Yun Zhou
300442776

School of Engineering and Computer Science

SWEN304 **Database System Engineering**

Assignment 3

Due date: 23:59, **Monday 23 May**

The objective of this assignment is to test your understanding of functional dependencies, normal forms, database normalization, The assignment is worth **5%** of your final grade. It will be marked out of 100.

Submission instructions:

- Submit your assignment in **pdf** via the submission system

Note: Assignments not in **pdf** will incur a deduction of 3 marks.

Question 1. Normal Forms [16 marks]

Consider a relation schema N (R, F) where R = {A, B, C, D}. For each of the following sets F of functional dependencies, determine which normal form (1NF, 2NF, 3NF, BCNF) the relation schema N is in. Justify your answer.

Hint: Note that in all four cases AB is the only key for N.

1) $F = \{AB \rightarrow C, C \rightarrow D\}$

It is in the 2 NF. First, it is not 3NF since we can see that AB functionally defines and determines C and C functionally defines D, so in this case, D is transitively dependent on AB via C that AB determines D via C. Accordingly, it does not satisfy the condition of 3NF that non-prime attribute is transitively functionally dependent on any relation schema key. Since it is not 3NF, so it further illustrates that it is not BCNF, also the 2nd one C->D does not contain the key in LHS.

Then, it is 2NF because all non-prime attributes(i.e. C and D) are fully(not partially) functional dependent on the primary key. First, we can easily see AB determines C, then, for the D, as mentioned above, the key AB determines D via C transitively. Therefore, we can see all non-prime attributes are fully functionally dependent on the key AB, so it is 2NF.

2) $F = \{AB \rightarrow D, B \rightarrow C\}$

It is in the 1NF. It is not 2NF since 2NF requires all non-prime attributes(C and D) are fully functional dependent on the primary key. However, we can see that D satisfy the condition that it is fully depended on the key AB, but C is partially depended on the key, which is B only. Therefore, it is not 2NF, which further illustrates that it is not 3NF or BCNF. Hence, it is 1NF.

3) $F = \{AB \rightarrow C, AB \rightarrow D\}$

It is in the BCNF since AB is on the LHS of each FD in F where AB is the only relation schema key that satisfy the condition that LHS contains a relation schema key, and for each FD, they are all non-trivial functional dependency.

4) $F = \{AB \rightarrow CD, C \rightarrow B\}$

It is in the 3NF wrt F. First, it is not BCNF since the LHS of c -> B does not contains the relation schema key. Then, from both slides and geeksforgeeks, we can get that it is 3NF because for the 1st FD: AB -> CD, the left-hand-side AB is the super key of F,

and for the 2nd one: C -> B, the right-hand-side B is the part of key AB, which means B is the prime attribute, so, it is the in the 3NF.

Question 2. Functional Dependency [9 marks]

Consider a relation schema $N(R, F)$ where $R=\{A, B, C, D, E\}$ with the set of functional dependencies

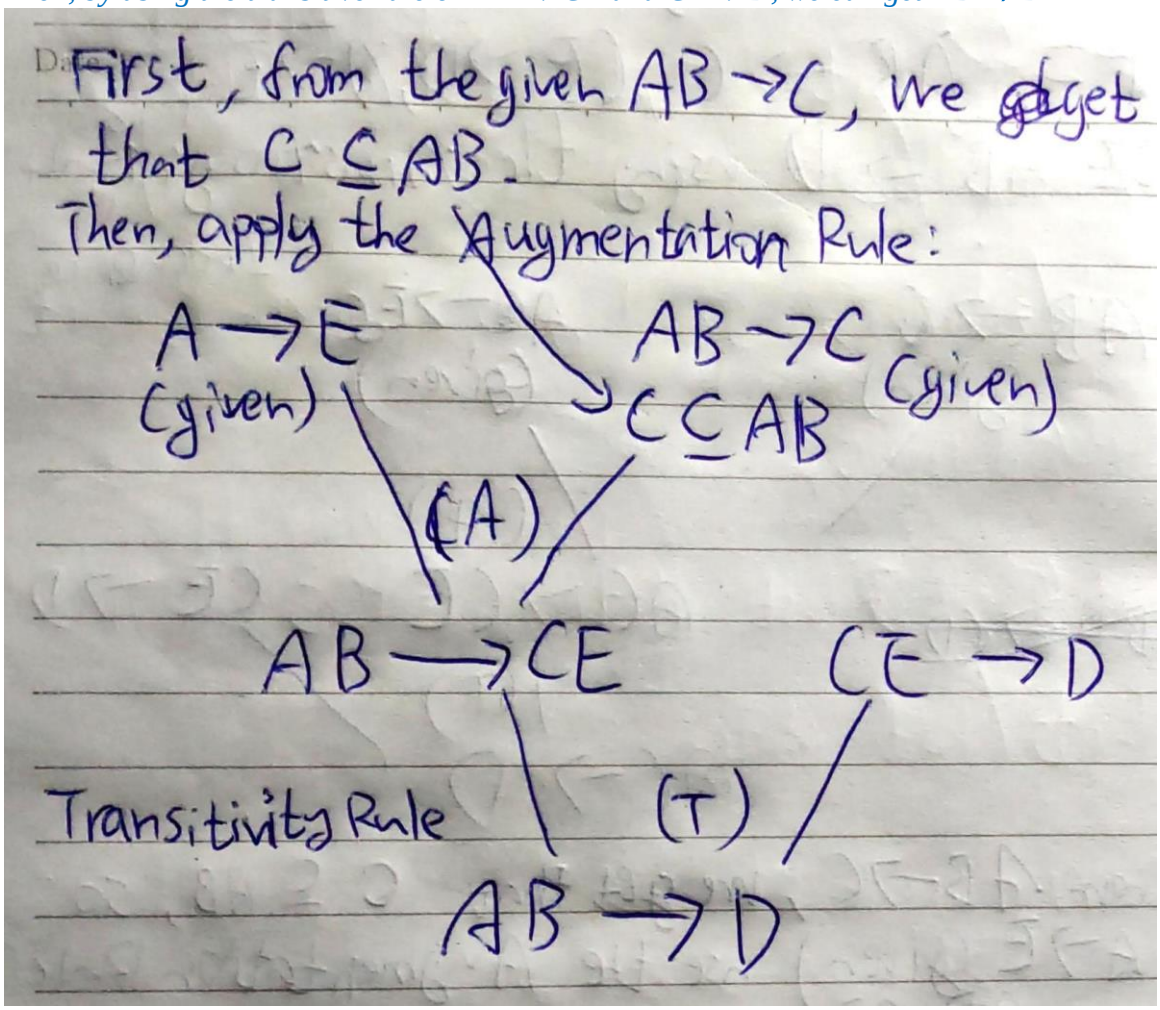
$$F = \{AB \rightarrow C, CE \rightarrow D, A \rightarrow E\}$$

Show that $AB \rightarrow D$ can be inferred from F using Armstrong's inference rules.

First, from $AB \rightarrow C$, we can get that $C \subseteq AB$.

So, take it and apply it with the Augmentation rule on the given $A \rightarrow E$, we can get $AB \rightarrow CE$

Then, by using the transitive rule on $AB \rightarrow CE$ and $CE \rightarrow D$, we can get ***$AB \rightarrow D$***



Question 3. Minimal Cover of a set of Functional Dependencies [20 marks]

Consider the set of functional dependencies $F = \{A \rightarrow B, B \rightarrow CD, D \rightarrow A, AC \rightarrow D\}$. Compute a minimal cover of F. Justify your answer.

Based on slides.

- Set $G = F = \{A \rightarrow B, B \rightarrow CD, D \rightarrow A, AC \rightarrow D\}$
- Apply the Decomposition Inference Rules

G1 = {A → B, B → C, B→D, D → A, AC→ D}

- Do Left Reduction

  Only the functional dependencies having more than one attribute on their LHS may be reduced, so we need to consider AC→ D.

  To test whether there is a superfluous attribute on the LHS, we try to remove each of the LHS attributes and apply attribute closure algorithm to see if the RHS still functionally depends on the remainder of the LHS.

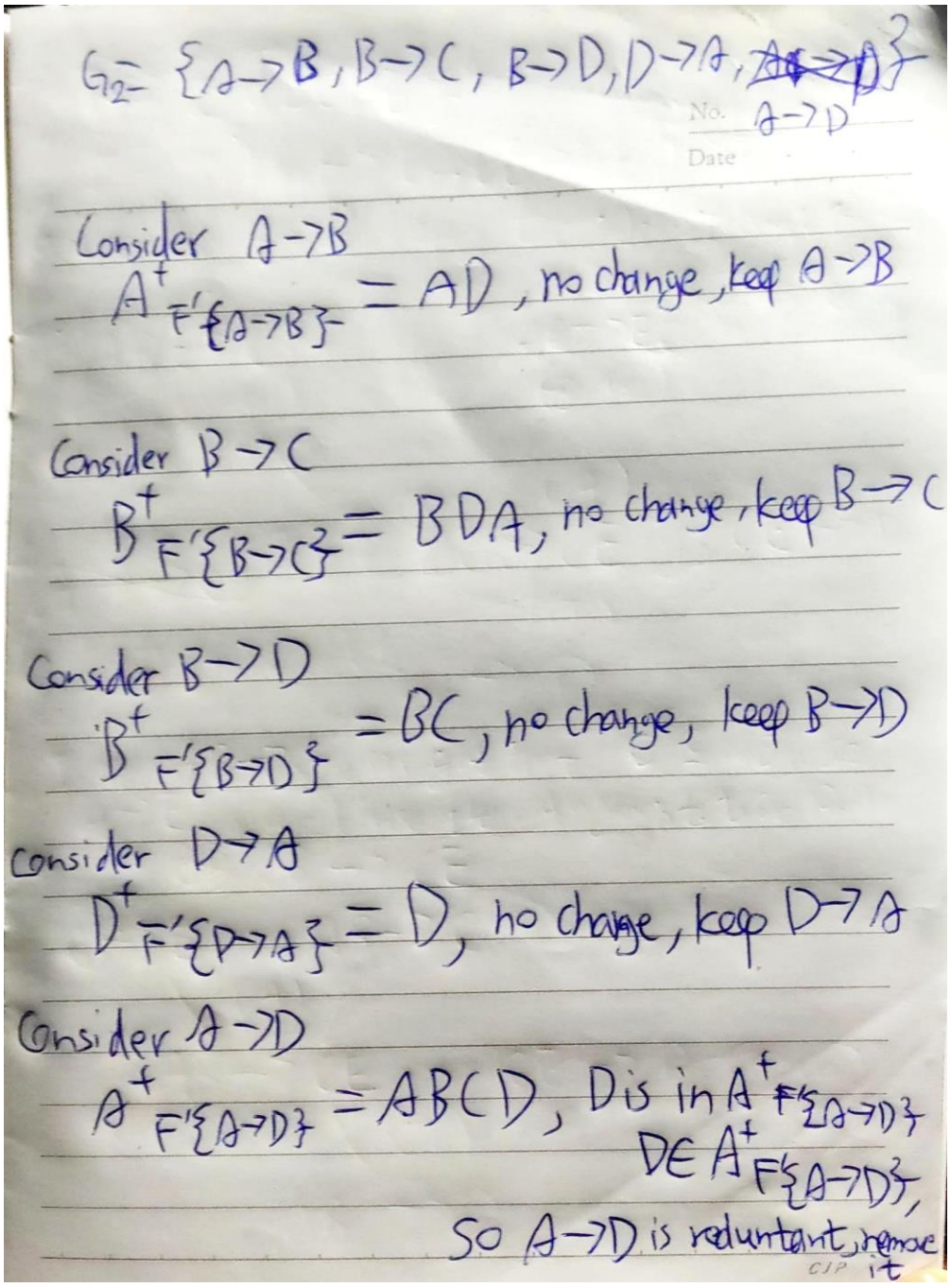
  (AC-A)+ = C+ = C, we cannot find any extraneous in there

  (AC-C)+ = A+ = ABCD, due to C is on the RHS, so C is extraneous and is removed. A→D

  Therefore, for this step, we can get the minimal cover:

  G2={A → B, B → C, B→D, D → A, A→ D}

- Eliminate all redundant FDs



Finally, we can see that A→ D is redundant, so remove it from G2.

Hence, the final minimal cover of *F*, G= {A → B, B → C, B→D, D → A}

Question 4. 3NF Normalization [25 marks]

Consider a relation schema $N(R, F)$ where $R = \{A, B, C, D\}$ and $F = \{A \rightarrow B, C \rightarrow D\}$. Perform the following tasks. Justify your answers.

1) [5 marks] Identify all keys for N. Show your process.

A+=AB

B+=B

C+=CD

D+=D

AB+=AB

AC+=ABCD

AD+=ABD

BC+=BCD

BD+=BD

CD+=CD

ABC+=ABCD

ABD+=ABD

ACD+=ABCD

BCD+=BCD

As we can see from above, AC, ABC and ACD covers all attributes so they can be seen as keys. Among them, AC should be the chosen key since it has minimal number of attributes.

2) [5 marks] Identify the highest normal form (1NF, 2NF, 3NF, BCNF) that N satisfies.

1NF is the highest normal form that N satisfies. It is because 2NF requires all non-key attributes are fully functional dependent on the key AC, but for both A->B and C->D, they are just partially dependent on the key AC, so the highest can only be the 1NF.

3) [10 marks] If N is not in 3NF, compute a lossless transformation into a set of 3NF relation schemas using the Synthesis algorithm.

1. Find a minimal cover G of F
   The given one is already the minimal cover,
   which is G={A->B, C->D}
2. Group FDs from G according to the same LHS, and make a schema for each group of FDs:
   For each Group of FDS: (A→B)
   (C→D)
   Relation Schema S = {({A, B }, {A }), ({C,D }, {C })}
3. check if any relation schema contains a key of schema U
   A + = AB
   C + = CD
   K (U, F) = AC
   create a relation schema that contains a Key
   S = {({A, B }, {A }), ({C, D }, {C }), ({A, C }, {AC })}

4. [5 marks] Verify explicitly that your result has the lossless property, satisfies 3NF, and that all functional dependencies are preserved.

- Attribute preservation & Lossless join decomposition
  By union all attributes, U=(A,B)U(C,D)U(A,C),we get{A,B,C,D}, which meet the attribute preservation as well as the lossless property.

- Functional dependency preservation
  As we can see from the union of F, FDs are: A -> B, C->D, A -> C, which preserve all Functional Dependencies.

- Satisfies at least third normal form (i.e. 3NF)
  As mentioned previously, we get R={A,B,C,D}, S={({A, B }, {A }), ({C, D }, {C }), ({A, C }, {AC })} with the set of FDs: { A -> B, C->D, A -> C }.
  For the 1st FD: A -> B, the LHS A is the super key of the relation {A, B}.
  For the 2nd one: C -> D, the LHS C is the super key of the relation {C, D},

For the 3rd one: A->C, the RHS C is the part of key AC, which means C is the prime attribute.
Hence, it is 3NF.

Question 5. BCNF Normalization [30 marks]

Consider a relation schema N (R, F), where R = {A, B, C, D} and F = {$A \rightarrow C$, $D \rightarrow B$, $BC \rightarrow A$, $BC \rightarrow D$}. Perform the following tasks. Justify your answers.

1) [5 marks] Identify all keys for N. Show process.

A+=AC

B+=B

C+=C

D+=BD

AB+=ABCD

AC+=AC

AD+=ADCB

BC+=BCAD

BD+=BD

CD+=CDBA

ABC+=ABCD

ABD+=ABCD

BCD+=ABCD

As we can see, there are lots of keys that cover all attributes, which are AB, AD, BC, CD, ABC, ABD and BCD. Among them, AB or AD or BC or CD should be the chosen key since it has minimal number of attributes.

2) [4 marks] Identify the highest normal form (1NF, 2NF, 3NF, BCNF) that N satisfies.

For this question, I choose BC as the schema key since it can let N satisfies the highest **3NF**.

First it's at least 2NF because all non-prime attributes(i.e. AD) are fully functional dependent on the key BC which means it is at least 2NF. Then, it is NOT the BCNF since A->C does not contain the relation schema key BC.

Then, as the slide mention, it is the 3NF since for each non-trivial FD X->A holds in F, either X is a superkey OR A is a prime attribute.

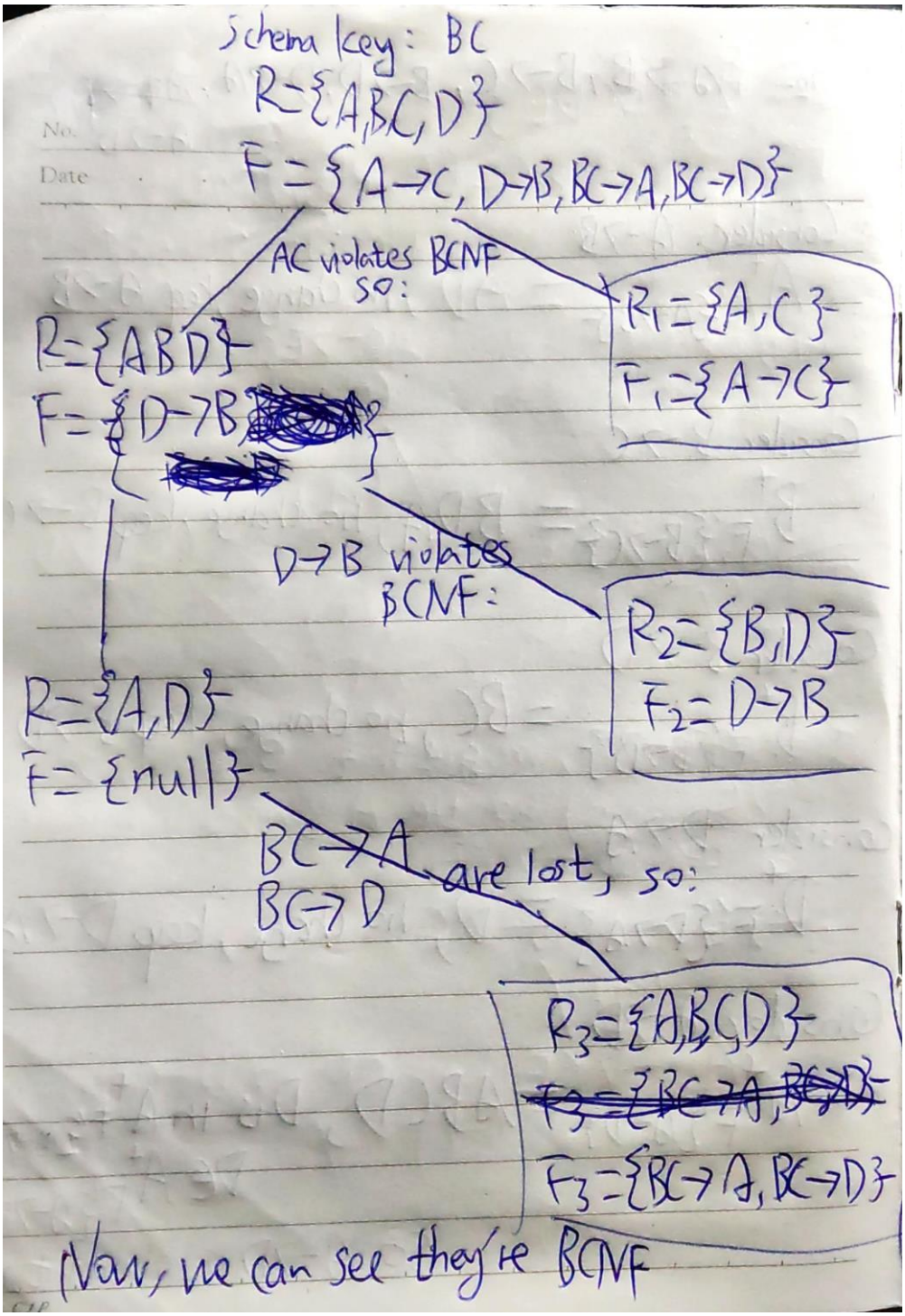
For A->C, the RHS C is the part of the key BC, which means it is the prime attribute.

For D->B, the RHS B is the part of the key BC, which means it is the prime attribute.

For BC->A and BC->D, they all have the same LHS BC which is the super key of N.

Hence, it is the 3NF.

3) [16 marks] If *N* is not in BCNF, compute a lossless decomposition into a set of BCNF relation schemas using the BCNF decomposition algorithm.



4) [5 marks] Verify explicitly whether your result satisfies BCNF, and all functional dependencies are preserved.

- Attribute preservation & Lossless join decomposition
  By union all attributes, U=(A,D)U(A,C)U(B,D)U(A,B,C,D),we get={A,B,C,D}, which meet the attribute preservation as well as the lossless property.

- Functional dependency preservation
  As we can see from the union of F, FDs are: A -> C, D->B, BC -> A, BC->D, which preserve all Functional Dependencies.

- Satisfies BCNF
  As we can see previously, we get:
  R = {A, D}, F = {null}
  R1 = {A,C} F1 = {A -> C}
  R2 = {B, D} F2 = {D -> B}
  R3 = {A,B,C,D} F3 = {BC -> A, BC -> D}

  For R1 F1, since A is the key of R1{A,C}, so, it is BCNF
  For R2 F2, since D is the key of R2{B,D}, so it is BCNF
  For R3 F3, since BC is the key of {A,B,C,D}, so it is the BCNF

  Hence, overall, we can see that they all satisfy BCNF, and also all FDs are preserved.

****