

Postwork Sesión 7

Team 22

1/28/2021

Objetivo

Realizar el alojamiento del fichero de un fichero `.csv` a una base de datos (BDD), en un local host de MongoDB a través de R

Desarrollo

Utilizando el manejador de BDD *Mongodb Compass* (previamente instalado), deberás de realizar las siguientes acciones:

1. Alojar el fichero `data.csv` en una base de datos llamada `match_games`, nombrando al `collection` como `match`
2. Una vez hecho esto, realizar un `count` para conocer el número de registros que se tiene en la base
3. Realiza una consulta utilizando la sintaxis de **Mongodb**, en la base de datos para conocer el número de goles que metió el Real Madrid el 20 de diciembre de 2015 y contra que equipo jugó, ¿perdió ó fue goleada?
4. Por último, no olvides cerrar la conexión con la BDD

Solución: Primero se importa, como `data.frame`, el conjunto de datos del fichero `data_ps07.csv` con la función `read.csv()`. Para establecer la conexión al servidor local de MongoDB se utiliza la librería `mongolite`. En el siguiente bloque de código se declara una instancia del objeto `mongolite::mongo` mediante su constructor, estableciendo los parámetros como: nombre de base de datos, colección, URL del host, etc. Después se insertan los datos mediante la función `inserty` finalmente se comprueba la operación realizando un conteo de los documentos mediante `count`.

```
library(mongolite)

# connection to local mongodb
m <- mongo(
  collection = 'match',
  db = 'match_games',
  url = 'mongodb://localhost',
  verbose = TRUE,
  options = ssl_options()
)

# read local dataset
match <- read.csv('../datasets/data_ps07.csv')

# insert into mongodb
m$insert(match)
```

```
## Processed 1000 rows...Complete! Processed total of 1140 rows.
```

```
## List of 5
```

```
## $ nInserted : num 1140
## $ nMatched  : num 0
## $ nRemoved  : num 0
## $ nUpserted : num 0
## $ writeErrors: list()
```

```
# check records
```

```
m$count()
```

```
## [1] 3420
```

Para resolver la pregunta de “¿Cuál es el número de goles que metió el Real Madrid el 20 de diciembre de 2015?” se utiliza la función `find` que recibe como parámetro una *query* con la sintaxis de MongoDB. La consulta es la siguiente:

```
# game math of Real Madrid on 2015-12-20
```

```
res <- m$find(
  '{
    "$and": [
      {"$or": [{"HomeTeam": "Real Madrid"}, {"AwayTeam": "Real Madrid"}]},
      {"Date": "2015-12-20"}
    ]
  }'
```

```
)

## Imported 0 records. Simplifying into dataframe...
```

Resultando con 0 registros que cumplan con el criterio de juegos jugados por el equipo **Real Madrid** (ya sea como local o visitante) en la fecha solicitada del 20/12/2015 dentro del conjunto de datos `data.csv` del repositorio.

Finalmente el cierre de conexión se puede lograr, mediante la destrucción de la instancia del objeto `monogolite::mongo` y después llamando al *garbage collector* de R con:

```
# close dataset
```

```
rm(m)
```

```
gc()
```

```
##          used (Mb) gc trigger (Mb) max used (Mb)
## Ncells  525826 28.1   1175234 62.8   654217 35.0
## Vcells 1016089  7.8    8388608 64.0   1770660 13.6
```