DASC 5333 Database Systems for Data Science
CSCI 4333 Design of Database Systems
Fall 2024
Homework #7 Normalization Theory

[1] (20%) Conduct normalization analysis on LabPro (HW #2 and #3) 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 #3). Complete the task by filling in the row "Normalization Analysis".

[2] (10%) Use Armstrong's axioms and rules to prove that

F = {A->B, AB->C, CD->E, AE->F} |- AD->F

[3] (20%) List the candidate keys and the highest normal forms for the following relations.

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

[4] (20%) Consider R(P, Q, R, S, T, U) with

F = {P->Q, R->S, PQ->ST, U->RS, UP->S}

(a) What are P+, Q+, R+, S+, T+, U+?

(b) What are the candidate keys?

(c) Show all prime attributes and non-prime attributes.

(d) Provide a canonical cover of F.

(e) What is the highest normal form (up to BCNF) of R? Why?

(f) If R is not in BCNF, provide a lossless FD preserving decompositions of R into BCNF relations?

[5] (25%) Consider the following relation Employee:

Employee(EId, EFName, ELName, EPhone, DeptId, DeptName, PositionId, RateScale)

The relation stores information about employees. Each row stores information about an employee.

EId	EFName	ELName	EPhone	DeptId	DeptName	PositionId	RateScale
E1	Mary	Gonzalez	832-811-2207	D639	IT	P23	4
E2	Joe	Gonzalez	281-222-1111	null	null	P23	4
E2	Joe	Gonzalez	832-255-9090	null	null	P23	4
E3	Jackie	Johnson	281-222-1124	D639	IT	P38	5
E4	Bryan	Williams	201-222-1922	D291	Accounting	P67	5
E4	Bryan	Williams	201-222-1922	D639	IT	P67	5

For example, the first row stores the information about the employee Mary Gonzalez, her EId (E1), and phone (832-811-2207). An employee can work for more than one department. In the case of Mary Gonzalez, she works only for department D639, with the department name 'IT'. Her position's Id is P23 (which may mean 'Summer Super Intern'). An employee has one positionId, even if she works for many departments. The rate scale of Mary is 4 (which may mean that the salary range is from $50,000 to $60,000). The rate scale of a position does not depend on the department. A position (positionId) always have the same rate scale no matter who the employee is.

An employee may have multiple phones (e.g. Joe Gonzalez has two phones) and an employee may work for multiple departments (e.g. Bryan Williams work for departments D291 and D639)

(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 the relation Employee be decomposed losslessly into component relations in BCNF while preserving functional dependencies? If yes, how? If not, why?

[6] (5%) It is known that for R(A,B,C,D,E)

1. There are three prime attributes: A, B and C.
2. There are two non-prime attributes: D and E.

What are the maximum and minimum number of superkeys R may have?

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