Database Systems, CSCI-GA.2433-011, New York University, Fall 2020

Instructor: Professor X. Sean Wang (x.sean.wang@nyu.edu)

Homework Assignment #4

Assignment Due time: Dec 14 **MONDAY, 2020 @ 23:00 (CST)**

There are **7 problems**. You need to submit **one file** to the Classes site. **Read the whole assignment carefully.**

What to SUBMIT: One **typed or a photocopy of handwritten (legibility is your responsibility!) in PDF file format**. **Naming for the file:** Assuming that your Net ID is abc123 and you are submitting your solution to Homework due 2034-02-15, your files should be named 20340215abc123.pdf, (of course you need to specify the correct date and the correct Net ID). **Repeated submission:** repeated submission on Classes site is allowed as long as it's before due time (and while the submission site remains open), and the **last submission** will be looked at for grading.

Team work with another student is allowed for this assignment. If the work is done collaboratively by two students, one of the students should submit the script file as above but with the name and netID of both students in the files (use # at the beginning of a line to add comments to script file), and the other student submit simply a text file, indicating the team work by listing the name and netID of both students. Note in this case, both students will receive the same score.

Special notes: (a) Unless specified otherwise, you have to show all of your work. We cannot look into your mind and determine your thinking process unless it is written out. You may refer to the numbers of the FDs to simplify your explanation. **(b)** Please start each problem on a new page. **(c)** In answers asking for an explanation, such as Problem 2, be very concise. Do not add unnecessary text as you may lose points for that. Address the question but not the topic. **(d)** You have to follow the procedures we covered in class and not any other procedures for what's asked for. You must use only the algorithmic techniques described starting with slide 7/100. Any time you compute a minimal cover, you have to show it is a minimal cover, unless stated otherwise.

Complete Procedure Review: For your convenience a summary is provided.

Input: A relational schema (informally, a relation) R and a set of FDs α.
Output: A set of relational schemas (informally, relations) such that the decomposition (a formal term for the set) satisfies the conditions

- It is lossless join
- It preserves dependencies
- The resulting tables are in 3NF.

Procedure:

1. Find a minimal cover for α, say ω
2. Produce a relation from each FD in ω by combining the attributes from both the LHS and the RHS
3. One by one, remove a relation that is a subset of another relation
4. If (at least) one of the relations contains a key of R, you are done

5. Otherwise, add a relation that is a key for R

Example 1. Assume that I am supposed to do the above for R = ABC with FDs

1. AB → C
2. A → B

I attempt to remove B from LHS of (1). The new set of FDs is

1. A → C
2. A → B

The new set can only be stronger than the old set to test for equivalence, I attempt to prove (1) of the new set from all of the old set. Doing that I get:

A + = ABC

and as I get the RHS of the new (1), the new set is equivalent to the old set.

Nothing else can be attempted, so my last set (here the second) is a minimal cover.

I now create relations

AB
AC

Nothing can be removed.

I now test whether at least one of the relations contains a key for R. Using my FDs, I compute

AB + = ABC.

I conclude that AB contains a key for R and I am finished.

Now 7 Problems for you to work on (10 points each with a total of 70 points):

1. **(10 points)** You are given the application described in slide 7/39. Do not assume any knowledge, such that the business rules are already in a nice format. You are just given the schema and the set of FDs that it satisfies. Follow the procedure in the above "**Complete Procedure Review**". Do not skip any steps. This means that you have to compute every closure of the sets of attributes that you need and you need to show that your minimal cover is indeed a minimal cover, as otherwise you do not know that it is a minimal cover.
2. **(10 points)** You are given a relational schema satisfying some FDs, among them A → B. Explain why ABC cannot be a key of the schema.
3. **(10 points)** Using the heuristics described on slides 7/134–7/135 and providing full details and explanations, compute all the keys of R = ABCDEFGH given the FDs
    - A → AB
    - A → G
    - C → A
    - D → G
    - F → GH
    - H → AF
4. **(10 points)** You are given a relational schema R = ABC, satisfying the following FDs
    - A → B

- A → C
- B → C

(a) Which normal forms are satisfied? **(Ignore 2NF)**

(b) A decomposition into AB and AC is proposed.

- Which normal forms are satisfied by AB and AC?
- Show by a small example that the dependencies are not preserved.

5. **(10 points)** You are given a relational schema R = ABC, satisfying the following FDs
   - A → B
   - A → C
   - B → C

   (a) Which normal forms are satisfied? **(Ignore 2NF)**

   (b) A decomposition into AC and BC is proposed.
   - Which normal forms are satisfied by AC and BC?
   - Show by a small example that the decomposition is not a lossless-join decomposition.
6. **(10 points)** You are given a relational schema R = ABCDEFGH, satisfying α, the set of FDs
   - A → D
   - A → E
   - AG → H
   - BF → H
   - C → H
   - E → D
   - F → B

   (a) Compute ω, a minimal cover for α.

   Follow the procedure in the above "**Complete Procedure Review**". Do not skip any steps. This means that you have to compute every closure of the sets of attributes that you need and you need to show that your minimal cover is indeed a minimal cover, as otherwise you do not know that it is a minimal cover.
7. **(10 points)** You are given a relational schema R = ABCDEFGH, satisfying α, the set of FDs
   - AB → CD
   - A → E
   - B → FG
   - E → AH

   You are told that α is a minimal cover, and you may believe that. (a) Find a decomposition satisfying the conditions in the above "**Complete Procedure Review**".