Relational Database Keys

Copyright 2009 Peter R. Douglass

Functional Dependencies

- We say that there exists a ***functional dependency*** relationship between two variables X and Y, if for a given value of X, there is a unique value of Y corresponding to X

Functional Dependencies : Example

- Example: Social Security Numbers and Birth Dates.

- There may exist fake personal identities

- However, if we assume that each social security number corresponds to a unique personal identity, and that each personal identity has a unique birth date, then there should be a unique birth date corresponding to each social security number.

Functional Dependencies : Assumptions About Data

- When we assert that there is a functional dependency between two variables, we are making assumptions about the data involved.

- Typically, we are modeling something in real life, and our model may be more or less precise, but may never correspond to reality *exactly*.

Assumptions about Data: Example…

- In the previous example, we made the assumption that each social security number corresponded to a single personal identity.

- However, we can conceive of the possibility that two fake personal identities could have the same social security number but different birth dates…

Assumptions about Data: …Example

- In the database of social security numbers kept by the Social Security Administration, it may be true that a single social security number corresponds to a single birth data.

- In a database kept by a law enforcement agency, there may be cases where a single social security number is associated with more than one birth date, or even no birth date.

Table Superkey

- A superkey is a set of columns in a table such that we assume that there is a function dependency between those columns and the rest of the columns in the table.

- There may be more than one superkey in a table.

Table Superkey – alternative definition

Another way of defining a superkey is this:

- A superkey is a set of columns in a table, such that it is an assumption about the data that might fill that table, whatever that data is, if we projected our table onto only those columns, then each row would be unique.

Primary Key

- If a database management system is responsible for ensuring that the values in a given set of columns (usually one) are unique (form a superkey) then that column (set of columns) is called a primary key.

Important:
Projections containing superkeys

- An arbitrary projection of a table can contain duplicate rows
- If a projection of a table contains a superkey of that table, then the projection does not contain duplicate rows.