ICS 321 Fall 2009

Schema Refinement & Normal Forms (ii)

Asst. Prof. Lipyeow Lim
Information & Computer Science Department
University of Hawaii at Manoa

Two More Rules

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

- **Union**
  - If $X \rightarrow Y$ and $X \rightarrow Z$, then $X \rightarrow YZ$
  - Eg. FLD → A and FLD → T, then FLD → AT
- **Decomposition**
  - If $X \rightarrow YZ$, then $X \rightarrow Y$ and $X \rightarrow Z$
  - Eg. FLD → AT , then FLD → A and FLD → T
- **Trivial FDs**
  - Right side is a subset of Left side
  - Eg. F → F, FLD → FD

Closure

- **Implication**: An FD *f* is *implied by* a set of FDs *F* if *f* holds whenever all FDs in *F* hold.
  - f=A $\rightarrow$C is implied by F={ A$\rightarrow$B, B $\rightarrow$C} (using Armstrong's transitivity)
- **Closure F^+** : the set of all FDs implied by F
  - Algorithm:
    - start with F^+ =F
    - keep adding new implied FDs to F^+ by applying the 5 rules (Armstrong's Axioms + union + decomposition)
    - Stop when F^+ does not change anymore.

Example: Closure

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

- Given FLD is the primary key and C $\rightarrow$ Z
- Find the closure:
  - Start with { FLD $\rightarrow$ FLDSCZT, C$\rightarrow$Z }
  - Applying reflexivity, { FLD $\rightarrow$ F, FLD $\rightarrow$L, FLD $\rightarrow$ D, FLD $\rightarrow$ FL, FLD $\rightarrow$ LD, FLD $\rightarrow$DF, FLDSCZT $\rightarrow$ FLD, …}
  - Applying augmentation, { FLDS $\rightarrow$ FS, FLDS $\rightarrow$ LS, …}
  - Applying transitivity …
  - Applying union …
  - Applying decomposition …
  - Repeat until F^+ does not change

Attribute Closure

- Computing the closure of a set of FDs can be expensive. (Size of closure is exponential in # attrs!)
- Typically, we just want to check if a given FD $X \rightarrow Y$ is in the closure of a set of FDs *F*. An efficient check:
  - Compute ***attribute closure*** of X (denoted X^+) wrt *F:*
    - Set of all attributes A such that $X \rightarrow A$ is in F^+
    - There is a linear time algorithm to compute this.
  - Check if Y is in X^+
- Does F = $\{A \rightarrow B,\ B \rightarrow C,\ C\,D \rightarrow E\}$ imply $A \rightarrow E$?
  - i.e, is $A \rightarrow E$ in the closure F^+ ? Equivalently, is E in A^+ ?

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

The only non-trivial FDs that hold are key constraints

- **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

Examples: BCNF

- Are the following in BCNF ?

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

F= { FLD $\rightarrow$ FLDAT}

<u>Firstname</u>	<u>Lastname</u>	<u>DOB</u>	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 ?

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

F= { FLD $\rightarrow$ FLDAT}

<u>Firstname</u>	<u>Lastname</u>	<u>DOB</u>	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 }