	1
2	5	3
3	3	3

This relation is same as the original relation R. Thus, we conclude that the above decomposition is lossless join decomposition.

NOTE-

Lossless join decomposition is also known as **non-additive join decomposition.**

- This is because the resultant relation after joining the sub relations is same as the decomposed relation.
- No extraneous tuples appear after joining of the sub-relations.

2. Lossy Join Decomposition-

- Consider there is a relation R which is decomposed into sub relations R_1 , R_2 , …. , R_n.
- This decomposition is called lossy join decomposition when the join of the sub relations does not result in the same relation R that was decomposed.
- The natural join of the sub relations is always found to have some extraneous tuples.
- For lossy join decomposition, we always have-

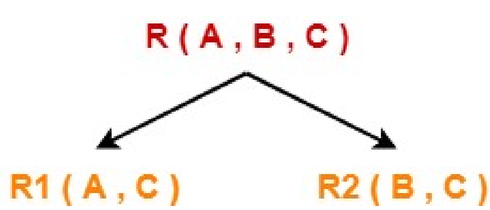
$$\boxed{R_1 \bowtie R_2 \bowtie R_3 \ldots\ldots \bowtie R_n \supset R}$$

where ⋈ is a natural join operator

Example-

Consider the above relation R(A , B , C)-

Consider this relation is decomposed into two sub relations as R_1(A , C) and R_2(B , C)-



The two sub relations are-

A	C
1	1
2	3
3	3

R_1(A , B)

B	C
2	1
5	3
3	3

R_2(B , C)

Now, let us check whether this decomposition is lossy or not.

For lossy decomposition, we must have-

$$R_1 \bowtie R_2 \supset R$$

Now, if we perform the natural join ($\bowtie$) of the sub relations R_1 and R_2 we get-

A	B	C
1	2	1
2	5	3
2	3	3
3	5	3
3	3	3

This relation is not same as the original relation R and contains some extraneous tuples.

Clearly, $R_1 \bowtie R_2 \supset R$. Thus, we conclude that the above decomposition is lossy join decomposition.

NOTE-

- Lossy join decomposition is also known as **careless decomposition.**
- This is because extraneous tuples get introduced in the natural join of the sub-relations.
- Extraneous tuples make the identification of the original tuples difficult.

Determining Whether Decomposition Is Lossless Or Lossy-

Consider a relation R is decomposed into two sub relations R_1 and R_2. Then,

- If all the following conditions satisfy, then the decomposition is lossless.
- If any of these conditions fail, then the decomposition is lossy.

Condition-01:

Union of both the sub relations must contain all the attributes that are present in the original relation R. Thus,

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

Condition-02:

Intersection of both the sub relations must not be null. In other words, there must be some common attribute which is present in both the sub relations. Thus,

$$\mathbf{R1 \cap R2 \neq \emptyset}$$

Condition-03:

Intersection of both the sub relations must be a super key of either R_1 or R_2 or both. Thus,

$$\mathbf{R_1 \cap R_2 = \text{Super key of } R_1 \text{ or } R_2}$$

Solved Examples to know whether a decomposition is lossy or losseless

Problem-01:

Consider a relation schema R (A , B , C , D) with the functional dependencies $A \to B$ and $C \to D$. Determine whether the decomposition of R into R_1 (A , B) and R_2 (C , D) is lossless or lossy.

Solution-

Condition-01:

According to condition-01, union of both the sub relations must contain all the attributes of relation R. So, we have-

$$R_1 (A , B) \cup R_2 (C , D)$$

$$= R (A , B , C , D)$$

Clearly, union of the sub relations contain all the attributes of relation R. Thus, condition-01 satisfies.

Condition-02:

According to condition-02, intersection of both the sub relations must not be null.

So, we have-

$$R_1 (A , B) \cap R_2 (C , D)$$

$$= \Phi$$

Clearly, intersection of the sub relations is null.

So, condition-02 fails.

Thus, we conclude that the **decomposition is lossy**.

Problem-02:

Consider a relation schema R (A , B , C , D) with the following functional dependencies-

$$A \rightarrow B$$

$$B \rightarrow C$$

$$C \rightarrow D$$

$$D \rightarrow B$$

Determine whether the decomposition of R into R_1 (A , B) , R_2 (B , C) and R_3 (B , D) is lossless or lossy.

Solution-

Strategy to Solve

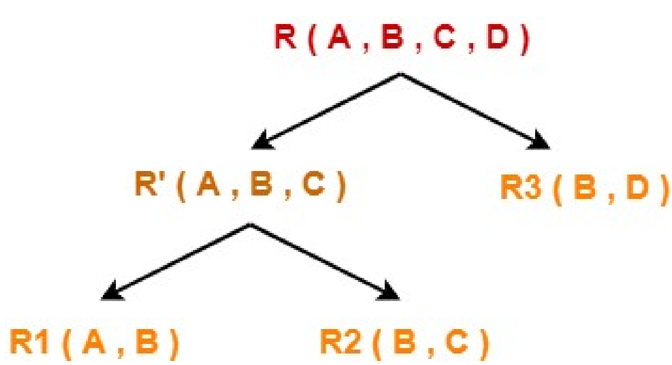
When a given relation is decomposed into more than two sub relations, then-

- Consider any one possible ways in which the relation might have been decomposed into those sub relations.
- First, divide the given relation into two sub relations.
- Then, divide the sub relations according to the sub relations given in the question.

As a thumb rule, remember-

Any relation can be decomposed only into two sub relations at a time.

Consider the original relation R was decomposed into the given sub relations as shown-



Decomposition of R(A, B, C, D) into R'(A, B, C) and R_3(B, D)-

Condition-01:

According to condition-01, union of both the sub relations must contain all the attributes of relation R.

So, we have-

$$R'(A, B, C) \cup R_3(B, D)$$

$$= R(A, B, C, D)$$

Clearly, union of the sub relations contain all the attributes of relation R.

Thus, condition-01 satisfies.

Condition-02:

According to condition-02, intersection of both the sub relations must not be null.

So, we have-

$$R'(A, B, C) \cap R_3(B, D)$$

$$= B$$

Clearly, intersection of the sub relations is not null.

Thus, condition-02 satisfies.

Condition-03:

According to condition-03, intersection of both the sub relations must be the super key of one of the two sub relations or both.

So, we have-

$$R'(A, B, C) \cap R_3(B, D)$$

$$= B$$

Now, the closure of attribute B is-

$$B^+ = \{ B, C, D \}$$

Now, we see-

- Attribute ‘B’ can not determine attribute ‘A’ of sub relation R’.
- Thus, it is not a super key of the sub relation R’.
- Attribute ‘B’ can determine all the attributes of sub relation R_3.
- Thus, it is a super key of the sub relation R_3.

Clearly, intersection of the sub relations is a super key of one of the sub relations.

So, condition-03 satisfies.

Thus, we conclude that the decomposition is lossless.

Decomposition of R'(A, B, C) into R_1(A, B) and R_2(B, C)-

Condition-01:

According to condition-01, union of both the sub relations must contain all the attributes of relation R’.

So, we have-

$R_1 (A , B) \cup R_2 (B , C)$

$= R' (A , B , C)$

Clearly, union of the sub relations contain all the attributes of relation R’.

Thus, condition-01 satisfies.

Condition-02:

According to condition-02, intersection of both the sub relations must not be null.

So, we have-

$R_1 (A , B) \cap R_2 (B , C)$

$= B$

Clearly, intersection of the sub relations is not null.

Thus, condition-02 satisfies.

Condition-03:

According to condition-03, intersection of both the sub relations must be the super key of one of the two sub relations or both.

So, we have-

$R_1 (A , B) \cap R_2 (B , C)$

$= B$

Now, the closure of attribute B is-

$B^+ = \{ B , C , D \}$

Now, we see-

- Attribute 'B' can not determine attribute 'A' of sub relation R_1.
- Thus, it is not a super key of the sub relation R_1.
- Attribute 'B' can determine all the attributes of sub relation R_2.
- Thus, it is a super key of the sub relation R_2.

Clearly, intersection of the sub relations is a super key of one of the sub relations.

So, condition-03 satisfies.

Thus, we conclude that the decomposition is lossless.

Conclusion-

Overall decomposition of relation R into sub relations R_1, R_2 and R_3 is lossless.