DASC 5333 Database Systems for Data Science
CSCI 4333 Design of Database Systems
Fall 2023
Homework #7

Functional Dependency and Normalization Analysis

[1] (25%) Conduct normalization analysis on GW (HW #3 and #4) by listing the FDs and the highest normal form of each relation. You must use h7q1_template.docx (which is based on the suggested solution of HW #4). Complete the task by filling in the rows "Normalization Analysis."

[2] (25%) List the candidate keys and the highest normal forms for the following relations.

[a] R(A,B,C,D) {D->C, C->B}
[b] R(A,B,C,D) {AB->C, C->D}
[c] R(A,B,C,D) {A->B, B->ACD}
[d] R(A,B,C,D) {AB->C, AD->C}
[e] R(A,B,C,D) {A->B, B->A, AC->D}

[3] (25%) (25%) Consider the following relation GO:

GO(GroupId, GroupName, GroupEMail, GroupChairId, GroupChairLName, GroupChairFName, GroupMemberId, GroupMemberMajor)

The relation stores information about student groups, their chairpersons and members. Chairpersons and members are students with unique student ids (stored as values in GroupChairId and GroupMemberId respectively). GroupId uniquely identifies a group. A group has a unique name, and an email address (that may not be unique.) A group has a single chairperson but a student can serve as a chairperson of multiple groups. For example, four tuples are shown below.

GroupId	GroupName	GroupEMail	GroupChairId	GroupChairLName	GroupChairFName	GroupMemberId	GroupMemberMajor
G1	Biology	bio@uhcl.edu	12345	Lee	Bryan	23323	BIOL
G1	Biology	bio@uhcl.edu	12345	Lee	Bryan	24990	BIOL
G1	Biology	bio@uhcl.edu	12345	Lee	Bryan	38879	PHYS
G1	Biology	bio@uhcl.edu	12345	Lee	Bryan	38879	CSCI

Bryan Lee is the chair student of the group G1 Biology. The four tuples also store information about three members. Note that the member 38879 has dual majors in PHYS and CSCI.

(a) List all applicable functional dependencies. (Make reasonable assumptions if necessary.)

(b) What are the candidate keys?

(c) What is the highest normal form? Why?

(d) If the highest normal form is not BCNF, can you decompose the relation TD losslessly into component relations in BCNF while preserving functional dependencies? If yes, how. If no, why?

[4] (20%) Consider the following relation

R(A,B,C,D,E) {A->B, AB->D, AD->E, C->D}

(a) Show all candidate keys.

(b) What is the highest normal form (up to BCNF)? Why?

(c) If it is not in BCNF, can you losslessly decompose R into component relations in BCNF while preserving functional dependencies?

[5] (5%) It is known that for R(A,B,C,D,E):

1. R has exactly two candidate keys
2. A is a candidate key.
3. D and E are non-prime attributes.

How many superkeys can R have?

Your solution filename should be h7_<<Your name>>_<<Your Student ID>>.docx. Submit your homework through Blackboard.