

Bases de Datos NoSQL

Tema 3: Bases de Datos Orientadas a Documentos. MongoDB

Ejercicio 2: Aggregation Framework

El documento que se adjunta ha sido obtenido a partir de datos disponibles en la página Web del Instituto Nacional de Estadística. Este documento contiene 52 documentos JSON con la información relativa al número de los habitantes desde el año 1971 al 2015 de las diferentes provincias de España.

- A. Importe en la base de datos BDProvincias el documento disponible en el Aula Virtual “provincias.json” utilizando el siguiente comando.

```
mongoimport --db BDProvincias --collection provincias --file  
"PATH/provincias.json"
```

Conéctese a la base de datos Mongo *BDProvincias* y compruebe que se han insertado 52 documentos en la colección *provincias* los 52 documentos (use la función *count()*).

- B. Utilizando el Aggregation Framework defina las siguientes consultas:

Parte I

1. Obtener tres provincias al azar.
2. Obtener cinco provincias de Andalucía (CA: “Andalucía”).
3. Obtener el nombre (que pasará a ser el identificador *_id*) y la superficie de 3 ciudades de Andalucía.
4. Obtener el nombre (identificador *_id*) y la superficie de 5 ciudades de Andalucía al azar ordenadas por superficie.
5. Obtener una nueva colección *provinciasCLM* que contenga documentos con el nombre y la superficie de las provincias de Castilla-La Mancha.

Parte II

6. Obtener la superficie y el número de provincias total.
7. Obtener una nueva colección *CAconProvincias* con un documento por cada Comunidad, que contenga el número de provincias, el total de superficie y un array con las provincias.
8. Obtener una colección con un documento por provincia que contenga la población del último año guardado (sin asumir que es el 2015).
9. Obtener el año con mayor población en España.
10. Obtener un documento por provincia, con la población del primer año almacenado, el último y la diferencia (*\$subtract*).
11. Obtener la comunidad autónoma que menos ha aumentado su densidad de población (o que más lo ha disminuido).

Parte III

12. Obtener el número de provincias de cada comunidad autónoma usando *sortByCount*.

13. Obtener las provincias incluyendo un campo que contenga el nombre de la provincia y el de la CA (\$concat).
14. Modificar la consulta anterior para que en el caso de las Ciudades Autónomas (sin Comunidad Autónoma) no salga null.
15. Obtener el número total de provincias utilizando \$count.
16. Generar las colecciones Comunidades y soloProvincias:

```
db.provincias.aggregate ([
  {$group:{"_id":"$CA",
    "numProvincias":{$sum:1}}},
  {$out: "Comunidades"})
```

```
db.provincias.aggregate([
  {$project:{Nombre:1,CA:1}},
  {$out:"soloProvincias"})
```

Y realizar un \$lookup entre ambas colecciones.