CS585: Big Data Management

Assignment 4

(MongoDB)

Total Points: 30
Release Date: 07/10/2022
Due Date: 07/20/2022 (@ 11:59pm)

References: Lecture notes & MongoDB Manual (in canvas under "Books" directory)

Question 1 [15 Points, 3 Points each sub-question]

Your task is to design a MongoDB database and apply some CRUD (Create/Read/Update/Delete) operations as follows.

Create a collection named "test", and insert into this collection the documents found in this link (10 documents): http://docs.mongodb.org/manual/reference/bios-example-collection/

1) Write a CRUD operation(s) that inserts the following new records into the collection:

```
{
  "_id" : 20,
  "name" : {
    "first" : "Alex",
    "last" : "Chen"
  },
  "birth" : ISODate("1933-08-27T04:00:00Z"),
  "death" : ISODate("1984-11-07T04:00:00Z"),
  "contribs" : [
    "C++",
    "Simula"
  ],
  "awards" : [
    {
      "award" : "WPI Award",
      "year" : 1977,
      "by" : "WPI"
    }
  ]
}
```

```
{
  "_id" : 30,
  "name" : {
    "first" : "David",
    "last" : "Mark"
  },
  "birth" : ISODate("1911-04-12T04:00:00Z"),
  "death" : ISODate("2000-11-07T04:00:00Z"),
  "contribs" : [
    "C++",
    "FP",
    "Lisp",
  ],
  "awards" : [
    {
      "award" : "WPI Award",
      "year" : 1963,
      "by" : "WPI"
    },
    {
      "award" : "Turing Award",
      "year" : 1966,
      "by" : "ACM"
    }
  ]
}
```

2) Report all documents of people who got less than 3 awards or have contribution in "FP"

3) Insert a new filed of type array, called "comments", into the document of "Alex Chen" storing the following comments: "He taught in 3 universities", "died from cancer", "lived in CA"

4) For each contribution by "Alex Chen", say X, list the peoples' names (first and last) who have contribution X. E.g., Alex Chen has two contributions in "C++" and "Simula". Then, the output should be similar to:
   a. {Contribution: "C++",
      People: [{first: "Alex", last: "Chen"}, {first: "David", last: "Mark"}]},
      { Contribution: "Simula",
        ….}

5) Report the distinct organization that gave awards. This information can be found in the "by" field inside the "awards" array. The output should be an array of the distinct values, e.g., ["wpi', "acm', …]

Question 2 [6 Points, 3 Points each sub-question]

As a continuation over the dataset used in Question 1, answer the following aggregation queries:

1) Write an aggregation query that group by the award name, i.e., the “award” field inside the “awards” array and reports the count of each award.

2) Write an aggregation query that groups by the birth year, i.e., the year within the “birth” field, and report an array of _ids for each birth year.

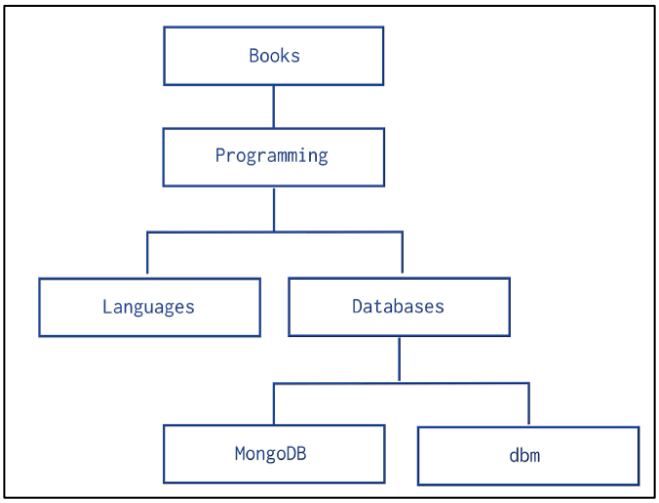

Figure 1: Tree Structure Relationships

Question 3 [9 Points, 3 Points each sub-question]

1) Assume we model the records and relationships in Figure 1 using the Parent-Referencing model (Slide 4 in MongoDB-3). Write a query to report the ancestors of "MongoDB". The output should be an array containing values [{Name: "Databases", Level: 1},
{Name: "Programming", Level: 2},
{Name: "Books", Level: 3}]

Note: *"Level" is the distance from "MongoDB" node to the other node. It should be computed in your code*

2) Assume we model the records and relationships in Figure 1 using the Parent-Referencing model (Slide 4 in MongoDB-3). You are given only the root node, i.e., _id = "Books", write a query that reports the height of the tree. (It should be 4 in our case).

3) Assume we model the records and relationships in Figure 1 using the Child-Referencing model (Slide 9 in MongoDB-3). Write a query to report the descendants of "Books". The output should be an array containing values ["Programming", "Languages", "Databases", "MongoDB", "dbm"]

What to Submit (for each student)

- You will submit a single zip file containing all problems and the queries' answers
- The zip file should also include a "Readme.pdf" file. In this file include:
    - Any assumptions that you have made
    - Create a table as shown below and fill it up

Question	Status (Select one) Fully Working/ Partially Working/ Not Working	Comment
Q1		
Q2		
Q3		

    - Any comments you would like to provide regarding your code.

How to Submit

- Use the Canvas system to submit your files.