5 Normalization: Quality of relational designs

5.1 Functional Dependencies

5.1.1 Design quality

• What is a "good" conceptual model?
  – Usually many alternatives
  – No clear criteria for comparison guidelines
  – Wanted: formal methods for comparing designs

• Informal guidelines
  – Avoid redundancies: (mid, format) in one relation?
    Movie (mid, format, director, ...) repeats the title, director information for each format;
    if tape with the format exists
  – Avoid to model more than one object from reality in one entity / relation
    Data about director in Movie relation?

5.1.2 Update Anomalies

– Basic idea: constraints must be modeled explicitly
  e.g. a tape is loaned by zero or one customer,
  each person has a unique birthday
  Movie (mid, title, director, birthdate, livesInCity,...)

– Redundancies may cause "anomalies"

  • Deletion of a row may delete all data about a different object
  • Update of an attribute may cause update on many rows
  • Insertion may be difficult / impossible, since data are missing

Examples….

"update anomaly" = deletion, update, insertion anomaly

UPDATE Movie
SET livesInCity = 'SF'
WHERE director = 'Spielberg'

Update anomalies: Examples

CREATE TABLE Experiment (id SERIAL PRIMARY KEY,
responsible_Person VARCHAR(40),
institute VARCHAR(30),
phone INT,
purpose VARCHAR(100),
start TIMESTAMP,
endTime TIMESTAMP,
result INT)

DELETE FROM Experiment WHERE result < 10

Consequence: data on experimentator might be lost

Update anomaly: example

Delete (43, 'Amistad', 'Spielberg', 12.10.47, 'LA',..)
from table Movie (mid, title, director,birthdate,livesInCity,...)
effect: data about director are lost if this is the only movie
with this director ('Spielberg') → deletion anomaly

Update anomaly: example

update Movie set livesInCity = 'SF'
WHERE director = 'Spielberg'
all those movie tuples having director = 'Spielberg' have to be changed (update anomaly)

• what is an insertion anomaly?
5.1.3 Functions and Functional dependencies

- Important formal concept: Functions
  - Used to formalize integrity constraints on relationships
  - General approach:
    find functions among attributes in Relation R

Examples:
- Experiment (id, responsible_Person, institute, phone,...result) is a (partial) function
- Movie (mId, title, director, birthdate, livesInCity,...) is a function

Functional Dependencies (FD)

- Keys and Functional dependencies
  - Property of a key: at most one row for each value k
  - Let the key of Relation R be composed of attributes
    K={a1,...,ak}
  - Then the attributes Σ(R) \ {a1,...,ak} are functionally
    dependent from K

  This means:
  - There is a function which maps keys to values of attributes
  - Function is represented by table
  - Table may be changed, but functional property is
time invariant

- Primary key: one of the candidate keys
- Prime attributes: attributes belonging to a candidate key

Functional Dependencies

- Generalization: Functional dependency between non-key attributes or sets of attributes

Example:
- Movie (mId, title, format, director, birthdate, livesInCity,...)
- director -> birthday
  - A particular director has one and only one birthday

- No FD between director and title:
  - A particular person may have directed many films (titles)

- Functional dependencies:
  - are constraints (invariants) of the application domain

Functional Dependency: Definition

Functional Dependencies (FDs)
Let A = Σ(R)* = {a1,b,c,...} be the attribute set of a relation E, e, e' tuples of R,
let X, Y ⊆ A
Y is functionally dependent on X (written: X → Y)
iff
(∀ x1 ∈ X) e1.x1 = e'.x1 ⇒ (∀ yj ∈ Y) ej.yj = e'.yj
- Important: invariants are independent of the particular
database state
- They must hold at all times,
i.e. they restrict the valid states of the database.

5.1.3 Properties of Functional Dependencies

- Trivial functional dependency
  - Trivial: if values of attributes yi ∈ Y are given, then the values of attributes in every subset of Y

- Augmentation
  - Z ⊆ A=Σ(R), X → Y ⊃ XZ → YZ

- Transitivity
  - X,Y,Z ⊆ A=Σ(R), X → Y, Y → Z ⇒ X → Z

  Notation XY → Z means X ⊆ Y → Z
Armstrong inference rules

Given a set of FDs, find all implied FD’s

A sound, complete, minimal set (Armstrong axioms):

\begin{align*}
Y \subseteq X & \Rightarrow X \rightarrow Y \quad (I: \text{inclusion}) \\
(X \rightarrow Y, Y \rightarrow Z) & \Rightarrow X \rightarrow Z \quad (T: \text{transitivity}) \\
(X \rightarrow Y) & \Rightarrow XZ \rightarrow YZ \quad (A: \text{augmentation})
\end{align*}

sound:
only logically implied FDs are produced by the inference rules

complete:
every logically implied FD will be produced by finite many inferences

Means:
- given a set $F$ of FDs. Every FD implied by $F$ will be produced by a finite number of inferences I, T or A
- No FD will be inferred, which is not implied by $F$

We will use this result to calculate normal forms

Functional Dependencies and keys

Non-key attributes functionally dependent on part of the key

$(\text{mId, format})$ is the key, but e.g. $(\text{mId}) \rightarrow (\text{title})$ holds

- Bad: key properties are checked by the DB system, other functional dependencies are NOT e.g. more than one title or director for ONE mId cannot be prevented by the DBS

Different kinds of FDs of a relation:

1. Partial dependences on one of the candidate keys
   $(\text{mId}) \rightarrow (\text{title})$, since key is $(\text{mId, format})$
2. Dependencies among non-key attributes
   $(\text{director}) \rightarrow (\text{birthday})$
3. Dependencies among attributes of different candidate keys