

## NoSQL – Práctica 1 – Carga de datos

### Preparación Mongo

- Descargar los ficheros minitweet.json y miniuser.json
- Crear una carpeta vacía para que se alojen los datos del servidor. Supongamos que esta carpeta se llama c:\hlocal\datos

Abrir 2 consolas (consola de Mongo en el laboratorio), terminal si hablamos de Linux

- En la primera arrancamos el servidor con

**mongod -dbpath c:\hlocal\datos**

este paso no es necesario si decidimos utilizar Atlas. Si lo hacemos debemos dejar esta consola aparte (no cerrarla, destruiríamos nuestro servidor)

- En la segunda vamos a importar los datos. Para ellos utilizamos mongoimport (o lo hacemos desde Compass). Por simplicidad suponemos que estamos en la carpeta que tiene los dos ficheros .json

**mongoimport -d usa -c tweet --drop --file minitweet.json**  
**mongoimport -d usa -c user --drop --file miniuser.json**

El primer import crea la colección tweet dentro de la base de datos usa con 3820 tweets de la campaña electoral americana 2020. El segundo carga datos de los 3789 usuarios que han emitido estos tweets.

- En la segunda consola, en la que hemos importado los datos, podemos ahora arrancar la shell de mongo como cliente con **mongo usa**

ya dentro de la shell podemos comprobar que todo ha ido bien mirando el número de documentos de cada colección:

```
> db.tweet.count()
3820
> db.user.count()
3789
```

### Examinando las colecciones

Antes de tratar los datos es conveniente “conocerlos” un poco.

#### Usuarios

Para los usuarios el esquema es muy sencillo, echar un vistazo y preguntar cualquier duda.

#### Tweets

En el caso de los tweets es más complejo porque además hay 2 tipos de documentos que se distinguen por el valor de la clave RT

- RT:true. Es un retweet. En este caso es muy importante la clave RT\_source que contiene el \_id del tweet original
- RT:false. No es un retweet. En este caso contiene mucha información extra destacando el lenguaje del tweet (lang) el propio texto del tweet (texto), un array con los usuarios que se mencionan y mucha más información