

**DASC 5333 Database Systems for Data Science**  
**CSCI 4333 Design of Database Systems**  
**Spring 2024**  
**Homework #7**

### Functional Dependency and Normalization Analysis

[1] (20%) Conduct normalization analysis on TR (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] (10%) Use Armstrong's axioms and rules to prove that

$$F = \{A \rightarrow B, CD \rightarrow E, AC \rightarrow D, E \rightarrow F\} \vdash AC \rightarrow F$$

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

[a]  $R(A,B,C,D)$   $\{B \rightarrow C, C \rightarrow DA\}$

[b]  $R(A,B,C,D)$   $\{B \rightarrow C, BC \rightarrow DA\}$

[c]  $R(A,B,C,D)$   $\{B \rightarrow C, A \rightarrow C, AC \rightarrow D\}$

[d]  $R(A,B,C,D)$   $\{B \rightarrow C, C \rightarrow AB, AB \rightarrow D\}$

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

$$F = \{RS \rightarrow PQ, QR \rightarrow S, S \rightarrow P, U \rightarrow ST, STU \rightarrow R, Q \rightarrow U\}$$

(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) Give 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, can you provide a lossless FD preserving decompositions of  $R$  into BCNF relations?

[5] (25%) Consider the following relation Tutor (which stands for tutoring assignments):

Tutor(TutorId, TLName, TFName, StudentId, SLName, SFName, SubjectId, SubjectName, StartDate, Level).

A tuple in Tutor may be:

(`'t1'`, `'Fong'`, `'Jenny'`, `'st11'`, `'Paul'`, `'Smith'`, `'su21'`, `'Data Science'`, `'2024-05-13'`, 2),

meaning that the tutor with id `'t1'` and name `'Jenny Fong'` is assigned as the tutor for the student with id `'st11'` and name `'Paul Smith'` on the subject `'Data Science'` (with the subject id `'su21'`) starting on 2024-05-13 at Level 2.

It is known that `TutorId`, `StudentId` and `SubjectId` are unique identifiers for tutors, students and subjects respectively. Furthermore, `StartDate` is an optional field (the only optional field).

Furthermore, no two `subjectId` should be associated with the same subject name. For example, `Data Science` uses the `subjectId` of `su21` only. A tutor can tutor many students, and a student can have many tutors. However, given a specific subject for a specific student, the relation stores all tutoring information, current or past. For example, the table may also contain:

(`'t1'`, `'Fong'`, `'Jenny'`, `'st11'`, `'Paul'`, `'Smith'`, `'su21'`, `'Data Science'`, `'2023-11-2'`, 1),  
and

(`'t9'`, `'Lopez'`, `'Jane'`, `'Paul'`, `'Smith'`, `'su21'`, `'Data Science'`, `'2023-11-5'`, 1)

In this case, Paul Smith has three tutoring assignments on `Data Science`. Jenny Fong served as his tutor in levels 1 and 2. Jane Lopez serves as his tutor in level 1.

(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?

[6] (5%) It is known that  $R(A,B,C,D,E)$  has exactly two candidate keys. Furthermore, one of the candidate key is known to be  $AB$ . 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.