Functional Dependencies

Yufei Tao

Department of Computer Science and Engineering
Chinese University of Hong Kong

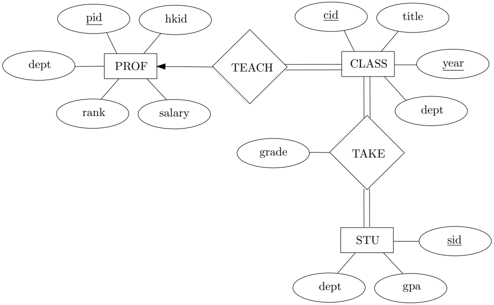
A primary goal of database design is to decide what tables to create. Usually, there are two principles:

1. Capture all the information that needs to be captured by the underlying application.
2. Achieve the above with little redundancy.

The first principle is enforced with an entity relationship (ER) diagram, while the second with normalization.

This and the next few lectures are devoted to normalization.

Tables created from an ER diagram may contain redundancy.



- PROF (pid, hkid, dept, rank, salary), candidate key pid
- CLASS (cid, title, year, dept), candidate key (cid, year)
- STU (sid, dept, gpa), candidate key sid
- TEACH (pid, cid, year), candidate key (cid, year)
- TAKE (cid, year, sid, grade), candidate key (cid, year, sid)

Where is redundancy?

Answer:
CLASS (cid, title, year, dept), candidate key (cid, year)

Why? Because every time the same course is offered again, its title and department are duplicated.

cid	title	year	dept
$c1$	database	2010	cs
$c1$	database	2011	cs
$c1$	database	2012	cs

The red values are redundant. Note that the "database" and "cs" of the first tuple are not redundant. Why?

cid	title	year	dept
$c1$	database	2010	cs
$c1$	database	2011	cs
$c1$	database	2012	cs

Observations: cid implies title, dept. This is written as:

$$\text{cid} \rightarrow \text{title}$$
$$\text{cid} \rightarrow \text{dept}$$

which are called functional dependencies.

Definition

A functional dependency (FD) has the form of $X \to Y$ (reads: X implies Y), where X and Y are sets of attributes. It means that whenever two tuples are identical on all the attributes in X, they must also be identical on all the attributes in Y.

Alternatively, you can interpret $X \to Y$ as: each possible value of X can correspond to exactly one value of Y.

Functional dependencies are constraints that are required by the underlying application.

Example

Assume that the following FDs hold:

$$\text{cid} \rightarrow \text{title}$$
$$\text{title} \rightarrow \text{dept}$$
$$\text{cid, year} \rightarrow \text{dept}$$
$$\text{cid, year} \rightarrow \text{cid, dept}$$

Can the following tuples co-exist?

cid	title	year	dept
$c1$	database	2010	cs
$c1$	database	2011	cs
$c2$	database	2012	ee

How about the following?

cid	dept
$c1$	cs
$c1$	ee

Armstrong's Axioms

1. (Reflexivity) $X \to Y$ for any $Y \subseteq X$.
2. (Augmentation) If $X \to Y$, then $XZ \to YZ$.
3. (Transitivity) If $X \to Y$ and $Y \to Z$, then $X \to Z$.

You can prove the above axioms easily by yourself.

Closure of an Attribute Set

Let F be the set of functional dependencies required by the underlying application.

Definition

The closure of a set X of attributes is the set of all such attributes A that $X \rightarrow A$ can be deduced from F.

Example: Let F be the set of following FDs:

$$\text{cid} \rightarrow \text{title}$$
$$\text{title} \rightarrow \text{dept}$$
$$\text{cid, year} \rightarrow \text{dept}$$
$$\text{cid, year} \rightarrow \text{cid, dept}$$

Then, “dept” is in the closure of $X = \{\text{cid}\}$. Is “year” in the closure of X?

Finding the Closure of an Attribute Set

algorithm (F, X)
/* F is a set of FDs, and X is an attribute set */
1. $C = X$
2. **while** F has a FD $A \rightarrow B$ such that $A \subseteq C$ **do**
3. $\quad C = C \cup B$
4. $\quad$ remove $A \rightarrow B$ from F
5. **return** C /* closure of X */

Example: Let F be the set of following FDs:

$$\begin{aligned} \text{cid} &\rightarrow \text{title} \\ \text{title} &\rightarrow \text{dept} \\ \text{cid, year} &\rightarrow \text{dept} \\ \text{cid, year} &\rightarrow \text{cid, dept} \end{aligned}$$

What is the closure of $X = \{\text{cid}\}$? We apply the algorithm on the previous slide:

1. $C = \{\text{cid}\}$
2. By "cid $\rightarrow$ title", $C = \{\text{cid, title}\}$
3. By "title $\rightarrow$ dept", $C = \{\text{cid, title, dept}\}$

Think:

Is "year" in the closure of X?

Can "cid $\rightarrow$ year" be deduced from F?

Example: Let F be the set of following FDs:

$$\begin{aligned} \text{cid} &\rightarrow \text{title} \\ \text{title} &\rightarrow \text{dept} \\ \text{cid, year} &\rightarrow \text{dept} \\ \text{cid, year} &\rightarrow \text{cid, dept} \end{aligned}$$

What is the closure of $X = \{\text{cid, year}\}$?

1. $C = \{\text{cid, year}\}$
2. By "cid $\rightarrow$ title", $C = \{\text{cid, year, title}\}$
3. By "title $\rightarrow$ dept", $C = \{\text{cid, year, title, dept}\}$