

2024 - 2025

**Máster en Big Data y Data Science**

**01MBID – Fundamentos de la Tecnología de Big Data**

**Nombre: Gonzalo Antonio Delgado Rubio**

**Fecha: 19-05-2024**

**Curso 2022 – Ed. Abril**

**ACTIVIDAD1**

Contenido

[1. Creación de BD MongoDB 3](#_Toc167017618)

[2. Creación de Colecciones 4](#_Toc167017619)

[2.1 Carga de Datos Tweeter 5](#_Toc167017620)

[2.2 Carga de Datos Geoespecial 7](#_Toc167017621)

[3. Creación de Cuenta Google Colab 8](#_Toc167017622)

[4. Consulta y Manipulación de Datos 9](#_Toc167017623)

[4.1 Añadir Campos Amigos y Tweets enviados 9](#_Toc167017624)

[4.2 Calcular Nivel Frescura del tweet 10](#_Toc167017625)

[4.3 Calcular Nivel Madurez tweet 11](#_Toc167017626)

[5. visualización de Datos 13](#_Toc167017627)

[5.1 Cuenta de Tweet con Mayor cantidad de tweets enviados 14](#_Toc167017628)

[5.2 Cuenta de Tweet con Mayor cantidad de Amigos 14](#_Toc167017629)

[5.3 Cantidad Total de Tweets 15](#_Toc167017630)

[5.4 Cantidad de Tweets por cuenta 15](#_Toc167017631)

[5.5 Cuenta de tweet vs cantidad de amigos 16](#_Toc167017632)

[5.6 Cuenta de tweet vs tweets enviados 16](#_Toc167017633)

[5.7 Tweets vs Días de la semana 16](#_Toc167017634)

[5.8 Ordenar tweets por nivel de madurez. Cuenta de tweet más madura 17](#_Toc167017635)

[5.9 Ordenar los Tweets según su Frescura. Indicar también que cuenta de Twitter es la más Fresca. 18](#_Toc167017636)

[5.10 Mapa mundial según precios hundidos 18](#_Toc167017637)

# Creación de BD MongoDB

Utilizar el servicio cloud MongoDB Atlas

1. Ingresar a mongo DB Atlas y crearse una cuenta ([link](https://www.mongodb.com/products/platform/atlas-database))
2. Una vez creada la cuenta, procederemos a crear un proyecto.

Captura de pantalla de computadora

Descripción generada automáticamente

1. Con nuestro proyecto ya creado, procederemos a crear un cluster.

Interfaz de usuario gráfica, Aplicación, Sitio web

Descripción generada automáticamente

1. Podremos dar las siguientes configuraciones a nuestro cluster

Interfaz de usuario gráfica, Aplicación, Word

Descripción generada automáticamente

1. Una vez completado, se nos brindará la cadena de conexión a nuestro cluster mongo db sobre la cual estaremos trabajando.

Para la actividad, se esta empleando la siguiente cadena de conexión asociada al proyecto1 en mongo db atlas.

mongodb+srv://gonzalodelgador:oCepoCGGsI1bVlkU@mongocluster.46ahjqy.mongodb.net/

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

# Creación de Colecciones

Mediante la herramienta MongoDBCompas realizar la carga de la data fuente.

Para este apartado, descargaremos y emplearemos Mongo db Compas [link descarga](https://www.mongodb.com/products/tools/compass).

Antes de iniciar, deberemos emplear la cadena de conexión (obtenida en el apartado1) para poder conectarnos a nuestro cluster.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

## 2.1 Carga de Datos Tweeter

Cargar los datasets proporcionados por el profesor en formato JSON: 1) json de cuentas de twitter y 2) json de tweets.

Para esta sección realizaremos los siguientes pasos:

1. Sobre mongoDB Compas, crearemos una base de datos sobre la cual cargaremos nuestra colección tweet.

Interfaz de usuario gráfica, Aplicación, Teams

Descripción generada automáticamente

1. Ahora comenzaremos importando la data “tweets\_Actividad\_R.json”. Para esto debemos dar click en importar JSON.

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

1. Ahora nos dirigiremos a donde tenemos nuestro archivo almacenado.

Captura de pantalla de un celular

Descripción generada automáticamente

1. Daremos click a importar y esperamos a que se termine de cargar los datos

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

1. Repetir los mismos pasos para el archivo “tweets\_Actividad\_Cuentas\_R.json” para lo cual hemos creado una colección llamada tweets\_cuentas.

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

## 2.2 Carga de Datos Geoespecial

Cargar la colección de ejemplo de MongoDB Atlas cond datos de geolocalización (sample\_geospatial >> shipwrecks).

Para esta parte, debemos desplegar las bases de datos ejemplo sobre nuestro cluster. Para lo cual, realizaremos el siguiente paso.

1. Dirigirse a mongo DB Atlas
2. Sobre la web dirigirnos a nuestro proyecto, apartado Database y daremos click en Load Sample Databases.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

1. Luego de unos minutos, podremos ver las nuevas bases de datos cargadas sobre la herramienta MongoDB Compass.

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

# Creación de Cuenta Google Colab

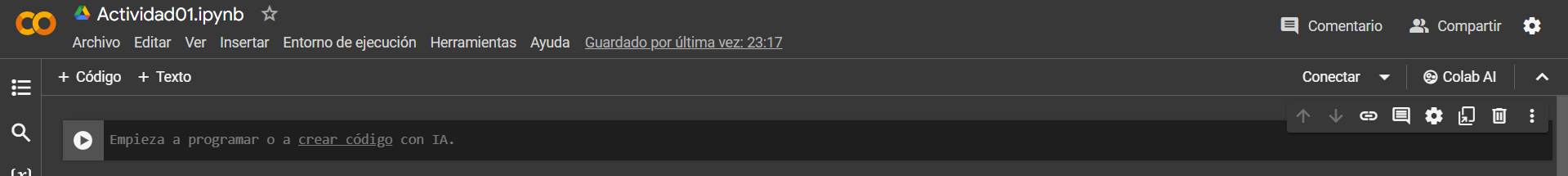
Para este apartado, emplearemos una cuenta de Google ya establecida y generaremos un archivo Google colab.

1. Podemos ingresar al siguiente link <https://colab.research.google.com/?hl=es>
2. Sobre el link crearemos un nuevo cuaderno de Google donde empezaremos a desarrollar las tareas solicitadas.

Captura de pantalla de un celular

Descripción generada automáticamente

1. Debemos tener algo similar a esto. Aquí empezaremos a codear.



# Consulta y Manipulación de Datos

Realizar los siguientes cálculos mediante consultas Python + MongoDb.

Para esto primero debemos conectarnos a la bd mediante el siguiente código

Link Consulta: <https://drive.google.com/file/d/1r0xV-kRhBno_PPHY8p7V02FRA1CczBa0/view?usp=sharing>

from pymongo import MongoClient

def conexion\_mongo\_db(cadena\_conexion:str, db\_name:str):

    client = MongoClient(cadena\_conexion)

    db = client[db\_name]

    return db

# Usar la función para crear una conexión

db = conexion\_mongo\_db(cadena\_conexion='mongodb+srv://gonzalodelgador:oCepoCGGsI1bVlkU@mongocluster.46ahjqy.mongodb.net/', db\_name='actividad01')

## 4.1 Añadir Campos Amigos y Tweets enviados

En la colección de twitters, añadir los campos amigos y tweets enviados: cargar los datos correspondientes mediante consulta mongodb + código python.

El siguiente método se encarga de añadir la cuenta de tweets enviados y amigos.

Texto

Descripción generada automáticamente

## 4.2 Calcular Nivel Frescura del tweet

En la colección de tweets, calcular la antigüedad del tweet en función de la fecha actual considerando antigüedad 0 el día de hoy y sumando +1 por cada día transcurrido. Este nuevo campo se llamará Frescura.

El siguiente código emplea la librería datetime para hacer la diferencia entre la fecha de creación del tweet y la fecha actual y poder realizar el cálculo del nivel de frescura solicitado.

from datetime import datetime

def calcular\_nivel\_frescura\_tweet(db):

    """Funcion encargada de calcular el nivel de frescura para la coleccion tweets"""

    collection = db['tweets']

    # obtenemos datos de fecha sistema

    now = datetime.now()

    now\_str = now.strftime('%Y-%m-%dT%H:%M:%S')

    # Definir el pipeline de agregación

    pipeline = [

        {

            '$addFields': {

                'created\_at\_date': {

                    '$toDate': '$created\_at'  # Convertir la cadena 'created\_at' a una fecha

                },

                'to\_date': now\_str  # Usar la fecha actual

            }

        },

        {

            '$addFields': {

                'Frescura': {

                    '$floor': [  # Redondear hacia abajo al entero más cercano

                        {

                            '$divide': [

                                {

                                    '$subtract': [

                                        {'$toDate': '$to\_date'},  # Convertir la cadena 'to\_date' a una fecha

                                        '$created\_at\_date'  # Restar la fecha de creación del tweet

                                    ]

                                },

                                1000 \* 60 \* 60 \* 24  # Convertir la diferencia de milisegundos a días

                            ]

                        }

                    ]

                }

            }

        },

        {

            '$merge': 'tweets'  # Actualizar los documentos existentes

        }

    ]

    # Ejecutar la operación de agregación

    collection.aggregate(pipeline)

## 4.3 Calcular Nivel Madurez tweet

En la colección de tweets, calcular la antigüedad del tweet relativa con la fecha de creación de la cuenta. Considerando antigüedad 0 si fue enviado el mismo día de creación de la cuenta y sumando +1 por cada día transcurrido desde entonces en función de la fecha del tweet. Este nuevo campo se llamará Madurez.

De igual forma, el código muestra los días de diferencia entre la fecha de creación de la cuenta y la fecha de creación del tweet para obtener el nivel de madurez solicitado.

Texto

Descripción generada automáticamente

Para hacer las validaciones de los cambios realizados se desplego código sobre pandas o también se puede visualizar sobre mongo DB Compass.

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

# visualización de Datos

Para la siguiente sección haremos uso de MongoDB Atlas para poder realizar los gráficos solicitados. Para ellos, sobre nuestro proyecto, nos dirigiremos a la sección de charts y crearemos un nuevo dashboard.

Los datos podrán ser visualizados según el link

<https://charts.mongodb.com/charts-project-0-fqvitlc/public/dashboards/6646e2b7-09fa-49f9-865c-06478ee4ff31>

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Deberemos obtener un espacio de trabajo como el siguiente

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

## 5.1 Cuenta de Tweet con Mayor cantidad de tweets enviados

Cuenta de Twitter con mayor cantidad de tweets enviados. Es un dato disponible dentro de los datos de cada tweet (Developer Plataform – Docs – Object Model).

Se emplea el tipo de gráfico top item para señalar el item con mayor cantidad de tweets enviados de acuerdo al dato statususes\_count

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

## 5.2 Cuenta de Tweet con Mayor cantidad de Amigos

Algo similar se emplea para la cuenta de Twitter con la mayor cantidad de amigos. Empleamos la colección tweets\_cuentas.

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente

## 5.3 Cantidad Total de Tweets

EL numero total de tweets lo obtenemos como la cuenta del campo id de la colección tweets.

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente

## 5.4 Cantidad de Tweets por cuenta

Genero una tabla a partir de la colección tweets mostrando el nombre de usuario y el promedio de la llave statususes\_count que lleva la cuenta de tweets de la cuenta.

Interfaz de usuario gráfica, Aplicación, Correo electrónico, Sitio web

Descripción generada automáticamente

Otra forma de hacerlo sería tomar la cantidad de tweets en la colección por usuario

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

## 5.5 Cuenta de tweet vs cantidad de amigos

Realizaremos el grafico agrupado de cuenta de tweeter según la cantidad de amigos, dato obtenido por medio de Python.

Gráfico

Descripción generada automáticamente

## 5.6 Cuenta de tweet vs tweets enviados

De forma similar hacemos el grafico agrupado de cantidad de tweets enviados agrupando el twiteer\_handle y statuses\_count.

Interfaz de usuario gráfica

Descripción generada automáticamente

## 5.7 Tweets vs Días de la semana

Para obtener la cantidad de tweet según el día de semana, se debe agrupar la información según el campo created\_at\_date originado mediante Python.

Interfaz de usuario gráfica

Descripción generada automáticamente

## 5.8 Ordenar tweets por nivel de madurez. Cuenta de tweet más madura

Obtengo la cuenta de tweet más madura agrupando por el valor máximo de nivel de madurez del tweet.

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

## 5.9 Ordenar los Tweets según su Frescura. Indicar también que cuenta de Twitter es la más Fresca.

Realizo un gráfico de barras donde agrupo la cuenta de tweet según el mínimo valor del nivel de frescura.

Interfaz de usuario gráfica, Aplicación, Sitio web

Descripción generada automáticamente

## 5.10 Mapa mundial según precios hundidos

Visualizar en el mapa mundial todos los pecios hundidos hasta 20 metros de profundidad. Pintando en verde los hundidos hasta 10 metros, amarillo hasta 15 metros y en rojo hasta 20 metros.

Se genera un gráfico GeoStatte donde mostraremos los valores de Depth según las coordenadas en el mapa y filtramos los valores entre 5 y 20 con un bin de 5.

Interfaz de usuario gráfica, Aplicación, Sitio web

Descripción generada automáticamente