The National School of Artificial Intelligence
المدرسة الوطنية العليا للذكاء الاصطناعي
Intelligent Systems Engineering Department
2023/2024

Advanced DataBase Labs
Third Year Semester 6
Lab 9

MongoDB: Exploring Data Manipulation with Robomongo and MapReduce

MongoDB is designed to handle large volumes of data and is widely used in modern web applications. Within the *MongoDB* environment, there are two key concepts that enhance its functionality: *Robomongo* and *MapReduce*.
Robomongo provides a user-friendly *GUI* for interacting with *MongoDB* databases. For simplifying database management tasks (*querying, inserting, updating, and deleting*) by offering this intuitive interface.
MapReduce, on the other hand, is a powerful data processing technique used in *MongoDB* for tasks *(Data Aggregation, Transformation, and Analysis)*. It breaks down large data sets into smaller parts, processes them in parallel across multiple nodes, and then combines the results to generate the final output.
This Lab aims to analyze data management using *Robomongo* and *MapReduce*.

(1) Useful Notions

(a) Robomongo

i) *Robomongo* is a *GUI* tool for managing MongoDB databases. It allows users to interact with *MongoDB* databases visually, browse collections, view documents, and execute queries more easily.
Some features of *Robomongo*:

- Which is known as well as *Robo 3T*, is a free, open-source *GUI*, now *Robo 3T* is extended and replaced by *Studio 3T*.
- It implants the same *JavaScript* engine that powers *MongoDB's shell*, providing a familiar experience while developing.
- A *cross-platform*, it works on *Windows*, *MacOS*, and *Linux.Installation and Documentation* are available from the link https://robomongo.org/.

(b) MapReduce

i) *MapReduce* is a programming (*Map-Reduce Paradigm*) model used for processing and generating large datasets. It enables parallel processing of data across multiple nodes in a cluster, making it efficient for tasks like data aggregation, transformation, and analysis in MongoDB. It is one of *MongoDB*'s most complex query mechanisms.

ii) It is based on the specification of two functions, *Map* and *Reduce (written in JavaScript)*. Both functions are *user-defined functions*.

iii) In *MongoDB*, the *Map* function is used to generate *key-value* documents and pass them as *input* to *Reduce*. The function takes no parameters, and accesses the analyzed object via the this operator. The function can emit pairs via emit*(key, value)* as many times as required in the function.

iv) The *Reduce* function returns the result aggregated from these input documents. The function takes two parameters *key and values (array of key values)*. It can be called several times for the same key, so it must return a value of the same type as those in the array.

v) A *third parameter*, a *JSON object* (*JSON literal*) , represents the function's options. The main option *(out)* is the collection in which the result will be placed. If you want to see the result without storing it, you can specify out: { *inline: 1* }

Important Remark

Restoring a Mongo database:

- *From the command prompt, access the Mongodb bin folder:* `cd D:\MongoDB\bin`.
- *The restore command is: mongorestore -d databaseName pathToDatabase*
  Where: mongorestore is the DB restore tool.
- *In our Lab Run:* `mongorestore -d gym D:\MongoDB\gym`

(2) Lab Work

i) Download the Mongo base *gym*, and extract the archive into the `D:\MongoDB\` folder.

ii) Restore Mongo *gym* database.

1- Download *Studio 3T*. Extract the archive and install it. The *RoboMongo* virtual tool lets you connect to *Mongo* databases on any server, in our case, the server is local.

2- Run *Studio 3T* and connect to the local server. It will display your *Mongo* databases.

3- Connect to the *gym* database.

iii) Answer the following queries using *RoboMongo*:

1- Which athletes (username, surname and first name) are aged between 20 and 30?

2- Which gyms in *Villetaneuse* or *Sarcelles* have a surface area of over 400 m^2?

3- Which athletes (ID and name) play handball?

4- In which gyms and on which days are handball sessions held?

5- In which gyms can field hockey be played on Wednesdays after 3 p.m.?

6- Which athletes (ID and name) do not play any sports?

7- Which gyms have no Sunday sessions?

8- Which gyms only offer basketball or volleyball?

9- Which coaches are also players?

10- Which athletes are advisors?

11- For the athlete *Kervadec*, what is the name of his advisor?

12- Which coaches train handball and basketball?

13- What is the average age of basketball players?

14- Who are the youngest athletes?

15- Which gymnasiums in *Stains* or *Montmorency* have the largest surface area?

16- Which coaches only train handball or basketball?

17- Which athletes have no advisors?

18- For each Stains gym, give the opening and closing times.

19- Which gyms have more than 15 sessions on Wednesdays?

iv) Answer using the *Map-Reduce paradigm*:

1- Calculate the number of gyms for each city.

2- Calculate the number of sessions for each day of the week.

3- Do the same for each sport.

4- Calculate the average surface area of gyms for each city. To do this, you need to calculate the sum of the surface areas and the number of gyms (to be included in the same object and reduced to take account of this double aspect).

(3) Homework

By the end of each Lab, a report should be done, that contains all the details of the elaborate work during the Lab session. The report is going to be submitted later on. **You can work in groups of three maximum**.