CSE4701 - Databases
HW3 - 2020-04-07

Problem 1

Prove or disprove that the following inference rules for functional dependencies. A proof can be made by using inference rules IR1 through IR3. A disproof should be done by showing a relational instance that refutes the rule. (20 points)

a. If {A → B, C → D}, then {AC → BD}
    1. If A → B, then AC → BC (IR2 Augmentation)
    2. If C → D, then BC → BD (IR2 Augmentation)
    3. If AC → BC, and BC → BD, then AC → BD (IR3 Transitive)

b. If {AB → C, C → D}, then {A → D}
    - Counterexample of this would be the following:
        - A = NAME, B = SSN, C = DEPT_# , D = DEPT_NAME
        - If SSN, NAME → DEPT_# , and DEPT_# → DEPT_NAME
        - NAME does not → DEPT_NAME

Problem 2

Show that AB → D is in the closure of {AB → C, CE → D, A → E} (10 points)

1. If A → E, then AC → CE (IR2 Augmentation)
2. If AB → C, and AC → CE, then ABA → CE (IR6 Pseudo Transitivity)
3. ABA → CE = AB → CE *Union of Attribute A twice, same as once
4. If AB → CE, and CE → D, then AB → D (IR3 Transitivity)

Problem 3

Consider the relation schema R = {A, B, C, D, E, F, G, H, I, J} and the set of functional dependencies F = {AB → C, A → DE, B → F, F → GH, D → IJ } that hold true for R. Determine whether each of the decompositions given below has (1) the dependency preservation property, and (2) the lossless join property, with respect to F. Also determine which normal form each relation in the decomposition is in. (20 points)

a. D_1 = {R1, R2, R3} where R_1 = {A, B, C, D, E}, R_2 = {B, F, G, H}, and R_3 = {D, I, J},
    - Dependency Preservation Property
        - Based on the definition, the decomposition D_1 is indeed dependency-preserving with respect to F, since each of the functional dependencies here are satisfied
        - AB → C using R_1
        - A → DE using R_1
        - B → F using R_2
        - F → GH using R_2
        - D → IJ using R_3
    - Lossless Join Property
        - Using the following table, which represents a natural join between all R_i (R_1, R_2 … etc.), we can determine if it is lossless or lossy by finding out if a row contains all columns filled with the "α". Cells that have a green background represent "α" that have been added in afterwards that aren't

found in the set of functional dependencies using that specific decomposition

	A	B	C	D	E	F	G	H	I	J
R_1	α	α	α	α	α	α	α	α	α	α
R_2		α				α	α	α		
R_3				α					α	α

Given we row R_1 contains "α" in all columns, it is determined that this is a lossless decomposition

b. D_2 = {R1, R2, R3, R4, R5} where R_1 = {A, B, C, D}, R_2 = {D, E}, R_3 = {B, F}, R_4 = {F, G, H}, and R_5 = {D, I, J}.

- Dependency Preservation Property
  - Based on the definition, the decomposition D_2 is not dependency-preserving with respect to F, one of the functional dependencies that is not preserved is provided:
  - A → DE cannot be determined using R_1 or R_2 or any projections
- Lossless Join Property

	A	B	C	D	E	F	G	H	I	J
R_1	α	α	α			α	α	α		
R_2										
R_3		α				α	α	α		
R_4						α	α	α		
R_5				α					α	α

Given that none of the rows contain all the "α" variables, it is determined that this to be a lossy decomposition.

Problem 4

Consider the relation schema R = {A, B, C, D, E, F} and the set of functional dependencies F = {ABC → D, D → E, BC → E, D → F} hold in R. Assume all domains of the attributes in R is atomic. (20 points)

a. Can you decompose R into R_1 and R_2 in which this decomposition is lossless and R_2 is 1NF at best and R_1 is BCNF?

- R_1 = {D, E, F}
  - R_1 is in BCNF if it is in 3NF and all functional dependencies applicable to this relation have the left hand side as superkey of the relation. D → E and D → F are the only applicable FDs, D^+ = {D, E, F}, making it a candidate key, or super key. Therefore, R_1 is in BCNF
- R_2 = {A, B, C, D, E}

- ○ R_2 is in 1NF, and has candidate key ABC, $ABC^+ = \{A, B, C, D, E\}$. However BC → E is also a FD, meaning it has practical dependency, R_2 is not in 2NF with atomic values in domain, meaning it is in 1NF

b. Now, can you decompose R_2 into R_3 and R_4 in which this decomposition is lossless and both R_3 and R_4 are BCNF?

- $R_3 = \{A, B, C, D\}$
  - The candidate key in R_3 is $ABC^+ = \{A, B, C, D\}$, with applicable FD ABC → D. Left hand side of all applicable FDs is a superkey on R_3, this makes R_3 in BCNF
- $R_4 = \{B, C, E\}$
  - Candidate key in R_4 is $BC^+ = \{B, C, E\}$, with applicable FD BC → E. Left hand side of all applicable FDs is a superkey on R_4, this makes R_4 in BCNF

R : $R_1 = \{D, E, F\}$, $R_3 = \{A, B, C, D\}$, $R_4 = \{B, C, E\}$