

Relational Database Design Theory Part II

CPS 116
Introduction to Database Systems

Announcements (October 12)

- ◆ Midterm graded; sample solution available
- ◆ Homework #2 graded; sample solution already distributed
- ◆ Please verify all your grades on Blackboard
- ◆ Project milestone #1 due today

Review

- ◆ Functional dependencies
 - $X \rightarrow Y$: If two rows agree on X , they must agree on Y
 - ☞ A generalization of the key concept
- ◆ Non-key functional dependencies: a source of redundancy
 - Non-trivial $X \rightarrow Y$ where X is not a superkey
 - ☞ Called a BCNF violation
- ◆ BCNF decomposition: a method for removing redundancies
 - Given $R(X, Y, Z)$ and a BCNF violation $X \rightarrow Y$, decompose R into $R_1(X, Y)$ and $R_2(X, Z)$
 - ☞ A lossless join decomposition
- ◆ Schema in BCNF has no redundancy due to FD's

Next

- ✦ 3NF (BCNF is too much)
- ✦ Multivalued dependencies: another source of redundancy
- ✦ 4NF (BCNF is not enough)

Motivation for 3NF

- ✦ *Address* (*street_address*, *city*, *state*, *zip*)
 - $street_address, city, state \rightarrow zip$
 - $zip \rightarrow city, state$
- ✦ Keys
- ✦ BCNF?

To decompose or not to decompose

- Address*₁ (*zip*, *city*, *state*)
- Address*₂ (*street_address*, *zip*)
- ✦ FD's in *Address*₁
- ✦ FD's in *Address*₂
- ✦ Hey, where is $street_address, city, state \rightarrow zip$?
 - Cannot check without joining *Address*₁ and *Address*₂ back together
- ✦ Problem: Some lossless join decomposition is not dependency-preserving
- ✦ Dilemma: Should we get rid of redundancy at the expense of making constraints harder to enforce?

3NF

- ◆ R is in Third Normal Form (3NF) if for every non-trivial FD $X \rightarrow A$ (where A is single attribute), either
 - X is a superkey of R , or
 - A is a member of at least one key of R
- ☞ Intuitively, BCNF decomposition on $X \rightarrow A$ would "break" the key containing A
- ◆ So *Address* is already in 3NF
- ◆ Tradeoff:
 - Can enforce all original FD's on individual decomposed relations
 - Might have some redundancy due to FD's

BCNF = no redundancy?

- ◆ *Student* (SID , CID , $club$)
 - Suppose your classes have nothing to do with the clubs you join
 - FD's?
 - BCNF?
 - Redundancies?

SID	CID	club
102	CP0110	hockey
102	CP0110	sumo
102	CP0110	hockey
102	CP0111	sumo
102	CP0110	chess
102	CP0110	golf
...

Multivalued dependencies

- ◆ A multivalued dependency (MVD) has the form $X \twoheadrightarrow Y$, where X and Y are sets of attributes in a relation R
- ◆ $X \twoheadrightarrow Y$ means that whenever two rows in R agree on all the attributes of X , then we can swap their Y components and get two new rows that are also in R

X	Y	Z
a	y1	z1
a	y2	z2
a	y1	z2
a	y2	z1
...

} Must be in R too
