Diplomado en Big Data Gestión de Datos Masivos

LABORATORIO MIGRACIONES DE DATOS CON MONGODB

Docente Evaluador: Ing. Alexander Ramírez Camargo Presenta:

Ing. Jorge Alberto Gómez López

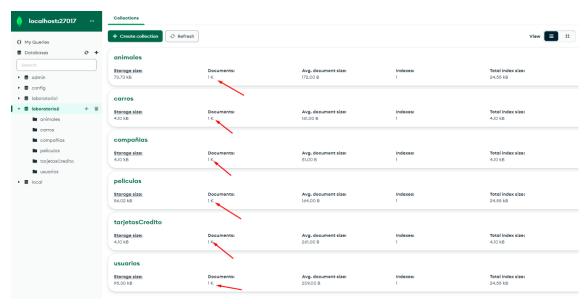
Ing. Guillermo Alexander Cornejo Argueta

- 1. Usando los modelos utilizados en el taller1 del modulo 2, hacer una inserción masiva de datos con Insert Many (al menos 1000 registros).
- Utilizando la pagina de https://www.mockaroo.com/ se crean los mil registros para los siguientes modelos:
- Usuarios
- Películas
- Carros
- Compañias
- Animales
- TarjetasCredito

En donde carros y películas estan relacionadas, a continuación, se muestra parte del código de creación de dichas inserciones:

```
## discustration of the control of t
```

Una vez ingresados todos los datos en la base de datos se mostrarán de la siguiente manera en mongodb compass:



En donde se puede apreciar que cada modelo tiene ya 1000 registros cada uno.

2. Crear al menos 3 cursores en donde al menos cada cursor consulte 100 registros:

En las capturas mostradas a continuación solo se muestra una parte de los datos consultados.

Filtrar usuarios por nacionalidad y ordenar por edad y mostrar su username, género y edad:

```
var usuarios = db.usuarios.find({
    nacionalidad: "China"
}).sort({ edad: 1 }).pretty();

while (usuarios.hasNext()) {
    var usuario = usuarios.next();
    print("username: " + usuario.username + ", Genero: " +
    usuario.genero+ ", Edad: " + usuario.edad );
}
```

```
var usardios = db.usardios.find(
... naionalidad: 'China
... print("servama: "a usuario.usemase ", Genero: " * usuario.genero ", Edad: " * usuario.edad );
... print("servama: "a usuario.usemase ", Genero: " * usuario.genero ", Edad: " * usuario.edad );
... print("servama: "a usuario.usemase ", Genero: " * usuario.genero ", Edad: " * usuario.edad );
... print("servama: "a usuario.usemase ", Genero: F, Edad: 1 * usuario.edad );
... print("servama: "a usuario.usemase "a usuario.genero ", Edad: " * usuario.edad );
... print("servama: "a usuario.usemase ", Genero: F, Edad: 5 * usuario.edad );
... print("servama: "a usuario.usemase );
... print("servama: "a
```

Filtrar películas por genero múltiple y mostrar el nombre, fecha de estreno y género:

```
var peliculas = db.peliculas.find({
    genero: { $in: ["Comedy", "Drama"] }
}).pretty();
while (peliculas.hasNext()) {
    var pelicula = peliculas.next();
    print("Nombre: " + pelicula.nombre + ", Fecha de estreno: " +
pelicula.fechaEstreno+ ", Genero: " + pelicula.genero );
}
```

```
responding to the control of the con
```

Mostrar las tarjetas de crédito que poseen sobregiro activado y mostrar tipo, empresa, limite de sobregiro y país:

```
var tarjetas = db.tarjetasCredito.find({
    sobregiro: true
}).pretty();
while (tarjetas.hasNext()) {
    var tarjeta = tarjetas.next();
    print("Tipo: " + tarjeta.tipo + ", Empresa: " + tarjeta.empresa+ ",
Limite Sobregiro: " + tarjeta.limiteSobregiro+ ", Pais: " + tarjeta.pais
);
}
```

```
var berjets of ditriptatoredito.rim()

singeric true

.); port()

dile (*prich absolut()) {

singeric true

.); port()

print(Tipo: * tarjeta.ripo *, Empresa: *, tarjeta.empresa *, Limite Schregiro: * tarjeta.limiteschregiros *, Pais: * * tarjeta.pais );

print(Tipo: * tarjeta.ripo *, Empresa: * tarjeta.empresa *, Limite Schregiro: undefined, Pais: Nolmed

print(Tipo: * tarjeta.ripo *, Empresa: * tarjeta.empresa *, Limite Schregiro: undefined, Pais: Nolmed

print(Tipo: * tarjeta.ripo *, Excelled and Accologiro: undefined, Pais: Nolmed

print(Tipo: * tarjeta.ripo *, Excelled and Accologiro: undefined, Pais: Nolmed

print(Tipo: * tarjeta.ripo *, Excelled and Accologiro: undefined, Pais: Norman

print(Tipo: * tarjeta.ripo *, Excelled and Accologiro: undefined, Pais: Norman

print(Tipo: * tarjeta.ripo *, Excelled and Accologiro: undefined, Pais: Norman

print(Tipo: * tarjeta.ripo *, Excelled and Accologiro: undefined, Pais: Norman

print(Tipo: * tarjeta.ripo *, Excelled and Accologiro: undefined, Pais: Norman

print(Tipo: * tarjeta.ripo *, Excelled and Accologiro: undefined, Pais: Norman

print(Tipo: * tarjeta.ripo *, Excelled and Accologiro: undefined, Pais: Norman

print(Tipo: * tarjeta.ripo *, Excelled and Accologiro: undefined, Pais: Norman

print(Tipo: * tarjeta.ripo *, Excelled and Accologiro: undefined, Pais: Norman

print(Tipo: * tarjeta.ripo *, Excelled and Accologiro: undefined, Pais: Norman

print(Tipo: * tarjeta.ripo *, Excelled and Accologiro: undefined, Pais: Norman

print(Tipo: * tarjeta.ripo *, Excelled and Accologiro: undefined, Pais: Norman

print(Tipo: * tarjeta.ripo *, Excelled and Accologiro: undefined, Pais: Norman

print(Tipo: * tarjeta.ripo *, Excelled and Accologiro: undefined, Pais: Norman

print(Tipo: * tarjeta.ripo *, Excelled and Accologiro: undefined, Pais: Norman

print(Tipo: * tarjeta.ripo *, Excelled and Accologiro: undefined, Pais: Undefined, Pais:
```

3. Cree un índice para cada colección y visualícelo.

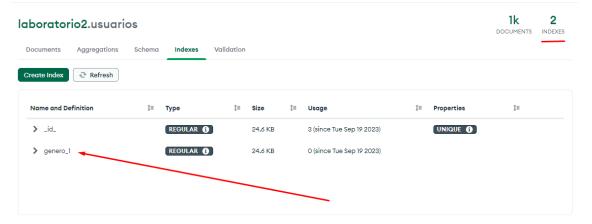
Se demostrará como se realizó el índice para una colección ya que el proceso es el mismo para las demás colecciones.

1. Para crear un índice a una colección se hace con el comando db.usuarios.createIndex();

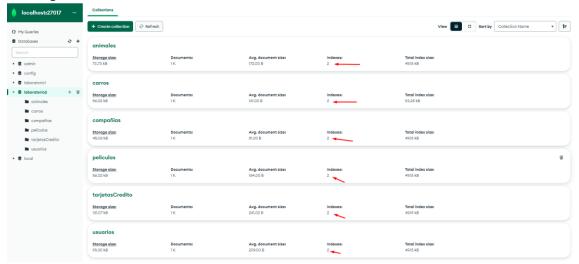
```
> db.usuarios.createIndex({genero: 1});
{
         "createdCollectionAutomatically" : false,
         "numIndexesBefore" : 1,
         "numIndexesAfter" : 2,
         "ok" : 1
}
```

2. Para poder ver el índice de la función utilizamos el comando db.usuarios.getIndexes();

3. También se pueden ver dichos Índices creados en mongodo compass en la opción de Indexes.



Una vez creados todos los índices en los modelos en mongoDB compas nos aparecerá de la siguiente manera:



4. Implemente el método limit en consultas.

Consulta 1:

Consultar 10 usuarios de nacionalidad china en orden de la menor edad a la mayor:

Consulta 2:

Consultar 10 películas que sean del género de drama o comedia

Consulta 3:

Consultar 16 carros que sean del fabricante Honda.

Consulta 4:

Consultar 15 animales que tengan estatus de rescatado pero que no estén en peligro:

5. Cree una nueva colección implementando un método validador

Para el siguiente ejemplo, se ha creado una nueva colección llamada "empleados", dicha colección contiene documentos con las siguientes características:

Documento empleado:

- usuario: nombre del usuario para el empleado
- **password:** contraseña para la cuenta de usuario, la contraseña esta restringida por las siguientes características:
 - La contraseña debe estar compuesta de mínimo 8 caracteres
 - La contraseña debe contener por lo menos una letra
 - o La contraseña debe contener por lo menos un número
 - o La contraseña debe contener por lo menos un carácter especial
- **tipo:** tipo de usuario registrado en la colección, este campo está restringido por las siguientes características:
 - Solo existen 2 tipos de usuarios válidos: Admin, Usuario

Para implementar una colección que contenga un documento con dichas características, se hizo uso de validadores en mongodb

```
db.createCollection("empleados", {
   validator: {
      $jsonSchema: {
         bsonType: "object",
         required: [ "usuario", "password", "tipo" ],
         properties: {
            usuario: {
               bsonType: "string",
description: "debe ser una cadena de caracteres (requerido)"
            password: {
               bsonType: "string",
               minLength: 8,
               pattern: "^(?=.*[A-Za-z])(?=.*\\d)(?=.*[@$!%*#?&])[A-Za-z\\d@$!%*#?&]{8,}$",
               description: "debe ser una cadena de caracteres de por lo menos 8 caracteres,
               que contenga letras, numeros, caracteres especiales
                bsonType: "string",
                enum: ['Admin','Usuario'],
                description: "Solo 2 tipos aceptados: Admin o Usuario"
```

A continuación, se realizan algunas observaciones acerca de la colección (y sus validadores) creados:

- required: En este campo se puede especificar la lista de campos que son requeridos al momento de querer insertar un documento.
- password -> minLength: propiedad donde se especifica el tamaño mínimo para el campo en cuestión (en este caso, una contraseña de minimo 8 caracteres)
- password -> pattern: propiedad donde se especifica el regex para verificar si el campo cumple con las reglas definidas por dicho regex.
- tipo -> enum: propiedad donde se especifica todos los valores que son válidos para un campo, en este ejemplo, se especifica que solamente Admin y Usuario son los dos unicos valores validos.

A continuación, se muestra un ejemplo cuando se quiere insertar un registro que no cumple con las condiciones estipuladas por el validador (contraseña no es por lo menos 8 caracteres, y no contiene una letra, un número, un carácter especial al mismo tiempo):

A continuación, se muestra un ejemplo de inserción que cumple con todas las reglas definidas por el validador:

6. Implemente un método Explain para alguna de las colecciones y sustente su uso

A continuación, se muestra un ejemplo donde se ejecuta el método explain para la siguiente query: Obtener el listado de películas cuyo país de origen son china o japon

```
db.peliculas.explain().find({$or: [ {pais:'Japan'}, {pais:'China'}]});
                                    "queryPlanner" : {
                                                                                 "plannerVersion" : 1,
"namespace" : "taller2.peliculas",
"indexFilterSet" : false,
                                                                                   "parsedQuery"
"$or"
                                                                                                                                                                                                                                                                                                           : "Japan"
                                                                                                                                                                                                                            "pais"
                                                                                                                                                                                                                                                                            "$eq"
                                                                                                                                                                                                                                                                                                                      "China"
                                                                                  "inputStage" : {
    "stage" : "FETCH",
                                                                                                                                                                             "inputStage" : {
    "stage"
                                                                                                                                                                                                                              "keyPattern"
                                                                                                                                                                                                                                                                            "pais"
                                                                                                                                                                                                                                mame" : "pais_1",
fmultiKey" : false,
fmultiKeyPaths" : {
    "pais" : false,
    "sUnion" : false,
    "pais" : false,

                                                                                                                                                                                                                              "indexName"
                                                                                                                                                                                                                            "isMultiKey"
                                                                                                                                                                                                                           },
"isUnique" : false,
"isSparse" : false,
"isPartial" : false,
                                                                                                                                                                                                                              ndexVersion" : 2,
"direction" : "forward",
"indexBounds" : {
                                                                                                                                                                                                                                                                           nds" : {
"pais"
                                                                                                                                                                                                                                                                                                               .
| "[\"China\", \"China\"]",
| "[\"Japan\", \"Japan\"]"
                                                                                 },
"rejectedPlans" : [ ]
                                },
"serverInfo" :
"host"
                                                                             nfo" : {
   "host" : "lap-033",
   "port" : 27017,
   "version" : "4.2.24",
   "citVersion" : "5e4ec1d24431fcdd28b579a024c5c801b8cde4e2"
```

El uso de explain nos permite comprender la manera en la que mongodb, de manera interna, busca ejecutar la query que en este caso nosotros hemos definido, de este modo nos permite conocer cualquier punto de optimización en caso de que se busque mejorar el tiempo de ejecución de dicha query. En el ejemplo para la consulta anterior, se pueden hacer las siguientes observaciones:

- La query se ejecutará en la colección "peliculas" (detallado en la sección namespace de la respuesta)
- Stage = FETCH nos indica que la query en cuestión debe devolver documentos de la colección como parte de su respuesta

- Stage = IXSCAN nos indica que, dentro del Stage = FETCH, para determinar si un
 documento debe ser incluido como parte de la respuesta, se hará a traves un indice, en
 este caso el indice "pais_1" que es un indice que ordena de manera ascendente
 (keyPattern = pais : 1) los registros en base al campo "pais"
- IndexBounds = pais indica que, los valores de indices a buscar, son los valores "China" o "Japan"
- RejectedPlan = [] nos indica que, para ejecutar dicha consulta, no existen ningunas variantes que hayan resultado "menos optimas" para ejecutar dicha consulta.

7. Implemente un ejemplo del uso de réplicas con mongodb

Para el siguiente ejemplo de implementación de replicas en mongodb, se detallan las características / condiciones de dicha implementación:

- El sistema de replicas estará conformado por 2 servidores mongodb, uno fungiendo como db primaria (donde se podrán realizar operaciones de lectura y escritura) mientras que el segundo servidor fungirá como replica (que recibirá todos los cambios que provengan del servidor primario)
- Ambos servidores mongodb, se encontrarán dentro de la misma computadora física (localhost)
- A continuación, se detallan los aspectos técnicos de las instancias de mongodb:
 - o Instancia "master":

Servidor: localhostPuerto: 27018

Ruta física de almacenamiento de datos:

C:/mongodb_instance/primary

Nombre de replica: replicaUsuarios

o Instancia "slave"

Servidor: localhostPuerto: 27019

Ruta física de almacenamiento de datos:

C:/mongodb instance/slave1

Nombre de replica: replicaUsuarios

Paso 1: Levantamiento de instancias

Se ejecuta el siguiente comando para levantar la instancia mongodo primaria (instancia master)

mongod --replSet replicaUsuarios --dbpath "C:/mongodb_instance/primary" --port 27018

```
C:Users\jorge.gomez>mongod --replSet replicaUsuarios --dbpath "C:/mongodb_instance/primary" --port 27018
2023-09-30720:00:21.209-0000 I CONTROL [main] Automatically disabling TL5 1.0, to force-enable TL5 1.0 specify --sslDi sabledProtocols 'none'
2023-09-30720:00:21.213-0600 I CONTROL [initandlisten] MongoDB starting : pid=30448 port=27018 dbpath=C:/mongodb_instance/primary 64-bit host=lap-033
2023-09-30720:00:21.215-0600 I CONTROL [initandlisten] MongoDB starting : pid=30448 port=27018 dbpath=C:/mongodb_instance/primary 64-bit host=lap-033
2023-09-30720:00:21.215-0600 I CONTROL [initandlisten] by version v4.2.24
2023-09-30720:00:21.215-0600 I CONTROL [initandlisten] db version v4.2.24
2023-09-30720:00:21.216-0600 I CONTROL [initandlisten] git version: Se4ectd24431fcdd28b579a024c5c801b8cde4e2
2023-09-30720:00:21.216-0600 I CONTROL [initandlisten] modules: none
2023-09-30720:00:21.216-0600 I CONTROL [initandlisten] build environment:
2023-09-30720:00:21.216-0600 I CONTROL [initandlisten] distandd: 2012plus
2023-09-30720:00:21.218-0600 I CONTROL [initandlisten] distandd: 2012plus
2023-09-30720:00:21.218-0600 I CONTROL [initandlisten] build environment:
2023-09-30720:00:21.218-0600 I CONTROL [initandlisten] build environment:
2023-09-30720:00:21.218-0600 I CONTROL [initandlisten] distandd: 2012plus
2023-09-30720:00:21.218-0600 I CONTROL [initandlisten] build environment:
2023-09-30720:00:21.216-0600 I CONTROL [initandlisten] build en
```

Se ejecuta el siguiente comando en otra consola cmd, para levantar la instancia secundaria de mongodb (instancia "slave")

mongod --replSet replicaUsuarios --dbpath "C:/mongodb_instance/slave1" --port 27019

```
C:\Users\jorge.gomez>mongod --replSet replicaUsuarios --dbpath "C:/mongodb_instance/slave1" --port 27019
2023-09-30720:01:19.715-0600 I CONTROL [main] Automatically disabling TLS 1.0, to force-enable TLS 1.0 specify --sslDi
sabledProtocols 'none'
2023-09-30720:01:19.717-0600 W ASIO [main] No TransportLayer configured during NetworkInterface startup
2023-09-30720:01:19.720-0600 I CONTROL [initandlisten] MongoDB starting: pid=37104 port=27019 dbpath=C:/mongodb_instance/slave1 64-bit host=lap-033
2023-09-30720:01:19.720-0600 I CONTROL [initandlisten] db version v4.2.24
2023-09-30720:01:19.720-0600 I CONTROL [initandlisten] db version v4.2.24
2023-09-30720:01:19.720-0600 I CONTROL [initandlisten] git version: SedecId24431fcdd28b579a024c5c801b8cde4e2
2023-09-30720:01:19.720-0600 I CONTROL [initandlisten] abild environment:
2023-09-30720:01:19.720-0600 I CONTROL [initandlisten] modules: none
2023-09-30720:01:19.720-0600 I CONTROL [initandlisten] distancd: 2012plus
2023-09-30720:01:19.720-0600 I STORAGE [initandlisten] potions: { net: { port: 27019 }, replication: { replSet: "replicalsuarios" }, storage engine, so setting the active storage engine to 'wiredTiger'.
2023-09-30720:01:19.722-0600 I STORAGE [initandlisten] wiredTiger one rofig: create, cache_size-7566M, cache_overflow-
(file_max=0M), session_max=33000,eviction=(threads_min-4, threads_max=4),config_base=false,statistics=(fast),log=(enabled=true_archive=true_path=inormal_compressor=nappy),file_manager=(close_idle time=1000000,close_scan_interval=10,close_hand
le_minimum=250),statistics_log=(wait=0),verbose=[recovery_progre
```

Paso 2: Configurando el replicaSet dentro de la instancia maestra

Se procede a conectarse a la instancia primaria de mongodb

mongo --port 27018

Dentro de la instancia primaria, se especifica la configuración del sistema de replicas, para este caso, se ha elegido la siguiente configuración:

Esta configuración representa lo siguiente:

- El nombre del sistema de replicas se llama "replicaUsuarios"
- El sistema de replicas está conformado por 2 instancias, uno que se encuentra en el puerto 27018, otro que se encuentra en el puerto 27019

Se ejecuta la configuración anterior y se procede a ejecutar el comando: rs.initiate(config)

Se observa que, al momento de ejecutar el ultimo comando, la línea de comandos cambia el sufijo de la instancia:

```
}
> rs.initiate(config)
{ "ok" : 1 }
replicaUsuarios:SECONDARY>
```

Nos volvemos a conectar de nuevo a la instancia primaria, para que aparezca de manera correcta el sufijo "PRIMARY" (representando que estamos conectados a la instancia primaria de la réplica):

```
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()
---
replicaUsuarios:PRIMARY>
```

Paso 3: Verificaciones finales de la réplica creada

Si nos conectamos a la instancia secundaria, e intentamos ejecutar un comando (como por ejemplo "show dbs"), obtenemos el siguiente resultado:

Si leemos el mensaje con atención, comenta que la instancia en la que estamos conectadas "Not master and slaveOk=false" indicando que esta instancia es nada más una réplica, y por ende no se pueden ejecutar operaciones dentro de dicha instancia (solo se pueden realizar operaciones en la instancia master).

Para terminar de corroborar que la réplica se ha configurado correctamente, realizamos una conexión a la instancia master, y creamos una colección, para poder observar el output de cada una de las instancias que conforman el replica set:

Dentro de la consola de la instancia primaria, obtenemos el log de que una nueva colección ha sido creada:

```
2023-09-30T20:16:08.828-0600 I STORAGE [conn16] createCollection: test.usuarios with generated UUID: 0ac8c94b-a517-478
7-af30-22427bc6dfba and options: {}
2023-09-30T20:16:08.836-0600 I INDEX [conn16] index build: done building index _id_ on ns test.usuarios
2023-09-30T20:17:09.487-0600 I QUERY [clientcursormon] Cursor id 294648868759700879 timed out, idle since 2023-09-30
T20:07:05.514-0600
```

Mientras que, dentro de la consola de la instancia secundaria, obtenemos el mensaje que ha replicado la creación de la colección dentro de su instancia, de esta manera, comprobando que el sistema de replicas ha sido configurado correctamente:

```
2023-09-30T20:16:08.838-0600 I STORAGE [repl-writer-worker-0] createCollection: test.usuarios with provided UUID: 0ac8
c94b-a517-4787-af30-22427bc6dfba and options: { uuid: UUID("0ac8c94b-a517-4787-af30-22427bc6dfba") }
2023-09-30T20:16:08.844-0600 I INDEX [repl-writer-worker-0] index build: done building index _id_ on ns test.usuario
s
```