

## **ACTIVIDAD 1:**

Uso de MongoDB

Alumno: Cuenca Roa, Leonard Jose

Grupo: 1001

Equipo 01C

# ÍNDICE

Carga de Datos	4
Evidencia carga de datos	4
Explora las colecciones	5
Identifica todas las distintas categorías (categories) de la colección books	5
Identifica los distintos estados (status) de la colección books	6
Describe brevemente qué arroja la siguiente consulta	6
Utiliza la condición de la consulta anterior para recuperar aquellos libros que posean exactamente 2 autores y que estén publicados	7
Describe brevemente qué ocurre si a la consulta del punto anterior le añades al final la siguiente instrucción	7
Qué ocurre si también le añades lo siguiente	8
Consulta la colección 1	9
Sobre la colección books realiza las siguientes consultas	9
¿Cuál es el tamaño de la colección (en bytes)?	9
¿Cuántos libros tiene la colección?	10
¿Cuántos libros tienen 200 o más páginas?	10
¿Cuántos libros tienen entre 300 y 600 páginas? [300, 600]	10
¿Cuántos libros tienen 0 páginas y cuántos no?	10
¿Cuántos libros han sido publicados y cuántos no?	12
Consulta la colección 2	13
¿Cuál es el tamaño de la colección (en bytes)?	13
¿Cuántas compañías tiene la colección?	13
¿Cuántas compañías se fundaron en los años 1996, 1997, 2001 y 2005 respectivamente	∍?

Lista las compañías que se dedican a «web» o «mobile» y recupera: nombre, descripenúmero de empleados, e-mail, año, mes y día de su fundación.	ción, 15
Lista las compañías que se dedican a videojuegos y muéstralas en orden descendente según el año en que fueron fundadas.	e 16
¿Cuántas compañías tienen 600 o más empleados?	17
Recupera el nombre, la URL, el usuario de Twitter y el número de empleados de las compañías fundadas entre los años 2001 y 2005 incluidos, que cuenten con 500 o más empleados y que se dediquen a los videojuegos o a la música.	17
¿Alguna empresa se dedica a videojuegos y a la música a la vez?	18
Lista las empresas que cuentan con única y exclusivamente 2 oficinas en la ciudad de Francisco.	San 19
Lista el nombre, el mes y día de adquisición de las empresas de videojuegos que haya sido adquiridas en el año 2007 por un precio igual o superior a los 10 millones de dólares y tengan oficinas en la ciudad de Culver City.	

## Carga de Datos

De la carga de datos, deseo resaltar lo que pude aprender y comprender qué el objetivo es organizar y almacenar datos sobre libros y compañías en una base de datos NoSQL, utilizando MongoDB y su herramienta de importación.

Resaltando la satisfacción y la ejecución de las siguientes actividades:

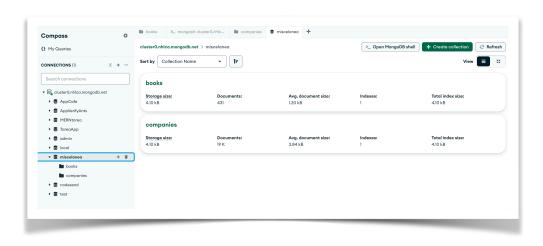
**Creación de una base de datos**: Se debe crear una nueva base de datos con el nombre "miscelanea".

**Creación de colecciones**: Dentro de la base de datos, se creó las dos colecciones: "books" y "companies".

**Importación de datos**: Los datos se cargó desde archivos JSON a las colecciones correspondientes utilizando la herramienta.

Herramienta a utilizar: Se adoptó una estrategia híbrida que integra la línea de comandos de MongoDB con las capacidades visuales de MongoDB Compass. Esta metodología fue seleccionada para llevar a cabo la importación y validación de datos, aprovechando las herramientas de última generación ofrecidas por Atlas. Al combinar la precisión de la línea de comandos con la facilidad de uso de la interfaz gráfica, se logró un proceso más eficiente y controlado.

#### Evidencia carga de datos



Evidencia de base de datos y sus colecciones creadas.

## **Explora las colecciones**

Identifica todas las distintas categorías (categories) de la colección books

Para resolver este ejercicio, se implementó una serie de funciones que ofrece MongoDB. En primer lugar, identifico la colección con la que quiero trabajar utilizando **getCollection**(), especificando la colección **books**. Luego, empleo el método **aggregate**(), que permite realizar operaciones complejas sobre una colección. Luego, utilizo el operador \$unwind para descomponer el campo 'categories' de cada documento en registros individuales. Una vez hecho esto, procedo a agrupar los resultados utilizando el operador \$group para agrupar por el campo deseado utilizando la propiedad \$addToSet ya que, toma el valor del campo categories de cada documento y lo agrega a un nuevo conjunto (set), con el propósito de que solo agrega un valor al conjunto si este aún no existe en caso contrario si un mismo valor de categoría aparece en múltiples documentos, solo se incluirá una vez en el conjunto final y para finalizar, utilizo el operador \$project para seleccionar los campos que quiero incluir en el resultado final, ocultando aquellos que no son necesarios.

Código	Evidencia
db.getCollection('books').aggregate([	<pre>&gt;_MONGOSH &gt; db.getCollection('books').aggregate([</pre>
{ \$unwind: '\$categories' },	<pre>( Sproject: (_sid: 0, categories: 1 ) )); &lt; (     categories: [     'NNL',</pre>
{ \$group: {_id: null, categories: { \$addToSet: '\$categories' }}},	'internet', 'Software Development', 'Algorithmic Art', 'Object-Technology Programming', 'Computer Graphica', 'Citent Server', 'Motife', 'Motife',
{ \$project: { _id: 0, categories: 1 } }	'Java', 'Mobits Technology', 'java', 'Theory', 'Object-or-insted Programming',
1);	'PowerBullder', 'Miscelleneous', 'Programing', 'Quen Source', 'Client-Server', 'Next Generation Databases', 'Perl', 'S', 'Microsoft/, HET', 'SOA', 'Computer Graph', 'In Action', 'Bus inses', 'Microsoft',

### Identifica los distintos estados (status) de la colección books

En este ejercicio, identificamos tres posibles soluciones para obtener registros únicos a partir de una colección. Utilizando el operador distinct: Este operador nos permite listar todos los campos de una colección, eliminando los duplicados y mostrando únicamente los valores únicos. Consultando directamente la colección de manera similar al primer caso, podemos obtener los mismos resultados accediendo directamente a la colección utilizando db.getCollection() otra opción es agrupando los resultados con \$group, para agrupar los datos por un campo específico para este caso el estatus y obtener valores únicos, podemos emplear el operador \$group dentro de una etapa de agregación.

Código	Evidencia
//Opción Uno	>_MONGOSH
db.books.distinct("status")	<pre>&gt; db.getCollection('books').aggregate(     [{ \$group: { _id: '\$status' } }] );</pre>
//Opción Dos	< { _id: 'MEAP'
db.getCollection('books').distinct("status")	} {
//Opción Tres	_id: 'PUBLISH' }
<pre>db.getCollection('books').aggregate(   [{ \$group: { _id: '\$status' } }] );</pre>	Atlas atlas-5lywcl-shard-0 [primary] miscelanea≯

## Describe brevemente qué arroja la siguiente consulta

Esta consulta busca en la colección **books** todos los libros cuyo campo **longDescription** comience por una letra que esté entre la letra "A" y la letra "B" sin incluir la letra "B". De todos los **books** que cumplan esta condición, solo se mostrarán los campos **title** y **longDescription**.

```
Evidencia

| Specification | Codigo | Evidencia | Specification | Code |
```

Utiliza la condición de la consulta anterior para recuperar aquellos libros que posean exactamente 2 autores y que estén publicados

Muestra solo los campos: title, longDescription, status y authors.

```
Código

db.getCollection('books').find({
    status: "PUBLISH",
    authors: { $size: 2 }
    }, {
        title: 1,
        longDescription: 1,
        status: 1,
        authors: 1
    }
}, {
    inguerential content books | find(1 fixes | 2 )
        authors: | fixes | 2 |
        authors: | fixes | 2 |
        authors: | fixes | 2 |
        authors: | fixes | fixes | 2 |
        authors: | fixes | fixes | 2 |
        authors: | fixes | fi
```

Describe brevemente qué ocurre si a la consulta del punto anterior le añades al final la siguiente instrucción

.toArray()

Al añadirle la propiedad este genera un arreglo de objetos con todos los resultados, en cambio la consulta anterior genera un cursor para iterar sobre los resultados, esta pequeña diferencia depende de la necesidades que se desees trabajar, por ejemplo si necesitas procesar los resultados de forma incremental ó cuando tienes un gran número de resultados y quieres

evitar cargarlos todos a la memoria a la vez ó cuando quieres iterar sobre los resultados varias veces **usas el cursor** pero cuando necesitas trabajar con todos los resultados a la vez o cuando quieres pasar los resultados a otra función o biblioteca debes usar **un arreglo de objetos**.



### Qué ocurre si también le añades lo siguiente

.forEach(function(valor, indice, array){print("Titulo: " + valor.title + "Autho 1: " + valor.authors[0]
+ " Author 2: " + valor.authors[1] + " Registro No. " + indice);}).

¿Para qué crees que sería útil esto último?

Al convertir una colección en un arreglo mediante toArray(), podemos aprovechar métodos de iteración para construir listas personalizadas. En este caso, ordenaremos los resultados por título, autores y un índice generado durante la iteración. Esta técnica es versátil, ya que permite generar salidas formateadas en consola, realizar cálculos y validar datos de diversas maneras.

```
Código

db.getCollection('books').find({
    status: "PUBLISH",
    authors: { $size: 2 }
}, {
    title: 1,
    longDescription: 1,
    status: 1,
    authors: 1

    status: 1,
    status: 1,
    status: 1,
    status: 1,
    authors: 1

    status: 1,
    authors: 1

    status: 1,
    authors: 1

}).toArray().forEach(function(valor, indice, array)[crist("long) is less for fagtire to. 2

**Title: status: totalism is status factor for status for factor factor for status factor factor for status factor for status factor factor for status factor factor for status factor for status factor factor for status factor factor for status factor factor for status factor for stat
```

## Consulta la colección 1

Sobre la colección books realiza las siguientes consultas ¿Cuál es el tamaño de la colección (en bytes)?

Código	Evidencia
//Total de 517474 bytes	
db.getCollection('books').dataSize()	>_MONGOSH
	> use miscelanea
	<pre>switched to db miscelanea</pre>
	<pre>&gt; db.getCollection('books').dataSize()</pre>
	<b>&lt;</b> 517474
	<pre>&gt; db.getCollection('companies').dataSize()</pre>
	<b>&lt;</b> 72236994

## ¿Cuántos libros tiene la colección?

Código	Evidencia
//Total de 431 libros	>_MONGOSH
db.getCollection('books').count()	<pre>&gt; db.getCollection('books').count() &lt; 431</pre>
	Atlas atlas-5lywcl-shard-0 [primary] miscelanea >

## ¿Cuántos libros tienen 200 o más páginas?

Código	Evidencia
//Total 264 libros con mayor o igual a 200 páginas db.getCollection('books').find({pageCount: {\$gte:200}}).count()	>_MONGOSH  > db.getCollection('books').find({pageCount:{\$gte:200}}).count()  < 264  Atlas atlas-5lywcl-shard-0 [primary] miscelanea>

## ¿Cuántos libros tienen entre 300 y 600 páginas? [300, 600]

Código	Evidencia
//Total 215 libros db.getCollection('books').find({pageCount: {\$gte:300, \$lte:600}}).count()	>_MONGOSH  > db.getCollection('books').find({pageCount:{\$gte:300, \$lte:600}}).count()  < 215  Atlas atlas-5lywcl-shard-0 [primary] miscelanea >

¿Cuántos libros tienen 0 páginas y cuántos no?

## Descripción de la consulta:

Con el objetivo de ilustrar este ejercicio, mostraré diversos ejemplos. Estos ejemplos me permitirán determinar, por un lado, la cantidad de libros sin páginas y, por otro, la cantidad de libros con páginas. Además, emplearé la función de agregación para obtener ambos resultados en una única operación. Finalmente, presentaré los resultados de cada ejemplo para su comprobación

## Código **Evidencia** // Libro 0 Páginas total de 166 db.getCollection('books').find({pageCount: 0}).count() // Libros Diferente a 0 Páginas total 265 db.getCollection('books').find({pageCount: {\$ne:0}}).count() db.getCollection('books').find({pageCount:{\$ne:0}}).count() //Consulta usando agregación db.getCollection('books').aggregate( \$group: { hasPages: { \$group: { { \$eq: ['\$pageCount', 0] }, 'Libros 0 páginas', \_id: { 'Libros con páginas' hasPages: { \$cond: [ { \$eq: ['\$pageCount', 0] }, 'Libros 0 páginas', 'Libros con páginas' count: { \$sum: 1 } hasPages: 'Libros 0 páginas' } );

#### Descripción de la consulta:

Mostraré diversos ejemplos para hallar el resultado, estos ejemplos me permitirán determinar, por un lado, la cantidad de libros que **sí** se han podido publicar, por otro, la cantidad de libros que **no** se han podido publicar. Además, emplearé la función de agregación para obtener ambos resultados en una única operación.

## Código **Evidencia** // Libros SI publicados total de 363 db.getCollection('books').find({status:'PUB LISH'}).count() db.getCollection('books').find({status:'PUBLISH'}).count() // Libros SI publicados total de 68 db.getCollection('books').find({status: db.getCollection('books').find({status:{\$ne:'PUBLISH'}}).count() {\$ne:'PUBLISH'}}).count() db.getCollection('books').aggregate( //Consulta usando agregación \$group: { db.getCollection('books').aggregate( hasPublish: { \$cond: [ 'Total Libros SI Publicados', \$group: { 'Total Libros NO Publicados' id: { hasPublish: { count: { \$sum: 1 } \$cond: [ { \$eq: ['\$status', 'PUBLISH'] }, 'Total Libros SI Publicados', 'Total Libros NO Publicados' count: { \$sum: 1 } } );

## Consulta la colección 2

¿Cuál es el tamaño de la colección (en bytes)?

Código	Evidencia
// total 72236994 Bytes	>_MONGOSH
db.getCollection('companies').dataSize()	> use miscelanea
	<pre>&lt; switched to db miscelanea</pre>
	<pre>&gt; db.getCollection('books').dataSize()</pre>
	< 517474
	<pre>&gt; db.getCollection('companies').dataSize()</pre>
	<b>&lt;</b> 72236994

¿Cuántas compañías tiene la colección?

Código	Evidencia
// total de registros en companies 18801	>_MONGOSH
db.getCollection('companies').count()	<pre>&gt; db.getCollection('companies').count()</pre>
	< 18801
	Atlas atlas-5lywcl-shard-0 [primary] miscelanea≯

¿Cuántas compañías se fundaron en los años 1996, 1997, 2001 y 2005 respectivamente?

#### Descripción de la consulta:

Para este ejercicio usaremos el método de **agregación** en combinación de **match** y **group** para mostrar los datos en una sola consulta implementando métodos matemáticos el **\$sum** y creando una columna para mostrar esa suma.

## Código Evidencia

```
//
db.getCollection('companies').aggregate(

{
    $match: {
        founded_year: {
            $in: [1996, 1997, 2001, 2005]
        }
    }
},

{
    $group: {
        _id: '$founded_year',
        count: { $sum:1 }
    }
}

, {$sort: {_id: 1}}
])
```

Lista las compañías que se dedican a «web» o «mobile» y recupera: nombre, descripción, número de empleados, e-mail, año, mes y día de su fundación.

#### Código **Evidencia** >\_MONGOSH db.getCollection('companies').aggregate( > db.getCollection('companies').aggregate( \$match: { category\_code: { category code: { \$in: ['web', 'mobile'] \$in: ['web', 'mobile'] } } \$project: { }, description: 1, \$project: { number\_of\_employees: 1, id: 0, email\_address: 1, name: 1, founded\_year: 1, founded\_month: 1, description: 1, founded\_day: 1 number of employees: 1, email address: 1, founded year: 1, founded month: 1, name: '1000MIKES'. founded day: 1 } founded vear: 2007. founded month: 11. founded day: 1, , { \$sort: { name: 1 } } email\_address: 'info@1000mikes.com', ])

#### Descripción de la consulta:

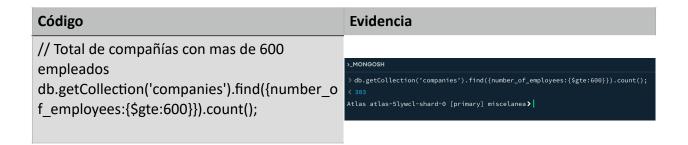
Decidí emplear el operador **\$aggregate** debido a su eficacia y claridad para resolver este caso. El criterio de búsqueda se centra en el atributo **category\_code**. Considero muy útil que la etapa **\$match** permita utilizar operadores como **\$and**, **\$or** y, en este caso específico, **\$in** para filtrar los documentos que coincidan con los valores "web" y "mobile". Para presentar los resultados de la forma deseada, utilicé las etapas **\$project** y **\$sort** para proyectar los campos relevantes y ordenar los documentos, respectivamente.

Lista las compañías que se dedican a videojuegos y muéstralas en orden descendente según el año en que fueron fundadas.

## Código Evidencia

```
db.getCollection('companies').aggregate(
   $match: {
    category_code: {
     $in: ['games video']
    }
   }
  },
   $project: {
    _id: 0,
    name: 1,
    description: 1,
    number of employees: 1,
    email address: 1,
    founded year: 1,
    founded_month: 1,
    founded_day: 1
   }
 , { $sort: { founded_year: -1 } }
 ])
```

¿Cuántas compañías tienen 600 o más empleados?



Recupera el nombre, la URL, el usuario de Twitter y el número de empleados de las compañías fundadas entre los años 2001 y 2005 incluidos, que cuenten con 500 o más empleados y que se dediquen a los videojuegos o a la música.

#### Código **Evidencia** db.getCollection('companies').aggregate( db.getCollection('companies').aggregate( \$match: { number\_of\_employees: { \$gte: 500 }, founded year: { \$gte: 2001, \$lte: 2005 }, category\_code: { number of employees: { \$gte: 500 }, \$in: ['games\_video', 'music'] category code: { \$in: ['games\_video', 'music'] \$project: { } homepage\_url: 1, }, number\_of\_employees: 1 \$project: { , { \$sort: { name: 1 } } name: 1, homepage url: 1, twitter username: 1, id: ObjectId('52cdef7f4bab8bd67529c178'), number of employees: 1 , { \$sort: { name: 1 } } \_id: ObjectId('52cdef7e4bab8bd67529a8b4'), ] );

### **Evidencia** Código db.getCollection('companies').aggregate( \$match: { founded year: { \$gte: 2001, \$lte: 2005 }, db.getCollection('companies').aggregate( number of employees: { \$gte: 500 }, \$and: [ { category code: "Music" }, \$match: { founded\_year: { \$gte: 2001, \$lte: 2005 }, { category code: "games video" } ] number\_of\_employees: { \$gte: 500 }, \$and: [ { category\_code: "Music" }, { category\_code: "Game" } ] } }, name: 1, \$project: { name: 1, twitter\_username: 1, homepage url: 1, twitter username: 1, number of employees: 1 } Atlas atlas-5lywcl-shard-0 [primary] miscelanea> , { \$sort: { name: 1 } } );

#### Descripción de la consulta:

Opté por utilizar el operador **\$aggregate** debido a su claridad y eficiencia para este caso; El criterio de búsqueda es el siguiente: año de fundación entre 2001 y 2005 (inclusive), número de empleados mayor o igual a 500, y empresas que se dediquen tanto a la **música** como a los **videojuegos**.

Para combinar estas condiciones, empleé el operador **\$and** dentro de la etapa **\$match**. Finalmente, utilicé **\$project** para seleccionar los campos de interés, Sin embargo, no se encontraron resultados que cumplan con todos estos criterios de búsqueda.

Lista las empresas que cuentan con única y exclusivamente 2 oficinas en la ciudad de San Francisco.

```
Código
                                                     Evidencia
db.getCollection('companies').aggregate(
                                                     >_MONGOSH
                                                              offices: {
                                                                 $filter: {
   $match: { 'offices.city': 'San Francisco' }
                                                                   input: '$offices',
  { $match: { 'offices': { $size: 2 } } },
                                                                   cond: {
                                                                    $eq: [
   $project: {
                                                                      '$$office.city',
    name: 1,
                                                                      'San Francisco'
     offices: {
      $filter: {
       input: '$offices',
       as: 'office',
       cond: {
        $eq: [
          '$$office.city',
                                                             $match: { offices: { $size: 2 } }
         'San Francisco'
                                                             $project:{
                                                              name:1,
  },
   $match: { offices: { $size: 2 } }
  { $project:{
                                                         name: 'GoGrid'
     id:0,
     name:1,
  }
]
                                                     Atlas atlas-5lywcl-shard-0 [primary] miscelanea≯
);
```

#### Descripción de la consulta:

Con el objetivo de resolver esta consulta, implementé una estrategia de **agregación** y **proyección**, al explorar diferentes enfoques para resolver esta consulta, descubrí que la combinación de los operadores de agregación \$match y \$filter en MongoDB era la más efectiva.

A través de varias pruebas, comprendí cómo utilizar \$filter para iterar sobre el arreglo de oficinas y crear un nuevo arreglo con los elementos que cumplen una determinada condición, para este caso filtrar por ciudad "San Francisco", Este proceso me permitió no solo obtener los resultados esperados sino también profundizar en mis conocimientos sobre el manejo de datos estructurados. Se debe mencionar que se tuvo el problema de que me listaba 3 empresas "Constant Contact", esta empresa en efecto tiene oficinas en "San Francisco", pero tiene un total de 4 empresas rompiendo con el criterio inicial que exclusivamente se requiere dos oficinas, por lo que se procedió realizar otro \$match logrando con el objetivo y bueno para mostrar los datos más concisos otro \$project, para tener la satisfacción de mostrar las dos únicas empresas que si logra con el criterio que son "GoGrid" y "Notifixious"

Lista el nombre, el mes y día de adquisición de las empresas de videojuegos que hayan sido adquiridas en el año 2007 por un precio igual o superior a los 10 millones de dólares y que tengan oficinas en la ciudad de Culver City.

#### Descripción de la consulta:

Para finalizar, utilizaremos la función aggregate() debido a su sencillez y facilidad para aprender de manera práctica. Sin embargo, me gustaría explorar otras opciones para obtener los datos a medida que adquiera más experiencia. En este caso particular, buscamos registros que cumplan con las siguientes condiciones: Categoría: Videojuegos, Precio de adquisición: Superior a 10 millones de dólares, Año de adquisición: 2007, Ciudad de la oficina: Culver City Para ello, emplearemos el operador \$match para filtrar los datos y \$project para seleccionar los campos de interés. Finalmente, utilizaremos \$sort para ordenar los resultados por nombre. Con esta consulta, se encontró un único resultado: la empresa "Flektor".

#### Código **Evidencia** >\_MONGOSH // Resultado > db.getCollection('companies').aggregate([ db.getCollection('companies').aggregate([ \$match:{ category\_code:'games\_video', \$match:{ 'acquisition.price\_amount': { category\_code:'games\_video', \$gte: 10000000 'acquisition.price\_amount': { 'acquisition.acquired\_year': 2007, \$gte: 10000000 'offices.city': 'Culver City' 'acquisition.acquired\_year': 2007, 'offices.city': 'Culver City' \$project:{ } 'acquisition.acquired\_month': 1, 'acquisition.acquired\_day': 1 \$project:{ \_id:0, name:1, {\$sort:{name:1}} 'acquisition.acquired\_month': 1, 'acquisition.acquired\_day': 1 }

Atlas atlas-5lywcl-shard-0 [primary] miscelanea≯

{\$sort:{name:1}}

])