**CSC 355 Database Systems**

**Fall Quarter 2018**

**Assignment 4 (10/3)**

**Due Sunday, October 14th @ 11:59pm.**

**Reading:** Chapter 15 ( Sec 1: Database Design Rules; Sec 2: Functional Dependencies; Sec 3 & 4: 1NF, 2NF, and 3NF; Sec 5: BCNF; (skip 4NF and 5NF),

Chapter 16 (Sec 1: Canonical Cover, Sec 2 & 3: Testing for Lossless Join Property and Dependency Preservation),

Chapter 21: Serializability and Transactions

Please note that there is a total of 6 questions.

**1. Canonical Cover Algorithm (11 pts).**

Given R(ABCD),

A. Find a canonical cover for the following set F of functional dependencies.

A🡪BC

AB🡪D

C🡪AD

D🡪B

Remove extraneous attributes

AB🡪D

{A}+ = BCAD

{B}+ = B

{A}+ has B so we can rewrite as A🡪D

Remove redundant

A🡪BC

{A}+ = ADB, yes we need this FD

A🡪D

{A}+ = BCAD, no need for this FD

C🡪AD

{C}+ = C, yes we need this FD

D🡪B

{D}+ = D, yes we need this FD

Final Canonical Cover

A🡪BC

C🡪AD

D🡪B

No LHS match others so no union needed.

B. Find all candidate keys.

{A}+ = ABCD

{C}+ = ABCD

{D}+ = B

A and C are candidate keys.

**2. Dependency Preservation (15 pts).**

For the universal relation R(w,x,y,z), consider the decomposition D consisting of R1(w,y,z) and R2(x,y), and the set F of functional dependencies { y🡪xz ; yz🡪w ; x🡪w }.

Recall that the projection of set of functional dependences G on relation Rx consists of every functional dependency in (G)+ that contains only attributes from Rx.

**a.** Compute the projection of F on R1.

yz🡪w

**b.** Compute the projection of F on R2.

None

**c.** Does the decomposition D preserve the set of dependencies F? Why or why not?

No. When taking the union of the projections we do not get the original set of FD. We’re missing y🡪xz and x🡪w.

**3. Nonadditive (Lossless) Join Property (24 pts).**

R = {ABCDEFGHIJ}

F = { AB -> C

A -> DE

B -> F

F -> GH

D -> IJ

}

Determine whether each decomposition into R1 and R2 has the lossless join property. Be sure to state why or why not.

**a.** R1(ABC) and R2(ABDEFGHIJ)

|  |  |
| --- | --- |
| Proj(R1) | Proj(R2) |
| AB -> C | A -> DE |
|  | B -> F |
|  | F -> GH |
|  | D -> IJ |

Yes, it has the lossless join property because all FDs are preserved with a union.

**b.** R1(ADE) and R2(EBCFGHIJ)

|  |  |
| --- | --- |
| Proj(R1) | Proj(R2) |
| A -> DE | B -> F |
|  | F -> GH |

No, it does not have the lossless join property. AB -> C and D -> IJ are missing when unioned.

**c.** R1(BFDIJ) and R2(ABCDEFGH)

|  |  |
| --- | --- |
| Proj(R1) | Proj(R2) |
| B -> F | AB -> C |
| D -> IJ | A -> DE |
|  | F -> GH |

Yes, it has the lossless join property because all FDs are preserved with a union.

**d.** R1(BDFGHIJ) and R2(ABCDE)

|  |  |
| --- | --- |
| Proj(R1) | Proj(R2) |
| B -> F | AB -> C |
| F -> GH | A -> DE |
| D -> IJ |  |

Yes, it has the lossless join property because all FDs are preserved with a union.

**4. Normalization and Decomposition (20 pts).**

R(ABCDEG)

FDs {AB=>C, AC=>B, AD=>E, B=>D, BC=>A, E=>G}

**a.** Identify candidate keys of R. Hint: There are three candidate keys.

{AB}+ = ABCDEG

{AC}+ = ABCDEG

{AD}+ = ADEG

{B}+ = DB

{BC}+ = ABCDEG

{E} + = G

AB, AC, and BC are candidate keys.

**b.** Construct a decomposition of R into relations in 2NF.

R1(A, B, C)

R2(B, D)

R3(A,D, E, G)

**c.** Construct a decomposition of R into relations in 3NF.

R1(A,B, C)

R3(B,D)

R4(A,D,E)

R5(E, G)

**d.** Are all of the relations in your decomposition in BCNF? Either explain why they are, or identify one that is not and explain why it is not. (Note that for a relation to be in BCNF, the determinants of all functional dependencies in the relation must be superkeys *of that relation* – not superkeys of the original universal relation.)

No they are all not in BCNF form. For example, R5(E, G) violates BCNF because the LHS of E->G is not a super key.

**5. Transactions and Serialization (15 pts).**

Consider the three transactions T1, T2, and T3, and the schedules S1 and S2 given below. Draw the serializability (precedence) graphs for S1 and S2, and state whether each schedule is serializable or not. If a schedule is serializable, write down the equivalent serial schedule(s).

T1: R1(X), R1(Z), W1(X)

T2: R2(Z), R2(Y), W2(Z), W2(Y)

T3: R3(X), R3(Y), W3(Y)

S1: R1(X), R2(Z), R1(Z), R3(X), R3(Y), W1(X), W3(Y), R2(Y), W2(Z), W2(Y)

|  |  |  |
| --- | --- | --- |
| T1 | T2 | T3 |
| R1(X) |  |  |
|  | R2(Z) |  |
| R1(Z) |  |  |
|  |  | R3(X) |
|  |  | R3(Y) |
| W1(X) |  |  |
|  |  | W3(Y) |
|  | R2(Y) |  |
|  | W2(Z) |  |
|  | W2(Y) |  |

Equivalent Serial Schedule: T3, T1, T2

S2: R1(X), R2(Z), R3(X), R1(Z), R2(Y), R3(Y), W1(X), W2(Z), W3(Y), W2(Y)

|  |  |  |
| --- | --- | --- |
| T1 | T2 | T3 |
| R1(X) |  |  |
|  | R2(Z) |  |
|  |  | R3(X) |
| R1(Z) |  |  |
|  | R2(Y) |  |
|  |  | R3(Y) |
| W1(X) |  |  |
|  | W2(Z) |  |
|  |  | W3(Y) |
|  | W2(Y) |  |

Not serializable because T2 reads Y and does not modify it before T3 reads and modifies Y, and T2 modifies Y afterwards

**6. Regular Expressions/SQL (15 pts).**

Write SQL for the following problems on the university database.

**A.** Find students whose name is properly capitalized (both first and last name start from an upper-case letter, rest of the name is lower-case). Hint: you can use [A-Z] to represent any upper case letter in regular expressions.

SELECT \*

FROM STUDENT

WHERE REGEXP\_LIKE(FirstName,'^[A-Z][a-z]+$')

AND REGEXP\_LIKE(LastName, '^[A-Z][a-z]+$');

**B.** Find course titles that consist of exactly 3 words (e.g., “Theory of Computation”).

SELECT \*

FROM Course

WHERE REGEXP\_LIKE(CourseName, '^\w+\s\w+\s\w+$');

**C.** Find students whose SSN numbers have a pattern of 321 repeated twice (sequentially).

SELECT \*

FROM STUDENT

WHERE REGEXP\_LIKE(SSN, '(321){2}');