

# Data lake

# Índice

- 01 [Creando un Data lake](#)
- 02 [Conectando el Data lake al MongoDB Shell](#)
- 03 [Consultando el Data lake a través de Mongosh](#)



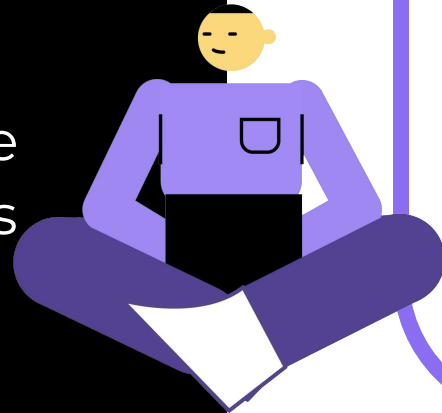
01

# Creando un data lake



Repasando..

El data lake es un tipo de repositorio que almacena grandes y variados conjuntos de datos en bruto en formato nativo. Con los data lakes, se obtiene una visión no refinada de los datos. ¿Vamos a conocerlo?

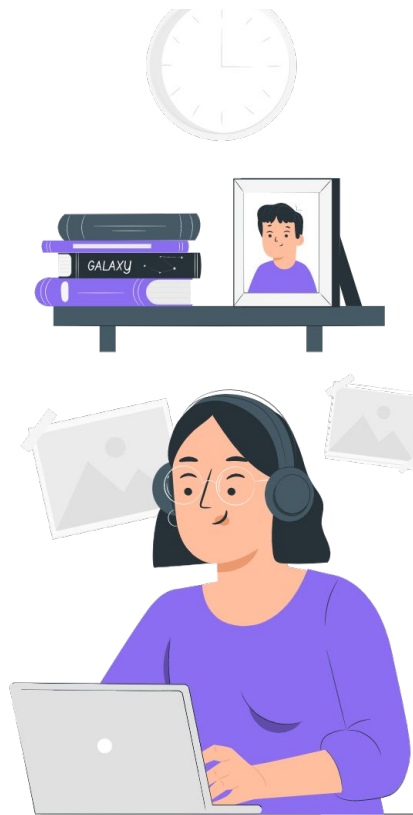


# Creando un Data lake

El Data lake permite consultar, transformar y mover datos de forma nativa en clusters **AWS S3** e **MongoDB Atlas**.

Se trata de datos disponibles para su análisis y manipulación, ricamente estructurados y almacenados en los siguientes formatos **JSON**, **BSON**, **CSV**, **TSV**, **Avro**, **ORC** e **Parquet**.


Pueden trabajar con estos datos utilizando la función **mongoshell**, **MongoDB Compass** o cualquier driver MongoDB.



El primer paso es hacer clic en el botón **Create Data Lake**.

MULTIPLIC'S ORG - 2022-02-13 > PROJECT 0Configure a New Data Lake

## Data Lake



### Create a Data Lake

Natively query and analyze data across AWS S3, Atlas Clusters and HTTP(S) Stores in-place using MQL

Create Data Lake

En el Data Stores, tenemos dos opciones:

**Connect Data** > direcciona para conectarse a un almacenamiento en la nube.

**Connect Simple Dataset** > acceder a las muestras de datos.

Seleccionen la segunda opción, **Connect Simple Dataset**.

## Data Stores



Connect your Amazon S3, Atlas  
Clusters, and HTTP(S) Stores

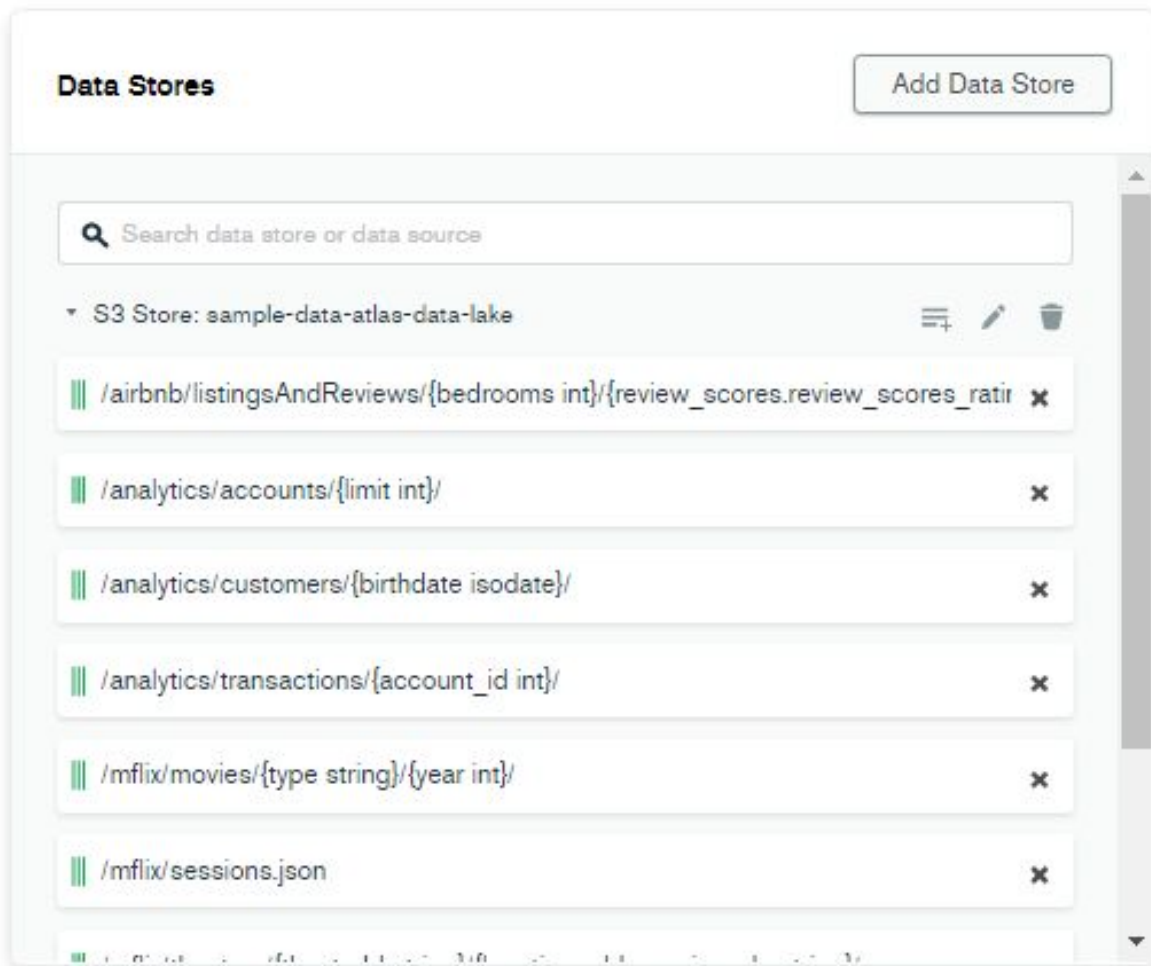
Connect Data

or

Connect Sample Dataset

Las mismas bases de datos que se cargaron en Cluster0, están disponibles aquí.

Seleccionen una base de datos de Data Stores y arrástrenlo hasta el Data Lake.





En el ejemplo siguiente, hemos seleccionado una muestra de datos de películas. Hagan clic en el botón Save.

The image shows a user interface for managing data lakes. On the left, under the heading "Data Stores", there is a list of data sources with a scrollbar. Each item consists of a green icon, a path, and a close button (x). The paths are:

- /airbnb/listingsAndReviews/{bedrooms int}/{review\_scores.revie x
- /analytics/accounts/{limit int}/ x
- /analytics/customers/{birthdate isodate}/ x
- /analytics/transactions/{account\_id int}/ x
- /mflix/movies/{type string}/{year int}/ x
- /mflix/sessions.json x
- /mflix/theaters/{theaterId string}/{location.address.zipcode string x
- /mflix/users.json x
- /nyc-yellow-cab-trips/{trip\_start isodate isodate}/{passenger.c x

At the top right of this panel is a button labeled "Add Data Store".

In the center, a green arrow points from the data stores panel to the data lake panel, with the text "Drag the data source to your Data Lake" below it.

On the right, under the heading "Data Lake: DataLake0", there is a "Create Database" button. Below this is a search bar labeled "Search database or collection name". Under the search bar, there is a section for "Database0" which contains a "Collection0". Inside "Collection0", there is a data source path: /mflix/movies/{type string}/{year int}/ x. Below this is a dashed box with the text "Drag and drop here from Data Stores". At the bottom of the panel, there is a "Data Lake Summary: 1 Database" and a green "Save" button.

El Data Lake está listo para ser explorado. Hagan clic en el botón CONNECT para definir de qué forma accederán a estos datos.

MULTIPLIC'S ORG - 2022-02-13 > PROJECT 0

## Data Lake

Configure a New Data Lake



DataLake0

1 DBS

1 COLLECTIONS



CONNECT

CONFIGURATION

...



Your data lake was created successfully!



You successfully connected to your data lake

3

Finally, write and run a [query](#)

02

## Conectando el Data lake al MongoDB Shell

Tienen la posibilidad de acceder a los datos a través de MongoDB Shell, a través de una aplicación o a través de MongoDB Compass.

Ahora accederemos a los datos a través del MongoDB Shell.

## Connect to DataLake0

✓ Setup connection security

Choose a connection method

Connect

Choose a connection method [View documentation](#)

Get your pre-formatted connection string by selecting your tool below.



**Connect with the MongoDB Shell**

Interact with your cluster using MongoDB's interactive Javascript interface



**Connect your application**

Connect your application to your cluster using MongoDB's native drivers



**Connect using MongoDB Compass**

Explore, modify, and visualize your data with MongoDB's GUI



Go Back

Close

Hagan clic en la opción **I have the MongoDB Shell Installed** y seleccionen la versión **Mongosh**.

Copien la **string de conexión** y cierren la ventana.

✓ Setup connection security > ✓ Choose a connection method > Connect

I do not have the MongoDB Shell installed

I have the MongoDB Shell installed

1 Select your mongo shell version

mongosh ▼

(To check your shell version, run `mongosh --version` or `mongo --version`)

2 Run your connection string in your command line

Use this connection string in your application:

```
mongosh "mongodb://datalake0-9ibcs.a.query.mongodb.net/myFirstDatabase" --tls --  
authenticationDatabase admin --username aluno01
```

Replace **myFirstDatabase** with the name of the database that connections will use by default. You will be prompted for the password for the Database User, **aluno01**. When entering your password, make sure all special characters are [URL encoded](#).

Having trouble connecting? [View our troubleshooting documentation](#)

Abran el prompt del sistema. Para verificar que ya tienen instalado MongoDB Shell, escriban:


**mongosh --version**

El resultado debe ser:

**1.1.9**

Si no tienen instalado MongoDB Shell, instalen el programa **MSI**.


Observen la versión y la plataforma. Hagan clic en **Download** para bajarlo.

Available Downloads 

Version


1.1.9

Platform

Windows 64-bit (8.1+) (MSI) 

Package

msi

 **Download**

Copy Link

Una vez completada la instalación, abran el prompt del sistema.

Confirмен la instalación escribiendo el comando **mongosh -version**.

Después de confirmar la versión, peguen el string de conexión en el prompt, teniendo cuidado de escribir el nombre de la base de datos. En nuestro caso, la base de datos es Database0.

```
C:\Users>mongosh "mongodb://datalake0-9ibcs.a.query.mongodb.net/Database0" --tls
--authenticationDatabase admin --username aluno01
```

```
C:\Users>mongosh "mongodb://datalake0-9ibcs.a.query.mongodb.net/Database0" --tls
--authenticationDatabase admin --username aluno01
Enter password: *****
```

Observen que ahora está conectado a su base de datos.

El mensaje en amarillo sólo indica que se ha encontrado un archivo MongoDB heredado, pero que no se cargará. Ahora pueden consultar su colección de películas, que en este ejemplo se llama Colección0.

```
C:\Users>mongosh "mongodb://datalake0-9ibcs.a.query.mongodb.net/Database0" --tls
--authenticationDatabase admin --username aluno01
Enter password: *****
Current Mongosh Log ID: 6212bf2058b79a1ef55b9eca
Connecting to:          mongodb://datalake0-9ibcs.a.query.mongodb.net/Database0?di
rectConnection=true&appName=mongosh+1.1.9
Using MongoDB:          5.0.3
Using Mongosh:          1.1.9

For mongosh info see: https://docs.mongodb.com/mongodb-shell/

Warning: Found ~/.mongorc.js, but not ~/.mongoshrc.js. ~/.mongorc.js will not be
loaded.
  You may want to copy or rename ~/.mongorc.js to ~/.mongoshrc.js.
AtlasDataLake Database0> 
```



03

## Consultando el Data lake a través de Mongosh

Comprobemos cuántos documentos existen en la colección ejecutando el comando `db.Collection0.find().count()`

El resultado debería ser:

```
AtlasDataLake Database0> db.Collection0.find().count()  
23493  
AtlasDataLake Database0>
```

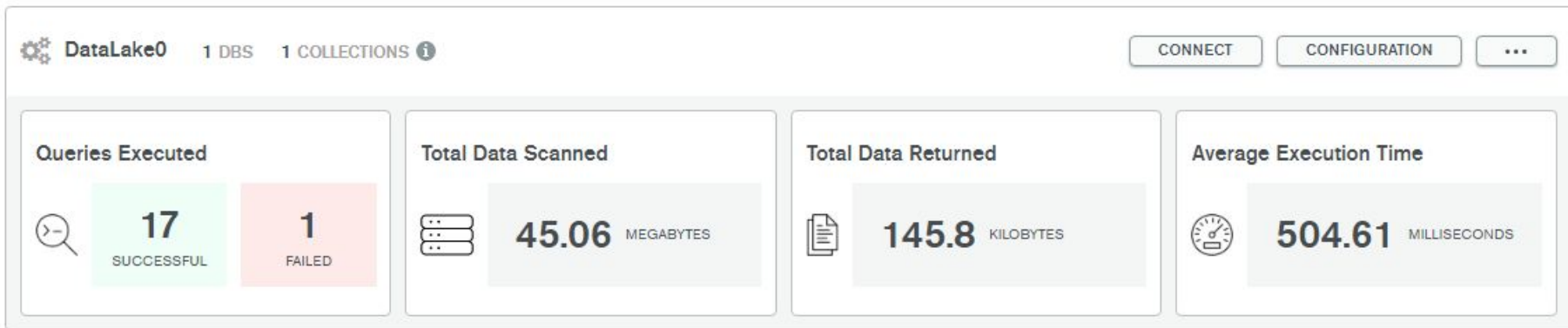
Aprovechen la oportunidad para practicar y consultar en su Data lake.

Nuevamente en el Atlas, en la pantalla del Data lake, se ve:

- Cantidad de consultas ejecutadas con éxito y las que fallaron.
- Número de datos leídos.
- Número de datos devueltos.
- Tiempo medio de ejecución en milisegundos.

## Data Lake

Configure a New Data Lake



# Conclusiones

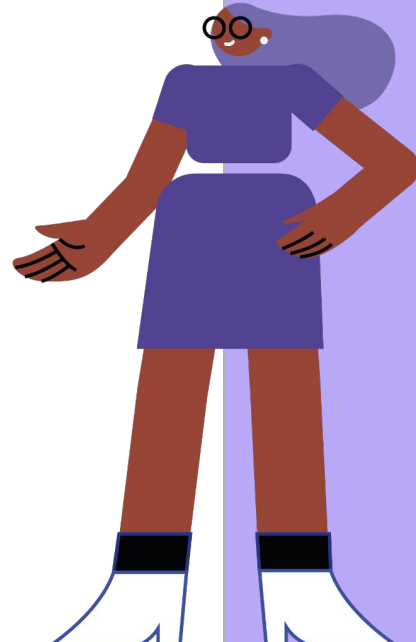
---

¡Bien hecho! Aprendieron a crear un Data lake.

Es ideal para almacenar grandes y variados conjuntos de datos en bruto en formato nativo.

Disfruten ahora de la creación de otros Data lakes.

¡Felicitaciones!



¡Muchas gracias!