Normalisation – Part 2

3NF

Australian National University

From BCNF to 3NF

- **Facts**

> (1) There exists an algorithm that can generate a **lossless decomposition into BCNF**.
>
> (2) However, a BCNF-decomposition that is **both lossless and dependency-preserving** does not always exist.

- 3NF is **a less restrictive normal form** such that a lossless and dependency preserving decomposition can always be found.

3NF - Definition

- A relation schema R is in **3NF** if whenever a non-trivial FD $X \rightarrow A$ holds in R, then X is a **superkey** or A is a **prime attribute**.

- 3NF allows data redundancy but excludes relation schemas with certain kinds of FDs (i.e., partial FDs and transitive FDs).

Australian National University

Normalisation to 3NF

- Consider the following FDs of ENROL:
  - {StudentID, CourseNo, Semester} → {ConfirmedBy_ID, StaffName},
  - {ConfirmedBy_ID} → {StaffName}.

ENROL

<u>StudentID</u>	<u>CourseNo</u>	<u>Semester</u>	ConfirmedBy_ID	StaffName
123456	COMP2400	2010 S2	u12	Jane
123458	COMP2400	2008 S2	u13	Linda
123458	COMP2600	2008 S2	u13	Linda

- **Is ENROL in 3NF?**
  - {StudentID, CourseNo, Semester} is the only key.
  - ENROL is not in 3NF because {ConfirmedBy_ID} → {StaffName}, {ConfirmedBy_ID} is not a superkey and {StaffName} is not prime attribute.

Normalisation to 3NF

- **Algorithm** for a dependency-preserving and lossless 3NF-decomposition

Input: a relation schema R and a set Σ of FDs on R.

Output: a set $\mathcal{S}$ of relation schemas in 3NF, each having a set of FDs

- Compute a **minimal cover** Σ' for Σ and start with $\mathcal{S} = \phi$
- Group FDs in Σ' by their left-hand-side attribue sets
- For each distinct left-hand-side X_i of FDs in Σ' that includes $X_i \to A_1, X_i \to A_2, \ldots, X_i \to A_k$:
  - Add $R_i = X_i \cup \{A_1\} \cup \{A_2\} \cdots \cup \{A_k\}$ to $\mathcal{S}$
- Remove all redundant ones from $\mathcal{S}$ (i.e., remove R_i if $R_i \subseteq R_j$)
- if $\mathcal{S}$ does not contain a superkey of R, add a key of R as R_0 into $\mathcal{S}$.
- Project the FDs in Σ' onto each relation schema in $\mathcal{S}$

Normalisation to 3NF

R

$R_1 = X_1A_1 \ldots A_K$ … $R_n = X_nA$

$X_1 \rightarrow A_1$

…

$X_1 \rightarrow A_K$

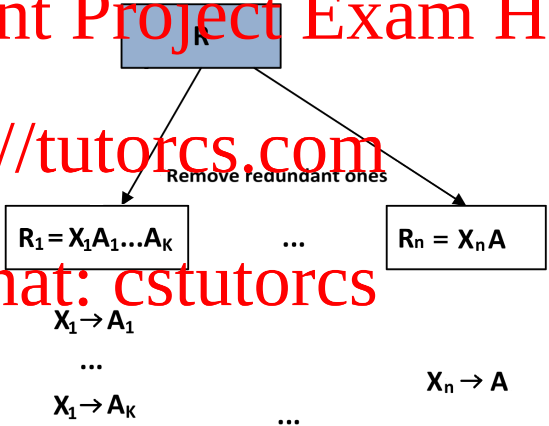
A minimal cover

…

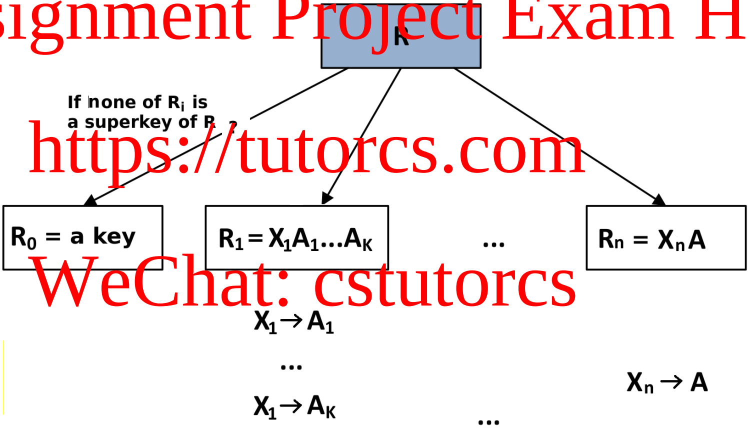
$X_n \rightarrow A$

Australian National University

Normalisation to 3NF



Normalisation to 3NF



Minimal Cover – The Hard Part!

- Let Σ be a set of FDs. A **minimal cover** Σ_m of Σ is a set of FDs such that

1. Σ_m is equivalent to Σ, i.e., start with $\Sigma_m = \Sigma$;
2. **Dependent:** each FD in Σ_m has only a single attribute on its right hand side, i.e., replace each FD $X \rightarrow \{A_1, \ldots, A_k\}$ in Σ_m with $X \rightarrow A_1, \ldots, X \rightarrow A_k$;
3. **Determinant:** each FD has as few attributes on the left hand side as possible, i.e., for each FD $X \rightarrow A$ in Σ_m, check each attribute B of X to see if we can replace $X \rightarrow A$ with $(X - B) \rightarrow A$ in Σ_m;
4. Remove a FD from Σ_m if it is redundant.

Minimal Cover

- **Theorem:**

> The minimal cover of a set of functional dependencies Σ always exists but is not necessarily unique.

- **Examples:** Consider the following set of functional dependencies:

$$\Sigma = \{A \to BC, B \to C, B \to A, C \to AB\}$$

Σ has two different minimal covers:

- $\Sigma_1 = \{A \to B, B \to C, C \to A\}$
- $\Sigma_2 = \{A \to C, C \to B, B \to A\}$

Minimal Cover - Examples

- The set $\{A \rightarrow B, B \rightarrow C, A \rightarrow C\}$ can be reduced to $\{A \rightarrow B, B \rightarrow C\}$, because $\{A \rightarrow C\}$ is implied by the other two.
- Given the set of FDs $\Sigma = \{B \rightarrow A, D \rightarrow A, AB \rightarrow D\}$, we can compute the minimal cover of Σ as follows:

  1. start from Σ;
  2. check whether all the FDs in Σ have only one attribute on the right hand side (look good);

  3. determine if $AB \rightarrow D$ has any redundant attribute on the left hand side ($AB \rightarrow D$ can be replaced by $B \rightarrow D$);
  4. look for a redundant FD in $\{B \rightarrow A, D \rightarrow A, B \rightarrow D\}$ ($B \rightarrow A$ is redundant);

  Therefore, the minimal cover of Σ is $\{D \rightarrow A, B \rightarrow D\}$.

Normalisation to 3NF – Example

- Consider ENROL again:
  - {StudentID, CourseNo, Semester} → {ConfirmedBy_ID, StaffName}
  - {ConfirmedBy_ID} → {StaffName}

StudentID	CourseNo	Semester	ConfirmedBy_ID	StaffName
...	...	...	...	...

- **Can we normalise ENROL into 3NF by a lossless and dependency preserving decomposition?**

Normalisation to 3NF – Example

- Consider ENROL again:
  - {StudentID, CourseNo, Semester} $\rightarrow$ {ConfirmedBy_ID, StaffName}
  - {ConfirmedBy_ID} $\rightarrow$ {StaffName}

StudentID	CourseNo	Semester	ConfirmedBy_ID	StaffName
...	...	...	...	...

- A **minimal cover** is {{StudentID, CourseNo, Semester} $\rightarrow$ {ConfirmedBy_ID}, {ConfirmedBy_ID} $\rightarrow$ {StaffName}}.

- Hence, we have:
  - R_1={StudentID, CourseNo, Semester, ConfirmedBy_ID} with {StudentID, CourseNo, Semester} $\rightarrow$ {ConfirmedBy_ID}
  - R_2={ConfirmedBy_ID, StaffName} with {ConfirmedBy_ID} $\rightarrow$ {StaffName}
  - Omit R_0 because R_1 is a superkey of ENROL.

3NF - Exercises

- Let us do some exercises for the 3NF-decomposition algorithm.
  - **Exercise 1:** $R = \{A, B, C, D\}$ and $\Sigma = \{A \rightarrow B, B \rightarrow C, AC \rightarrow D\}$:

  - **Exercise 2:** $R = \{A, B, C, D\}$ and $\Sigma = \{AD \rightarrow B,\ AB \rightarrow C,\ C \rightarrow B\}$:

3NF - Exercises

- Let us do some exercises for the 3NF decomposition algorithm.

  - **Exercise 1:** $R = \{A, B, C, D\}$ and $\Sigma = \{A \rightarrow B, B \rightarrow C, AC \rightarrow D\}$:
    - $\{A \rightarrow B, B \rightarrow C, A \rightarrow D\}$ is a minimal cover.
    - $R_1 = ABD$, $R_2 = BC$ (omit R_0 because R_1 is a superkey of R)
    - The 3NF-decomposition is $\{$ABD, BC$\}$.

  - **Exercise 2:** $R = \{A, B, C, D\}$ and $\Sigma = \{AD \rightarrow B,\ AB \rightarrow C,\ C \rightarrow B\}$:
    - Σ is its own minimal cover.
    - $R_1 = ABD$, $R_2 = ABC$, $R_3 = CB$ (omit R_3 because $R_3 \subseteq R_2$ and omit R_0 because R_1 is a superkey of R)
    - The 3NF-decomposition is $\{$ABD, ABC$\}$.