

INF1343, Winter 2011

Data Modeling and Database Design

Yuri Takhteyev
University of Toronto



This presentation is licensed under Creative Commons Attribution License, v. 3.0. To view a copy of this license, visit <http://creativecommons.org/licenses/by/3.0/>. This presentation incorporates images from the Crystal Clear icon collection by Everaldo Coelho, available under LGPL from <http://everaldo.com/crystal/>.

Week 12

“NoSQL” Databases
(A 30 Minute Introduction)
+ a Review Session

Backup

mysqldump

```
mysqldump -u kenobiol -p starwars > starwars.sql  
mysql < starwars.sql
```

(Add "create database starwars; use starwars;")

hot backup

replication

off-site copies

The Relational Model

What's not to like?

Scaling

Facebook

.5 billion users

30 billion pieces of content per month

20 billion events *per day*

= 200,000 events per second

```
select url, count(*) from status join  
liking on status.id = liking.status_id  
group by status.url;
```

Instead:

[http://www.facebook.com/video/video.php?
v=707216889765&oid=9445547199&comments](http://www.facebook.com/video/video.php?v=707216889765&oid=9445547199&comments)

“NoSQL” Storage

Basic ideas:

Key-Value Storage & Retrieval

Documents or sparse relations

Schema-free

Distributed storage

“Wide Column Store”:

BigTable, HBase, Cassandra

“Document Store”:

CouchDB, MongoDB

MapReduce

Map

split the task into many parts

all parts should be equivalent

e.g.: count “likes” by URL for batches
of 1 mln hits

Reduce

put together the results

e.g: merge the batch counts into a total
count

Cf. with Rel. Model

Pros:

horizontal scaling

Cons:

more work

“you are not Google”

The Final Exam

Date: Monday, April 4

Place: Bissel **Rm. 114** (basement)

Time: 9:10 am (arrive by 9 am)

Duration: 2 hours 50 minutes

Materials: pencils, erasers

See the practice exam on the course website.

Kensington Airlines

Design a database to record flight bookings, including seat assignments. Assume a single airline.

Level 1: Just handle direct flights.

Level 2: Handle connections.

Level 3: Several people on 1 itinerary.

Hint: 4, 5, and 6 entities respectively.

Normalization

2NF: 1NF + “all non-key attributes are functionally dependent on the entire primary key”.

3NF: 2NF + “no transitive dependencies” [of non-key attributes via non-key attributes].

BCNF (for comparison): 3NF + “all determinants are candidate keys”.

Normalization

film (film_id,
imdb_id,
english_title, 1NF?
original_title, 2NF?
year,
~~directors,~~ 3NF?
genre,
duration,
poster_image_file_small,
poster_image_file_large)

Normalization

ingredient (ingredient_id,
english_name,
french_name,
description,
vegan?,
vegetarian?)

Normalization

```
rating ( url,  
         userid,  
         username,  
         rating )
```

```
user (   userid,  
        username,  
        password )
```

Normalization

```
rating ( url,  
        userid,  
        username,  
        admin?  
        rating )
```

Normalization

```
rating ( rating_id,  
         url,  
         userid,  
         username,  
         admin?  
         rating )
```


Q & A