ICS 321 Fall 2010

Normal Forms

Asst. Prof. Lipyeow Lim

Information & Computer Science Department

University of Hawaii at Manoa

The Problem with Redundancy

Hourly_Emps

SSN	Name	Lot	Rating	Hourly_wages	Hours_worked
123-22-2366	Attishoo	48	8	10	40
231-31-5368	Smiley	22	8	10	30
131-24-3650	Smethurst	35	5	7	30
434-26-3751	Guldu	35	5	7	32
612-67-4134	Madayan	35	8	10	40

- Suppose hourly wages are determined by rating
- **Redundant storage** : (8,10) stored multiple times
- **Update anomaly** : change hourly wages in row 1
- **Insertion anomaly** : requires knowing hourly wages for the rating
- **Deletion anomaly** : deleting all (8,10) loses info

Using Two Smaller Tables

Hourly_Emps

SSN	Name	Lot	Rating	Hours_ worked
123-22-2366	Attishoo	48	8	40
231-31-5368	Smiley	22	8	30
131-24-3650	Smethurst	35	5	30
434-26-3751	Guldu	35	5	32
612-67-4134	Madayan	35	8	40

RatingWages

Rating	Hourly_ wages
5	7
8	10

- *Notation*: denote relation schema by listing the attributes SNLRWH
- **Update anomaly** : Can we change W for Attishoo?
- **Insertion anomaly** : What if we want to insert an employee and don't know the hourly wage for his rating?
- **Deletion anomaly** : If we delete all employees with rating 5, do we lose the information about the wage for rating 5?

Decomposition

Hourly_Emps

SSN	Name	Lot	Rating	Hours_ worked
123-22-2366	Attishoo	48	8	40
231-31-5368	Smiley	22	8	30
131-24-3650	Smethurst	35	5	30
434-26-3751	Guldu	35	5	32
612-67-4134	Madayan	35	8	40

RatingWages

Rating	Hourly_ wages
5	7
8	10

- Remove redundancy by decomposition
  - Since hourly wage is completely determined by rating, factor out hourly wage.
- Pros: less redundancy less anomalies
- Cons: retrieving the hourly wage of an employee requires a join

Normal Forms

- Helps with the question: do we need to refine the schema ?
- If a relation is in a certain *normal form* (BCNF, 3NF etc.), it is known that certain kinds of problems are avoided/minimized. This can be used to help us decide whether decomposing the relation will help.
- Role of FDs in detecting redundancy:
  - Consider a relation R with 3 attributes, ABC.
    - No FDs hold: There is no redundancy here.
    - Given $A \rightarrow B$: Several tuples could have the same A value, and if so, they'll all have the same B value!

Boyce-Codd Normal Form (BCNF)

- Let R denote a relation, X a set of attributes from R, A an attribute from R, and F the set of FDs that hold over R.
- R is in **BCNF** if for all $X \rightarrow A$ in F^+,
  - $A \in X$ (trivial FD) or
  - X is a superkey
- **Negation**: R is not in BCNF if there exists an $X \rightarrow A$ in F^+, such that $A \notin X$ (non-trivial FD) AND X is not a key

The only non-trivial FDs that hold are key constraints

Examples: BCNF

- Are the following in BCNF ?

Firstname	Lastname	DOB	Address	Telephone
John	Smith	Sep 9 1979	Honolulu,HI	808-343-0809

F= { FLD $\rightarrow$ FLDAT}

Firstname	Lastname	DOB	Street	CityState	Zipcode	Telephone
John	Smith	Sep 9 1979	1680 East West Rd.	Honolulu,HI	96822	808-343-0809

F= { FLD $\rightarrow$ FLDSCZT, C$\rightarrow$Z }

Third Normal Form (3NF)

- Let R denote a relation, X a set of attributes from R, A an attribute from R, and F the set of FDs that hold over R.
- R is in **3NF** if for all $X \rightarrow A$ in F^+,
  - $A \in X$ (trivial FD) or
  - X is a superkey or
  - A is part of some key
- **Negation**: R is not in 3NF if there exists an $X \rightarrow A$ in F^+, such that $A \notin X$ (non-trivial FD) AND X is not a key AND A is not part of some key
- If R is in BCNF, obviously in 3NF.
- If R is in 3NF, some redundancy is possible. It is a compromise, used when BCNF not achievable (e.g., no ``good'' decomp, or performance considerations).

Example: 3NF

- Which of the following is in 3NF and which in BCNF ?

Firstname	Lastname	DOB	Address	Telephone
John	Smith	Sep 9 1979	Honolulu,HI	808-343-0809

F= { FLD $\rightarrow$ FLDAT}

Firstname	Lastname	DOB	Street	CityState	Zipcode	Telephone
John	Smith	Sep 9 1979	1680 East West Rd.	Honolulu,HI	96822	808-343-0809

F= { FLD $\rightarrow$ FLDSCZT, C$\rightarrow$Z }

Student	Course	Instructor
Smith	OS	Mark

F= { SC $\rightarrow$ I, I$\rightarrow$C }