31/10/17

Database management Systems

DHRUV GARG
16BCE1190

MULTIVALUED DEPENDENCY AND FOURTH NORMAL FORM

To keep a relation state consistent and to avoid any spurious relationship between two independent attributes, we must have a separate tuple to represent every combination for a record in the relation. If a relation is an all-key relation (with key made up of all ~~relation~~ attributes) and therefore has no f.d's and as such qualifies to be a BCNF relation. But, in some cases there is an obvious redundancy in the relation.

To address the above situation, we have the concept of multivalued dependency (MVD).

- **Multivalued dependencies:**

If the column headings in a relational database table are divided into three disjoint groupings X, Y and Z, then, in context of a row, we can refer to data beneath each group of headings as x, y, z. A multivalued dependency $X \twoheadrightarrow Y$ signifies that if we choose any x actually occuring in the table, and compile a list of all $x_c yz$ combinations that occur in the table, we will find that x_c is associated with same y entries regardless of z. So essentially the presence of z provides no useful information to constrain the possible values of y.

To illustrate normalization to 4NF, we take the example as given below.

EXAMPLE

Restaurant	Pizza variety	Delivery area
A1 Pizza	Thick Crust	Springfield
A1 Pizza	Thick Crust	Shelbyville
A1 Pizza	Thick Crust	Capital City
A1 Pizza	Stuffed Crust	Springfield
A1 Pizza	Stuffed Crust	Shelbyville
A1 Pizza	Stuffed Crust	Capital City
Elite Pizza	Thin Crust	Capital City
Elite Pizza	Stuffed Crust	Capital City
Vincenzo's Pizza	Thick Crust	Springfield
Vincenzo's Pizza	Thick Crust	Shelbyville
Vincenzo's Pizza	Thin Crust	Springfield
Vincenzo's Pizza	Thin Crust	Shelbyville

Each row indicates that a given restaurant can deliver a given variety of pizza to a given area. Table has no non-key attributes, therefore it forms up to BCNF. However pizza varieties offered by a restaurant are not affected by delivery area, then it does not meet 4NF.

The dependences are: $\{Restaurant\} \twoheadrightarrow \{Pizza\ variety\}$
$\{Restaurant\} \twoheadrightarrow \{Delivery\ area\}$

These non-trivial multivalued dependences leads to a redundancy in the table. To eliminate the possibility of these anomalies, we must separate varieties data from delivery areas data, yielding two tables that are both in 4NF.

Restaurant	Pizza variety
A1 Pizza	Thick Crust
A1 Pizza	Stuffed Crust
Elite Pizza	Thin Crust
Elite Pizza	Stuffed Crust
Vincenzo's Pizza	Thick Crust
Vincenzo's Pizza	Thin Crust

Restaurant	Delivery area
A1 Pizza	Springfield
A1 Pizza	Shelbyville
A1 Pizza	Capital City
Elite Pizza	Capital City
Vincenzo's Pizza	Springfield
Vincenzo's Pizza	Shelbyville

DHRUV GARG
16BCE1190

JOIN DEPENDENCIES AND FIFTH NORMAL FORM

Achieving 4NF typically involves eliminating MVDs by repeated binary decompositions. In some cases, there may be no non-additive join decomposition of R into two relational schemas. Moreover there may be no functional dependency in R that violates any normal form upto BCNF, and there may be no nontrivial MVD present in R either that violates 4NF.

We then move on to another dependency called the join dependency, and, if it is present, carry out multiway decomposition into fifth normal form (5NF).
IMPORTANT NOTE: Such a dependency is a peculiar semantic constraint that is difficult to detect in practice, therefore normalization into 5NF is rarely done in practise.

- Join dependency :

It is a constraint on the set of legal relations over a database. A table is subject to a join dependency if it can be always re-created by joining multiple tables, each having a subset of ~~values~~ attributes of the table. If one of the tables has all the attributes of the table, the join dependency is called trivial.

A join dependency $*\{A, B, C \ldots Z\}$ on R is implied by the candidate keys of R if and only if each of A, B, ... Z is a superkey of R. To illustrate 5NF, we take the example as given below.

EXAMPLE

Salesman	Brand	Product Type
Jack John	Acme	Vacuum Cleaner
Jack John	Acme	Breadbox
Mary Jones	Robusto	Pruning Shears
Mary Jones	Robusto	Vacuum Cleaner
Mary Jones	Robusto	Breadbox
Mary Jones	Robusto	Umbrella stand
Louis Ferguson	Robusto	Vacuum Cleaner
Louis Ferguson	Robusto	Telescope
Louis Ferguson	Acme	Vacuum Cleaner
Louis Ferguson	Acme	Lava Lamp
Louis Ferguson	Nimbus	Tie Rack

The primary Key is the composite of all 3 columns. Also note that the table is in 4NF since there are no MVDs — no column is determinant for other two columns.
In absence of rules restricting valid combinations of salesman, Brand and Product Type, the table above is necessary to model the situation correctly.

Suppose however, a following rule applies: " A salesman has certain Brands and Product Types in his repertoire. If Brand B1 and B2 are in his repertoire, and Product type P is in his repertoire, then, the salesman must offer products of Product type P those made by brand B1 and B2."

∴ It is possible to split the table into three:

Salesman	Brand
Jack John	Acme
Mary Jones	Robusto
Louis Ferguson	Robusto
Louis Ferguson	Acme
Louis Ferguson	Nimbus

Salesman	Product Type
Jack John	Vacuum Cleaner
Jack John	Breadbox
Mary Jones	Pruning Shears
Mary Jones	Vacuum Cleaner
Mary Jones	Breadbox
Mary Jones	Umbrella Stand
Louis Ferguson	Telescope
Louis Ferguson	Vacuum Cleaner
Louis Ferguson	Lava Lamp
Louis Ferguson	Tie Rack

Brand	Product Type
Acme	Vacuum Cleaner
Acme	Breadbox
Acme	Lava lamp
Robusto	Pruning Shears
Robusto	Vacuum Cleaner
Robusto	Breadbox
Robusto	Umbrella Stand
Robusto	Telescope
~~Robusto~~	—
Nimbus	Ties Rack

Note how this setup helps to remove redundancy.

Suppose that Jack John starts selling Robusto's products Breadboxes and vacuum cleaners. In the previous setup, we would have had to add 2 new entries for each product type:

< Jack John, Robusto, Breadboxes>

< Jack John, Robusto, Vacuum Cleaners>

With the new setup, we just need to add a single entry < Jack John, Robusto>

A table is in 5NF if it cannot have lossless decomposition into any number of smaller tables.