



M03-DC02 MongoDB

EJERCICIO EVALUABLE MongoDB

Programa: **Máster Executive en Big Data, Cloud & Analytics**

Periodo académico: **2019 – 2020**

Autor/es: **CARLOS ALFONSEL JAÉN**

1. ENUNCIADO

Se requiere la entrega de un documento PDF con las soluciones y comentarios propuestos de cada ejercicio.

1. Lanza los servicios correspondientes para ejecutar **MongoDB** en tu máquina local.
 2. Utiliza el método **mongoimport** para cargar el archivo "restaurants_dataset.json" alojado en la carpeta "BBDD_JSON_MongoDB" en el apartado "Material de Trabajo" disponible en Moodle en la asignatura de **MongoDB-HBase**. Llama a la base de datos "restaurants" y a la colección "dataset".
 3. Ejecuta en la Shell una consulta que devuelva 5 documentos aleatorios de la base de datos recién cargada.
 4. Ejecuta en la Shell una consulta que devuelva 5 documentos aleatorios que contengan el string "Caterer" en el campo "name" (consulta documentación sobre aplicación de **RegEx**).
 5. Ejecuta en la Shell una consulta que devuelva 5 documentos aleatorios donde el resultado contenga el string "Caterer" en el campo "name" y el campo "borough" no sea "Brooklyn".
 6. Ejecuta en la Shell una consulta que devuelva 5 documentos aleatorios que contengan en el campo "street" (campo embebido en "address") un string "Barrow Street".
 7. ¿Cuántos documentos existen que cumplan la condición anterior?. ¿Cuántos documentos existen en total?.
 8. Ejecuta en la Shell una consulta que devuelva los campos "borough" distintos existentes en la base de datos.
 9. ¿Cuántos documentos hay de cada campo "borough" obtenidos en la consulta anterior?.
 10. ¿Cuántos restaurantes hay que tengan "cuisine" de tipo "Jewish/Kosher"?.
¿Cuántos de esos restaurantes están en cada "borough"?.
-

2. SOLUCIONES

- 2.1. Lanza los servicios correspondientes para ejecutar MongoDB en tu máquina local.
- 2.2. Utiliza el método mongoimport para cargar el archivo "restaurants-dataset.json" alojado en la carpeta "BBDD_JSON_MongoDB" en el apartado "Material de Trabajo" disponible en Moodle en la asignatura de MongoDB-HBase. Llama a la base de datos "restaurants" y a la colección "dataset".

Desde un terminal de Windows se ejecuta el siguiente comando:

```
C:\Program Files\MongoDB\Server\4.2\bin>mongoimport --db restaurants -  
-collection dataset --file "D:\MBIT Big Data\M03-DC02  
MongoDB_HBase\Material de  
Trabajo\BBDD_JSON_MongoDB\Data_MongoDB\restaurants-dataset.json"
```

```
C:\Program Files\MongoDB\Server\4.2\bin>mongoimport --db restaurants --collection dataset --file "D:\MBIT Big Data\M03-DC02 MongoDB_HBase\Material de Trabajo\BBDD_JSON_MongoDB\Data_MongoDB\restaurants_dataset.json"  
2020-03-25T18:32:33.955+0100    connected to: mongodb://localhost/  
2020-03-25T18:32:34.850+0100    25359 document(s) imported successfully. 0 document(s) failed to import.  
C:\Program Files\MongoDB\Server\4.2\bin>_
```

Resultado: 25359 documentos importados.

- 2.3. Ejecuta en la Shell una consulta que devuelva 5 documentos aleatorios de la base de datos recién cargada.

```
>db.dataset.find({}).limit(5).pretty()
```

```
MongoDB
> db.dataset.find({}).limit(5).pretty()
{
  "_id" : ObjectId("5e7b95b20a9d497663d6ea60"),
  "address" : {
    "building" : "97-22",
    "coord" : [
      -73.8601152,
      40.7311739
    ],
    "street" : "63 Road",
    "zipcode" : "11374"
  },
  "borough" : "Queens",
  "cuisine" : "Jewish/Kosher",
  "grades" : [
    {
      "date" : ISODate("2014-11-24T00:00:00Z"),
      "grade" : "Z",
      "score" : 20
    },
    {
      "date" : ISODate("2013-01-17T00:00:00Z"),
      "grade" : "A",
      "score" : 13
    },
    {
      "date" : ISODate("2012-08-02T00:00:00Z"),
      "grade" : "A",
      "score" : 13
    },
    {
      "date" : ISODate("2011-12-15T00:00:00Z"),
      "grade" : "B",
      "score" : 25
    }
  ],
  "name" : "Tov Kosher Kitchen",
  "restaurant_id" : "40356068"
}
{
  "_id" : ObjectId("5e7b95b20a9d497663d6ea61"),
  "address" : {
    "building" : "1007",
    "coord" : [
```

- 2.4. Ejecuta en la Shell una consulta que devuelva 5 documentos aleatorios que contengan el string "Caterer" en el campo "name" (consulta documentación sobre aplicación de \$Regex).

```
>db.dataset.find({"name":{"regex":".*Caterer"}}).limit(5).pretty()
```

```
MongoDB
> db.dataset.find({"name": {$regex:". *Caterer"}}).limit(5).pretty()
{
  "_id" : ObjectId("5e7b95b20a9d497663d6ea66"),
  "address" : {
    "building" : "2780",
    "coord" : [
      -73.98241999999999,
      40.579505
    ],
    "street" : "Stillwell Avenue",
    "zipcode" : "11224"
  },
  "borough" : "Brooklyn",
  "cuisine" : "American",
  "grades" : [
    {
      "date" : ISODate("2014-06-10T00:00:00Z"),
      "grade" : "A",
      "score" : 5
    },
    {
      "date" : ISODate("2013-06-05T00:00:00Z"),
      "grade" : "A",
      "score" : 7
    },
    {
      "date" : ISODate("2012-04-13T00:00:00Z"),
      "grade" : "A",
      "score" : 12
    },
    {
      "date" : ISODate("2011-10-12T00:00:00Z"),
      "grade" : "A",
      "score" : 12
    }
  ],
  "name" : "Riviera Caterer",
  "restaurant_id" : "40356018"
}
{
  "_id" : ObjectId("5e7b95b20a9d497663d6ea6d"),
  "address" : {
    "building" : "6409",
    "coord" : [
```

- 2.5. Ejecuta en la Shell una consulta que devuelva 5 documentos aleatorios donde el resultado contenga el string "Caterer" en el campo "name" y el campo "borough" no sea "Brooklyn".

```
>db.dataset.find({$and:[{"name":{$regex:". *Caterer"}},{"borough":{$ne:"Brooklyn"}}]}).limit(5).pretty()
```

```
MongoDB
> db.dataset.find({$and:[{"name":{"regex":".*Caterer"}}, {"borough":{"ne":"Brooklyn"}}]}).limit(5).pretty()
{
  "_id" : ObjectId("5e7b95b20a9d497663d6ed97"),
  "address" : {
    "building" : "205",
    "coord" : [
      -73.9890601,
      40.7538643
    ],
    "street" : "West 38 Street",
    "zipcode" : "10018"
  },
  "borough" : "Manhattan",
  "cuisine" : "American",
  "grades" : [
    {
      "date" : ISODate("2014-07-31T00:00:00Z"),
      "grade" : "A",
      "score" : 0
    },
    {
      "date" : ISODate("2013-07-12T00:00:00Z"),
      "grade" : "A",
      "score" : 5
    },
    {
      "date" : ISODate("2012-07-20T00:00:00Z"),
      "grade" : "A",
      "score" : 2
    },
    {
      "date" : ISODate("2012-03-13T00:00:00Z"),
      "grade" : "A",
      "score" : 5
    },
    {
      "date" : ISODate("2011-09-02T00:00:00Z"),
      "grade" : "A",
      "score" : 4
    }
  ],
  "name" : "Cambridge Caterers",
  "restaurant_id" : "40392244"
}
```

- 2.6. Ejecuta en la Shell una consulta que devuelva 5 documentos aleatorios que contengan en el campo "street" (campo embebido en "address") un string "Barrow Street".**

```
>db.dataset.find({"address.street":{"regex":".*Barrow Street"}}).limit(5).pretty()
```



```
MongoDB
> db.dataset.find({"address.street":{"$regex":".*Barrow Street"}}).limit(5).pretty()
{
  "_id" : ObjectId("5e7b95b20a9d497663d6eb30"),
  "address" : {
    "building" : "17",
    "coord" : [
      -74.002592999999999,
      40.7324849
    ],
    "street" : "Barrow Street",
    "zipcode" : "10014"
  },
  "borough" : "Manhattan",
  "cuisine" : "Seafood",
  "grades" : [
    {
      "date" : ISODate("2014-08-19T00:00:00Z"),
      "grade" : "B",
      "score" : 23
    },
    {
      "date" : ISODate("2014-01-14T00:00:00Z"),
      "grade" : "A",
      "score" : 11
    },
    {
      "date" : ISODate("2013-04-02T00:00:00Z"),
      "grade" : "A",
      "score" : 10
    },
    {
      "date" : ISODate("2012-06-12T00:00:00Z"),
      "grade" : "A",
      "score" : 13
    },
    {
      "date" : ISODate("2011-11-03T00:00:00Z"),
      "grade" : "A",
      "score" : 0
    }
  ],
  "name" : "One If By Land Two If By Sea",
  "restaurant_id" : "40367797"
}
```

**2.7. ¿Cuántos documentos existen que cumplan la condición anterior?
¿Cuántos documentos existen en total?**

```
>db.dataset.find({"address.street":{"$regex":".*Barrow Street"}}).count()
```

Respuesta: 7

```
>db.dataset.count()
```

Respuesta: 25359

```
> db.dataset.find({"address.street":{"regex":".*Barrow Street"}}).count()
7
> db.dataset.count()
25359
>
```

2.8. Ejecuta en la Shell una consulta que devuelva los campos "borough" distintos existentes en la base de datos.

```
>db.dataset.distinct("borough")
```

```
> db.dataset.distinct("borough")
[
  "Bronx",
  "Brooklyn",
  "Manhattan",
  "Missing",
  "Queens",
  "Staten Island"
]
>
```

Respuesta: Bronx, Brooklyn, Manhattan, Missing, Queens, Staten Island

2.9. ¿Cuántos documentos hay de cada campo "borough" obtenidos en la consulta anterior?.

```
>db.dataset.aggregate({'$group':{'_id':"borough",counter:{$sum:1}}})
```

```
> db.dataset.aggregate({'$group':{'_id':"borough",counter:{$sum:1}}})
{ "_id" : "Brooklyn", "counter" : 6086 }
{ "_id" : "Manhattan", "counter" : 10259 }
{ "_id" : "Missing", "counter" : 51 }
{ "_id" : "Queens", "counter" : 5656 }
{ "_id" : "Bronx", "counter" : 2338 }
{ "_id" : "Staten Island", "counter" : 969 }
>
```

Respuestas:

- Bronx: 2338 restaurantes.
- Brooklyn: 6086 restaurantes.
- Manhattan: 10259 restaurantes.
- Missing: 51 restaurantes.
- Queens: 5656 restaurantes.
- Staten Island: 969 restaurantes.

Si sumamos todos los restaurantes de cada barrio, nos da los 25359 del apartado 2.7.

2.10. ¿Cuántos restaurantes hay que tengan "cuisine" de tipo "Jewish/Kosher"? ¿Cuántos de esos restaurantes están en cada "borough"?

```
> db.dataset.find({"cuisine": "Jewish/Kosher"}).count()
316
```

```
> db.dataset.find({"cuisine": "Jewish/Kosher"}).count()
316
```

Respuesta: 316 restaurantes con estilo de cocina Jewish/Kosher.

Usando de nuevo la función **aggregate**:

```
> db.dataset.aggregate([{$match: {"cuisine": "Jewish/Kosher"}},
{$group: {'_id': "$borough", counter: {$sum: 1}}}]])
```

```
> db.dataset.aggregate([{$match: {"cuisine": "Jewish/Kosher"}}, {$group: {'_id': "$borough", counter: {$sum: 1}}}]])
{ "_id" : "Brooklyn", "counter" : 167 }
{ "_id" : "Manhattan", "counter" : 73 }
{ "_id" : "Missing", "counter" : 1 }
{ "_id" : "Queens", "counter" : 62 }
{ "_id" : "Staten Island", "counter" : 5 }
{ "_id" : "Bronx", "counter" : 8 }
>
```

Respuestas:

- Bronx: 8 restaurantes con estilo de cocina Jewish/Kosher.
- Brooklyn: 167 restaurantes con estilo de cocina Jewish/Kosher.
- Manhattan: 73 restaurantes con estilo de cocina Jewish/Kosher.
- Missing: 1 restaurantes con estilo de cocina Jewish/Kosher.
- Queens: 62 restaurantes con estilo de cocina Jewish/Kosher.
- Staten Island: 5 restaurantes con estilo de cocina Jewish/Kosher.

Nuevamente, si sumamos estos resultados nos dan los 316 restaurantes Jewish/Kosher en total de la consulta anterior.