CS4221: Database Design

Tutorial 2: The Relational Model - Normalization

12 February 2015

Note: Due to the time constraint, we will only discuss some of the questions.

1. Describe the criteria for choosing the primary key for a given relation. How to modify the set of relations generated by Bernstein's Algorithm by choosing the primary keys for relations which have more than one key?

2. Consider the following two relational schemas:

   Schema 1: R (A, B, C, D)

   Schema 2: R1 (A, B, C), R2 (B,D)

   (a) Consider Schema 1 and suppose that the only functional dependencies that hold on the relations in this schema are $A \rightarrow B$, $C \rightarrow D$. Is relation R in Boyce-Codd Normal Form (BCNF)?

   (b) Consider Schema 2 and suppose that the only functional dependencies that hold on the relations in this schema are $A \rightarrow B$, $A \rightarrow C$, $B \rightarrow A$, $A \rightarrow D$. Are relations R1 and R2 in BCNF?

   (c) Suppose we omit dependency $A \rightarrow D$ from part (b). Are relations R1 and R2 in BCNF?

   (d) Consider Schema 1 and suppose that the only functional and multivalued dependencies that hold on the relations in this schema are $A \rightarrow BC$, $B \rightarrow D$, $B \twoheadrightarrow CD$. Is relation R in Fourth Normal Form (4NF)?

   Hint: Find keys of relations and use Coalescence Rule.

3. Show whether the following 3 decompositions are lossless and dependency preserving decompositions. Show the keys of the relations. Justify your answers.

(a) Let R (A, B, C, D, E) be a relation with the following FDs:

$A \rightarrow BC$
$CD \rightarrow E$
$B \rightarrow D$
$E \rightarrow A$

Decompose it into:

R1 (A, B, C)
R2 (A, D, E)

(b) Let R (A, B, C, D) be a relation with the following FDs:

$A \rightarrow B$
$B \rightarrow C$
$C \rightarrow D$
$D \rightarrow A$

Decompose it into:

R1 (A, B)
R2 (B, C)
R3 (C, D)

(c) Let R (A, B, C, D) be a relation with the following FD and MVD:

$A \twoheadrightarrow B$
$CD \rightarrow A$

Decompose it into:

R1 (A, B)
R2 (A, C, D)

Hints:
1. The 4NF Decomposition Theorem (Slide 46).
2. The following is a sufficient condition which can guarantee the losslessness of a binary decomposition:

A decomposition {R1, R2} of R is a lossless-join decomposition if
$R1 \cap R2 \rightarrow R1$ or $R1 \cap R2 \rightarrow R2$
which means $R1 \cap R2$ forms a superkey of either R1 or R2.

4. Consider a relation R (A, B, C, D, E) with the following FDs:

$A \rightarrow B$,
$A \rightarrow C$,
$BD \rightarrow A$

(a) What are the keys of R?

(b) Which FDs violate BCNF (which FDs make R unable to be BCNF)?

(c) Which FDs are 3NF violations (which FDs make R unable to be 3NF)?

(d) Decompose R into:

R1 (A, B)
R2 (A, C, D, E)

Which FD cannot be preserved?

5. Give a data structure for storing functional dependencies in order to implement the FD membership test problem efficiently.

6. Give an algorithm to implement step 1 of Bernstein's Algorithm efficiently.

   Hint: Let $ABC \rightarrow D$ be a functional dependency in G. Are the below 3 statements correct?

   (a) Attribute A is an extraneous in the FD $ABC \rightarrow D$ if and only if $BC \rightarrow ABC \in G^+$?

   (b) Attribute A is an extraneous in the FD $ABC \rightarrow D$ if and only if $BC \rightarrow D \in (G - \{ABC \rightarrow D\})^+$?

   (c) Attribute A is extraneous in the FD $ABC \rightarrow D$ if and only if $B \rightarrow A \in G^+$?

7. Give an algorithm to implement step 4 (Merge equivalent keys) of Bernstein's Algorithm efficiently.

   Hint: How to represent the FDs in **J** efficiently, especially when there are many groups can be merged?