Fourth Normal Form

- Relation has redundant data
- In BCNF (since there are no non-trivial FDs)
- Redundancy is due to set valued attributes (in the E-R sense), not because of the FDs
Multi-Valued Dependency

- **Problem**: multi-valued (or binary join) dependency
  - **Definition**: If every instance of schema $R$ can be (losslessly) decomposed using attribute sets $(X, Y)$ such that:
    $$ r = \pi_X(r) \bowtie \pi_Y(r) $$
  - then a multi-valued dependency
    $$ R = \pi_X(R) \bowtie \pi_Y(R) $$ holds in $r$
  - Ex: $\text{Person} = \pi_{\text{SSN},\text{PhoneN}}(\text{Person}) \bowtie \pi_{\text{SSN},\text{ChildSSN}}(\text{Person})$

Fourth Normal Form (4NF)

- A schema is in **fourth normal form** (4NF), if for every MVD $R = X \bowtie Y$ in that schema is either:
  - $X \subseteq Y$ or $Y \subseteq X$ (trivial case); or
  - $X \cap Y$ is a superkey of $R$ (i.e., $X \cap Y \rightarrow R$)
Fourth Normal Form (Cont’d)

- **Intuition**: if \( X \cap Y \rightarrow R \), there is a unique row in relation \( r \) for each value of \( X \cap Y \) (hence no redundancy)
  - Ex: SSN does not uniquely determine PhoneN or ChildSSN, thus Person is not in 4NF.

- **Solution**: Decompose \( R \) into \( X \) and \( Y \)
  - Decomposition is lossless – but not necessarily dependency preserving (since 4NF implies BCNF – next)

4NF Implies BCNF

- Suppose \( R \) is in 4NF and \( X \rightarrow Y \) is a FD.
  - Assume \( X \) and \( Y \) are disjoint
  - \( R_1 = XY \), \( R_2 = R - Y \) is a lossless decomposition of \( R \)
  - Thus \( R \) has the MVD: \( R = R_1 \Join R_2 \)

- Since \( R \) is in 4NF, one of the following must hold:
  - \( XY \subseteq R - Y \)
    - (an impossibility)
  - \( R - Y \subseteq XY \)
    - (i.e., \( R = XY \) and \( X \) is a superkey)
  - \( XY \cap R - Y = X \) is a superkey
  - Hence, \( X \rightarrow Y \) satisfies BCNF condition
4NF Decomposition Algorithm

For simplicity, assume A and B are disjoint for FDs $A \rightarrow B$ in $R$

Input: $R = (\bar{R}; \bar{D})$ /* $\bar{D}$ is a set of FDs and MVDs; FDs are treated as MVDs */
Output: A lossless decomposition of $R$ where each schema is in 4NF.

Decomposition := [R] /* Initially decomposition consists of only one schema */
while there is a schema $S = (\bar{S}; \bar{D}')$ in Decomposition that is not in 4NF do
/* Let $\bar{X} \rightarrow \bar{Y}$ be an MVD in $\bar{D}$ * such that $\bar{X} \bar{Y} \subseteq \bar{S}$ and */
it violates 4NF in $S$. Decompose using this MVD */
Replace $S$ in Decomposition with schemas $S_1 = (\bar{X} \bar{Y}; \bar{D}_1')$ and
$S_2 = ((\bar{S} - \bar{Y}) \cup \bar{X}; \bar{D}_2')$, where $\bar{D}_1' = \pi_{\bar{X}\bar{Y}}(\bar{D}')$ and $\bar{D}_2' = \pi_{\bar{S} - \bar{Y}}(\bar{D}')$
end
return Decomposition

The algorithm is not correct. $S_1$ and $S_2$ should be
$S_1 = (X; D_1')$
$S_2 = (Y; D_2)$;

Otherwise, $X$ join $Y$ should be replaced to $X\rightarrow\!\!\!\!\!> Y$. (See slide 88)
If $X\rightarrow\!\!\!\!\!> Y$, $R = XY$ join $X(R-Y)$

Projection of MVD on a Set of Attributes

- Projection of MVD $R = V \bowtie W$ on a set of attributes $X$
  - $X = (X \cap V) \bowtie (X \cap W)$, if $V \cap W \subseteq X$
  - Undefined, otherwise.

- Example
  - Projection of MVD: $ABCD = AB \bowtie BCD$ on $ABC$
    - $AB \cap BCD = B \subseteq ABC$. So, the projection is $AB \bowtie BC$
  
  - Projection of MVD: $ABCD = ACD \bowtie BD$ on $ABC$
    - $ACD \cap BD = D \subseteq ABC$. So, the projection is undefined.
4NF Decomposition Example

- Example
  - Attributes = \( \{ABCD\} \)
  - MVDs
    - MVD1. \( ABCD = AB \bowtie BCD \)
    - MVD2. \( ABCD = ACD \bowtie BD \)
    - MVD3. \( ABCD = ABC \bowtie BCD \)
  - From MVD1, decomposed to \( AB, BCD \)
    - Projection of remaining MVDs on \( AB \) is not defined
    - Projection of remaining MVDs on \( BCD \) is:
      - For MVD2, \( BCD = CD \bowtie BD \)
      - For MVD3, \( BCD = BC \bowtie BCD \) (trivial)
  - Finally, \( AB, BD, CD \)

3NF Synthesis, BNCF, and 4NF Decomposition

- Example
  - Attributes = \( \{ABCDEFG\} \)
  - FDs = \( \{AB \rightarrow C, C \rightarrow B, BC \rightarrow DE, E \rightarrow FG\} \)
  - MVDs: \( R = BC \bowtie ABDEFG, R = EF \bowtie FGABCD \)
  - 3NF Synthesis result
    - \( R_1 = (ABC; (AB \rightarrow C, C \rightarrow B)) \)
    - \( R_2 = (CBDE; (C \rightarrow BDE)) \)
    - \( R_3 = (EF; (E \rightarrow FG)) \)
    - \( R_1 \) is not in BCNF due to \( C \rightarrow B \)
      - \( R_{11} = (BC; (C \rightarrow B)), R_{12} = (AC; \{\}) \)
3NF Synthesis & 4NF Decomposition (cont’)

Example

- BCNF Synthesis result
  - \( R_{11} = (AC; \{\}) \), \( R_{12} = (BC; \{ C \rightarrow B \}) \)
  - \( R_2 = (CBDE; \{ C \rightarrow BDE \}) \), \( R_3 = (EFG; \{ E \rightarrow FG \}) \)

- MVDs: \( R = BC \bowtie ABDEFG \), \( R = EF \bowtie FGABCD \)
  - The first MVD can be projected to \( R_2 \) (here, \( B = V \cap W \subseteq CBDE \))
    - then, “projected \( R_2'' = BC \bowtie BDE \). Is \( R_2 \) in 4NF?
      - No! because \( BC \cap BDE = B \) and \( B \) is not the key
    - \( R_{21} = (BC; \{ C \rightarrow B \}) \), \( R_{22} = (BDE; \{\}) \)
  - Similarly, the second MVD can be projected to \( R_3 \)
    - here, \( F = V \cap W \subseteq EFG \)
    - then, “projected \( R_3'' = EF \bowtie FG \). Is \( R_3 \) in 4NF?
      - No! because \( EF \cap FG = F \) and \( F \) is not the key
    - \( R_{31} = (EF; \{ E \rightarrow F \}) \), \( R_{32} = (GF; \{\}) \)

Customary Representation of MVDs

- Customary representation of MVDs
  - \( MVD \ R = \ V \bowtie W \) over \( R = (R; D) \), where
    - \( X = V \cap W \)
    - \( X \cup Y = V \) or \( X \cup Y = W \)
    - are represented as \( X \rightarrow Y \)
    - i.e., \( R = XY \bowtie X(R-Y) \)

- Another way of defining MVD in a relation
  - \( X \rightarrow Y \) then,
    - \( \forall \) tuple \( t, u \in R: t[X] = u[X] \). then \( \exists \) tuple \( v \in R \) where
      - \( v[X] = t[X] \) and
      - \( v[Y] = t[Y] \) and
      - \( v[rest] = u[rest] \)
Examples

- Apply (SSN, college, hobby)
  - SSN → college

- Apply (SSN, college, date, major)
  - Requirements
    - Apply once to each college
    - May apply to multiple majors
  - We can derive...
    - SSN, college → date, major / date → college
    - SSN → college, date
      - What is the real world constraint encoded by the MVD above?
      - A student must apply to the same set of majors at all colleges.

4NF Decomposition Algorithm (Rewritten)

**Input:** relation R + FDs for R + MVDs for R

**Output:** decomposition of R into 4NF relations with “lossless join”

**Compute keys for R**

Repeat until all relations are in 4NF:
- Pick any R’ with nontrivial A → B that violates 4NF
- Decompose R’ into R₁(A, B) and R₂(A, rest)
- Compute FDs and MVDs for R₁ and R₂
- Compute keys for R₁ and R₂