

Internal Assignment 1

Student Id: 2048494

Student Name: Dhaniswar B.K

Module Leader: Jnaneshwar Bohara

Tutor: Sujan Shrestha

Section: L6CG5

Submitted on : 17/03/2022

Table of Contents

| | |
|---|----|
| 1. Dividing my university id by 3..... | 3 |
| 2. Examine your dataset and carry out the following tasks: | 3 |
| 2.1. Starting the mongo server | 3 |
| 2.2. Import the data into MongoDB database:..... | 4 |
| 2.3. Successfully importing the dataset into MongoDB | 4 |
| 2.4. Write a command to show how many documents are in your collection | 4 |
| 2.5. Analyse the data write a command to: | 5 |
| 2.6. Show one document | 5 |
| 2.7. Show the unique values in one field..... | 6 |
| 2.8. Shows a set of documents based on some criteria. Output just two fields from the document..... | 6 |
| 2.9. Use a regular expression to search for some criteria. The search should be case insensitive..... | 7 |
| 2.10. Reshape the collection Write a command to:..... | 7 |
| 2.11. Create a new collection based on a subset of the dataset..... | 8 |
| 2.12. Include a query to show a document from the new collection | 9 |
| 3. Name one advantage to using this approach for handling Big Data and include brief explanation of why you think this is an advantage..... | 10 |
| 4. Name one disadvantage to using this approach for handling Big Data and include brief explanation of why you think this is a disadvantage. | 11 |

1. Dividing my university id by 3.

Answer: calculating the remainder of the university id using short trick

Step 1 $\Rightarrow 2048494 = 2+0+4+8+4+9+4 \Rightarrow 31 \Rightarrow 3+1 = 4$

Now, $4\%3 = 1$ So, I have to use Nepal_Cricket.json data set for this assessment.

2. Examine your dataset and carry out the following tasks:

2.1. Starting the mongo server

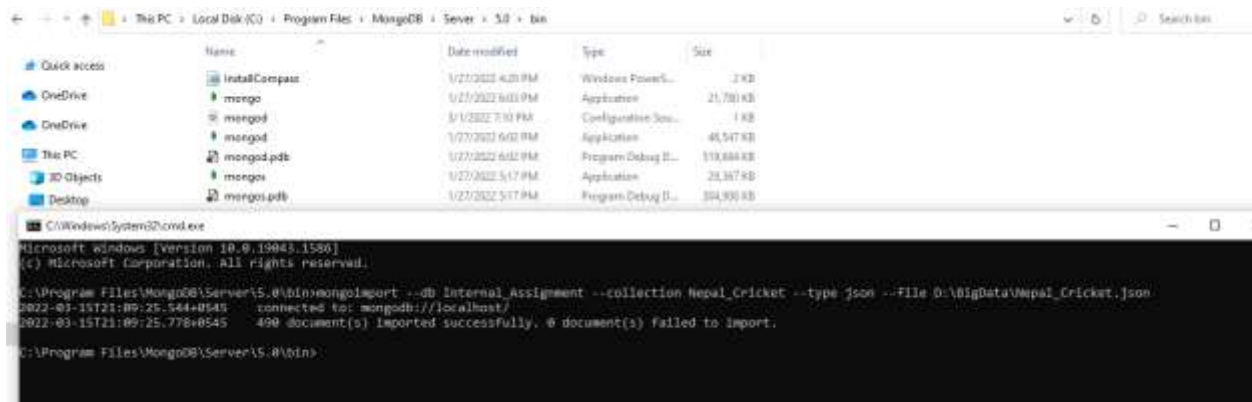
```
Microsoft Windows [Version 10.0.19043.1586]
(c) Microsoft Corporation. All rights reserved.

C:\Users\Acer>mongo
MongoDB shell version v5.0.6
connecting to: mongodb://127.0.0.1:27017/?compressors=disabled&gssapiServiceName=mongodb
Implicit session: session { "id" : UUID("ef51d35a-602f-45c3-a561-fd7831f0bdb3") }
MongoDB server version: 5.0.6
=====
Warning: the "mongo" shell has been superseded by "mongosh",
which delivers improved usability and compatibility. The "mongo" shell has been deprecated and will be removed in
an upcoming release.
For installation instructions, see
https://docs.mongodb.com/mongodb-shell/install/
=====
---
The server generated these startup warnings when booting:
  2022-03-15T14:14:51.325+05:45: Access control is not enabled for the database. Read and write access to data and
configuration is unrestricted
---
---
  Enable MongoDB's free cloud-based monitoring service, which will then receive and display
metrics about your deployment (disk utilization, CPU, operation statistics, etc).

  The monitoring data will be available on a MongoDB website with a unique URL accessible to you
and anyone you share the URL with. MongoDB may use this information to make product
improvements and to suggest MongoDB products and deployment options to you.

  To enable free monitoring, run the following command: db.enableFreeMonitoring()
  To permanently disable this reminder, run the following command: db.disableFreeMonitoring()
---
>
```

2.2. Import the data into MongoDB database:



2.3. Successfully importing the dataset into MongoDB



2.4. Write a command to show how many documents are in your collection



2.5. Analyse the data write a command to:

2.6. Show one document

```
> db.Nepal_Cricket.findOne()
{
  "_id" : ObjectId("6230efa96ad2b9ba1d065e65"),
  "contributors" : null,
  "truncated" : true,
  "text" : "Today, England is playing against S. Africa in the 1st ODI match & England Cricket Fans from Nepal are going to Don. http://t.co/ue7jHMLN8",
  "is_quote_status" : false,
  "in_reply_to_status_id" : null,
  "id" : NumberLong("1224633582758480083"),
  "favorite_count" : 28,
  "entities" : {
    "symbols" : [ ],
    "user_mentions" : [ ],
    "hashtags" : [ ],
    "urls" : [
      {
        "url" : "https://t.co/ue7jHMLN8",
        "indices" : [
          121,
          144
        ],
        "expanded_url" : "https://twitter.com/i/web/status/1224633582758480083",
        "display_url" : "twitter.com/i/web/status/1.."
      }
    ]
  },
  "retweeted" : false,
  "coordinates" : null,
  "source" : "<a href='\"http://twitter.com/download/iphone\"' rel='\"noFollow\"'>Twitter for iPhone</a>",
  "in_reply_to_screen_name" : null,
  "in_reply_to_user_id" : null,
  "retweet_count" : 2,
  "id_str" : "1224633582758480083",
  "favorited" : false,

```

```
  "user" : {
    "follow_request_sent" : false,
    "has_extended_profile" : false,
    "profile_use_background_image" : false,
    "default_profile_image" : false,
    "id" : NumberLong("3141875413"),
    "profile_background_image_url_https" : "https://abs.twimg.com/images/themes/theme1/bg.png",
    "verified" : false,
    "translator_type" : "none",
    "profile_text_color" : "000000",
    "profile_image_url_https" : "https://pbs.twimg.com/profile_images/777188083445739521/tS0MGfAc_normal.jpg",
    "profile_sidebar_fill_color" : "000000",
    "entities" : {
      "description" : {
        "urls" : [ ]
      }
    },
    "followers_count" : 17610,
    "profile_sidebar_border_color" : "000000",
    "id_str" : "3141875413",
    "profile_background_color" : "000000",
    "listed_count" : 24,
    "is_translation_enabled" : false,
    "utc_offset" : null,
    "statuses_count" : 10232,
    "description" : "Official twitter account for Routine Of Nepal Banda. \"Helping Hand, Entertainment, Up to date information\" Every update matters. follow twitter @RONUpdates..",
    "friends_count" : 1,
    "location" : "",
    "profile_link_color" : "ABBC21",
    "profile_image_url" : "http://pbs.twimg.com/profile_images/777188083445739521/tS0MGfAc_normal.jpg",
    "following" : false,
    "geo_enabled" : true,
    "profile_banner_url" : "https://pbs.twimg.com/profile_banners/3141875413/1485377497",
    "profile_background_image_url" : "http://abs.twimg.com/images/themes/theme1/bg.png",
    "screen_name" : "RONUpdates",
    "lang" : null,
    "profile_background_tile" : false,
    "favourites_count" : 1368,
    "name" : "Routine of Nepal Banda",
  }
}
```

```

    "name": "Routine of Nepal banda",
    "notifications": false,
    "url": null,
    "created_at": "Mon Apr 06 12:30:55 +0000 2015",
    "contributors_enabled": false,
    "time_zone": null,
    "protected": false,
    "default_profile": false,
    "is_translator": false
  },
  {
    "geo": null,
    "in_reply_to_user_id_str": null,
    "lang": "en",
    "created_at": "Tue Feb 04 09:59:33 +0000 2019",
    "in_reply_to_status_id_str": null,
    "place": null,
    "metadata": {
      "iso_language_code": "en",
      "result_type": "recent"
    }
  }
}

```

2.7. Show the unique values in one field

```

> db.Nepal_Cricket.distinct("favorite_count")
{
  0,
  1,
  2,
  3,
  4,
  5,
  6,
  8,
  9,
  10,
  12,
  17,
  19,
  20,
  24,
  25,
  26,
  27,
  40,
  49,
  51,
  53,
  75,
  93,
  94,
  121,
  358,
  384
}

```

2.8. Shows a set of documents based on some criteria. Output just two fields from the document.

```

> db.Nepal_Cricket.find({"user.name": "Aacsh Subedi"}, {"_id": 0, "text": 1, "user.profile_background_image_url_https": 1}).pretty()
{
  "text": "RT @RONBupdtes: Special Announcement by Digree bro : After Sandip Lamichhane declared about donating Rs. 8,848 for every wicket he takes..",
  "user": {
    "profile_background_image_url_https": null
  }
},
{
  "text": "RT @ZamSandeep25: I want nepal cricket to be a force to reckon with. To invest in this dream, I will be donating 8,848 Rupees for every wic..",
  "user": {
    "profile_background_image_url_https": null
  }
}

```

2.9. Use a regular expression to search for some criteria. The search should be case insensitive

```
db.Nepal_Cricket.find({'text':/Sandip/i},{_id:0, text:1}).pretty()

  "text" : "RT @RONBupdates: Special Announcement by Jigree bro : After Sandip Lamichhane declared about donating
t he takes,..."

  "text" : "RT @RONBupdates: Special Announcement by Jigree bro : After Sandip Lamichhane declared about donating
t he takes,..."

  "text" : "RT @RONBupdates: Special Announcement by Jigree bro : After Sandip Lamichhane declared about donating
t he takes,..."

  "text" : "RT @RONBupdates: Special Announcement by Jigree bro : After Sandip Lamichhane declared about donating
t he takes,..."

  "text" : "RT @RONBupdates: Special Announcement by Jigree bro : After Sandip Lamichhane declared about donating
t he takes,..."

  "text" : "RT @RONBupdates: Special Announcement by Jigree bro : After Sandip Lamichhane declared about donating
t he takes,..."

  "text" : "RT @RONBupdates: Special Announcement by Jigree bro : After Sandip Lamichhane declared about donating
t he takes,..."
```

2.10. Reshape the collection Write a command to:

```
> db.Nepal_Cricket.find({'user.name':"Aacch Subedi"},{_id:0,text:1}).pretty()
{ "text" : "Hello World Python Programming" }
{ "text" : "Hello World Python Programming" }
> db.Nepal_Cricket.updateMany({'user.name':"Aacch Subedi"}, {$set: {text:"Hello World java programming"}})
{ "acknowledged" : true, "matchedCount" : 2, "modifiedCount" : 2 }
> db.Nepal_Cricket.find({'user.name':"Aacch Subedi"},{_id:0,text:1}).pretty()
{ "text" : "Hello World java programming" }
{ "text" : "Hello World java programming" }
```

2.11. Create a new collection based on a subset of the dataset.

```
> db.Nepal_Cricket_new.insertMany([
... {
...   "_id" : ObjectId("6230afa96ad2b9bd0065e65"),
...   "contributors" : null,
...   "truncated" : true,
...   "text" : "Today, England is playing against S. Africa in the 1st ODI match &esp; England Cricket Fans from Nepal are going to Don. https://t.co/us7jHwALNE",
...   "is_quote_status" : false,
...   "in_reply_to_status_id" : null,
...   "id" : NumberLong("1224633582758400003"),
...   "favorite_count" : 20,
...   "entities" : {
...     "symbols" : [ ],
...     "user_mentions" : [ ],
...     "hashtags" : [ ],
...     "urls" : [
...       {
...         "url" : "https://t.co/us7jHwALNE",
...         "indices" : [
...           121,
...           164
...         ]
...       },
...       {
...         "expanded_url" : "https://twitter.com/i/web/status/1224633582758400003",
...         "display_url" : "twitter.com/i/web/status/1..."
...       }
...     ]
...   },
...   "retweeted" : false,
...   "coordinates" : null,
...   "source" : "<a href='\"http://twitter.com/download/iphone\"' rel='\"nofollow\"'>Twitter for iPhone</a>",
...   "in_reply_to_screen_name" : null,
...   "in_reply_to_user_id" : null,
...   "retweet_count" : 2,
...   "id_str" : "1224633582758400003",
...   "favorited" : false,
...   "user" : {
...     "follow_request_sent" : false,
...     "has_extended_profile" : false,
...     "profile_use_background_image" : false,
...     "default_profile_image" : false,
...     "id" : NumberLong("3141875413"),
...     "profile_background_image_url_https" : "https://abs.twimg.com/images/themes/theme1/bg.png",
...     "default_profile" : false,
...     "is_translator" : false
...   },
...   "geo" : null,
...   "in_reply_to_user_id_str" : null,
...   "lang" : "en",
...   "created_at" : "Tue Feb 04 09:59:31 +0000 2020",
...   "in_reply_to_status_id_str" : null,
...   "place" : null,
...   "metadata" : {
...     "iso_language_code" : "en",
...     "result_type" : "recent"
...   }
... })
... ])
{
  "acknowledged" : true,
  "insertedIds" : [
    ObjectId("6230afa96ad2b9bd0065e65")
  ]
}
```

```
...   "profile_background_image_url_https" : "https://abs.twimg.com/images/themes/theme1/bg.png",
...   "verified" : false,
...   "translator_type" : "none",
...   "profile_text_color" : "000000",
...   "profile_image_url_https" : "https://pbs.twimg.com/profile_images/777188083445739521/t5QW6fAc_normal.jpg",
...   "profile_sidebar_fill_color" : "000000",
...   "entities" : {
...     "description" : {
...       "urls" : [ ]
...     }
...   },
...   "followers_count" : 37610,
...   "profile_sidebar_border_color" : "000000",
...   "id_str" : "3141875413",
...   "profile_background_color" : "000000",
...   "listed_count" : 24,
...   "is_translation_enabled" : false,
...   "utc_offset" : null,
...   "statuses_count" : 10232,
...   "description" : "Official twitter account for Routine Of Nepal banda. \"Helping Hand, Entertainment, Up to date Information\" Every update matters.fol
... },
...   "friends_count" : 1,
...   "location" : "",
...   "profile_link_color" : "A888C2",
...   "profile_image_url" : "http://pbs.twimg.com/profile_images/777188083445739521/t5QW6fAc_normal.jpg",
...   "following" : false,
...   "geo_enabled" : true,
...   "profile_banner_url" : "https://pbs.twimg.com/profile_banners/3141875413/1463377497",
...   "profile_background_image_url" : "http://abs.twimg.com/images/themes/theme1/bg.png",
...   "screen_name" : "RONBupdates",
...   "lang" : null,
...   "profile_background_tile" : false,
...   "favorites_count" : 1360,
...   "name" : "Routine of Nepal banda",
...   "notifications" : false,
...   "url" : null,
...   "created_at" : "Mon Apr 06 12:38:55 +0000 2015",
...   "contributors_enabled" : false,
...   "time_zone" : null,
...   "protected" : false,
...   "default_profile" : false,
```


2.12. Include a query to show a document from the new collection

```
> db.Nepal_Cricket_new.find().pretty()
{
  "_id" : ObjectId("6238afa96ad2d9bad865e63"),
  "contributors" : null,
  "truncated" : true,
  "text" : "Today, England is playing against S. Africa in the 1st ODI match &np England Cricket Fans from Nepal are going to Don. https://t.co/ue7jH9UJ8H",
  "is_quote_status" : false,
  "in_reply_to_status_id" : null,
  "id" : NumberLong("1224633582758400003"),
  "favorite_count" : 20,
  "entities" : {
    "symbols" : [ ],
    "user_mentions" : [ ],
    "hashtags" : [ ],
    "urls" : [
      {
        "url" : "https://t.co/ue7jH9UJ8H",
        "indices" : [
          121,
          144
        ],
        "expanded_url" : "https://twitter.com/i/web/status/1224633582758400003",
        "display_url" : "twitter.com/i/web/status/1..."
      }
    ]
  },
  "retweeted" : false,
  "coordinates" : null,
  "source" : "<a href='\"http://twitter.com/download/iphone\"' rel='\"nofollow\"'>Twitter for iPhone</a>",
  "in_reply_to_screen_name" : null,
  "in_reply_to_user_id" : null,
  "retweet_count" : 2,
  "id_str" : "1224633582758400003",
  "favorited" : false,
  "user" : {
    "follow_request_sent" : false,
    "has_extended_profile" : false,
    "profile_use_background_image" : false,
    "default_profile_image" : false,
    "id" : NumberLong("3141875413"),
    "profile_background_image_url_https" : "https://abs.twimg.com/images/themes/theme1/bg.png",
    "profile_background_image_url_https" : "https://abs.twimg.com/images/themes/theme1/bg.png",
    "verified" : false,
    "translator_type" : "none",
    "profile_text_color" : "000000",
    "profile_image_url_https" : "https://pbs.twimg.com/profile_images/777180003445739521/t5QK6f4c_normal.jpg",
    "profile_sidebar_fill_color" : "000000",
    "entities" : {
      "description" : {
        "urls" : [ ]
      }
    },
    "followers_count" : 37610,
    "profile_sidebar_border_color" : "000000",
    "id_str" : "3141875413",
    "profile_background_color" : "000000",
    "listed_count" : 24,
    "is_translation_enabled" : false,
    "utc_offset" : null,
    "statuses_count" : 18032,
    "description" : "Official twitter account for Routine Of Nepal bands. \"Helping Hand, Entertainment, Up to date information\" Every update matters..follow",
    "friends_count" : 1,
    "location" : "",
    "profile_link_color" : "A888C2",
    "profile_image_url" : "http://pbs.twimg.com/profile_images/777180003445739521/t5QK6f4c_normal.jpg",
    "following" : false,
    "geo_enabled" : true,
    "profile_banner_url" : "https://pbs.twimg.com/profile_banners/3141875413/1483377497",
    "profile_background_image_url" : "http://abs.twimg.com/images/themes/theme1/bg.png",
    "screen_name" : "RONHupdates",
    "lang" : null,
    "profile_background_tile" : false,
    "favorites_count" : 1360,
    "name" : "Routine of Nepal bands",
    "notifications" : false,
    "url" : null,
    "created_at" : "Mon Apr 06 12:30:55 +0000 2015",
    "contributors_enabled" : false,
    "time zone" : null,
    "protected" : false,
    "default_profile" : false,
    "default_profile_image" : false,
    "is_translator" : false
  },
  "geo" : null,
  "in_reply_to_user_id_str" : null,
  "lang" : "en",
  "created_at" : "Tue Feb 04 09:59:33 +0000 2020",
  "in_reply_to_status_id_str" : null,
  "place" : null,
  "metadata" : {
    "iso_language_code" : "en",
    "result_type" : "recent"
  }
}
```

3. Name one advantage to using this approach for handling Big Data and include brief explanation of why you think this is an advantage.

✓ Flexible Database

We know that MongoDB is a schema-less database. That means we can have any type of data in a separate document. This thing gives us flexibility and a freedom to store data of different types.

✓ Sharding

People can store a large data by distributing it to several servers connected to the application. If a server cannot handle such a big data then there will be no failure condition. The term we can use here is “auto-sharding”.

✓ Scalability

A great advantage of MongoDB is that it is a horizontally scalable database. When you have to handle a large data, you can distribute it to several machines.

MongoDB is much more than a database. It's a complete application data platform. With MongoDB Atlas, the cloud offering by MongoDB, people have access to a collection of services that all integrate nicely with their database. Amongst other things, people will have:

- The Performance Advisor, which provides you with recommendations to optimize your database.
- Atlas Search, a full-text search engine that uses the same MongoDB Query API as other queries.
- MongoDB Charts, an easy-to-use interface to create stunning dashboards and visualizations.
- Multi-cloud deployment, which is offered out-of-the-box on any major cloud provider.

4. Name one disadvantage to using this approach for handling Big Data and include brief explanation of why you think this is a disadvantage.

Nothing is fully complete or perfect. In spite of having so many advantages, MongoDB has some limitations. Let's discuss some of them here

a. Joins not Supported

MongoDB doesn't support joins like a relational database. Yet one can use joins functionality by adding by coding it manually. But it may slow execution and affect performance.

b. High Memory Usage

MongoDB stores key names for each value pairs. Also, due to no functionality of joins, there is data redundancy. This results in increasing unnecessary usage of memory.

c. Limited Data Size

You can have document size, not more than 16MB.

d. Limited Data Size

MongoDB allows document size, not more than 16MB.