

PRÁCTICA – NOSQL

0. Iniciar el servidor (mongod -dbpath datos)

1. Bajar del campus los ficheros minitweet.json y miniuser.json

2. Desde un terminal de linux teclear

```
mongoimport --db test --collection user --file user.json
```

```
mongoimport --db test --collection tweet --file tweet.json
```

1 Usando la colección tweet, para cada usuario con al menos 10 tweets, indicar el número total de tweets. El usuario corresponde al campo user.screen. La salida será de la forma:

```
....
{ "_id" : "FoxNews", "total" : 471 }
{ "_id" : "DanScavino", "total" : 51 }
{ "_id" : "KellyannePolls", "total" : 18 }
{ "_id" : "LouDobbs", "total" : 34 }
....
```

2 La misma consulta pero mostrando solo el usuario con más tweets.

3 Repetir de nuevo la consulta 1 pero mostrando además qué proporción de los tweets son retweets (campo RT)

4 Repetir una vez más la consulta 1, pero mostrando además si el usuario está verificado o no (campo verified de user)

5 En tweet, para cada usuario (user.screen), queremos saber cuándo han emitido su primer tweet (primero en orden cronológico, es decir, más antiguo según *created_at*). Además, los usuarios también deben mostrarse por orden alfabético. Las primeras líneas:

```
{ "_id" : "0HOUR1____", "fecha" : ISODate("2016-11-01T16:21:33Z") }
{ "_id" : "2ALAW", "fecha" : ISODate("2016-11-01T23:02:38Z") }
{ "_id" : "ABCPolitics", "fecha" : ISODate("2016-11-01T14:56:31Z") }
```

6 En tweet, para cada usuario (user.screen) queremos saber cuántos tweets (RT:false) y cuantos retweets (RT:true) ha emitido, pero usando map-reduce

7 En tweet, queremos saber qué usuarios han sido los más mencionados (campo mentions) hacer que se muestre por pantalla una lista con cada usuario y su número de menciones, ordenado de más a menos menciones.

8 En user, queremos obtener un listado de las zonas horarias (time_zone) en un array sin repeticiones. La salida será del estilo

```
["Tehran", "Auckland", "America/New_York", "Karachi", "New Delhi", "London", "Central Time (US & Canada)", "America/Chicago", null, "Eastern Time (US & Canada)", "Quito", "Mexico City", "Atlantic Time (Canada)", "Arizona", "Indiana (East)", "Pacific Time (US & Canada)", "Mountain Time (US & Canada)"]
```

9 En tweet, cuando un tweet tiene RT:true, se incluye una clave *source* con el ID del tweet original. Escribir una consulta, que genere como salida una colección tweet2, que es igual a tweet solo que los retweets incluyen además una nueva clave 'sourceUser' con el usuario que publicó el tweet original (Solo se puede usar tweet)

10 En tweet, queremos saber el número total de tweets agrupados por día y hora (el mes no influye) Por ejemplo, el resultado nos dirá que el día 13 a la hora 0 hubo 7 tweets.