Database Normalization Theory

Minimal Criteria for Good Design

Overview

• Series of Criteria for Good Relations
• First we will define criteria assuming single key in relation
• Then allow for several Candidate Keys
Why do theory twice?

- Primary key version came first.
- Primary key version easier to understand.
- Candidate key version different only when relation has > 1 key.
- Candidate key version needed for subtle problems.

Boyce Codd Normal Form (BCNF)

Aims of Normalization

Reduced Redundancy
- Eliminate Duplication of Information
- Duplication leads to errors/inconsistency

Dependency Preservation
- Dependencies are rules of the database
- Preserve them for data integrity

Lossless Join
- Restructuring of DB must allow recovery of original data
### Terminology

#### Superkey
- Subset of attributes of relation
- No two tuples have same superkey

#### Key
- Minimal Superkey
- Can’t remove attribute and keep uniqueness

#### Candidate Key
- Any key of a relation

#### Primary Key
- Key designated by designer for reference

### Terminology (cont.)

#### Two terms vary with version of theory
- Primary Key version
- Candidate Key version

#### Prime Attribute
- member of Primary Key
- member of any Candidate key

#### Nonprime Attribute
- not PK attribute
- not attribute of any Candidate key