CS422 Principles of Database Systems
Normalization

Chengyu Sun
California State University, Los Angeles

Adapted from Jeffrey Ullman's lecture notes at http://www-db.stanford.edu/~ullman/dscb.html

Bad Schema Design

name	addr	beersLiked	manf	favBeer
Janeway	Voyager	Bud	A.B.	WickedAle
Janeway	Voyager	WickedAle	Pete's	WickedAle
Spock	Enterprise	Bud	A.B.	Bud

drinkers

- Update anomaly
- Delete anomaly

Decomposition

beer	manf
Bud	A.B.
WickedAle	Pete's

beers

drinker	beer
Janeway	Bud
Janeway	WickedAle
Spock	Bud

likes

drinker	addr	favBeer
Janeway	Voyager	WickedAle
Spock	Enterprise	Bud

drinkers

- How do we decompose?
- When do we stop decomposing?

Boyce-Codd Normal Form (BCNF)

- A relation R is in BCNF if for every *nontrivial FD* **A** B in R, **A** is a *super key* of R.

Drinkers Is Not BCNF

name	addr	beersLiked	manf	favBeer
Janeway	Voyager	Bud	A.B.	WickedAle
Janeway	Voyager	WickedAle	Pete's	WickedAle
Spock	Enterprise	Bud	A.B.	Bud

drinkers

??

A Non-BCNF Example

name	addr	beersLiked	manf	favBeer
Janeway	Voyager	Bud	A.B.	WickedAle
Janeway	Voyager	WickedAle	Pete's	WickedAle
Spock	Enterprise	Bud	A.B.	Bud

drinkers

- So why *drinkers* is not BCNF??

Other Non-BCNF Examples

- Beers(name, manf, manfAddr)
- R(A, B, C, D) with FD's AB C, C D, and D A

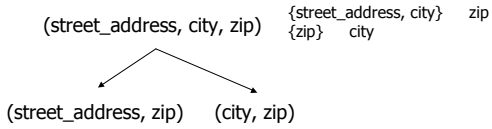
Decompose into BCNF

- Given relation **R** with FD's F
- Look among F for a BCNF violation **A** B
- Compute $\mathbf{A}^+$
- Decompose R into:
  - $\mathbf{R_1} = \mathbf{A}^+$
  - $\mathbf{R_2} = (\mathbf{R} - \mathbf{A}^+) \cup \mathbf{A}$
- *Continue decomposition with* $\boldsymbol{R_1}$ *and* $\boldsymbol{R_2}$ until all resulting relations are BCNF

Decomposition Examples

- Drinkers(name, addr, beerLiked, manf, favBeer)
- R(A, B, C, D) with FD's AB C, C D, and D A

Motivation for 3NF



- We lose the FD {street_address,city} zip after decomposition, or in other words, it becomes *unenforceable*.

An Unenforceable FD

Before decomposition:

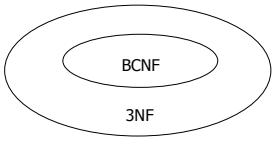
street	city	zip
545 Tech Sq.	Cambridge	02138
545 Tech Sq.	Cambridge	02139

FD violation

After decomposition:

street	zip
545 Tech Sq.	02138
545 Tech Sq.	02139

city	zip
Cambridge	02138
Cambridge	02139

Third Normal Form (3NF)

- A relation R is in 3NF if for every nontrivial *FD* **A** B in R,
  - **A** is a super key of R
  - or B is part of a key of R