CSCI 403 Database Management

19

Multivalued Dependencies

4^{th} Normal Form

This Lecture

A brief overview of multivalued dependencies and fourth normal form.

This will not be on the test (but will be extra credit on project 8).

MVDs and 4NF

- Previous lectures:
  - Functional dependencies (FD)
  - Boyce-Codd Normal Form (BCNF)
- This lecture:
  - Multivalued dependencies (MVD)
  - 4^{th} Normal Form (4NF)

We'll see that MVDs guide the way to 4NF the same way FDs guided us to BCNF!

Working Example

Consider this relation detailing:

- instructors
- courses they teach
- their hobbies

What FD's do you see?

What superkeys?

Is it in BCNF?

instructor	course	hobby
CPW	CSCI262	hiking
CPW	CSCI262	sci-fi
CPW	CSCI403	hiking
CPW	CSCI403	sci-fi
Baldwin	CSCI262	photography
Paone	CSCI261	hiking
Paone	CSCI261	board games
…	…	…

Recall: FDs

- Definition:

  Given a relation schema R and sets of attributes X and Y, then we say a functional dependency X → Y exists if, whenever tuples t_1 and t_2 are two tuples from any relation r(R) such that $t_1[X] = t_2[X]$, it is also true that $t_1[Y] = t_2[Y]$.
- The lingo:

  We say X functionally determines Y, or Y is functionally dependent on X.

Multivalued Dependencies

- Definition:

  A multivalued dependency X ↠ Y exists on a relation R if whenever there are two tuples t_1 and t_2 which agree on the attributes in X, then there must also exist a tuple t_3 (possibly the same as t_1 or t_2) such that the following are true:

  $t_3[X] = t_1[X] = t_2[X]$

  $t_3[Y] = t_2[Y]$

  $t_3[Z] = t_1[Z]$

  where Z contains all attributes of R not in X or Y.

Multivalued Dependencies 2

To recap: if $X \twoheadrightarrow Y$ and we have two tuples t_1 and t_2 such that
$t_1[X] = t_2[X]$
then t_3 (which may actually be t_1 or t_2) exists, and
$t_3[X] = t_1[X] = t_2[X]$
$t_3[Y] = t_2[Y]$
$t_3[Z] = t_1[Z]$
Furthermore, it turns out that by symmetry there must also exist t_4, with
$t_4[X] = t_1[X] = t_2[X] = t_3[X]$
$t_4[Y] = t_1[Y]$
$t_4[Z] = t_2[Z]$

CS@Mines

MVD Summary

This table may help (assume $X \twoheadrightarrow Y$)

		X	Y	Z
Agree on X	t_1	x	y_1	z_1
	t_2	x	y_2	z_1
Must exist	t_3	x	y_2	z_2
Must also exist	t_4	x	y_1	z_2

CS@Mines

Multivalued Dependencies 3

- If $X \twoheadrightarrow Y$, we say "X multi-determines Y".
- Note:
  - If $X \twoheadrightarrow Y$, with Z being everything else, then $X \twoheadrightarrow Z$ as well (follows from the definition/table above)
  - It is common to therefore write $X \twoheadrightarrow Y|Z$

CS@Mines

How Does This Happen?

Combining independent concepts in a relation.
Look at our example from the beginning:

What do courses have to do with hobbies?

What is the MVD?

instructor	course	hobby
CPW	CSCI262	hiking
CPW	CSCI262	sci-fi
CPW	CSCI403	hiking
CPW	CSCI403	sci-fi
Baldwin	CSCI262	photography
Paone	CSCI261	hiking
Paone	CSCI261	board games
…	…	…

CS@Mines

Discussion

Effectively, we have a cross-product of independent relations: instructors have courses; they also have hobbies.

- For every course I teach, I have all of my hobbies.
- For every hobby I enjoy, I have all of my courses!

Note this results in a subtle form of redundancy.

However, the relation has no non-trivial FDs, and the table is in BCNF!

CS@Mines

4^{th} Normal Form

- Definition:

  A relation R is in 4NF with respect to some set of multivalued dependencies if, for every non-trivial MVD $X \twoheadrightarrow Y$, X is a superkey of R.

  This looks familiar, and so will the next part…

CS@Mines

4NF Decomposition Algorithm

If we have a relation not in 4NF:

- Find a violating MVD $X \twoheadrightarrow Y|Z$
- Decompose R into
  - R1 = X ∪ Y
  - R2 = X ∪ Z

CS@Mines

Subsumption of BCNF by 4NF

- Any FD qualifies as an MVD
  - By definition, although you may have to convince yourself of this…
- Therefore, if we eliminate all violating MVDs, this includes all violating FDs ⇒ BCNF as well as 4NF

CS@Mines

Next Time

- Disks, database file organization, B-Trees

CS@Mines