CIS*3530 Data Base Systems and Concepts
Fall 2020
Instructor: Fangju Wang

Assignment 4 (20%)
Due: Wednesday Nov 25, 2020
Submission By: Mitchell Van Braeckel, Student ID: 1002297

NOTE: remember non-trivial means Y is not a subset of X

Question 1: (3%)
Mark True or False for each of the following:

1. $\{ A \rightarrow B, A \rightarrow D, BD \rightarrow C \} \Rightarrow \{ A \rightarrow C \}$. **True**
    a. $\Rightarrow \{A \rightarrow BD, BD \rightarrow C\} \Rightarrow \{A \rightarrow C\}$
2. $\{ A \rightarrow B, BD \rightarrow C \} \Rightarrow \{ A \rightarrow C \}$. **False**
3. $\{ A \rightarrow B \}, C \supseteq D, C \subseteq B \Rightarrow \{ A \rightarrow D \}$. **True**
    a. $C \supseteq D \Rightarrow C \rightarrow D$
    b. $C \subseteq B \Rightarrow B \rightarrow C$ (since $C \subseteq B$ rewritten as $B \supseteq C$)
    c. So, $(A \rightarrow B)$ & $(B \rightarrow C)$ & $(C \rightarrow D) \Rightarrow A \rightarrow D$
4. $\{ A \rightarrow B, D \rightarrow C, B \rightarrow D \} \Rightarrow \{ A \rightarrow C \}$. **True**
    a. $\Rightarrow \{A \rightarrow D, D \rightarrow C\} \Rightarrow \{A \rightarrow C\}$
5. $\{ D \rightarrow B, A \rightarrow C \} \Rightarrow \{ AD \rightarrow B \}$. **True**
    a. Since $X = \{A,D\}$ gives us $X^+ = \{A,B,C,D\}$ (where $B \in X^+$)

Question 2: (3%)
Consider the following relation:

A	B	C
10	b1	c1
10	b2	c2
11	b4	c1
12	b3	c4
13	b1	c1
14	b3	c4

1. Which of the following dependencies may hold in the above relation?
   $A \rightarrow B$, $B \rightarrow C$, $C \rightarrow B$, $B \rightarrow A$, $C \rightarrow A$
   So, only $B \rightarrow C$ may hold in the above relation.

2. Does the relation schema have a candidate key? If it does, what is it? If it does not, why not?
   We just determined $F = \{B \rightarrow C\}$, so necessary attributes are A and B, while C is the only useless attribute and there are no middle-ground attributes.

So, if we have X = {A,B}, then we get X+ = {A,B,C} = R (from FD B → C), therefore the relation schema has a candidate key which is X = {A,B}, because it is a minimal superkey for R(A,B,C).

Question 3: (3%)
Answer the following questions:

1. Consider relation schema R(A, B, C, D, E) with the following dependencies:
   AB → C, CD → E, DE → B

   Find a candidate key for R:

   A, D are necessary attributes
   There are no useless attributes
   B, C, E are middle-ground attributes

   Try X = { A, D }, where initial X^+ = X

   No FD's apply
   X^+ = {A,D}, but X is NOT a superkey of R, thus not a candidate key either

   So, add X to a subset of middle-ground attributes: Try X = { A, D, B }, where initial X^+ = X

   AB → C : X^+ = {A,B,C,D}
   CD → E : X^+ = {A,B,C,D,E} = R, therefore a candidate key for R is X = {A,B,D}

2. Consider relation schema S(A, B, C, D, E, Q, G, H, I, J) with the following dependencies:
   AB → C, A → DE, B → Q, Q → GH, D → IJ

   Find a candidate key for S:

   A, B are necessary attributes
   C, E, G, H, I, J are useless attributes
   D, Q are middle-ground attributes

   Try X = { A, B }, where initial X^+ = X

   AB → C : X^+ = {A,B,C}
   A → DE : X^+ = {A,B,C, D,E}
   B → Q : X^+ = {A,B,C, D,E,Q}
   Q → GH : X^+ = {A,B,C, D,E,Q,G,H}
   D → IJ : X^+ = {A,B,C, D,E,Q,G,H,I,J} = S, thus a candidate key of S is X = {A,B}

Question 4: (3%)
Consider relation schema S(A, B, C, D, E, Q, G, H, I, J) with the following dependencies:

AB → C, BD → EQ, AD → GH, A → I, H → J

Assume the domains of the attributes are single values of atomic data types.

1. Which normal forms does the relation schema violate? Why?

If we have X = {A,B,D}, then we get X+ = {A,B,C,D,E,Q,G,H,I,J} = S as the only candidate key, so we'll use this as the primary key for checking which normal forms the relation schema violates and why:

- 1NF: not violated because we were given the assumption that the domains of the attributes are single values of atomic data types.

- 2NF: is violated because there exists a non-key attribute that depends on a proper subset of a key (ie. not fully dependent on the primary key)
    - Eg. I (non-key attribute) is partially dependent on A (proper subset of key) instead of the entire primary key {A, B, D}
- 3NF and BCNF: are both violated because 2NF is violated

2. If it violates the 3NF, decompose it into 3NF.

Above, we have showed that S is not in 3NF, so we will decompose it into 3NF:

i = 0; W := {}

i = 1; Use AB → C to make S1 ⇒ S1 = (A,B,C) & W = {S1(A,B,C)}

i = 2; Use BD → EQ to make S2 ⇒ S2 = (B,D,E,Q) & W = {S1(A,B,C), S2(B,D,E,Q)}

i = 3; Use AD → GH to make S3 ⇒ S3 = (A,D,G,H) & W = {S1(A,B,C), S2(B,D,E,Q), S3(A,D,G,H)}

i = 4; Use A → I to make S4 ⇒ S4 = (A,I) & W = {S1(A,B,C), S2(B,D,E,Q), S3(A,D,G,H), S4(A,I)}

i = 5; Use H → J to make S5 ⇒ S5 = (H,J) & W = {S1(A,B,C), S2(B,D,E,Q), S3(A,D,G,H), S4(A,I), S5(H,J)}

Since W does not contain a candidate key of the original schema S, make another schema to contain the candidate key of S ⇒ S6 = (A,B,D) & W = {S1(A,B,C), S2(B,D,E,Q), S3(A,D,G,H), S4(A,I), S5(H,J), S6(A,B,D}

So, S6 contains candidate key {A,B,D}, therefore W is the decomposition of S into 3NF.

Question 5: (5%)

Relation schema U(L, A, T, R, P, M) has the following function dependencies F:

TR → L

TA → R

LP → M

TP → R

L → A

Where L is flight, A is airline, T is time, R is route, P is passenger, and M is airfare.

1. Find a candidate key for U:

    T, P are necessary attributes

    M is the only useless attribute

    L, A, R are middle-ground attributes

    Try X = { P, T }, where initial X^+ = X

    TP → R : X^+ = {P,R,T}

    TR → L : X^+ = {L,P,R,T}

    LP → M : X^+ = {L,M,P,R,T}

    L → A : X^+ = {A,L,M,P,R,T} = U, thus a candidate key for U is X = {P,T}

2. Is U in BCNF? Why?

    U is not in BCNF because there is nontrivial FD TR → L in which the determinant is not a superkey of U ($\{T,R\}^+$ = {T,R,L,A} ≠ U).

3. If U is NOT in BCNF, decompose U into BCNF relation schemas (show the steps). Analyze if the decomposition is lossless. Analyze if the decomposition is dependency preserving.

Above, we have showed that U is not in BCNF, so we will decompose it into BCNF:

Iteration 0: W := {U}

Iteration 1: Use TR → L to decompose U

W := {W-U} ∪ (U-L) ∪ (T,R,L) = {ø} ∪ U1(A,M,P,R,T) ∪ U2(L,R,T)
= {U1(A,M,P,R,T), U2(L,R,T)}

Check if decomposition of U into {U1, U2} is lossless:

U1 ∩ U2 → U2: RT → LRT, which is in F^+, therefore the decomposition is lossless.

So, W = { U1(A,M,P,R,T), U2(L,R,T) }, F1 = {TA → R, TP → R}, F2 = {TR → L}

Check if U1 and U2 are in BCNF:

- U1 is not in BCNF: Non-trivial FD TA → R on U1 - determinant TA not a superkey of U1
- U2 is in BCNF: only non-trivial FD TR → L on U2 - determinant TR is a superkey of U2

Iteration 2: Use TA → R to decompose U1

W := {W-U1} ∪ (U1-R) ∪ (A,R,T) = {U2(L,R,T)} ∪ U3(A,M,P,T) ∪ U4(A,R,T)
= {U2(L,R,T), U3(A,M,P,T), U4(A,R,T)}

Check if decomposition of U1 into {U3, U4} is lossless:

U3 ∩ U4 → U4: AT → ART, which is in $F1^+$, therefore the decomposition is lossless

So, W = { U2(L,R,T), U3(A,M,P,T), U4(A,R,T) }, F2 = {TR → L}, F3 = {}, F4 = {TA → R}

Check if U3 and U4 are in BCNF:

- U3 is in BCNF: no FD on U3
- U4 is in BCNF: only non-trivial FD TA → R on U4 - determinant TA is a superkey of U4

STOP: All the relation schemas in W are in BCNF, so the decomposition process stops.

Check if the decomposition of U into {U2, U3, U4} is dependency preserving:

The FDs on U : F = { TR → L, TA → R, LP → M, TP → R, L → A }.
TR → L, TA → R have new homes U2, U4 respectively (but no FD for U3).
Let F' = { TR → L, TA → R }, LP → M, TP → R, L → A have no new home.
They all cannot be logically implied by the FDs in F'.
Therefore, the decomposition is not dependency preserving.

In conclusion W = { U2(L,R,T), U3(A,M,P,T), U4(A,R,T) } is the decomposition of U into BCNF, which is lossless, but not dependency preserving.

Question 6: (3%)
Consider relation schema U and dependency set F on the relation schema in Question 5.

1. Is U in 3NF? Why?

    U is not in 3NF because there is nontrivial FD TR → L in which the determinant is not a superkey of U ($\{T,R\}^+ = \{T,R,L,A\} \neq U$), so it fails to satisfy either condition for BCNF, and also attribute L (in Y but not in X) is not in the only candidate key X = {P,T} (which we found earlier in 5.1).

2. If U is NOT in 3NF, decompose U into 3NF relations schemas.

Above, we have showed that U is not in 3NF, so we will decompose it into 3NF:

i = 0; W := {}
i = 1; Use TR → L to make U1 ⇒ U1 = (L,R,T) & W = {U1(L,R,T)}
i = 2; Use TA → R to make U2 ⇒ U2 = (A,R,T) & W = {U1(L,R,T), U2(A,R,T)}
i = 3; Use LP → M to make U3 ⇒ U3 = (L,M,P) & W = {U1(L,R,T), U2(A,R,T), U3(L,M,P)}
i = 4; Use TP → R to make U4 ⇒ U4 = (P,R,T) & W = {U1(L,R,T), U2(A,R,T), U3(L,M,P), U4(P,R,T)}
i = 5; Use L → A to make U5 ⇒ U5 = (A,L) & W = {U1(L,R,T), U2(A,R,T), U3(L,M,P), U4(P,R,T), U5(A,L)}

Since U4(P,R,T) contains candidate key {P,T}, no more relation schema are created and we stop. Therefore, W is the decomposition of U into 3NF.