

TALLER V: MONGODB

Objetivos

- Introducirse en los SGBD orientados a documentos (MongoDB) como base de datos remota (Mongo Atlas) y su organización.
- Familiarizarse con el uso de objetos JSON y aprender a realizar las operaciones CRUD (create, read, update & delete) sobre MongoDB.
- Comprender el funcionamiento del pipeline de agregación de *MongoDB*.

PREPARACIÓN

Instalación Cliente

Se deberá instalar un cliente de la base de datos, se dejan las dos opciones principales oficiales:

- **MongoDB Compass**: Una interfaz gráfica oficial de MongoDB muy práctica.
- **MongoDB Shell**: es la interfaz básica de MongoDB para interactuar con un servidor.

Conexión

Ya sea utilizando `mongosh` u otro IDE, se debe realizar la conexión a la base de datos, utilizando la siguiente uri (estará disponible sólo en los días cercanos a la clase):

```
mongodb+srv://estudiante:bddfiuba  
@taller.zxyqk.mongodb.net/tweets?readPreference=secondaryPreferred
```

Set de datos reducido: (siempre disponible)

```
mongodb+srv://estudiante:bddfiuba@tweets.cbxukni.mongodb.net/tweets
```

- (*Instalación Local de un servidor*) No será necesario pero si se desea indagar sobre la instalación del gestor de forma local **aquí**, o bien utilizando **docker**
- (*Herramientas Útiles*) No serán utilizadas durante el taller pero tenemos referencia a **varias herramientas** para facilitar la utilización de MongoDB.

ABM - DEMO

1. Agregar un tweet.
2. Modificar el tweet agregado para que tenga más retweets.
3. Agregar el campo de ubicación para el tweet agregado.
4. Eliminar el tweet agregado.

CONSULTAS BÁSICAS

1. Hallar los tweets del usuario con `userid` '818839458'.
2. Hallar aquellos tweets que tengan más de 500000 retweets.
3. Mostrar la cantidad de retweets de los tweets que se hayan hecho desde Argentina o Brasil.
4. Hallar los usuarios que tengan tweets con 200000 o más retweets y sean en idioma español.
5. Mostrar la cantidad de retweets para los tweets que no se hayan hecho en Argentina ni Brasil, pero sí tengan un lugar definido y sean en español.
6. Mostrar los `screen_name` de aquellos usuarios que tengan "Juan" como parte de su nombre.
7. Mostrar de los 10 tweets con más retweets, su usuario y la cantidad de retweets.

CONSULTAS DE AGREGACIÓN

1. Mostrar de los 10 tweets con más retweets, su usuario y la cantidad de retweets. Ordenar la salida de forma ascendente.
2. Encontrar los 10 hashtags más usados.
3. Encontrar a los 5 usuarios más mencionados. (les hicieron @)
4. Hallar la cantidad de retweets promedio para los tweets que se hayan hecho desde Argentina y aquellos que no.
5. Por cada usuario obtener una lista de ids de tweets y el largo de la misma.
6. Hallar la máxima cantidad de retweets *totales* que tuvo algún usuario.
7. Hallar para cada intervalo de una hora cuántos tweets realizó cada usuario.

ANEXO

Le sugerimos leer las siguientes entradas de la documentación de *MongoDB*:

- <https://docs.mongodb.com/manual/core/document/>
(Introducción a documentos JSON en MongoDB)
- <https://docs.mongodb.com/manual/reference/method/db.collection.find/>
(Búsquedas básicas con find)
- <https://docs.mongodb.com/manual/crud/>
(ABM)
- <https://docs.mongodb.com/manual/reference/operator/aggregation-pipeline/>
(Etapas del pipeline de agregación (aggregate))
- <https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/object-model/tweet>
(Documentación sobre el esquema de datos de tweets)