Relational Model 1: Tables and Keys

Yufei Tao

Department of Computer Science and Engineering
Chinese University of Hong Kong

The relational model is the *de facto* standard implemented in all the major database systems. It defines:

1. the format by which data should be stored;
2. the operations for querying the data.

We will focus on the first aspect in this lecture, leaving the second aspect to the next lecture.

A database conforming to the relational model is called a relational database.

Table, a.k.a. Relation

In a relational database, data are stored in tables.

PROF

pid	name	dept	rank	sal
$p1$	Adam	CS	asst	6000
$p2$	Bob	EE	asso	8000
$p3$	Calvin	CS	full	10000
$p4$	Dorothy	EE	asst	5000
$p5$	Emily	EE	asso	8500
$p6$	Frank	CS	full	9000
		...		

- Each row is also called a tuple.
- Each column is also called an attribute.
- The relation schema of a table is the set of its attribute names.
  - E.g., the schema of the above table is {pid, name, dept, rank, sal}.

Candidate Key

Definition

In a table, a candidate key is a minimal set K of attributes such that no two tuples are allowed to be equivalent on all the attributes in K.

E.g., in the PROF table of the previous slide, if we set {pid} as a candidate key, then no two tuples can have the same pid.

- A candidate key is designated when the table is created.
- There can be multiple candidate keys.
  - E.g., if you want, you can specify {name} as another candidate key, but do you think it makes sense?
  - How about {dept, rank}?

Discussion

CLASS

cid	**title**	**dept**	**year**
$c1$	database	CS	2011
$c2$	signal processing	EE	2012
	...		

How would you set a candidate key?

As a good practice, every table should have at least a candidate key, a convention that will be enforced in the rest of the course. This implies that no two tuples in the table can be entirely equivalent to each other (think: why?).

Super Key

Definition

In a table, if K is a candidate key, any super set of K is called a super key.

E.g., in the PROF table (pid, name, dept, rank, sal) in Slide 3, {pid} is a candidate key. Hence, all the following are super keys:

- {pid}
- {pid, name}
- {pid, dept}
- {pid, rank, sal}
- ...

Lemma

In a table, no two tuples can be equivalent on all the attributes of a super key.

The proof is easy and left to you.

Foreign Key

Definition

Let T and T' be two tables, and K a candidate key in T. If T' also contains K, then K is a foreign key of T' referencing T.

See the next slide for an example.

PROF

pid	**name**	**dept**	**rank**	**sal**
$p1$	Adam	CS	asst	6000
$p2$	Bob	EE	asso	8000
$p3$	Calvin	CS	full	10000
$p4$	Dorothy	EE	asst	5000
$p5$	Emily	EE	asso	8500
$p6$	Frank	CS	full	9000
		...		

CLASS

cid	**title**	**dept**	**year**
$c1$	database	CS	2011
$c2$	signal processing	EE	2012
$c1$	database	CS	2012
	...		

TEACH

pid	**cid**	**year**
$p1$	$c1$	2011
$p2$	$c2$	2012
$p1$	$c1$	2012
	...	

Suppose that PROF has a candidate key {pid}, and CLASS has a candidate key {cid, year}. Then:

- {pid} is a foreign key of TEACH referencing PROF.
- {cid, year} is a foreign key of TEACH referencing CLASS.

Discussion

PROF

pid	**name**	**dept**	**rank**	**sal**
$p1$	Adam	CS	asst	6000
$p2$	Bob	EE	asso	8000
$p3$	Calvin	CS	full	10000
$p4$	Dorothy	EE	asst	5000
$p5$	Emily	EE	asso	8500
$p6$	Frank	CS	full	9000
		...		

CLASS

cid	**title**	**dept**	**year**
$c1$	database	CS	2011
$c2$	signal processing	EE	2012
$c1$	database	CS	2012
	...		

TEACH

pid	**cid**	**year**
$p1$	$c1$	2011
$p2$	$c2$	2012
$p1$	$c1$	2012
	...	

How would you designate a candidate key for TEACH?