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

[1] See h7q1_sol.pdf.

[2] Proof of F = {A->B, AB->C, CD->E, AE->F} |- AD->F

[1] A->B (given)
[2] AB->C (given)
[3] AA->C (pseudo-transitivity on [1] and [2])
[4] A-> C (simplification of [3])
[5] CD -> E (given)
[6] AD ->E (pseudo-transitivity on [4] and [5])
[7] AE -> F (given)
[8] ADA -> F (pseudo-transitivity on [6] and [7])
[9] AD -> F (simplification of [8])

[3]

[a] R(A,B,C,D) {C->D, AC->B}

CK:[1] AC
Canonical Cover (optional): {C->D, AC->B}; as is.
Highest NF: 1NF
Reason: C-> D violates 2NF since C is a proper subset of a CK, and D is non-prime.

[b] R(A,B,C,D) {C-> AD, D->AB, A->B, B->A}

CK:[1] C
Canonical Cover (optional): {C->D, D->A, A->B, B->A}
Highest NF: 2NF
Reason: The FD D->A, A->B, and B->A violate 3NF. The LHS are not superkeys and the RHS are non-prime attributes.

[c] R(A,B,C,D) R(A,B,C,D) {C->BD, BD->AC}

CK: [1] C, [2] BD
Canonical Cover (optional): {C->BD, BD->AC}; as is.
Highest NF: BCNF
Reason: The LHS of all non-trivial FD are superkeys.

[d] R(A,B,C,D) {C->B, B->C, BD->A, CD->A}

CK:[1] BD, [2] CD
Canonical Cover (optional): {C->B, B->C, BD->A}
Highest NF: 3NF
Reason: The FD C->B and B->C violate BCNF as B and C are both not superkeys.

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

(a)

P+=PQST
Q+= Q
R+ = RS
S+ = S
T+ = T
U+ = UPSR

(b) CK: [1] UP

L/NR: P, U
M: Q, R
R: S, T

(c) Prime: U,P; non-prime: Q, R, S, T

(d) Canonical Cover: {P->QST, R->S, U->R}

there are many, for examples, {RS-> Q, Q->RSTU, S-> P, U->Q}

(e) Highest NF: 1NF. S->P violates the 2NF.

(f) The following lossless decomposition results in all relations in BCNF.

R1(P,Q,R,S) {P->QST}
R2(R,S) {R->S}
R3(U,R) {U->R}
R4(P,U) {}

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

[a] Functional Dependencies (as in a canonical cover):

1. Eid -> EId, EFName, ELName, PositionId
2. DeptId -> DeptName
3. DeptName -> DeptId (likely)
4. PositionId -> RateScale

[b] The CKs are

1. EId, DeptId, EPhone
2. EId, DeptName, EPhone

[c] Thus, Highest normal form is 1NF as Eid -> EId, EFName, ELName, PositionId violate 2NF, for example.

[d] Decomposition:

Employee(<u>EId</u>, EFName, ELName, PositionId)

1. FD: [1] Eid -> EId, EFName, ELName, PositionId
2. FK: [1] PostionId references Position(PositionId)
3. Highest NF: BCNF

EmployeePhone(<u>Eid, EPhone</u>)

1. FD: {}
2. FK: [1] Eid references Employee(Eid)
3. Highest Normal Form: BCNF

Department(<u>DeptId</u>, DeptName)

1. FD: [1] DeptId -> DeptName, [2] DeptName -> DeptId
2. FK: {}
3. Highest NF: BCNF

WorkFor(DeptId, Eid)

1. FD: {}
2. FK: [1] DeptId references Department(DeptId), [2] Eid references Employee(Eid)
3. Highest NF: BCNF

Position(<u>PositionId</u>, PayScale):

1. FD: [1] PositionId -> PayScale
2. FK: {}
3. Highest NF: BCNF

[6] Minimum: 4, when there is only one candidate key: ABC.

Maximum: 28 when there are three candidate keys: [1] A, [2] B and [3] C.