Project 2: Working with MongoDB

Data Science and Engineering
Data Base

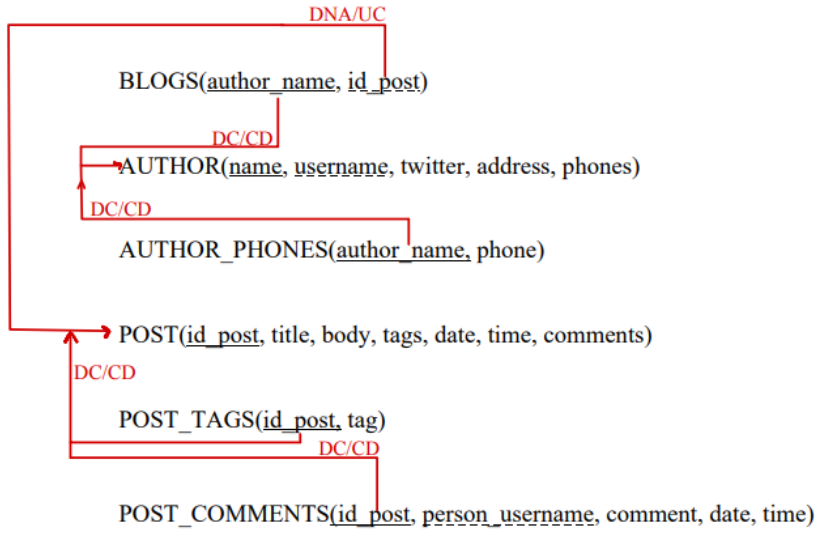
Group 3
Alejandro Leonardo García Navarro
Melania Guerra Ulloa

Table of contents

Exercise 1

Schema Description

- Relational model:



- Mongo structure:

```
({
  author: {name: “”, username: “”, twitter: “”, address: “” phones: [“”]},
  posts:
        {title: “”
         body: “”,
         date: ISODate(“”),
         tags: [“”],
         comments: [{
                  person_username: “”,
                  comment: “”,
                  date: ISODate(“”)}]}
        });
```

- Explanation:

As it was required in the statements, it is of high vitality to explain the scheme with some JSON documents. In our case, the best to do this is by comparing both the relational schema shown above and the mongo structure.

In addition, it is very important to note that, since MongoDB is a non-relational Database (a database that does not use the tabular schema of rows and columns found in most traditional database system) [1], it was decided to implement our database in a single collection, using a key feauture of MongoDB. This is a feature which allows related data to be embedded within a single document.

Once the necessary concepts of MongoDB are explained, it will be proceeded to explain the scheme with an example of an insertion:

```
db.blogs.insertOne({
  author: {name: "Mery Garcia", username: "Mery123", twitter: "@merygarcia#", address: "merygarcia@outlook.es", phones: 299258239},
  posts:
      {title: "Data Science Central",
       body: "Data Science Central does exactly what its name suggests and acts as an online resource hub for just about everything related to data science and big data. The site covers a wide array of data science topics regarding analytics, technology, tools, data visualization, code, and job opportunities. Industry experts contribute discussion and insights about key topics. The site updates frequently, nearly two blog posts a day from contributing writers, and it also offers a community forum for discussion or questions.",
       date: ISODate("2019-05-19T18:42:34Z"),
       tags: ['Branding', 'bread', 'Statistical Learning', 'Data Science', 'Finite Automata', 'Turing', 'pies', 'nosql'],
       comments: [{
            person_username: "chiyembekezo_mohinder",
            comment: "I was completely amazed at the quality of the information found in the blog. Extraordinary and amazing!!",
            date: ISODate("2021-02-14T12:37:18Z")}]}
      });
```

- As stated, the blog is composed of its author and a post. Because of this, one can notice the presence of **author** and **post**.
- With respect to **author**, there are some fields along the lines of:
  1. **name:** name of the author of the blog.
  2. **username**: identification used by an author to access the blog.
  3. **twitter:** represents the twitter username of the author.
  4. **address:** it is the author's email (as stated, there can only be one).
  5. **phones:** the telephone numbers of the author. Please note that the option of having multiple phone numbers was contemplated; that is, in some case you may notice that phones are described like **phones: value1** or **phones: [value1, value2, value3…valuex]**.

- Finally, regarding post, the fields are:
  1. **title:** representing the title of the blog.
  2. **body:** it is a description of an undefined length that serves to get an idea of that the blog is about.
  3. **date:** field to store when the blog was created.
  4. **tags:** a label attached to the blog for the purpose of identification or to give other information.
  5. **comments:** field made up of information related to a comment posted by a user. This field stores information like: **person_username** (allows to know who posted the comment), **comment** (stores the comment made by the user), **date** (it is the date when the comment was made).

Insert statements

The code of insertion is the following, inserting all of them individually:

```
db.blogs.insertOne({
   author: {name: "Mery Garcia", username: "Mery123", twitter: "@merygarcia#", address: "merygarcia@outlook.es", phones: 299258239},
   posts:
            {title: "Data Science Central",
             body: "Data Science Central does exactly what its name suggests and acts as an online resource hub for just about everything related to data science and big data. The site covers a wide array of data science topics regarding analytics, technology, tools, data visualization, code, and job opportunities. Industry experts contribute discussion and insights about key topics. The site updates frequently, nearly two blog posts a day from contributing writers, and it also offers a community forum for discussion or questions.",
             date: ISODate("2019-05-19T18:42:34Z"),
             tags: ['Branding', 'bread', 'Statistical Learning', 'Data Science', 'Finite Automata', 'Turing', 'pies', 'nosql'],
             comments: [{
                        person_username: "chiyembekezo_mohinder",
                        comment: "I was completely amazed at the quality of the information found in the blog. Extraordinary and amazing!!",
                        date: ISODate("2021-02-14T12:37:18Z")}]}
            });

db.blogs.insertOne ({
   author: {name: "Mery Garcia", username: "Mery123", twitter: "@merygarcia#", address: "merygarcia@outlook.es", phones: 299258239},
   posts:
            {title: "Creating a NOSQL Database",
             body: "NoSQL databases emerged in the late 2000s as the cost of storage dramatically decreased. Gone were the days of needing to create a complex, difficult-to-manage data model in order to avoid data duplication. Developers (rather than storage) were becoming the primary cost of software development, so NoSQL databases optimized for developer productivity.",
             date: ISODate("2019-05-06T19:12:24Z"),
             tags: ['database','Python', 'Turing', 'nosql'],
             comments: [{
                        person_username: "alex_123",
                        comment: "This was very useful!",
                        date: ISODate("2022-03-11T19:58:55Z")}]}
            });

db.blogs.insertOne({
   author: {name: "Vosgi Beren", username: "Vosgi_Beren_9", twitter: "@vosgiberen#", address: "vosgiberen@gmail.com", phones: [228837994, 228837995, 228837996]},
   posts:
            {title: "Exploring Visual Similarity with Modista",
             body: "SmartDataCollective.com SmartData Collective is a community site focused on trends in business intelligence and data management. Similar to Data Science Central, it also features insights into data science through contributions by industry experts. Where Data Science Central focuses directly on data science as a whole, SmartData Collective looks at the wider field and how data science can intersect with business.",
             date: ISODate("2016-07-03T13:31:47Z"),
             tags: ['cookies', 'Automata Theory and Compilers', 'Branding', 'Discrete Mathematics', 'mongoDB', 'nosql', 'Turing', 'Data Science', 'brownies'],
             comments: [{
                        person_username: "hui_darby",
                        comment: "I do not relate that much to the fact that good grades mean a good future. Totally disagree. Nonetheless, my experience was both amazing and extraordinary reading it. Keep it up.",
                        date: ISODate("2016-09-12T12:37:18Z")}]}
            });
```

```
db.blogs.insertOne({
   author:{name: "Devon Tulugaq", username: "Devon_Tulugaq#", twitter: "@devontulugaq45", address:
"devontulugaq@hotmail.com", phones: 440104566},
   posts:
          {title: "What's The Big Data?",
           body: "WhatsTheBigData.com What's The Big Data? takes a different approach to data science and focuses on
the impact of big data's growth into the digital behemoth it is today. The blog's founder, Gil Press, is intimately familiar
with big data and data science, having spent a career in data research and now running a consulting practice. In his blog,
Press explores how big data interacts with our lives and impacts everything from technology to business to government
and policy. He provides a source of news and commentary on the sphere of data.",
           date: ISODate("2018-07-03T13:31:47Z"),
           tags: ['pies', 'Branding'],
           comments: [{
                    person_username: "mudiwa_tivoli",
                    comment: "It is amazing how much pies can cause you such a degree of diabetes. I do not believe all
the information exposed here…any doctor in the comments?",
                    date: ISODate("2019-09-21T18:11:34Z")}]}
          });

db.blogs.insertOne({
   author:{name: "Dusty Keone", username: "Dusty_Keone#", twitter: "@dustykeone45", address:
"dustykeone@gmail.com", phones: [689878890, 689878891]},
   posts:
          {title: "insideBIGDATA",
           body: "This blog is slightly different than the others, offering a look directly into the minds of data scientists, as
well as tutorials and news. This is the blog of the data science website Kaggle, which hosts data science projects and
competitions that challenges data scientists to produce the best models for featured data sets. Organizations can post their
data problems with a prize amount and data professionals will enter to solve it. Crowdsourcing ensures that the
experiments are innovative and interesting—and offer a lot of perspectives to learn from. Over 200 competitions have run,
including high profile ones like improving Microsoft Kinect gesture recognition, improving the search for the Higgs
boson at CERN, and the notorious Heritage Health $3 million award for improving predictions regarding which patients
will need to visit hospitals. Kaggle's official blog goes deeper into these competitions, offering interviews with the
winners to discuss their approach to solving the data science problems. The blog also features news and tutorials for all
levels of data science enthusiasts.",
           date: ISODate("2019-02-04T17:32:55Z"),
           tags: ['Statistical Learning', 'nosql', 'Finite Automata', 'Discrete Mathematics', 'Data Science'],
           comments: [{
                    person_username: "pema_leilani",
                    comment: "The quality of the information I found in the blog absolutely astounded me. What an
extraordinary source of valid info!",
                    date: ISODate("2020-08-03T18:22:50Z")}]}
          });

db.blogs.insertOne({
   author:{name: "Zhi Tiriaq", username: "Zhi_Tiriaq45", twitter: "@zhitiriaq%", address: "zhitiriaq@gmail.com",
phones: 292783279},
   posts:
          {title: "No Free Hunch",
           body: "InsideBIGDATA focuses on the machine learning side of data science. It covers big data in IT and
business, machine learning, deep learning, and artificial intelligence. Guest features offer insight into industry
perspectives, while news and Editor's Choice articles highlight important goings-on in the field. All the articles are neatly
categorized by topic to zero in on any subject in particular. The blog also maintains a host of resources for events, jobs,
and research reports, and more. This is a resource for anyone wanting to stay up to date with machine learning.",
           date: ISODate("2017-03-04T13:31:47Z"),
           tags: ['Discrete Mathematics', 'Video Marketing', 'brownies'],
           comments: [{
                    person_username: "ejiro_ailbhe",
                    comment: "Statistics are like bikinis. What they reveal is suggestive, but what they conceal is vital",
                    date: ISODate("2018-04-08T18:22:50Z")}]}
          });
```

```
db.blogs.insertOne({
   author: {name: "Kajal Casey", username: "Kajal_Casey12", twitter: "@kajalcasey_9", address:
"kajalcasey@hotmail.com", phones: [885024055, 885024022]},
   posts:
            {title: "Learning MongoDB",
             body: "If you can't get enough of statistics, here's the blog for you. Run by three biostats professors, they blog
about an abundance of statistics in big data and how they are used by data scientists across all kinds of fields—including
their own. For any new statisticians looking to jump into the career, they also have interviews with data scientists about
their careers and roles in the industry.",
             date: ISODate("2021-02-15T18:48:49Z"),
             tags: ['nosql', 'Excel', 'Data Science', 'bread', 'Turing', 'Branding', 'Python', 'Signals and Systems'],
             comments: [{
                        person_username: "hinata_takondwa",
                        comment: "It is very common to find that the data to support many of the business information needs
is simply not available at the levels required, or that it is of such bad quality that it is impossible to use. Resolution of
these types of issues often requires fundamental changes to business processes.",
                        date: ISODate("2022-05-27T18:11:34Z")}]}
            });

db.blogs.insertOne({
   author: {name: "Kelsey İlkay", username: "Kelsey_İlkay%", twitter: "@kelseyilkay%", address:
"kelseyilkay@yahoo.com", phones: 667974752},
   posts:
            {title: "Start with SQL todaySimply Statistics",
             body: "Datafloq is run by Mark Van Rijmenam, author of "Think Bigger: Developing a Successful Big Data
Strategy for Your Business," and is a great resource for big data in data science. The blog focuses on the business aspects
of big data and how to make data science work for organizations. It also features information about trending tech topics
like blockchain and artificial intelligence. While it largely acts as a resource with articles and insights, Datafloq also seeks
to connect professionals via job postings, vendors, events, and training.",
             date: ISODate("2017-09-13T14:58:24Z"),
             tags: ['Video Marketing', 'R Studio', 'Discrete Mathematics', 'Finite Automata', 'brownies', 'Python'],
             comments: [{
                        person_username: "ime_ayanda",
                        comment: "The world is one extraordinary big data problem.",
                        date: ISODate("2020-08-10T12:15:59Z")}]}
            });

db.blogs.insertOne({
   author: {name: "Ophir Neely", username: "Ophir_Neely$", twitter: "@ophirneely_9", address:
"ophirneely@yahoo.com", phones: [607109523, 607109525]},
   posts:
            {title: "Do you know what is a Relational Database?",
             body: "For anyone looking to enter the field of data science, here is great—if dense—start. Ryan Swanstrom
has worked in data science for Microsoft, Wells Fargo, and government defense contractors. He currently consults as the
Director of Data Science for Unify Consulting. In this blog, he shares his valuable experience, tips, and advice on how to
be a successful data scientist. The blog extends back to 2012 with extensive archives, which are worth diving into for a
hands-on history of the last few years in data science discussion.",
             date: ISODate("2017-09-09T13:31:47Z"),
             tags: ['cakes', 'Finite Automata'],
             comments: [{
                        person_username: "tatum_chi",
                        comment: "Consumer data will be the biggest differentiator in the next two to three years. Whoever
unlocks the reams of data and uses it strategically will win.",
                        date: ISODate("2019-09-23T12:15:59Z")}]}
            });
```

```
db.blogs.insertOne({
   author: {name: "Lerato Esme", username: "Lerato_Esme$", twitter: "@leratoesme45", address:
"leratoesme@gmail.com", phones: 255939577},
   posts:
            {title: "Are you afraid of Pharmacies? Should you?",
             body: "Dataconomy is another resource for prospective data scientists. It features the usual big data news and
tech trends as well as editorials from industry experts. But what sets it apart from the other data science hubs is its
resources for building a career in data science. The site offers a free IT research library and beginner’s guides to get
started. For those already in the industry and looking to advance, it also has a job board and candidate database.",
             date: ISODate("2017-09-09T17:32:55Z"),
             tags: ['cookies', 'Signals and Systems', 'Branding', 'R Studio'],
             comments: [{
                        person_username: "kaulana_chisomo",
                        comment: "That which cannot be measured cannot be proven.",
                        date: ISODate("2020-04-05T16:18:22Z")}]}
            });

db.blogs.insertOne({
   author: {name: "Odell Socheat", username: "Odell_Socheat%", twitter: "@odellsocheat12", address:
"odellsocheat@gmail.com", phones: [572149807, 572149866]},
   posts:
            {title: "The Unremarkable Cookie Experience",
             body: "Speaking of in-depth resources, Data Science Report curates resources from all variety of formats to get
data science into your brain. The site collects free courses, articles, books, videos, and TED Talks to help any level of data
scientist. You can filter the topics to find select information regarding how to get started, salary negotiation, interviews,
technology, social media, marketing, and topics that are just “simply interesting.” It’s a resource hub for data scientists at
any point in their career and anyone with a mind to learn about data.",
             date: ISODate("2020-05-15T18:48:49Z"),
             tags: ['PowerPoint'],
             comments: [{
                        person_username: "franny_durga",
                        comment: "Hiding within those mounds of data is the knowledge that could change the life of a
patient, or change the world.",
                        date: ISODate("2021-10-16T16:18:22Z")}]}
            });

// CHECK THAT INSERTIONS WERE PERFORMED CORRECTLY
db.getCollection('blogs').find();
```

Queries statements

Query 1

1. Retrieve all blog posts that contain in the 'comment' the values 'amazing' and 'extraordinary' greater than equal to the publication date, the query result should return the fields from author (name, Twitter account), from the blog (title, publication date) and the result sort it ascending.
    - Code:

```
db.blogs.find({'posts.comments.comment': {$all: [/amazing/i, /extraordinary/i]}, $expr:{$gte:['$posts.date', '$posts.comments.date']}}, {'_id':0,"author.name": 1, "author.twitter": 1, "posts.title": 1, "posts.date": 1}).sort({"date": 1});
```

    - Result:

```
{
   "author" : {
      "name" : "Mery Garcia",
      "twitter" : "@merygarcia#"
   },
   "posts" : {
      "title" : "Data Science Central",
      "date" : ISODate("2019-05-19T18:42:34.000+0000")
   }
}
{
   "author" : {
      "name" : "Vosgi Beren",
      "twitter" : "@vosgiberen#"
   },
   "posts" : {
      "title" : "Exploring Visual Similarity with Modista",
      "date" : ISODate("2016-07-03T13:31:47.000+0000")
   }
}
```

Query 2

2. Retrieve at least all blog posts that contain in the values tags such 'Video Marketing', 'Online Marketing', 'Data Science', 'Branding' and the result sort it ascending.
    - Code:

```
db.blogs.find({ $or: [ {"posts.tags": "Video Marketing"}, {"posts.tags": "Online Marketing"}, {"posts.tags": "Data Science"}, {"posts.tags": "Branding"}]}, {'_id': 0}).sort( {"posts.date": 1});
```

    - Result:

```
{
   "author" : {
      "name" : "Vosgi Beren",
      "username" : "Vosgi_Beren_9",
      "twitter" : "@vosgiberen#",
      "address" : "vosgiberen@gmail.com",
      "phones" : [
         NumberInt(228837994),
         NumberInt(228837995),
         NumberInt(228837996)
      ]
   },
   "posts" : {
```

```
      "title" : "Exploring Visual Similarity with Modista",
      "body" : "SmartDataCollective.com SmartData Collective is a community site focused on trends in business intelligence and data management. Similar to Data Science Central, it also features insights into data science through contributions by industry experts. Where Data Science Central focuses directly on data science as a whole, SmartData Collective looks at the wider field and how data science can intersect with business.",
      "date" : ISODate("2016-07-03T13:31:47.000+0000"),
      "tags" : [
         "cookies",
         "Automata Theory and Compilers",
         "Branding",
         "Discrete Mathematics",
         "mongoDB",
         "nosql",
         "Turing",
         "Data Science",
         "brownies"
      ],
      "comments" : [
         {
            "person_username" : "hui_darby",
            "comment" : "I do not relate that much to the fact that good grades mean a good future. Totally disagree. Nonetheless, my experience was both amazing and extraordinary reading it. Keep it up.",
            "date" : ISODate("2016-09-12T12:37:18.000+0000")
         }
      ]
   }
}
{
   "author" : {
      "name" : "Zhi Tiriaq",
      "username" : "Zhi_Tiriaq45",
      "twitter" : "@zhitiriaq%",
      "address" : "zhitiriaq@gmail.com",
      "phones" : NumberInt(292783279)
   },
   "posts" : {
      "title" : "No Free Hunch",
      "body" : "InsideBIGDATA focuses on the machine learning side of data science. It covers big data in IT and business, machine learning, deep learning, and artificial intelligence. Guest features offer insight into industry perspectives, while news and Editor’s Choice articles highlight important goings-on in the field. All the articles are neatly categorized by topic to zero in on any subject in particular. The blog also maintains a host of resources for events, jobs, and research reports, and more. This is a resource for anyone wanting to stay up to date with machine learning.",
      "date" : ISODate("2017-03-04T13:31:47.000+0000"),
      "tags" : [
         "Discrete Mathematics",
         "Video Marketing",
         "brownies"
      ],
      "comments" : [
         {
            "person_username" : "ejiro_ailbhe",
            "comment" : "Statistics are like bikinis. What they reveal is suggestive, but what they conceal is vital",
            "date" : ISODate("2018-04-08T18:22:50.000+0000")
         }
      ]
   }
}
{
   "author" : {
      "name" : "Lerato Esme",
      "username" : "Lerato_Esme$",
      "twitter" : "@leratoesme45",
      "address" : "leratoesme@gmail.com",
      "phones" : NumberInt(255939577)
   },
   "posts" : {
```

```
      "title" : "Are you afraid of Pharmacies? Should you?",
      "body" : "Dataconomy is another resource for prospective data scientists. It features the usual big data news and tech trends as well as editorials from industry experts. But what sets it apart from the other data science hubs is its resources for building a career in data science. The site offers a free IT research library and beginner's guides to get started. For those already in the industry and looking to advance, it also has a job board and candidate database.",
      "date" : ISODate("2017-09-09T17:32:55.000+0000"),
      "tags" : [
         "cookies",
         "Signals and Systems",
         "Branding",
         "R Studio"
      ],
      "comments" : [
         {
            "person_username" : "kaulana_chisomo",
            "comment" : "That which cannot be measured cannot be proven.",
            "date" : ISODate("2020-04-05T16:18:22.000+0000")
         }
      ]
   }
}
…
```


Showing 3 of 8 results.

Query 3

3. Retrieve all blog posts that not contain in the values tags such 'Video Marketing', 'Online Marketing', 'Data Science', 'Branding' and the result sort it ascending.
    - Code:

```
db.blogs.find({"posts.tags": {$nin: ["Video Marketing", "Online Marketing", "Data Science", "Branding"]}}, {"_id": 0}).sort({"posts.date": 1});
```

    - Result:


```
{
   "author" : {
      "name" : "Ophir Neely",
      "username" : "Ophir_Neely$",
      "twitter" : "@ophirneely_9",
      "address" : "ophirneely@yahoo.com",
      "phones" : [
         NumberInt(607109523),
         NumberInt(607109525)
      ]
   },
   "posts" : {
      "title" : "Do you know what is a Relational Database?",
      "body" : "For anyone looking to enter the field of data science, here is great—if dense—start. Ryan Swanstrom has worked in data science for Microsoft, Wells Fargo, and government defense contractors. He currently consults as the Director of Data Science for Unify Consulting. In this blog, he shares his valuable experience, tips, and advice on how to be a successful data scientist. The blog extends back to 2012 with extensive archives, which are worth diving into for a hands-on history of the last few years in data science discussion.",
      "date" : ISODate("2017-09-09T13:31:47.000+0000"),
      "tags" : [
         "cakes",
         "Finite Automata"
      ],
      "comments" : [
         {
```

```
            "person_username" : "tatum_chi",
            "comment" : "Consumer data will be the biggest differentiator in the next two to three years. Whoever unlocks the reams of data and uses it strategically will win.",
            "date" : ISODate("2019-09-23T12:15:59.000+0000")
         }
      ]
   }
}
{
   "author" : {
      "name" : "Mery Garcia",
      "username" : "Mery123",
      "twitter" : "@merygarcia#",
      "address" : "merygarcia@outlook.es",
      "phones" : NumberInt(299258239)
   },
   "posts" : {
      "title" : "Creating a NOSQL Database",
      "body" : "NoSQL databases emerged in the late 2000s as the cost of storage dramatically decreased. Gone were the days of needing to create a complex, difficult-to-manage data model in order to avoid data duplication. Developers (rather than storage) were becoming the primary cost of software development, so NoSQL databases optimized for developer productivity.",
      "date" : ISODate("2019-05-06T19:12:24.000+0000"),
      "tags" : [
         "database",
         "Python",
         "Turing",
         "nosql"
      ],
      "comments" : [
         {
            "person_username" : "alex_123",
            "comment" : "This was very useful!",
            "date" : ISODate("2022-03-11T19:58:55.000+0000")
         }
      ]
   }
}
{
   "author" : {
      "name" : "Odell Socheat",
      "username" : "Odell_Socheat%",
      "twitter" : "@odellsocheat12",
      "address" : "odellsocheat@gmail.com",
      "phones" : [
         NumberInt(572149807),
         NumberInt(572149866)
      ]
   },
   "posts" : {
      "title" : "The Unremarkable Cookie Experience",
      "body" : "Speaking of in-depth resources, Data Science Report curates resources from all variety of formats to get data science into your brain. The site collects free courses, articles, books, videos, and TED Talks to help any level of data scientist. You can filter the topics to find select information regarding how to get started, salary negotiation, interviews, technology, social media, marketing, and topics that are just “simply interesting.” It’s a resource hub for data scientists at any point in their career and anyone with a mind to learn about data.",
      "date" : ISODate("2020-05-15T18:48:49.000+0000"),
      "tags" : [
         "PowerPoint"
      ],
      "comments" : [
         {
            "person_username" : "franny_durga",
            "comment" : "Hiding within those mounds of data is the knowledge that could change the life of a patient, or change the world.",
            "date" : ISODate("2021-10-16T16:18:22.000+0000")
```

```
        }
      ]
    }
}
```

Query 4

4. Retrieve title of the blog posts that have been written in May 2019.
    - Code:

```
db.blogs.find({"posts.date":{$gte:ISODate('2019-05-01'),$lt:ISODate('2019-05-31')}}, {'_id': 0,'posts.title':1});
```

    - Result:

```
{
  "posts" : {
    "title" : "Data Science Central"
  }
}
{
  "posts" : {
    "title" : "Creating a NOSQL Database"
  }
}
```

Query 5

5. Retrieve title of the blog posts that have been written by the user with username "Mery123" with any of the tags: "nosql", "database" or "mongoDB", and published in 2019.
    - Code:

```
db.blogs.find({$and: [{"author.username": "Mery123"}, {"posts.date": {$gte:ISODate('2019-01-01'),$lt:ISODate('2019-12-31')}}, {$or: [{"posts.tags": "nosql"}, {"posts.tags": "database"}, {"posts.tags": "mongoDB"}]}]},  {'_id': 0,'posts.title':1});
```

    - Result:

```
{
  "posts" : {
    "title" : "Data Science Central"
  }
}
{
  "posts" : {
    "title" : "Creating a NOSQL Database"
  }
}
```

Exercise 2

Import Code

After downloading the JSON file, the code was imported through this line of code:

Mongoimport --host localhost --port 27017 -d test -c city C:\data\db\zips.json

Queries statements

Query 1

1. Recover the minimum value of the field 'pop' for all cities.
    - Code:

```
db.city.aggregate([{$group:{_id: "$city", MinValuePop:{$min: "$pop"}}}]);
```

    - Result:

```
{
   "_id" : "REDLANDS",
   "MinValuePop" : NumberInt(29784)
}
{
   "_id" : "HEPLER",
   "MinValuePop" : NumberInt(229)
}
{
   "_id" : "WEST OLIVE",
   "MinValuePop" : NumberInt(6722)
}
{
   "_id" : "WINTER PARK",
   "MinValuePop" : NumberInt(24236)
}
{
   "_id" : "KILGORE",
   "MinValuePop" : NumberInt(213)
}
{
   "_id" : "READLYN",
   "MinValuePop" : NumberInt(1858)
}
{
   "_id" : "CUMBERLAND FORES",
   "MinValuePop" : NumberInt(2879)
}
...

Showing 7 of 16584 results.
```

Query 2

2. Retrieve all the zip codes in the city "Springfield".
   - Code:

```
db.city.find({"city": /^springfield$/i}, {"_id": 1, "city": 1});
```

   - Result:

```
{
   "_id" : "01104",
   "city" : "SPRINGFIELD"
}
{
   "_id" : "01107",
   "city" : "SPRINGFIELD"
}
{
   "_id" : "01103",
   "city" : "SPRINGFIELD"
}
…

Showing 3 of 41 results.
```

2.1. How many?
   - Code:

```
db.city.find({"city": /^springfield$/i}, {"_id": 1, "city": 1}).count();
```

   - Result:

```
41
```

Query 3

3. Retrieve the number of cities in state "GA" with less than equal 1500 population.
   - Code:

```
db.city.find({$and:[{"state":"GA"}, {"pop":{$lt:1500}}]}).count();
```

   - Result:

```
123
```

Query 4

4. Update all cities from Massachusetts (state: ”MA”), a new field { area : “27 340 km”}.
   - Code:

```
db.city.updateMany({state: "MA"}, {$set: {area: 27340}});
```

- Check that the operations were performed correctly:

```
db.city.find({"state": "MA"}, {"_id": "MA", "city": 1, "area": 1});
```

Query 5

5. Recover all cities that are located between -81.000001 and 32.915592.
   - Code:

```
db.city.find({$and:[{"loc.0": {$gt: -81.000001}}, {"loc.1": {$lt: 32.915592}}]});
```

   - Result:

```
{
   "_id" : "29401",
   "city" : "CHARLESTON",
   "loc" : [
      -79.937069,
      32.779506
   ],
   "pop" : NumberInt(12475),
   "state" : "SC"
}

{
   "_id" : "29403",
   "city" : "CHARLESTON",
   "loc" : [
      -79.949283,
      32.797575
   ],
   "pop" : NumberInt(24620),
   "state" : "SC"
}

{
   "_id" : "29405",
   "city" : "CHARLESTON",
   "loc" : [
      -79.976442,
      32.851206
   ],
   "pop" : NumberInt(30621),
   "state" : "SC"
}
…

Showing 3 of 278 results.
```

Query 6

6. Update the populations to 50000 for those cities that have a population between 45000 and 50000.
    - Code:

```
db.city.updateMany({$and: [{"pop": {$gt: 45000}}, {"pop": {$lt: 50000}}]}, {$set: {"pop": 75000}});
```

Bibliography

[1] Tejada, Z. (n.d.). *Non-relational data and NoSQL - Azure Architecture Center*. Microsoft Learn. Retrieved November 24, 2022, from https://learn.microsoft.com/en-us/azure/architecture/data-guide/big-data/non-relational-data