Relational Normalization Theory (supplemental material)

CSE 532, Theory of Database Systems
Stony Brook University
http://www.cs.stonybrook.edu/~cse532
Quiz 8

- Consider a schema S with functional dependencies:
  - \{A \rightarrow BC, C \rightarrow FG, E \rightarrow HG, G \rightarrow A\}.

(a) Compute the attribute closure of A with respect to S.
- ABCFG

(b) Suppose we decompose S so that one of the subrelations, say called R, contains attributes AFE only. Find a projection of the functional dependencies in S on R (i.e., find the functional dependencies between attributes AFE).
- \{A \rightarrow F, E \rightarrow AF\} or \{A \rightarrow F, E \rightarrow A, E \rightarrow F\} or \{A \rightarrow F, E \rightarrow A\}

(c) Is R in BCNF?
- No, A is not a superkey.

Use one cycle of the BCNF decomposition algorithm to obtain two subrelations, and check BCNF cond.
- \(R1 = \{AF, A \rightarrow F\}\) and \(R2 = \{EA, E \rightarrow A\}\). They are in BCNF.

(c) Pearson and P. Fodor (CS Stony Brook)
Quiz 8 (cont.)

- \{A \rightarrow BC, \ C \rightarrow FG, \ E \rightarrow HG, \ G \rightarrow A\}.
- Decomposition: \( R_1 = \{AF, A \rightarrow F\} \) and \( R_2 = \{EA, E \rightarrow A\} \).

(d) Show that your decomposition is lossless.
- \( \{AF\} \cap \{EA\} = \{A\} \), which is a key of \( R_1 \).

(e) Is your decomposition dependency preserving?
- Yes, because each of the FDs of \( R \) is entailed by FDs of \( R_1 \) and \( R_2 \).

(f) Is \( R \) in 3NF? Explain.
- No, \( F \) is not part of a key.

Use 3NF decomposition algorithm to obtain subrelations, and answer whether they are in 3NF.
- Obtain same \( R_1 = \{AF, A \rightarrow F\} \) and \( R_2 = \{EA, E \rightarrow A\} \). They are in 3NF.
Quiz 9

- Consider the following functional dependencies over the attribute set BCGHMVWY:

  \[
  \begin{align*}
  W & \rightarrow V &Wy & \rightarrow BV & WC & \rightarrow V & V & \rightarrow B \\
  BG & \rightarrow M & BV & \rightarrow Y & BYH & \rightarrow V & M & \rightarrow W \\
  Y & \rightarrow H & CY & \rightarrow W \\
  \end{align*}
  \]

- Find a minimal cover.

- Decompose into lossless 3NF.

- Check if all the resulting relations are in BCNF.
  - If not, decompose it into a lossless BCNF.
Quiz 9

W $\rightarrow$ V  
WY $\rightarrow$ BV  
WC $\rightarrow$ V  
V $\rightarrow$ B  
BG $\rightarrow$ M  
BV $\rightarrow$ Y  
BYH $\rightarrow$ V  
M $\rightarrow$ W  
Y $\rightarrow$ H  
CY $\rightarrow$ W

- **Minimal cover:**
  1. Split WY $\rightarrow$ BV into WY $\rightarrow$ B and WY $\rightarrow$ V.
Quiz 9

\[
\begin{align*}
W &\rightarrow V \\
WY &\rightarrow BV \\
BG &\rightarrow M \\
BV &\rightarrow Y \\
Y &\rightarrow H \\
CY &\rightarrow W \\
WC &\rightarrow V \\
V &\rightarrow B \\
BYH &\rightarrow V \\
M &\rightarrow W \\
\end{align*}
\]

- **Minimal cover:**
  1. Split \(WY \rightarrow BV\) into \(WY \rightarrow B\) and \(WY \rightarrow V\).
  2. Since \(W \rightarrow V\) and \(W \rightarrow B\) are implied by the given set of FDs, we can replace \(WY \rightarrow B\), \(WY \rightarrow V\), and \(WC \rightarrow V\) in the original set with \(W \rightarrow B\) (\(W \rightarrow V\) already exists in the original set, so we discard the duplicate).
Quiz 9

\[
\begin{align*}
W &\rightarrow V & WY &\rightarrow BV & WC &\rightarrow V & V &\rightarrow B \\
BG &\rightarrow M & BV &\rightarrow Y & BYH &\rightarrow V & M &\rightarrow W \\
Y &\rightarrow H & CY &\rightarrow W \\
\end{align*}
\]

• **Minimal cover:**

1. Split \( WY \rightarrow BV \) into \( WY \rightarrow B \) and \( WY \rightarrow V \).

2. Since \( W \rightarrow V \) and \( W \rightarrow B \) are implied by the given set of FDs, we can replace \( WY \rightarrow B \), \( WY \rightarrow V \), and \( WC \rightarrow V \) in the original set with \( W \rightarrow B \) (\( W \rightarrow V \) already exists in the original set, so we discard the duplicate).

3. \( V \rightarrow Y \) and \( BY \rightarrow V \) can be derived from the original set so we can replace \( BV \rightarrow Y \) and \( BYH \rightarrow V \) with \( V \rightarrow Y \) and \( BY \rightarrow V \).
Quiz 9

\[
\begin{align*}
W &\rightarrow V \\
WY &\rightarrow BV \\
WC &\rightarrow V \\
V &\rightarrow B \\
BG &\rightarrow M \\
BV &\rightarrow Y \\
BYH &\rightarrow V \\
M &\rightarrow W \\
Y &\rightarrow H \\
CY &\rightarrow W
\end{align*}
\]

- **Minimal cover:**
  1. Split \(WY \rightarrow BV\) into \(WY \rightarrow B\) and \(WY \rightarrow V\).
  2. Since \(W \rightarrow V\) and \(W \rightarrow B\) are implied by the given set of FDs, we can replace \(WY \rightarrow B\), \(WY \rightarrow V\), and \(WC \rightarrow V\) in the original set with \(W \rightarrow B\) (\(W \rightarrow V\) already exists in the original set, so we discard the duplicate).
  3. \(V \rightarrow Y\) and \(BY \rightarrow V\) can be derived from the original set so we can replace \(BV \rightarrow Y\) and \(BYH \rightarrow V\) with \(V \rightarrow Y\) and \(BY \rightarrow V\).
  4. We remove redundant FDs: \(W \rightarrow B\).
Quiz 9

W → V  
WY → BV  
WC → V  
V → B  
BG → M  
BV → Y  
BYH → V  
M → W  
Y → H  
CY → W

- **Minimal cover:**

1. Split WY → BV into WY → B and WY → V.
2. Since W → V and W → B are implied by the given set of FDs, we can replace WY → B, WY → V, and WC → V in the original set with W → B (W → V already exists in the original set, so we discard the duplicate).
3. V → Y and BY → V can be derived from the original set so we can replace BV → Y and BYH → V with V → Y and BY → V.
4. We remove redundant FDs: W → B.
5. The final result is:

W → V  
V → B  
BG → M  
V → Y  
BY → V  
M → W  
Y → H  
CY → W

(c) Pearson and P. Fodor (CS Stony Brook)
Quiz 9

- Lossless 3NF:

  (WV; \{W \rightarrow V\})
  (VBY; \{BY \rightarrow V; V \rightarrow Y; V \rightarrow B\})
  (BGM; \{BG \rightarrow M\})
  (MW; \{M \rightarrow W\})
  (YH; \{Y \rightarrow H\})
  (CYW; \{CY \rightarrow W\})

- Since BGM is a superkey of the original schema, we don't need to add anything to this decomposition.
Quiz 9

- **BCNF:**
  - The schema \((BGM; \{BG \rightarrow M\})\) is not in BCNF because \(M \rightarrow B\) is entailed by the original set of FDs.
    - Decompose \(BGM\) with respect to \(M \rightarrow B\) into \((BM; \{M \rightarrow B\})\) and \((GM; \{\})\), loosing the FD \(BG \rightarrow M\).
  - The schema \((CYW; \{CY \rightarrow W\})\) is not in BCNF because \(W \rightarrow Y\) is entailed by the original set of FDs.
    - Decompose \(CYW\) using \(W \rightarrow Y\) into \((WY; \{W \rightarrow Y\})\) and \((WC; \{\})\), loosing \(CY \rightarrow W\).
Extra problem 1

- Prove that the following sets of FDs are equivalent:
  
  F: $A \rightarrow B$, $C \rightarrow A$, $AB \rightarrow C$
  
  G: $A \rightarrow C$, $C \rightarrow B$, $CB \rightarrow A$
  
  $F \equiv G \iff F$ entails $G$ and $G$ entails $F$

1. F entails G:
   
   \[ A^+_F = ABC, \text{ so } F \text{ entails } A \rightarrow C \]
   
   \[ C^+_F = ABC, \text{ so } F \text{ entails } C \rightarrow B \]
   
   \[ CB^+_F = ABC, \text{ so } F \text{ entails } CB \rightarrow A \]

2. G entails F:
   
   \[ A^+_G = ABC, \text{ so } G \text{ entails } A \rightarrow B \]
   
   \[ C^+_G = ABC, \text{ so } G \text{ entails } C \rightarrow A \]
   
   \[ AB^+_G = ABC, \text{ so } G \text{ entails } AB \rightarrow C \]
Extra problem 2

- Consider the schema R over the attributes ABCDEFG with the following functional dependencies:
  \[ AB \rightarrow C, \quad C \rightarrow B, \quad BC \rightarrow DE, \quad E \rightarrow FG \]

  and the following multivalued dependencies:
  \[ R = BC \Join ABDEFG \]
  \[ R = EF \Join FGABCD \]

- Decompose this schema into 4NF while trying to preserve as many functional dependencies as possible: first use the 3NF synthesis algorithm, then the BCNF algorithm, and finally the 4NF algorithm.
  - Is the resulting schema after decomposition dependency-preserving?
Extra problem 2

AB → C, C → B, BC → DE, E → FG

1. 3NF: split the right-hand sides of the FDs and minimize the left-hand sides:

AB → C, C → B, C → D, C → E, E → F, E → G

Synthesize the 3NF schemas: R1 = (ABC; \{AB → C\}), R2 = (CBDE; \{C → BDE\}), R3 = (EFG; \{E → FG\})

2. BCNF:

R1 is not in BCNF, because C → B is implied by our original set of FDs (and C is not a superkey of R1): we decompose R1 further using C → B: R11 = (AC; \{\}) and R12 = (BC; \{C → B\}).
Extra problem 2

\[ R_{11} = (AC; \{\};), \quad R_{12} = (BC; \{C \rightarrow B\}), \]
\[ R_2 = (CBDE;\{C \rightarrow BDE\}), \quad R_3 = (EFG; \{E \rightarrow FG\}) \]

\[ R = BC \bowtie ABDEFG, \quad R = EF \bowtie FGABCD \]

4NF: \( R = BC \bowtie ABDEFG \) projects onto \( R_2 \) as \( R_2 = BC \bowtie BDE \)
and it violates 4NF because \( B = BC \cap BDE \) is not a superkey.

Therefore, we decompose \( R_2 \) into \((BC;\{C \rightarrow B\})\)and\((BDE;\{\})\).

MVD \( R = EF \bowtie FGABCD \) projects onto \( R_3 \) as \( R_3 = EF \bowtie FG \)
and it violates 4NF because \( F = EF \cap FG \) is not a superkey.

Therefore, we decompose \( R_3 \) into \((EF;\{E \rightarrow F\})\) and \((FG;\{\})\).
Extra problem 2

\[ \begin{align*}
AB & \rightarrow C, \ C \rightarrow B, \ C \rightarrow D, \ C \rightarrow E, \ E \rightarrow F, \ E \rightarrow G \\
R & = BC \Join ABDEFG, \ R = EF \Join FGABCD \\
R_{11} & = (AC; \emptyset), \ R_{12} = (BC; \{C \rightarrow B\}) \\
R_{21} & = (BC; \{C \rightarrow B\}), \ R_{22} = (BDE; \emptyset) \\
R_{31} & = (EF; \{E \rightarrow F\}), \ R_{32} = (FG; \emptyset)
\end{align*} \]

The decomposition is not dependency preserving:

- \( A \rightarrow B \), which was present in the original schema, is now not derivable from the FDs attached to the schemas in the decomposition.
Extra problem 3

- Consider the schema R over the attributes ABCDEFG:
  
  \[ \begin{align*}
  \text{DE} &\rightarrow \text{F}, \text{BC} \rightarrow \text{AD}, \text{FD} \rightarrow \text{G}, \text{F} \rightarrow \text{DE}, \text{D} \rightarrow \text{E} \\
  \text{R} &= \text{ABCDEFG} \bowtie \text{BCDE}
  \end{align*} \]

- Compute 4NF:

1. 3NF: minimal cover:

   **Split the LHS:**
   \[ \begin{align*}
   \text{DE} &\rightarrow \text{F}, \text{BC} \rightarrow \text{A}, \text{BC} \rightarrow \text{D}, \text{FD} \rightarrow \text{G}, \text{F} \rightarrow \text{D}, \text{F} \rightarrow \text{E}, \text{D} \rightarrow \text{E}
   \end{align*} \]

   **Reduce the RHS:**
   
   Since \( \text{D} \rightarrow \text{E} \), we can delete \( \text{E} \) in the LHS of \( \text{DE} \rightarrow \text{F} \):
   \( \text{D} \rightarrow \text{F} \)
   
   Since \( \text{F} \rightarrow \text{D} \), we can delete \( \text{D} \) in the LHS of \( \text{FD} \rightarrow \text{G} \):
   \( \text{F} \rightarrow \text{G} \)

   **Remove redundant FDs:**
   \[ \begin{align*}
   \text{D} &\rightarrow \text{F}, \text{BC} \rightarrow \text{A}, \text{BC} \rightarrow \text{D}, \text{F} \rightarrow \text{G}, \text{F} \rightarrow \text{D}, \text{D} \rightarrow \text{E}
   \end{align*} \]
Extra problem 3

D → F, BC → A, BC → D, F → G, F → D, D → E

R = ABCFG \Join BCDE

3NF: (DEF; \{D \rightarrow \ EF\})

(BCAD; \{BC \rightarrow AD\})

(FGD; \{F \rightarrow GD\})

This decomposition is lossless since BCAD⁺ = BCADEFG, i.e., the attribute set of the second schema is a superkey of the original schema.

BCNF: yes

4NF: We project the MVD on BCAD: BCAD = ABC \Join BCD.

ABC \cap BCD = BC is a superkey of the schema (BC \rightarrow AD), so, this MVD does not violate the 4NF.

(c) Pearson and P. Fodor (CS Stony Brook)
Extra problem 4

• Consider R=BCEGWWY with the MVDs:
  
  MVD1: R = WBCY \bowtie YEVG
  MVD2: R = WBCE \bowtie WBYVG
  MVD3: R = WBY \bowtie CYEVG

4NF 1: Use MVD1 to obtain the following decomposition: WBCY, YEVG.

Use MVD2 to WBCY to yield WBC, WBY.

Use MVD3 cannot be used to decompose WBC because the join attribute, Y, is not in this attribute set.

MVD3 cannot be used to decompose WBY or YEVG because it projects as a trivial dependency in these cases: WBY=WBY \bowtie Y and YEVG = Y \bowtie YEVG.

4NF 2: the decomposition 4NF 1 is not unique: we can apply MVD3 and then MVD1 then will obtain: WBY, CY, YEVG.