Database Normalization – Lossless Join Decomposition

Suppose a relational schema $R$ is decomposed into two relations $R_1$ and $R_2$, then this decomposition is considered as **lossless decomposition** if there is no loss of information after decomposition. This property is also known as **lossless join decomposition**.

More technically, we can say that the result of natural join of $R_1$ and $R_2$ are same as $R$ i.e. no extra or less tuple is generated in natural join of $R_1$ and $R_2$.

$$\pi_{\text{Attributes}(R_1)} \bowtie \pi_{\text{Attributes}(R_2)} = \pi_{\text{Attributes}(R)}$$

It is mandatory property and must always hold while decomposing of table. The decomposition of $R$ into $R_1$ and $R_2$ always will be lossless, if and only if, all the following three conditions holds true:

1. Attribute ($R_1$) $\bigcup$ Attribute ($R_2$) = Attribute ($R$)
2. Attribute ($R_1$) $\bigcap$ Attribute ($R_2$) $\neq \emptyset$
3. The common attribute(s) between $R_1$ and $R_2$ must be candidate key either in $R_1$ or in $R_2$ i.e.

$$\text{Attribute (R1) } \cap \text{ Attribute (R2) } \implies \text{ Attribute (R)}$$

$$\text{Attribute (R1) } \cap \text{ Attribute (R2) } \implies \text{ Attribute (R)}$$

If the decomposition does not follow above three conditions then it is called **lossy decomposition** or **lossy-join decomposition**.
Consider a relation \( R \) (emp_id, emp_name, emp_dob, dept_id, dept_name, dept_location)

Now, it is divided into:

\[
\begin{align*}
R_1 & \text{ (emp_id, emp_name, emp_dob, dept_id)} \\
R_2 & \text{ (dept_id, dept_name, dept_location)}
\end{align*}
\]

If it is lossless join decomposition then it has to follow all the above mentioned three conditions:

1. Union of attributes \( \text{R}_1 \) and attributes \( \text{R}_2 \) is equal to attributes of \( R \)
2. The intersection of attributes \( \text{R}_1 \) and attributes \( \text{R}_2 \) is not null. It is dept_id.
3. The common attribute between \( \text{R}_1 \) and \( \text{R}_2 \) (dept_id) is the candidate key in \( R \).

\[
\text{dept}_\text{id} \rightarrow \text{dept}_\text{id}, \text{dept}_\text{name}, \text{dept}_\text{location}
\]

Therefore, \textbf{it concludes that above decomposition of \( R_1(\text{emp}_\text{id}, \text{emp}_\text{name}, \text{emp}_\text{dob}, \text{dept}_\text{id}) \) and \( R_2(\text{dept}_\text{id}, \text{dept}_\text{name}, \text{dept}_\text{location}) \) is lossless join decomposition.}

\textbf{Exercise:}

A. Suppose \( R \) (v w x y z) and set of FDs

\[ F : \{ z \rightarrow y, \quad y \rightarrow z, \quad x \rightarrow yv, \quad vw \rightarrow x \} \]

Which of the following decomposition is lossless, justify your answer:

1. \( R_1(v w x), R_2( x y z) \)
2. \( R_1(v w), R_2( y z) \)
3. \( R_1(v w x), R_2( y z) \)
4. \( R_1(v w), R_2( w x y z) \)
B. Suppose a relation R

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>111</td>
<td>6</td>
<td>a</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>q</td>
<td>134</td>
<td>5</td>
<td>b</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td>p</td>
<td>338</td>
<td>6</td>
<td>c</td>
<td>z</td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>463</td>
<td>4</td>
<td>d</td>
<td>w</td>
<td></td>
</tr>
</tbody>
</table>

Which of the following decomposition is lossless, justify your answer:

1. R₁(A B), R₂( C D)
2. R₁(A B C), R₂(D E)
3. R₁(A B C), R₂(C D E)
4. R₁(A B C D), R₂( A C D E)
5. R₁(A B C D), R₂( D E)
6. R₁(A B C), R₂( B C D), R₃(D E)