

Laboratorio de: ANALÍTICA DE DATOS – BIG DATA

Estudiante: Danny Díaz Padilla

Práctica No.: 3

Tema: MongoDb Atlas (Cloud)

# **Objetivos:**

- Crear un usuario en MongoDb Atlas.
- Crear un cluster en MongoDb Atlas.
- Cargar un archivo csv en MongoDb Atlas y realizar 10 consultas sobre el mismo.

## Marco teórico:

# **MongoDB Atlas**

MongoDB Atlas es el servicio global de base de datos en la nube para aplicaciones modernas. Implemente MongoDB totalmente administrado en AWS, Azure o GCP. La mejor automatización y prácticas comprobadas de su clase garantizan la disponibilidad, la escalabilidad y el cumplimiento de los estándares de seguridad y privacidad de datos más exigentes. Use el robusto ecosistema de controladores, integraciones y herramientas de MongoDB para construir más rápido y pasar menos tiempo administrando su base de datos.

# **Funciones**

## **Pretty**

El método pretty () se usa principalmente para mostrar el resultado en un formato más fácil de leer. La sintaxis del método pretty () es la siguiente: db.collection.find (). pretty ()

## Data size

Muestra el tamaño en bytes de la colección. La compresión de datos no afecta este valor. La sintaxis del método dataSize () es la siguiente: db.collection.find (). dataSize ()

## **Storage Size**

Devuelve la cantidad total de almacenamiento asignado a esta colección para el almacenamiento de documentos. Si los datos de recopilación están comprimidos el tamaño de almacenamiento refleja el tamaño comprimido y puede ser menor que el valor devuelto por db.collection.dataSize (). Su sintaxis es la siguiente: db.collection.find (). storageSize ()



#### Find one

Devuelve un documento que satisface los criterios de consulta especificados en la colección o vista. Si varios documentos satisfacen la consulta, este método devuelve el primer documento de acuerdo con el orden natural que refleja el orden de los documentos en el disco. db.collection.find (). findOne ()

## Find

Selecciona documentos en una colección o vista y devuelve un cursor a los documentos seleccionados.

db.collection.find(). find({"tipo\_delito":"Robo (Con violencia)"})

#### Count

Devuelve el recuento de documentos que coincidirían con una consulta find () para la colección o vista. El método db.collection.count () no realiza la operación find () sino que cuenta y devuelve el número de resultados que coinciden con una consulta. db.collection.count(query, options)

## **Distinct**

Encuentra los valores distintos para un campo específico en una única colección o vista y devuelve los resultados en una matriz. db.collection.distinct(field, query, options)

# Sum: \$SUM

Calcula y devuelve la suma de valores numéricos. \$ sum ignora los valores no numéricos. { \$sum: <expression> }

## Average: \$AVG

Calcula y devuelve el promedio de los valores numéricos. \$ avg ignora los valores no numéricos.

{ \$avg: <expression> } ESCUELA POLITÉCNICA NACIONAL FACULTAD DE INGENIERÍA DE SISTEMAS INGENIERÍA DE SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN



# Greater than or equal to: \$gte

\$gte compara tanto el valor como el tipo, utilizando el orden de comparación BSON especificado para valores de diferentes tipos.

```
{ $gte: [ <expression1>, <expression2> ] }
```

## Group: \$group

Agrupa documentos de entrada por la expresión \_id especificada y para cada agrupación distinta, genera un documento. El campo \_id de cada documento de salida contiene el grupo único por valor. Los documentos de salida también pueden contener campos calculados que contienen los valores de alguna expresión de acumulador.

```
db.sales.aggregate( [ { $group : { _id : "$item" } } ] )
```

# OR (Operación lógica): \$or

El operador \$or realiza una operación OR lógica en una matriz de dos o más <expresiones> y selecciona los documentos que satisfacen al menos una de las <expresiones>. El \$or tiene la siguiente sintaxis:

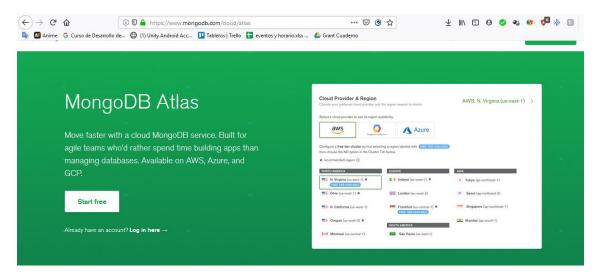
```
{ $or: [ {<expression1>},{ <expression2> },..., { <expressionN> } ] }
• Sort: $sort

Ordena los datos.
{$sort: { <field1>: <sort order>, <field2>: <sort order> ... } }
```

## Desarrollo de la práctica:

Primero se ingresa al enlace https://www.mongodb.com/cloud/atlas

Y nos mostrará una pantalla como esta:





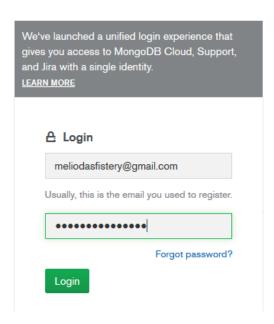
## • Crear un usuario

En mi caso ya me había creado un usuario con anterioridad.

# Get started free No credit card required This email address is already in use. Email Address meliodasfistery@gmail.com Danny Password Díaz Padilla I agree to the terms of service and privacy policy. Get started free

Por lo que ingreso directamente:



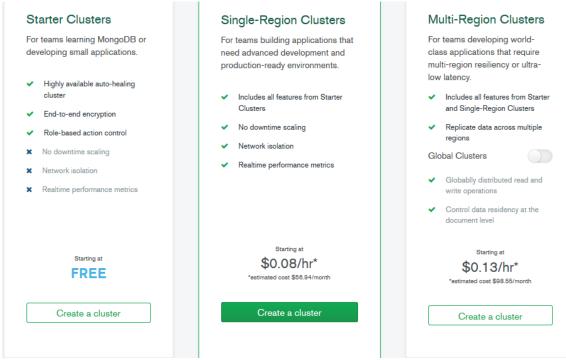


Si es la primera vez en iniciar, se va a demorar unos minutos mientras crea el espacio

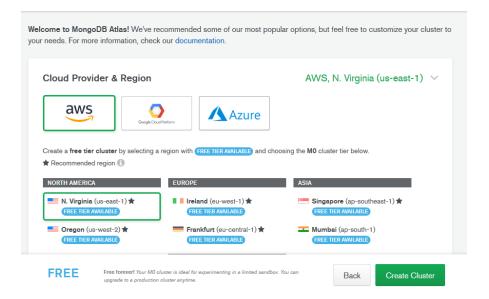


• Crear cluster: mongoclase6

## Elegiremos el gratis

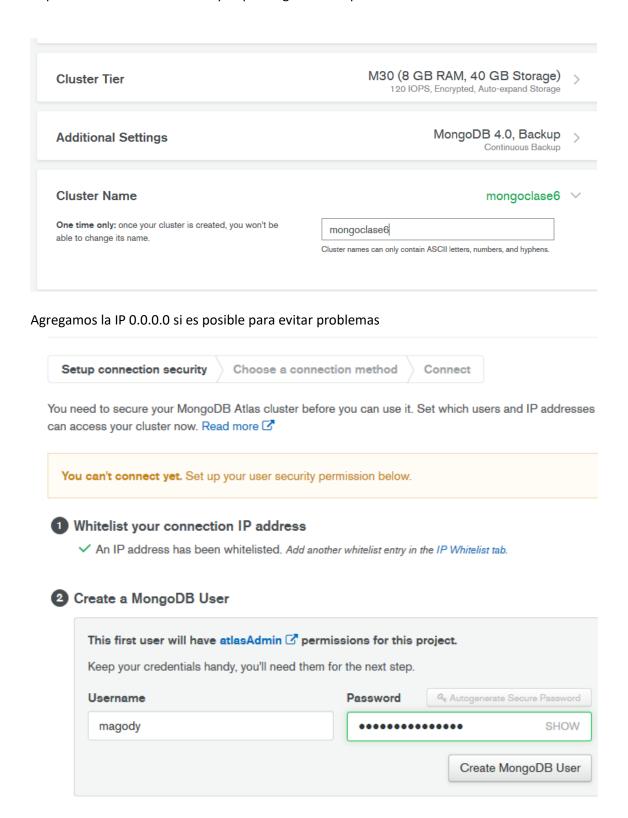


## Dejaremos los datos que vienen por defecto Create a Starter Cluster



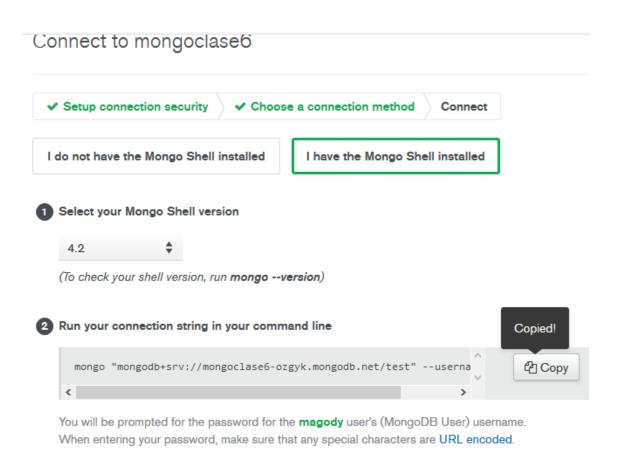


Importante cambiar el nombre porque luego no se lo podrá cambiar.



Nos conectamos por medio de consola. Es importante agregar al PATH el camino hacia MongoDB





Having trouble connecting? View our troubleshooting documentation

mongo "mongodb+srv://mongoclase6-ozgyk.mongodb.net/test" --username magody

```
C:\Users\COMPANY>mongo "mongodb+srv://mongoclase6-ozgyk.mongodb.net/test" --username magody

C:\Users\COMPANY>mongo "mongodb+srv://mongoclase6-ozgyk.mongodb.net/test" --username magody

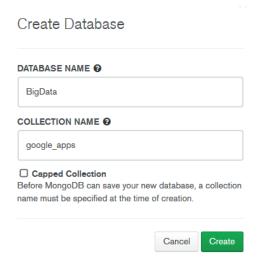
MongoDB shell version v4.2.0

Enter password:
connecting to: mongodb://mongoclase6-shard-00-00-ozgyk.mongodb.net:27017,mongoclase6-shard-00-01-ozgyk.mongodb.net:27017
,mongoclase6-shard-00-02-ozgyk.mongodb.net:27017/test?authSource=admin&compressors=disabled&gssapiServiceName=mongodb&re
plicaSet=mongoclase6-shard-00-02-ozgyk.mongodb.net:27017/test?authSource=admin&compressors=disabled&gssapiServiceName=mongodb&re
plicaSet=mongoclase6-shard-00-solsterue
2019-10-22109:30:01.351-0500 I NETWORK [js] Starting new replica set monitor for mongoclase6-shard-0/mongoclase6-shard
-00-00-ozgyk.mongodb.net:27017,mongoclase6-shard-00-01-ozgyk.mongodb.net:27017,mongoclase6-shard-00-02-ozgyk.mongodb.net
2019-10-22109:30:01.459-0500 I CONNPOOL [ReplicaSetMonitor-TaskExecutor] Connecting to mongoclase6-shard-00-02-ozgyk.mo
ngodb.net:27017
2019-10-22109:30:01.460-0500 I CONNPOOL [ReplicaSetMonitor-TaskExecutor] Connecting to mongoclase6-shard-00-02-ozgyk.mo
ngodb.net:27017
2019-10-22109:30:01.461-0500 I CONNPOOL [ReplicaSetMonitor-TaskExecutor] Connecting to mongoclase6-shard-00-01-ozgyk.mo
ngodb.net:27017
```



## Cargar csv

El .CSV a cargar es una base de datos de 32mil aplicaciones de la Google Play el cual está disponible en https://www.kaggle.com/gauthamp10/google-playstore-apps



Finalmente agregamos por medio de consola (oculto el password por seguridad).

mongoimport --host mongoclase6-shard-0/mongoclase6-shard-00-00-ozgyk.mongodb.net:27017,mongoclase6-shard-00-01-ozgyk.mongodb.net:27017,mongoclase6-shard-00-02-ozgyk.mongodb.net:27017 --ssl -- username magody --password \*\*\*\*\*\*\*\*\* --authenticationDatabase admin --db BigData -- collection google\_apps --type csv --headerline --file C:/Users/COMPANY/Downloads/Google-Playstore-32K.csv

```
Símbolo del sistema
                                                                                                                                                 error validating settings: must specify --fields, --fieldFile or
019-10-22T13:21:18.201-0500
019-10-22T13:21:21.575-0500
-00-01-ozgyk.mongodb.net:27017
019-10-22T13:21:24.584-0500
019-10-22T13:21:27.576-0500
019-10-22T13:21:30.576-0500
019-10-22T13:21:30.576-0500
019-10-22T13:21:36.576-0500
                                                                                                     492KB/3.58MB (13.4%)
708KB/3.58MB (19.3%)
940KB/3.58MB (25.7%)
1.14MB/3.58MB (32.0%
                                         ######.....
                                                                           BigData.google_apps
                                                                                                                        (32.0%)
(38.3%)
(44.6%)
019-10-22T13:21:39.586-0500
019-10-22T13:21:42.576-0500
                                         ########
                                                                           BigData.google_apps
BigData.google_apps
                                                                                                      1.37MB/3.58MB
1.60MB/3.58MB
                                         #########.
                                         ****************
                                                                                                      1.82MB/3.58MB
019-10-22T13:21:45.576-0500
019-10-22T13:21:48.577-0500
                                                                           BigData.google_apps
019-10-22T13:21:51.576-0500
019-10-22T13:21:54.575-0500
                                         ************
                                         ################
                                                                           BigData.google_apps
                                         **************
                                         BigData.google_apps
019-10-22T13:22:03.576-0500
019-10-22T13:22:06.576-0500
                                                                          BigData.google_apps
BigData.google_apps
                                                                                                      3.04MB/3.58MB
3.26MB/3.58MB
                                         ********************
                                         [##########################] BigData.google_apps 3.47MB/3.58MB (97.0%)
[######################### BigData.google_apps 3.58MB/3.58MB (100.0%)
32000 document(s) imported successfully. 0 document(s) failed to import.
019-10-22T13:22:09.575-0500
019-10-22T13:22:11.751-0500
     10-22T13:22:11.751-0500
```



Con show dbs desde el CMD de windows comprobamos la conexión.

```
MongoDB Enterprise mongoclase6-shard-0:PRIMARY> show dbs
BigData 0.004GB
admin 0.000GB
local 5.003GB
sample_airbnb 0.053GB
sample_geospatial 0.001GB
sample_mflix 0.042GB
sample_supplies 0.001GB
sample_training 0.068GB
sample_training 0.068GB
sample_weatherdata 0.004GB
MongoDB Enterprise mongoclase6-shard-0:PRIMARY> use BigData
switched to db BigData
```

## Realizar 10 consultas

1. Find One: Encuentra un dato de ejemplo (Vemos que esta base de datos le falta limpieza).

2. Data Size: Tamaño en bytes y KiloBytes

```
MongoDB Enterprise mongoclase6-shard-0:PRIMARY> db.google_apps.dataSize()
9215553
MongoDB Enterprise mongoclase6-shard-0:PRIMARY> db.google_apps.dataSize()/1024
8999.5634765625
```

3. Storage Size: Tamaño en bytes y KiloBytes incluso comprimido

```
MongoDB Enterprise mongoclase6-shard-0:PRIMARY> db.google_apps.storageSize()
4009984
MongoDB Enterprise mongoclase6-shard-0:PRIMARY> db.google_apps.storageSize()/1024
3916
```

4. Count: Counta todas las aplicaciones que tengan como categoría "FOOD AND DRINK".



5. \$AVG: Obtine el promedio de Reviews humanas por categoría db.google\_apps.aggregate([{\$group: {\_id: "\$Category", num\_homi: {\$avg: "\$Reviews"}}}])

```
MongoDB Enterprise mongoclase6-shard-0:PRIMARY> db.google_apps.aggregate([{$group: {_id : "$Category",
num_homi : {$avg : "$Reviews"}}}])
{ "_id" : ")", "num_homi" : 4 }
{ "_id" : "WEATHER", "num_homi" : 98674.465 }
{ "_id" : "DATING", "num_homi" : 32856.47572815534 }
{ "_id" : "GAME_WORD", "num_homi" : 102409.11029411765 }
{ "_id" : "ENTERTAINMENT", "num_homi" : 35935.00071942446 }
{ "_id" : "EVENTS", "num_homi" : 3633.7956089247314 }
{ "_id" : "GAME_MUSIC", "num_homi" : 88907.1649484536 }
{ "_id" : "EDUCATION", "num_homi" : 8851.615855435733 }
{ "_id" : "POdcasts", "num_homi" : 14059.036697247706 }
{ "_id" : "PARENTING", "num_homi" : 326612.8163934426 }
{ "_id" : "GAME_SPORTS", "num_homi" : 399888.2182320442 }
{ "_id" : "GAME_STRATEGY", "num_homi" : 117593.69335347431 }
{ "_id" : "BOOKS_AND_REFERENCE", "num_homi" : 18948.241610738256 }
{ "_id" : "GAME_BUCATIONAL", "num_homi" : 15669.494929006085 }
{ "_id" : "GAME_ARCADE", "num_homi" : 171736.9909090901 }
{ "_id" : "GAME_ARCADE", "num_homi" : 171736.99090909091 }
{ "_id" : "GAME_ROLE_PLAYING", "num_homi" : 119910.7434679335 }
{ "_id" : "GAME_ROLE_PLAYING", "num_homi" : 119910.7434679335 }
{ "_id" : "ART_AND_DESIGN", "num_homi" : 16056.815950920245 }

Type "it" for more
```

6. \$MIN: obtiene la minima cantidad de Reviews

db.google\_apps.aggregate([{\$group: {\_id: "\$Category", num\_homi: {\$min: "\$Reviews"}}}])

```
MongoDB Enterprise mongoclase6-shard-0:PRIMARY> db.google_apps.aggregate([{$group: {_id: "$Category", num homi: {$min: "$Reviews"}}}])
{ "_id": ")", "num_homi": 4 }
{ "_id": "WEATHER", "num homi": 1 }
{ "_id": "GAME_WORD", "num_homi": 1 }
{ "_id": "GAME_WORD", "num_homi": 1 }
{ "_id": "ENTERTAINMENT", "num_homi": 1 }
{ "_id": "EVENTS", "num_homi": 1 }
{ "_id": "EDUCATION", "num_homi": 1 }
{ "_id": "POCCASTS", "num_homi": 1 }
{ "_id": "POCCASTS", "num_homi": 1 }
{ "_id": "GAME_SPORTS", "num_homi": 1 }
{ "_id": "GAME_SRATEGY", "num_homi": 1 }
{ "_id": "GAME_SPORTS", "num_homi": 1 }
{ "_id": "GAME_ROLE_PLAYINGAL", "num_homi": 1 }
{ "_id": "GAME_ROLE_PLAYING", "num_homi": 1 }
}
```

7. \$MAX: obtiene la máxima cantidad de reviews por categoría

db.google\_apps.aggregate([{\$group: {\_id: "\$Category", num\_homi: {\$max: "\$Reviews"}}}])

```
MongoDB Enterprise mongoclase6-shard-0:PRIMARY> db.google_apps.aggregate([{$group: {_id : "$Category",
num_homi : {$max : "$Reviews"}}}])
( "_id" : ")", "num_homi" : 4 }
( "_id" : "WEATHER", "num_homi" : 2423736 }
( "_id" : "DATING", "num_homi" : 548465 }
( "_id" : "GAME_WORD", "num_homi" : 2168702 }
( "_id" : "GAME_WORD", "num_homi" : 8900879 }
( "_id" : "EVENTS", "num_homi" : 52507 }
( "_id" : "EVENTS", "num_homi" : 979700 }
( "_id" : "EVENTS", "num_homi" : 7375938 }
( "_id" : "POdcasts", "num_homi" : 7375938 }
( "_id" : "POdcasts", "num_homi" : 14642083 }
( "_id" : "PARENTING", "num_homi" : 811793 }
( "_id" : "GAME_SPORTS", "num_homi" : 48401470 }
( "_id" : "GAME_STRATEGY", "num_homi" : 48401470 }
( "_id" : "BOOKS_AND_REFERENCE", "num_homi" : 3182256 }
( "_id" : "GAME_CDUCATIONAL", "num_homi" : 1311175 }
( "_id" : "GAME_BDUCATIONAL", "num_homi" : 29834812 }
( "_id" : "GAME_ROLE_PLAYING", "num_homi" : 2649608 }
( "_id" : "GAME_ROLE_PLAYING", "num_homi" : 2649608 }
( "_id" : "ART_AND_DESIGN", "num_homi" : 960950 }

Type "it" for more
```

8. \$GTE: Obtiene las aplicaciones con Rating mayor a 4.8 Y con más de 1millón de reviews.

db. google\_apps.find({Rating:{\$gte:4.8}, Reviews:{\$gte:1000000}}).pretty()

9. OR: Obtiene los Ratings que sean mayores o iguales a 4.9 O en su defecto los que tengan un precio de 1.99 dólares

db. google\_apps.find( { \$or: [{Rating: {\$gte: 4.9}}, {Price: {\$eq: "\$1.99"}}]} ).pretty()

```
longoDB Enterprise mongoclase6-shard-0:PRIMARY> db. google_apps.find( { $or: [{Rating: {$gte: 4.9}}, {Pice: {$eq: "$1.99"}}]} ).pretty()
                     "_id" : ObjectId("5daf48a1a99d0e9ff023c975"),
"App Name" : "Candy Blast",
"Category" : "GAME_CASUAL",
"Rating" : 4.967391491,
"Reviews" : 644,
"Installs" : "10,000+",
"Size" : "55M",
"Price" : 0,
"Content Rating" : "Everyone",
"Last Updated" : "March 26, 2019",
"Minimum Version" : "4.1 and up",
"Latest Version" : "1.0.9"
                     "_id": ObjectId("5daf48a1a99d0e9ff023c9f2"),
"App Name": "123 Kids Fun MUSIC - Kids Music Educational Games",
"Category": "GAME_MUSIC",
"Rating": 4.083333492,
"Reviews": 36,
"Installs": "1,000+",
"Size": "22M",
"Price": "$1.99",
"Content Rating": "Everyone",
"Last Updated": "June 22, 2018",
"Minimum Version": "4.0.3 and up",
"Latest Version": 1.19
                        "Latest Version" : 1.19
```

10. Distinct: Obtiene los distintos precios de las aplicaciones.

```
db.google_apps.distinct("Price")
```

```
terprise mongoclase6-shard-0:PRIMARY> db.google_apps.distinct("Price")
"$5.99",
"$2.99",
 $10.00"
```



#### Análisis de resultados:

Todos los comandos se ejecutaron correctamente.

Para el comando MapReduce el clusted bloqueaba la ejecución:

Igual con otros comandos que incluyen "runCommand".

# **Conclusiones y recomendaciones:**

- De la practica: MongoDB Atlas tiene medidas de seguridad bastante estrictas.
- De la practica: Solo se puede tener un cluster, para un segundo cluster se debe pagar.
- De la practica: Solo se reserva una cantidad limitada de 500MB para memoria ram por lo que para procesamiento de BigData es limitante la versión gratis.
- De los objetivos: Se logró crear un usuario en MongoDb Atlas efectivamente.
- De los objetivos: Se logró crear un cluster en MongoDb Atlas efectivamente.
- De los objetivos: Se logró cargar un archivo csv en MongoDb Atlas y realizar 10 consultas sobre el mismo efectivamente.
- Se recomienda agregar el "headline" en la instrucción propuesta de MongoDB porque está desactualizada.
- Es recomendable agregar una ip 0.0.0.0 para la práctica y admitir todas las conexiones..

## Bibliografía

MongoDB. (s.f.). *MongoDB*. Obtenido de MongoDB Web site: https://www.mongodb.com/cloud/atlas