

Lab 12 (PA5)

MongoDB

Assigned: 12/4

Due: 12/10

1. Introduction

In this assignment, you will use a NoSQL database system called MongoDB.

More about MongoDB and JSON you can find [here](#) and [here](#).

2. Setup a database

1. Download and install MongoDB (see more [here](#) or [here](#).) Some of you may have difficulty on the data directory and you can take a look on [here](#)

2. Download a restaurants dataset at <https://raw.githubusercontent.com/mongodb/docs-assets/primer-dataset/primer-dataset.json>

3. Download a restaurants dataset at <http://media.mongodb.org/zips.json>

4. Use [mongoimport](#) to import the restaurants dataset into MongoDB:

```
mongoimport --db test --collection restaurants --drop --file primer-  
dataset.json
```

```
mongoimport --db test --collection zips --drop --file zips.json
```

(\ ⓘ may need to specify the full directory of the file)

You can find more instructions [here](#)

City.

5. You can use [MongoDB compass](#) to simplify your work

2. A small demo for MongoDB

Useful functions:

- Show all city names:

```
db.zips.distinct("city")
```

- Create an index on loc:

```
db.zips.createIndex({"loc":"2d"})
```

- Find all the cities that have the name GOSHEN:

```
db.zips.find({"city":"GOSHEN"})
```

- Return all the tuples that OST appears somewhere in the city name:

```
db.zips.find({"city": /. *OST. */})
```

- Show all the cities named BOSTON:

```
db.zips.aggregate([{$match: {"city":"BOSTON"}}])
```

3. Queries in MongoDB

After importing the data, write queries in MongoDB to answer the following questions (try to use [aggregation](#) pipelines when possible):

RESTAURANTS Table

(1) Find the restaurant ID of “Caffe Dante”.

(2) Find all restaurants whose name has “Steak” in it, return only the restaurant’s ids and names.

(3) Find the names of all restaurants that serve either Italian or American cuisine and are located in the Brooklyn borough. (the result of cuisine is just "Italian" or "American")

(4) Return the list of boroughs ranked by the number of American restaurants (cuisine with the word "America") in it. This is, for each borough, find how many restaurants serve American cuisine and print the borough and the number of such restaurants sorted by this number in descending order.

(6) Consider a rectangle area on the location field, in which the vertices are $[-74, 40.5]$, $[-74, 40.7]$, $[-73.5, 40.5]$ and $[-73.5, 40.7]$. Find the number of restaurants in this area that have received a grade score (at least one) more than 70.

Hint: Use the `$geoWithin` and `$box`.

ZIPs Table

(7) Find the top 10 zipcodes with the largest population, return the zipcode, the city name and the state

(8) Find the location of the largest city in each state, return the location, the city name and the state

(9) Find the states where the average population of cities is larger than 10000, return the population and the state

(10) Find the top 5 cities nearest to $[-70, 40]$, return only the city name

Hint: Use the `$near`

Write down your queries.

4. Submitting your assignment

Create a **javascript** file (ends with `.js`) that contains the queries in MongoDB and your comments (if any) but NOT the results of the queries.

Each query should follow the format `"cursor = db.* ;"`. You can type a single query into multiple lines but make sure a query should always end with a semicolon `'` (Just like how you code javascript)**

Gradescope submission will be open soon.

Course Evaluation

Please leave your feedback on this course on <https://bu.campuslabs.com/courseeval/>.

