COS 597A:
Principles of
Database and Information Systems

Managing Functional Dependencies and Redundancy

General functional constraints
(Review)

General form for relational model:

- Let α(R) denote the set of names of attributes in the schema for relation R
- Let X and Y be subsets of α(R)

The functional dependency X → Y holds if for any instance I of R and for any pair of tuples t_1 and t_2 of R,

$$\pi_X(t_1) = \pi_X(t_2) \Rightarrow \pi_Y(t_1) = \pi_Y(t_2)$$

- special cases: candidate keys, superkeys

Redundancy

- Functional dependencies capture redundancy in a relation
  - e.g. area code → state: why store state?
- Redundancy good for reliability
- Redundancy bad for
  - space to store
    - repetitions
  - must maintain on changes
  - representation of one relationship embedded in another

Example relation for a city elementary school system:

school_child: (name, st_addr, apt., birthday, school)
st_addr → school

consider a large apt. building

Solution: decompose

Example:

child: (name, st_addr, apt., birthday)
placement: (st_addr, school)

- child ◊◊ placement gives school_child
  because of *functional dependency*
- space gain larger than space cost
- functional dependency now primary key constraint
- st_addr, school correspondence explicitly maintained

General Form:

- for X, Y ⊆ α(R) and X → Y
- decompose R into
  R1: α(R) - (Y-X)
  R2: X U Y

Downside of decompose

Example:

school_child: (school, stuID, st_addr, apt., birthday)
st_addr → school
becomes
stu: (stuID, st_addr, apt., birthday)
placement: (st_addr, school)

General Form:
for X, Y ⊆ α(R)
and X → Y
decompose R into
•R1: α(R) - (Y-X)
•R2: X U Y

Constraint (school, stuID) → (st_addr, apt., birthday)

- was primary key constraint
- now split constraint
  to check requires ◊◊ - expensive
- primary key for stu?

Downside of decompose

Example:

school_child: (school, stuID, st_addr, apt., birthday)
st_addr → school
becomes stu: (stuID, st_addr, apt., birthday)
placement: (st_addr, school)

Constraint (school, stuID) → (st_addr, apt., birthday)

- was primary key constraint
- now split constraint
  to check requires ◊◊ - expensive
- primary key for stu?
  (stuID, st_addr) → (stuID, st_addr, school)
  (stuID, st_addr, school) → (stuID, st_addr, apt., birthday)
  so **stu: (stuID, st_addr, apt., birthday)**
- new primary key constraints do not imply old primary key constraint:
  (school, stuID) → (st_addr, apt., birthday)

Decomposition: Formal Properties

- Let Φ be a set of functional dependencies for a relational scheme R with attribute set $\alpha(R)$
- Let Φ^+ denote the set of all functional dependences implied by Φ
- Let $X, Y \subseteq \alpha(R)$, where $X \cap Y$ is not necessarily empty
- Decomposition of R into R_1: X and R_2:Y is
  - lossless if for every instance I of R that satisfies Φ

    $\pi_X(I) \lozenge\lozenge \pi_Y(I) = I$
    - guaranteed to get back R
  - dependency preserving if $(\Phi_x \cup \Phi_Y)^+ = \Phi^+$

    where Φ_x denotes the set of functional dependencies $V \rightarrow W$ in Φ^+ with $V \subseteq X$ and $W \subseteq X$
    - can check all functional dependencies for R by checking all for X and all for Y without doing JOIN

Implied functional dependencies

- Definition: a functional dependency $X \rightarrow Y$ is implied by Φ if $X \rightarrow Y$ holds whenever all functional dependences in Φ hold
- Armstrong's Axioms

  for attribute sets X, Y, Z
  1. if $X \subseteq Y$ then $Y \rightarrow X$ — reflexivity
  2. if $X \rightarrow Y$ then $\forall Z\ (XZ \rightarrow YZ)$ — augmentation
  3. if $X \rightarrow Y$ and $Y \rightarrow Z$ then $X \rightarrow Z$ — transitivity

- Theorem: The set of all functional dependences obtained from Φ by repeated application of Armstrong's Axioms gives Φ^+

Normal Forms

- How do we find "good " ("best"?) decomposition?
- Identify normal forms with desirable properties
- Decompose so resulting relations are in normal form

Boyce-Codd Normal Form (BCNF)

- Let R denote a relational scheme with attribute set $\alpha(R)$
- R is in BCNF with respect to a set Φ of func. dep.s if for all func. dep.s in Φ^+ of the form $X \rightarrow Y$ with $X, Y \subseteq \alpha(R)$, at least one of
  - $Y \subseteq X$ (trivial func. dep.)
  - X is a superkey for R
- very strong normal form
- can't always get dependency preserving decomposition into set of BCNF relations

Third Normal Form (3NF)

- Let R denote a relational scheme with attribute set $\alpha(R)$
- R is in 3NF with respect to a set Φ of func. dep.s if for all func. dep.s in Φ^+ of the form $X \rightarrow Y$ with $X, Y \subseteq \alpha(R)$, at least one of
  - $Y \subseteq X$ (trivial func. dep.)
  - X is a superkey for R
  - each attribute A in Y-X is contained in a candidate key for R
- can always get lossless, dependency preserving decomposition into 3NF relations
- cannot always remove all functional dependencies

Why allow right hand side part of some candidate key?

- consider decomposing R using $X \rightarrow A$
  - A an attribute
  - X not superkey
  - A not in X
- get R_1: $\alpha(R)$ - (A) and R_2: X U {A}
- if A not part of a candidate key then

  for any candidate key $K \subseteq \alpha(R)$

  check $K \rightarrow \alpha(R)$ -{A} in R_1

  including $K \rightarrow X$

  check $X \rightarrow A$ in R_2

  conclude $K \rightarrow A$

  } all checks local to R_1 or R_2 NO HARM DECOMPOSE
- if A is part of a candidate key K

  splitting key: K-A in $\alpha(R_1)$; $K \cap (X \cup \{A\})$ in $\alpha(R_2)$

  to check K is a candidate key need $R_1 \lozenge\lozenge R_2$ AVOIDING

Revisit example

Lossless-join decompositon?
Dependency preserving decomposition?
Normal forms? school_child in 3NF

school_child: (school, stuID, st_addr, apt., birthday)
st_addr → school
becomes
stu: (stuID, st_addr, apt., birthday)
placement: (st_addr, school)

Constraint (school, stuID) → (st_addr, apt., birthday)

- was primary key constraint
- now split constraint
  to check requires ◊◊ - expensive

Discussion

- Is polynomial-time algorithm for 3NF lossless dependency-preserving decomposition
- Using 3NF minimizes problems of general functional dependencies
  – does not eliminate
- Use BCNF if can get it
  – decomposition algorithm simpler too!