Models and Systems for Big Data
Document Database
MongoDB & Javascript-based Query Language
Pr. Nacéra Seghouani

The purpose of this practical work is to use MongoDB Community as NoSQL document-oriented database [1] server. Studio 3T [2] is required as a client to connect to a MongoDB server (instance) and to submit queries. We will use a dataset that consists of a collection of movies (the json file is available on Edunao).

1 HOW TO START WITH MONGODB, DATA IMPORT AND STUDIO3T

First, you need to install MongoDB Community. The instructions are available at:

- https://docs.mongodb.com/manual/tutorial/install-mongodb-on-os-x/ for macOS, all programs are available in /usr/local/.../mongodb/4.0.4_1/bin (if you follow carefully the installation the repository is explicitly displayed)

- and https://docs.mongodb.com/v3.2/tutorial/install-mongodb-on-windows/ for Windows.

Then, you install the client Studio 3T using the link https://studio3t.com/download/.
Here are the instructions to launch an instance of MongoDB server :

- Create a directory "data" in "/PATH" where "PATH" is the path you used.

- Open a terminal to run an instance of MongoDB server by typing the following command. **Keep this terminal open** otherwise the launched instance will be stopped.

  ```
  mongod --dbpath /PATH/data --port 27018
  ```

  The option `port` is used to specify the port the server listens to. The option `dbpath` is used to specify the directory where the data files will be stored. Each MongoDB instance manages its own data directory. For each MongoDB instance you must specify a new data directory.

---

[1] https://www.mongodb.com
[2] https://studio3t.com/features/

- Open another terminal to import the data `moviesEmbedded.json` using the following command.

  ```
  mongoimport --db movie --collection moviesEmbedded --drop
  --jsonArray --file  /PATH/moviesEmbedded.json --port 27018
  ```

  Where `PATH` is the path to access to `moviesEmbedded.json` file you downloaded. The option `db` creates a database named `movie`, the option `collection` creates the collection named `moviesEmbedded` which will store the documents imported from file `moviesEmbedded.json`. The option `drop` removes the collection if it already exists.

- To query MongoDB database, we will use Studio 3T's through IntelliShell. Launch Studio 3T and create a new connection to MongoDB server by specifying its location as `localhost` and the port number 27018. Then open intellishell to edit/execute queries.

- Here are some useful queries: to get the list of the databases `show databases`, to access to the target database `use movie`, to show the collections located in the current database `show collections`, to count the number of documents in the target collection `db.moviesEmbedded.count()`. You can also shutdown the server (as admin) using: `use admin` and `db.shutdownServer()`