

Lab 3: MongoDB

DSCI 551 – Spring 2024

Release: 3/6

Due: 11:59pm, March 13, 2024, Wednesday

Points: 20 (notes this lab is worth twice as much)

Using the three collections in the country database: county, city, and countrylanguage, write a MongoDB query for each of the following questions. You can find the collections in the country-mongodb.zip file under the Week 9 folder in the course website.

Importing Data:

Import the provided country datasets into the 'dsci551' database using the 'mongoimport' command. Make sure to use the correct file names and specify the collection names as 'country', 'city' and 'countrylanguage'.

1. Find out, **for each continent, the number of countries** whose GNP is at least 100,000 and has at least 1 million people.
2. Find out **how many** districts (states) there are in the United States.
3. Find out **how many** countries do not have capital cities, using the fact that the Capital field of such a country takes a value of 0.
4. Find out **how many** countries do not have capital cities, using the fact that such countries do not join with any city (i.e., their capital city ID does not show up in the city collection).
To use \$count in aggregate, check out the manual:
<https://www.mongodb.com/docs/manual/reference/operator/aggregation/count/>
5. Find out the **top-10 countries** ranked by the number of its cities whose population is more than 100,000. Return **country code and the number of such cities only**.
6. Find the most populated city in the United States, assuming there is only one such city. **Return the city name and its population**.
7. Find out which district (state) in the USA has the largest number of people (who live in the cities of the district). Assume there is only one such district. **Output name of district and its total population**.
8. Find out countries which have English as one of their official languages and spoken by at least 90% of its population. **Return the country code only**.

9. Find out languages spoken in the United States or Canada. Your output should look like:

```
[  
  { language: 'Chinese' },  
  { language: 'English' },  
  { language: 'French' },  
  ...
```

Make sure the languages are listed in the **alphabetical order**.

10. Find out the countries whose GNP is greater than the GNP of China. Note you should first find GNP of China. Use two separate queries. **Return the name of the country and its GNP.**

You are provided with the following sample script that uses a bit of Javascript for this question.

Consider the following person collection:

```
{_id: 1, name: "john", age: 26}  
{_id: 2, name: "bill", age: 27}  
{_id: 3, name: "david", age: 26}  
{_id: 4, name: "mary", age: 26}
```

- Find people at the same age as the person with `_id = 1`.
`var my_age = db.person.find({_id: 1}, {age: 1, _id: 0}).toArray()[0].age`
`db.person.find({age: my_age})` // note `my_age` is a variable

SUBMISSION INSTRUCTIONS:

1. Submit one **txt file** with all your MongoDB queries and the outputs
2. Clearly mark each question with its number. For each question, include the query and the output
3. Late submission will not be accepted
4. A 10% PENALTY will be applied if these submission instructions are not followed