

JSON Stores Lab (MongoDB)

Ioana Manolescu

INRIA Saclay

ioana.manolescu@inria.fr

<http://pages.saclay.inria.fr/ioana.manolescu/>

M2 Data and Knowledge

Université de Paris Saclay

MongoDB operations

1. Replication group

- Create working directories for 3 MongoDB servers
- Create a replication set for a collection named **social**
- Launch 3 MongoDB servers (in different shells). The server program is **mongod**. Leave those shells alone.
- Connect a client (**mongo**) to one server. Through the client, initialize the replication: add the other replica server, and the arbiter.
- Identify the master from the outputs in the servers' shell and by requesting replica set information from the servers.
- Continue the lab through a client connected to the master.

MongoDB operations

2. Create a set of JSON documents, where each document is a social network user (you may use a function). Each user must have:
 - name
 - a stream (array) of messages, some of which may have comments)
 - a set of interests (each interest is a string)
 - a city and country of residence (use some occurring in https://raw.githubusercontent.com/sevenweeks/databases/master/chap5-mongo/mongo_cities1000.json)
 - a set of friends (pointers to other users).
3. Write a function that returns all users having a certain interest

MongoDB operations

4. For every user x , find every user y having at least an interest in common with x , and make y a friend of x .
5. Import the city documents from https://raw.githubusercontent.com/sevenweeks/databases/master/chap5-mongo/mongo_cities1000.json in a **cities** collection (you may use mongoimport)
6. Based on the social and cities collections, find the 3 users in the social network nearest to a given user (identified by its name).

MongoDB geographic search; replication

7. Find the top-k cities closest to Paris, where k in {1, 5, 10, 20, 50, 100}; note down the time each search took (you may use system profiling)
8. Build a 2Dsphere index on the cities, redo the measures above. Report the two number series in a graph.
9. Stop (ctrl-c) the master; note down the output of the two other servers.
10. Start two shard servers; shard the cities by the country.