

Práctica - Agregación en MongoDB

Preparación:

- Descargar el fichero webs.json
- mongoimport -d sniff -c webs --drop < webs.json
- mongo sniff

La colección webs de la base de datos sniff tiene datos sobre páginas web visitadas por distintos usuarios. Cada documento es una *sesión* y se apunta el nombre del usuario, la fecha de inicio, el navegador utilizado y las páginas visitadas. Para cada página se indica la url y el número de segundos en los que la página ha estado activa.

Un ejemplo:

```
{
  "_id" : ObjectId("56b681556fde79920df96374"),
  "nombre" : "Bertoldo",
  "navegador" : "firefox",
  "fecha" : ISODate("2016-01-21T06:01:17.171Z"),
  "pags" : [
    {
      "url" : "http://www.vuelosparaAbuelos.es",
      "segs" : 500
    },
    {
      "url" : "https://www.mikokoloko.com",
      "segs" : 120
    },
    {
      "url" : "http://www.comomemongo.com",
      "segs" : 185
    },
    {
      "url" : "https://www.ucm.com",
      "segs" : 68
    }
  ]
}
```

1) Obtener el número de sesiones por cada navegador, ordenadas en orden ascendente.

La salida debe ser:

```
{ "_id" : "Chrome", "total" : 14 }
{ "_id" : "firefox", "total" : 17 }
{ "_id" : "Explorer", "total" : 20 }
```

Escribir en la solución la consulta.

Solución:

consulta:

```
db.webs.aggregate([{$group: {_id:'$navegador', total:{$sum:1}} },
  {$sort:{total:1}} ])
```

2) Persona con más sesiones en la colección.

Salida:

```
{ "_id" : "Herminia", "total" : 20 }
```

Copia la respuesta en la solución.

Solución

Consulta:

```
db.webs.aggregate([{$group: {_id:'$nombre', total:{$sum:1}} },
                    {$sort:{total:-1}},
                    {$limit:1},
                    ])
```

3) Para cada persona indicar el número de webs (urls) visitadas durante más de 100 segundos (cuentan las repeticiones).

Salida:

```
{ "_id" : "Aniceto", "num" : 35 }
```

```
{ "_id" : "Bertoldo", "num" : 51 }
```

```
{ "_id" : "Herminia", "num" : 40 }
```

Copia en la solución la consulta

Solución

Consulta:

```
db.webs.aggregate([
                    {$unwind:'$pags'},
                    {$match: {'pags.segs':{'$gt':100}}},
                    {$group: {_id:'$nombre', num:{$sum:1}} },
                    {$sort:{_id:1}}
                    ])
```

4) Nombre de la persona que visita más páginas de media por sesión

Pista1 : para cada sesión calcular su número de urls, y luego hacer la media de este dato por cada persona

Pista2: No hace falta \$unwind

Salida: { "_id" : "Bertoldo", "media" : 4.411764705882353 }

Copia en la solución la consulta

Solución

Consulta:

```
db.webs.aggregate([{$project:{nombre:1, totalurls:{$size:'$pags'}} },
                    {$group:{_id:'$nombre', media:{$avg:'$totalurls'}} },
                    {$sort:{media:-1}},
                    {$limit:1},
                    ])
```

5) Total de sesiones con Explorer en 2016.

Salida: { "_id" : 2016, "total" : 10 }

Ayuda: Para obtener el año de una fecha se puede usar {"año":{\$year:'\$fecha'}}

Solución:

Consulta:

```
db.webs.aggregate([{$match:{navegador:"Explorer"}},
  {$group:{_id:{$year:'$fecha'}, total:{$sum:1}}},
  {$match:{_id:2016}} ])
```

6) Número de sesiones por navegador y año.

Salida:

```
{ "navegador" : "Chrome", "año" : 2015 }
{ "navegador" : "Chrome", "año" : 2016 }
{ "navegador" : "Explorer", "año" : 2015 }
{ "navegador" : "Explorer", "año" : 2016 }
{ "navegador" : "firefox", "año" : 2016 }
```

Solución:

Consulta:

```
db.webs.aggregate([{$group:{_id:{'año':{$year:'$fecha'}, navegador:'$navegador' }, total:
{'$sum:1}}},
  {$project:{_id:0, navegador:'$_id.navegador', 'año':'$_id.año' } },
  {$sort:{navegador:1, 'año':1}} ])
```

7) Número de sesiones (documentos) en los que se ha visitado la web

“<https://www.mikokoloko.com>”.

Nota: si una sesión incluye dos visitas solo se contará una vez.

Copia en la solución la consulta.

Solución

```
db.webs.find( {"pags.url" : "https://www.mikokoloko.com" } ).count ( )
(27 documentos)
```

8) Número de webs distintas (urls diferentes) visitadas.

Pista1: las repeticiones no cuentan (conjuntos)

Pista2: En alguna etapa hay que agrupar todos los documentos

Salida:

```
{ "total" : 16 }
```

Copiar la consulta en la respuesta

Solución

Consulta:

```
db.webs.aggregate([
  {$unwind:'$pags'},
  {$group: {_id:null, urls:{$addToSet:'$pags.url'}} },
  {$project:{_id:0, total:{$size:'$urls'}} }
])
```

9) Para cada persona mostrar el tiempo total acumulado de sus urls correspondientes, en minutos.

Muestra la respuesta por orden alfabético de nombre.

Salida:

```
{ "nombre" : "Aniceto", "tiempomin" : 329.35 }
{ "nombre" : "Bertoldo", "tiempomin" : 263.33333333333333 }
{ "nombre" : "Herminia", "tiempomin" : 551.66666666666666 }
```

Copia la consulta en la solución.

Solución

Consulta:

```
db.webs.aggregate([{$unwind:'$pags'},
  {$group:{_id:"$nombre", tiempo:{$sum:'$pags.segs'}}},
  {$project:{nombre:'$_id', _id:0,
    tiempomin:{$divide:['$tiempo',60]}}},
  {$sort:{nombre:1}}
])
```

10) Nombre de las personas que han visitado alguna página que comienza por “https”.

Pista: usar el operador de proyección \$substr

Salida:

```
{ "nombre" : "Aniceto" }
{ "nombre" : "Bertoldo" }
```

Copiar la consulta en la solución.

Solución:

```
db.webs.aggregate([{$unwind:'$pags'},
  {$project:{nombre:1, _id:0,
    inicio:{$substr: [ "$pags.url", 0, 5 ]} }},
  {$match:{inicio:"https"}} ,
  {$group:{_id:null, nombre:{$addToSet:'$nombre'}}},
  {$unwind:'$nombre'},
  {$project:{nombre:1,_id:0}},
  {$sort:{nombre:1}}
])
```