CSc 337

HI, THIS IS YOUR SON'S SCHOOL. WE'RE HAVING SOME COMPUTER TROUBLE.

OH, DEAR – DID HE BREAK SOMETHING?

IN A WAY –

DID YOU REALLY NAME YOUR SON Robert'); DROP TABLE Students;-- ?

OH. YES. LITTLE BOBBY TABLES, WE CALL HIM.

WELL, WE'VE LOST THIS YEAR'S STUDENT RECORDS. I HOPE YOU'RE HAPPY.
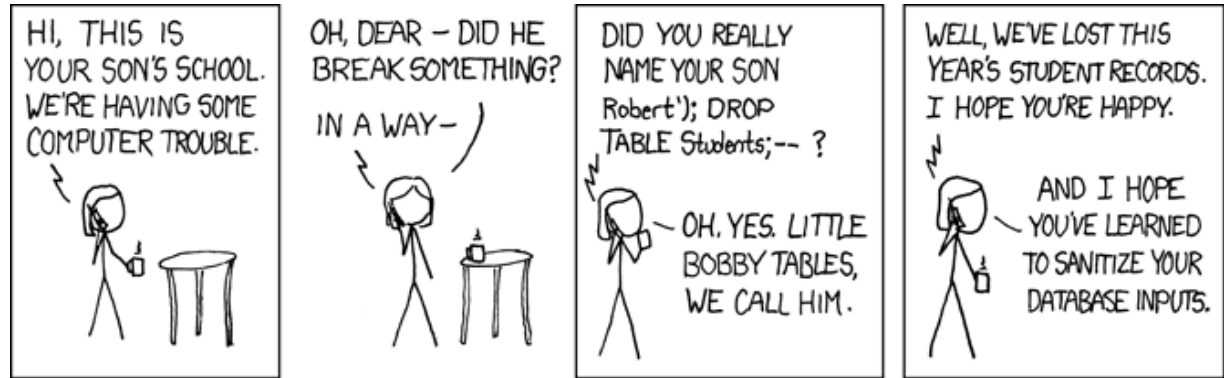
AND I HOPE YOU'VE LEARNED TO SANITIZE YOUR DATABASE INPUTS.

LECTURE 22: CREATING A DATABASE AND MORE JOINS

Creating a database

In the command line in mysql:

```
CREATE DATABASE name;
```

To get to your database:

```
USE name
```

Creating a Database

```
CREATE TABLE name (
    columnName type constraints,
    ...
    columnName type constraints
);
CREATE TABLE students (
    sid INTEGER UNSIGNED NOT NULL PRIMARY KEY,
    name VARCHAR(20),
    email VARCHAR(32)
);
```

- adds/deletes an entire new table from this database

- you can add constraints such as NOT NULL for a field that cannot be blank or PRIMARY KEY for a column that must be unique for every row

- related commands: CREATE DATABASE, DROP TABLE, ALTER TABLE

Inserting into a database

```
INSERT INTO table VALUES ('value1', 'value2', …);
```

example.sql

```sql
create table zipcodes (
   zip integer(5) primary key,
   city  varchar(30),
   State varchar(20));

create table employees (
   eno  varchar(10) primary key,
   ename  varchar(30),
   zip  integer(5) references zipcodes,
   hire_date date);

insert into zipcodes values (98225, 'Bellingham', 'WA');
insert into zipcodes values (95388, 'Winton', 'CA');
insert into zipcodes values (44242, 'Stow', 'OH');
insert into zipcodes values (61536, 'Hanna city', 'IL');
insert into zipcodes values (01254, 'Richmond', 'MA');
insert into zipcodes values (95124, 'San Jose', 'CA');
insert into zipcodes values (95382, 'Turlock', 'MA');
insert into zipcodes values (95380, 'Turlock', 'CA');


insert into employees values ('P0239400', 'Jones Hoffer',98225, '2000-12-12');
insert into employees values ('P0239401', 'Jeffrey Prescott',95388, '2006-01-01');
insert into employees values ('P0239402', 'Fred NcFaddeb',95124, '2008-09-01');
```

SQL data types

- BOOLEAN

- INTEGER

- FLOAT

- VARCHAR : a string

- DATE, TIME, DATETIME

- BLOB : binary data

- quick reference

Database Design 1

name	email	course	grade
Bart	bart@fox.com	Computer Science 142	B-
Bart	bart@fox.com	Computer Science 143	C
Milhouse	milhouse@fox.com	Computer Science 142	B+
Lisa	lisa@fox.com	Computer Science 143	A+
Lisa	lisa@fox.com	Computer Science 190M	A+
Ralph	ralph@fox.com	Informatics 100	D+

what's good and bad about this design?

Database Design 2

id	name	email
123	Bart	bart@fox.com
456	Milhouse	milhouse@fox.com
888	Lisa	lisa@fox.com
404	Ralph	ralph@fox.com

id	name
10001	Computer Science 142
10002	Computer Science 143
10003	Computer Science 190M
10004	Informatics 100

student_id	course_id	grade
123	10001	B-
123	10002	C
456	10001	B+
888	10002	A+
888	10003	A+
404	10004	D+

splitting data into multiple tables avoids redundancy

normalizing: splitting tables to improve structure and remove redundancy / anomalies

normalized tables are often linked by unique integer IDs

Related Tables and Keys

id	name	email
123	Bart	bart@fox.com
456	Milhouse	milhouse@fox.com
888	Lisa	lisa@fox.com
404	Ralph	ralph@fox.com

id	name
10001	Computer Science 142
10002	Computer Science 143
10003	Computer Science 190M
10004	Informatics 100

student_id	course_id	grade
123	10001	B-
123	10002	C
456	10001	B+
888	**10002**	A+
888	**10003**	A+
404	10004	D+

records of one table may be associated with record(s) in another table
- record in Student table with student_id of 888 is Lisa Simpson's student info
- records in Grade table with student_id of 888 are Lisa Simpson's course grades

primary key: a table column guaranteed to be unique for each record

Designing a query

- Figure out the proper SQL queries in the following way:

  - Which table(s) contain the critical data? (`FROM`)

  - Which columns do I need in the result set? (`SELECT`)

  - How are tables connected (`JOIN`) and values filtered (`WHERE`)?

- Test on a small data set (`imdb_small`).

- Confirm on the real data set (`imdb`).

- Try out the queries first in the query tool.

- Write the NodeJS code to run those same queries.

  - Make sure to check for SQL errors at every step!!

Example imdb database

id	first_name	last_name	gender
433259	William	Shatner	M
797926	Britney	Spears	F
831289	Sigourney	Weaver	F
...			

actors

id	name	year	rank
112290	Fight Club	1999	8.5
209658	Meet the Parents	2000	7
210511	Memento	2000	8.7
...			

movies

actor_id	movie_id	role
433259	313398	Capt. James T. Kirk
433259	407323	Sgt. T.J. Hooker
797926	342189	Herself
...		

roles

movie_id	genre
209658	Comedy
313398	Action
313398	Sci-Fi
...	

movies_genres

id	first_name	last_name
24758	David	Fincher
66965	Jay	Roach
72723	William	Shatner
...		

directors

director_id	movie_id
24758	112290
66965	209658
72723	313398
...	

movies_directors

- also available, `imdb_small` with fewer records (for testing queries)

IMDb table relationships / ids
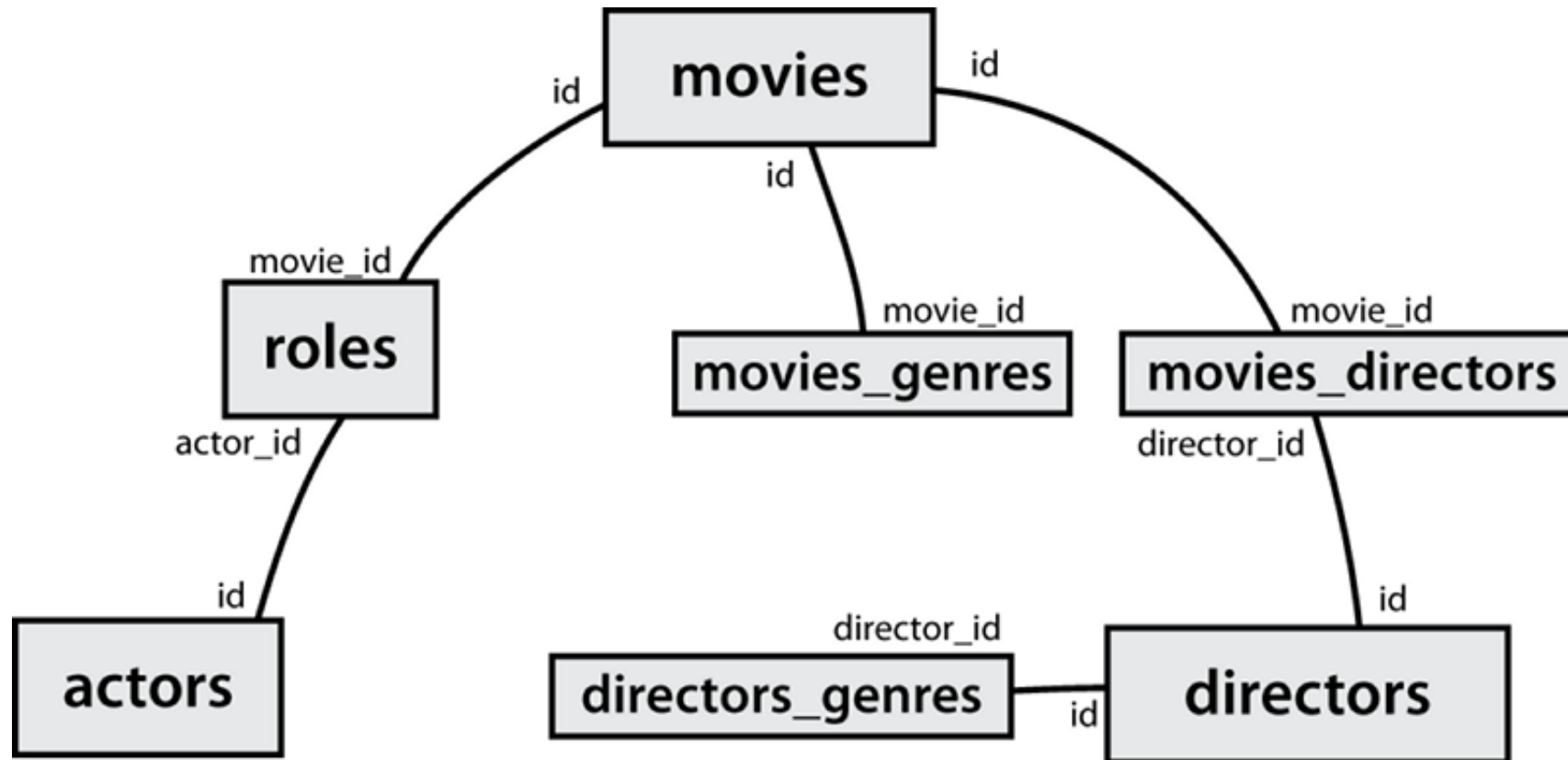
IMDb practice queries

- What are the names of all movies released in 1995?

- How many people played a part in the movie "Lost in Translation"?

- What are the *names* of all the people who played a part in the movie "Lost in Translation"?

- Who directed the movie "Fight Club"?

- How many movies has Clint Eastwood directed?

- What are the *names* of all movies Clint Eastwood has directed?

- What are the names of all directors who have directed at least one horror film?

- What are the names of every actor who has appeared in a movie directed by Christopher Nolan?