

# A Deeper Look at Data Modeling

Shan-Hung Wu & DataLab  
CS, NTHU

# Outline

- More about ER & Relational Models
  - Weak Entities
  - Inheritance
- Avoiding redundancy & inconsistency
  - Functional Dependencies
  - Normal Forms

# Outline

- More about ER & Relational Models
  - Weak Entities
  - Inheritance
- Avoiding redundancy & inconsistency
  - Functional Dependencies
  - Normal Forms

# Modeling Users Addresses

- Street, city, etc.
- Each user may have multiple addresses
  - Home, office, etc.

users

id	name	karma
729	Bob	35
730	John	0

posts

id	text	ts	authorId
33981	'Hello DB!'	1493897351	729
33982	'Show me code'	1493904323	812

# Modeling Users Addresses

- How to reflect:
  - Home and office addresses?
  - Address exists only when it owner (user) exists?

**users**

<u>id</u>	name	karma
729	Bob	35
730	John	0

**addresses**

<u>id</u>	userId	street	city
4356	729	'X Rd.'	'New York'
4357	729	'Y Rd.'	'LA'

**posts**

<u>id</u>	text	ts	authorId
33981	'Hello DB!'	1493897351	729
33982	'Show me code'	1493904323	812

# Modeling Users Addresses

- How to reflect:
  - *Home and office addresses?*
  - Address exists only when it owner (user) exists?

users

<u>id</u>	name	karma
729	Bob	35
730	John	0

addresses

<u>userId</u>	<u>type</u>	street	city
729	'home'	'X Rd.'	'New York'
729	'office'	'Y Rd.'	'LA'

# Modeling Users Addresses

- How to reflect:
  - Home and office addresses?
  - ***Address exists only when it owner (user) exists?***

```
CREATE TABLE addresses (  
  userId          serial NOT NULL,  
  type            text NOT NULL,  
  ...  
  PRIMARY KEY    (userId, type),  
  FOREIGN KEY     userId  
                REFERENCES users ON DELETE CASCADE  
);
```

# Outline

- More about ER & Relational Models
  - Weak Entities
  - Inheritance
- Avoiding redundancy & inconsistency
  - Functional Dependencies
  - Normal Forms



# Modeling Inheritance

- Suppose you have employees in your model
- How to model special types of employees?
  - Contracted: contractId
  - Hourly: wage, workHours

# Modeling Inheritance

## employees

<u>id</u>	name	department
729	Bob	'R&D'
730	John	'Sales'

## contractEmployees

<u>eld</u>	contractId
834	\$10
878	\$20

## hourlyEmployees

<u>eld</u>	wage	workHours
729	\$10	4
730	\$20	16

- If a superclass tuple is deleted, cascade delete the subclass tuple

# Outline

- More about ER & Relational Models
  - Weak Entities
  - Inheritance
- **Avoiding redundancy & inconsistency**
  - Functional Dependencies
  - Normal Forms

# How Good Are Your Data?

- Let's say, if you want to track the topics of a blog page
- Is this a good table?

## blog\_pages

blogId	url	created	authorId	topic	topicAdmin
33981	ms.com/...	2012/10/31	729	programming	5638
33981	ms.com/...	2012/10/31	729	db	5649
33982	apache.org/...	2012/11/15	4412	programming	5638
33982	apache.org/...	2012/11/15	4412	os	7423

# Insertion Anomaly

## blog\_pages

blogId	url	created	authorId	topic	topicAdmin
33981	ms.com/...	2012/10/31	729	programming	5638
33981	ms.com/...	2012/10/31	729	db	5649
33982	apache.org/...	2012/11/15	4412	programming	5638
33982	apache.org/...	2012/11/15	4412	os	7423

33983	apache.org/...	2013/02/15	7412		
-------	----------------	------------	------	--	--




- A blog cannot be inserted without knowing all fields of topics (except setting them to null)

# Update Anomaly

## blog\_pages

blogId	url	created	authorId	topic	topicAdmin
33981	ms.com/...	2012/10/31	729	<i>win prog.</i>	5638
33981	ms.com/...	2012/10/31	729	db	5649
33982	apache.org/...	2012/11/15	4412	programming	5638
33982	apache.org/...	2012/11/15	4412	os	7423




- If you forget to update all duplicated cells, you get inconsistent data

# Deletion Anomaly

## blog\_pages

blogId	url	created	authorId	topic	topicAdmin
33981	ms.com/...	2012/10/31	729	<i>programming</i>	<i>5638</i>
33981	ms.com/...	2012/10/31	729	db	5649
33982	apache.org/...	2012/11/15	4412	programming	5638
33982	apache.org/...	2012/11/15	4412	os	7423



- Deleting topics force you to delete the blog fields too

# Outline

- More about ER & Relational Models
  - Weak Entities
  - Inheritance
- Avoid redundancy & inconsistency
  - **Functional Dependencies**
  - Normal Forms



# Functional Dependency

- FD:  $X \rightarrow Y$ 
  - If two tuples agree in X, then they agree in Y
- What are the FDs for blog\_pages?
  - blogId  $\rightarrow$  ... (key-based)
  - *topic  $\rightarrow$  topicAdmin (non key-based)*

## blog\_pages

blogId	url	created	authorId	topic	topicAdmin
33981	ms.com/...	2012/10/31	729	programming	5638
33981	ms.com/...	2012/10/31	729	db	5649
33982	apache.org/...	2012/11/15	4412	programming	5638
33982	apache.org/...	2012/11/15	4412	os	7423

# Non Key-based FDs

- The root cause of anomalies
- Data redundancy
- Inconsistency

## blog\_pages

blogId	url	created	authorId	topic	topicAdmin
33981	ms.com/...	2012/10/31	729	<i>win prog.</i>	5638
33981	ms.com/...	2012/10/31	729	os	5649
33982	apache.org/...	2012/11/15	4412	programming	5638
33982	apache.org/...	2012/11/15	4412	os	7423



# Outline

- More about ER & Relational Models
  - Weak Entities
  - Inheritance
- Avoid redundancy & inconsistency
  - Functional Dependencies
  - Normal Forms

# Keys

- **Super key**: an attribute or set of attributes that uniquely identifies a tuple within a relation
- **Candidate key**: a super key such that no proper subset is a super key within the relation
  - An attribute that does not occur in any candidate key is called a non-prime attribute
- **Primary key**: the candidate key that is selected to identify tuples uniquely within the relation
  - Candidate keys which are not selected as PK are called alternate keys

# Example

- Candidate keys

blog\_pages

blogId	url	created	authorId	topic	topicAdmin
33981	ms.com/...	2012/10/31 8:59	729	programming	5638
33981	ms.com/...	2012/10/31 4:23	729	db	5649
33982	apache.org/...	2012/11/10 19:06	4412	programming	5638
33982	apache.org/...	2012/11/15 23:11	4412	os	7423

# Normal Forms

- 1<sup>st</sup> normal form:
  - Single-valued columns
- 2<sup>nd</sup> normal form:
  - All fields depends on the primary key
- BCNF normal form:
  - For every FD  $X \rightarrow Y$ ,  $X$  is a super key
- 3<sup>rd</sup> normal form:
  - For every FD  $X \rightarrow Y$ ,  $X$  is a super key *or  $Y$  is a prime attribute*
  - Weaker than BCNF

# 3<sup>rd</sup> Normal Form?

blog\_pages

blogId	url	created	authorId	topic	topicAdmin
33981	ms.com/...	2012/10/31 8:59	729	programming	5638
33981	ms.com/...	2012/10/31 4:23	729	db	5649
33982	apache.org/...	2012/11/10 19:06	4412	programming	5638
33982	apache.org/...	2012/11/15 23:11	4412	os	7423

- FD: topic → topicAdmin
  - Topic is not a superkey
  - TopicAdmin is not a prime attribute
- No!

# 3<sup>rd</sup> Normal Form

## blog\_pages

blogId	url	created	authorId	topicId
33981	ms.com/...	2012/10/31	729	13
33981	ms.com/...	2012/10/31	729	14
33982	apache.org/...	2012/11/15	4412	13
33982	apache.org/...	2012/11/15	4412	14

## topics

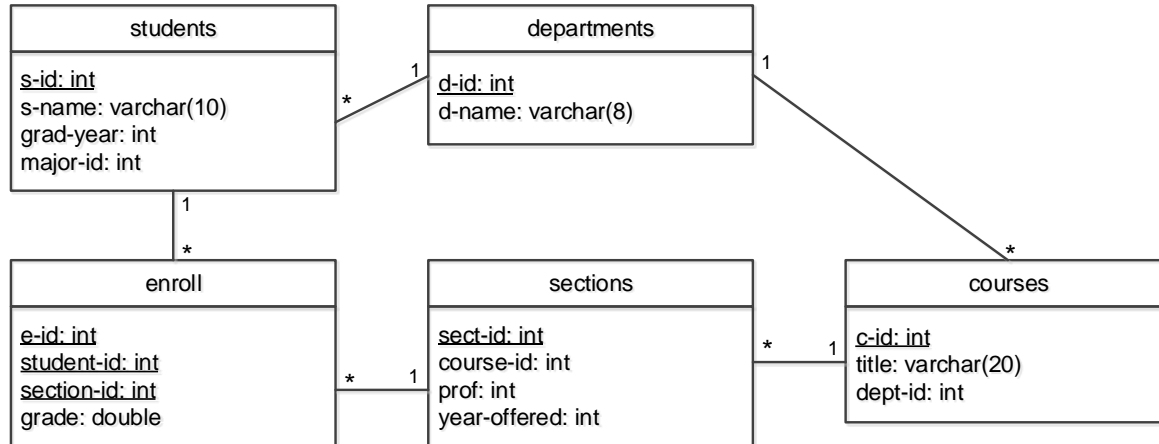
topicId	name	admin
13	programming	5638
14	os	5649
15	db	5638
16	alg	7423

- Move non key-based FDs to new tables
- Avoids redundancy & inconsistency



# BCNF Normal Form (1/2)

- Recall  
student DB:



- Suppose we have “sections” relation like this:

## sections

course-id	prof-id	Prof-name	...
13	271	“Brandon”	...
13	283	“Jerry”	...

# BCNF Normal Form (2/2)

sections

course-id	prof-id	prof-name	...
13	271	"Brandon"	...
13	283	"Jerry"	...

- “sections” is in 3<sup>rd</sup> normal form
- FDs:
  - prof-id  $\rightarrow$  prof-name
  - prof-name  $\rightarrow$  prof-id
- **Not** in BCNF normal form
  - prof-name is not a super key

# Normalized $\neq$ Well-Designed

- Norm forms help reducing redundancy & avoiding inconsistency
- At the cost of lowered query speed
  - Due to Joins
- If query speed is the system bottleneck, don't be afraid to used a *denormalized* schema

# Assigned Reading

- Chaps 2 and 3 on ER & relational models
- Chap 19 on FDs and normal forms

