

IT ACADEMY – DATA ANALYTICS

Alumna: Josselyn Maritza Jumpa Gordillo

SPRINT 9

Treballarem amb una base de dades que conté col·leccions relacionades amb una aplicació d'entreteniment cinematogràfic:

- users: Emmagatzema informació d'usuaris/es, incloent-hi noms, emails i contrasenyes xifrades.
- theatres: Conté dades de cinemes, com ID, ubicació (direcció i coordenades geogràfiques).
- sessions: Guarda sessions d'usuari, incloent-hi ID d'usuari i tokens JWT per a l'autenticació.
- movies: Inclou detalls de pel·lícules, com a trama, gèneres, durada, elenc, comentaris, any de llançament, directors, classificació i premis.
- comments: Emmagatzema comentaris d'usuaris/es sobre pel·lícules, amb informació de l'autor/a del comentari, ID de la pel·lícula, text del comentari i la data.

Duràs a terme algunes consultes que et demana el client/a, el qual està mesurant si seràs capaç o no de fer-te càrrec de la part analítica del projecte vinculat amb la seva base de dades.

NIVEL 1

Crea una base de dades amb MongoDB utilitzant com a col·leccions els arxius adjunts.

Se ha creado la base de datos “entertainment” que incluye 5 colecciones creadas a partir de los archivos json.

▼ entertainment

■ comments ...

■ movies

■ sessions

■ theaters

■ users

```
_id: ObjectId('5a9427648b0beeb69579cc')
name: "Andrea Le"
email: "andrea_le@fakegmail.com"
movie_id: ObjectId('573a1390f29313caabcd418c')
text: "Rem officiis eaque repellendus amet eos doloribus. Porro dolor volupta..."
date: 2012-03-26T23:20:16.000+00:00
```

```
_id: ObjectId('5a9427648b0beeb69579cf')
name: "Greg Powell"
```

EJERCICIO 1

Enunciado:

- Mostra els 2 primers comentaris que hi ha en la base de dades.
- Quants usuaris tenim registrats?
- Quants cinemes hi ha en l'estat de Califòrnia?
- Quin va ser el primer usuari/ària en registrar-se?
- Quantes pel·lícules de comèdia hi ha en la nostra base de dades?

Respuesta:

Muestra los 2 primeros comentarios

Si ordenamos los documentos de la colección comments por el field “date” de manera ascendente, estos serían los dos primeros comentarios:

Project { field: 0 }

Sort {date:1} Max Time MS 60000

Collation { locale: 'simple' } Skip 0 Limit 2

Index Hint { field: -1 }

ADD DATA EXPORT DATA UPDATE DELETE 1 - 2 of 2

```

_id: ObjectId('5a9427648b0beeb695e67e')
name: "Mercedes Tyler"
email: "mercedes_tyler@fakegmail.com"
movie_id: ObjectId('573a1399f29313caabcec3ce')
text: "Optio totam dolores magni. Enim ratione fuga tempora voluptatum est cu..."
date: 1970-01-01T01:07:09.000+00:00

```

```

_id: ObjectId('5a9427658b0beeb695efe')
name: "Jon Snow"
email: "kit_harington@gameofthron.es"
movie_id: ObjectId('573a13a4f29313caabd117df')
text: "Dolorem animi tempora ullam quas. Iusto nobis reprehenderit aspernatur..."
date: 1970-01-01T09:37:49.000+00:00

```

Si ordenamos por el field _id, estos serían los dos primeros comentarios:

Generate query Explain Reset Find

Project { field: 0 }

Sort {_id:1} Max Time MS 60000

Collation { locale: 'simple' } Skip 0 Limit 2

Index Hint { field: -1 }

ADD DATA EXPORT DATA UPDATE DELETE 1 - 2 of 2

```

_id: ObjectId('5a9427648b0beeb69579cc')
name: "Andrea Le"
email: "andrea_le@fakegmail.com"
movie_id: ObjectId('573a1390f29313caabcd418c')
text: "Rem officiis eaque repellendus amet eos doloribus. Porro dolor volupta..."
date: 2012-03-26T23:20:16.000+00:00

```

```


_id: ObjectId('5a9427648b0beeb69579cf')
name: "Greg Powell"
email: "greg_powell@fakegmail.com"
movie_id: ObjectId('573a1390f29313caabcd41b1')
text: "Tenetur dolorum molestiae ea. Eligendi praesentium unde quod porro. Co..."
date: 1987-02-10T00:29:36.000+00:00

```

¿Cuántos usuarios hay registrados?

Tenemos 185 usuarios registrados

Para ello, se ha utilizado la pestaña Aggregations y se ha utilizado el Stage \$count y se ha utilizado el field "name" de la colección users para realizar el recuento.


▼ Stage1 \$count 

```
1  /**
2   * Provide the field name for the count.
3   */
4   'name'
```

Output after \$count stage (Sample of 1 document)

name : 185

En la colección users, en la sección documents si no aplicamos ningún filtro, también vemos que hay 185 usuarios



1 – 20 of 185 

¿Cuántos cines hay en el estado de California?


Se han encontrado 169 cines en el estado de California.

La query utilizada es la siguiente:

```
{"location.address.state": "CA"}
```

⌵ {"location.address.state": "CA"}  [Generate query](#)  [Explain](#) [Reset](#)

[+ ADD DATA](#) [EXPORT DATA](#) [UPDATE](#) [DELETE](#)

1 – 20 of 169 

```
_id: ObjectId('59a47286cfa9a3a73e51e72e')
theaterId: 1008
location: Object
```

```
_id: ObjectId('59a47286cfa9a3a73e51e734')
theaterId: 1009
location: Object
```

```
_id: ObjectId('59a47286cfa9a3a73e51e737')
theaterId: 1018
location: Object
```

```
_id: ObjectId('59a47286cfa9a3a73e51e739')
theaterId: 1011
location: Object
```

¿Quién fue el primer usuario/a registrado?

El primero usuario registrado fue “Ned Stark”

Debido a que el ObjectId contiene información del timestamp, para averiguar el primer usuario registrado he ordenado los documentos de la colección users por el campo _id en orden ascendente y he limitado el resultado a 1.

The screenshot shows the MongoDB Compass interface. On the left, the 'Databases' sidebar is open, showing the 'users' collection under the 'entertainment' database. The main panel displays a query editor with the following query: `{ '_id': 1 }`. The query is executed, and the result shows a single document with the following fields: `_id` (ObjectId), `name` (Ned Stark), `email` (sean_bean@gameofthron.es), and `password` (a long alphanumeric string).

¿Cuántas películas de comedia hay en la base de datos?

Hay 7024 películas de comedia.

La query utilizada ha sido la siguiente:

`{genres:"Comedy"}`

The screenshot shows the MongoDB Compass interface. On the left, the 'Databases' sidebar is open, showing the 'movies' collection under the 'entertainment' database. The main panel displays a query editor with the following query: `{ 'genres': 'Comedy' }`. The query is executed, and the result shows a single document with the following fields: `_id` (ObjectId), `plot` (Cartoon figures announce, via comic strip balloons, that they will mov...), `genres` (Array of 3), `runtime` (7), `cast` (Array of 1), `num_mflix_comments` (1), `poster` (https://m.media-amazon.com/images/M/MV5BYzg2NjNhNTctMjUxMjU0ZWU4LWI3Zj...), `title` (Winsor McCay, the Famous Cartoonist of the N.Y. Herald and His Moving ...), `fullplot` (Cartoonist Winsor McCay agrees to create a large set of drawings that ...), `languages` (Array of 1), `released` (1911-04-08T00:00:00.000+00:00), `directors` (Array of 2), `writers` (Array of 2), `awards` (Object), `lastupdated` (2015-08-29 01:09:03.030000000), `year` (1911), `imdb` (Object), `countries` (Array of 1), `type` (movie), and `tomatoes` (Object).

EJERCICIO 2

Enunciado:

Mostra'm tots els documents de les pel·lícules produïdes en 1932, però que el gènere sigui drama o estiguin en francès.

Respuesta:

Se han encontrado 18 documentos.

La query utilizada es la siguiente:

```
{ $and: [{ year: 1932 }, { $or: [{ genres: "Drama" }, { languages: "French" }] } ] }
```

También realicé una comprobación de mayúsculas y minúsculas, pero el resultado era el mismo.

```
{ $and: [{ year: 1932 }, { $or: [{ languages: RegExp("french", 'i') }, { genres: RegExp("drama", 'i') } ] } ] }
```

The screenshot shows the MongoDB Compass interface. On the left, the 'Databases' sidebar lists 'admin', 'config', 'entertainment', and 'local'. The 'entertainment' database is selected, and the 'movies' collection is highlighted. The search bar contains the query: `{ $and: [{ year: 1932 }, { $or: [{ genres: 'Drama' }, { languages: 'French' }] }] }`. Below the search bar, there are buttons for 'ADD DATA', 'EXPORT DATA', 'UPDATE', and 'DELETE'. The results pane shows a single document with the following fields: `_id`, `plot`, `runtime`, `rated`, `cast`, `num_mflix_comments`, `poster`, `title`, `lastupdated`, `languages`, `released`, `directors`, `writers`, `awards`, `year`, `imdb`, `countries`, `type`, and `tomatoes`. The document represents a movie titled 'The Blood of a Poet' from 1932.

EJERCICIO 3

Enunciado:

Mostra'm tots els documents de pel·lícules estatunidenques que tinguin entre 5 i 9 premis que van ser produïdes entre 2012 i 2014.

Respuesta:

Se han encontrado 166 documentos. La query utilizada es la siguiente:

```
{ $and: [{ countries: "USA" }, { "awards.wins": { $gte: 5, $lte: 9 } }, { year: { $gte: 2012, $lte: 2014 } } ] }
```

🔍 `{ $and: [{countries: "USA"}, {"awards.wins": { $gte: 5, $lte: 9 }}, {year: 2013}] }` [Generate query](#) [Explain](#) [Reset](#) [Find](#)

[ADD DATA](#) [EXPORT DATA](#) [UPDATE](#) [DELETE](#) 1 - 20 of 166

```

_id: ObjectId('573a13acf29313caabd29366')
fullplot: "The manager of the negative assets sector of Life magazine, Walter Mit..."
imdb: Object
year: 2013
plot: "When his job along with that of his co-worker are threatened, Walter t..."
genres: Array (3)
rated: "PG"
metacritic: 54

```

NIVEL 2

EJERCICIO 1

Enunciado:

Compte quants comentaris escriu un usuari/ària que utilitza "gameofthron.es" com a domini de correu electrònic.

Respuesta:

Se han encontrado 22841 comentarios escritos por usuarios/usuarios que utilizan "gameofthron.es" como dominio en sus cuentas de correo electrónico.

En la sección aggregations de la colección comments se ha añadido un primer stage \$match para encontrar aquellos correos electrónicos con dominio "gameofthron.es".

La query ha sido {email: {\$regex:"gameofthron.es"}}

The screenshot shows the MongoDB Compass interface. On the left, the 'comments' collection is selected. The main area displays an aggregation pipeline with one stage: \$match. The query is: `{ "email": { "$regex": "gameofthron.es" } }`. The output shows a sample of 10 documents. Two documents are visible:

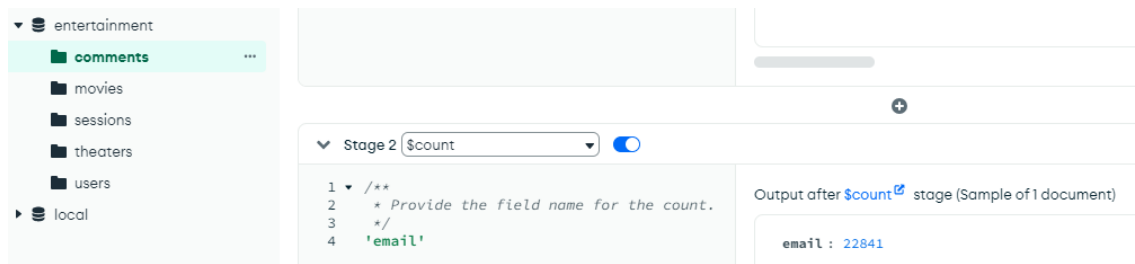
```

{
  "_id": "ObjectId('5a9427648b0beebe69579d0')",
  "name": "Talisa Maegyr",
  "email": "oona_chaplin@gameofthron.es",
  "movie_id": "ObjectId('573a1390f29313caabcc41b1')",
  "text": "Rem itaque ad sit rem voluptatibus. Ad fugiat maxime illum optio iure ...",
  "date": "1998-08-22T11:45:03.000+00:00"
}

{
  "_id": "ObjectId('5a9427648b0beebe69579d0')",
  "name": "Petyr Baelish",
  "email": "aidan_gillen@gameofthron.es",
  "movie_id": "ObjectId('573a1390f29313caabcc41b1')",
  "text": "Quo deserunt ipsam ipsum nemo nam sint praesentium",
  "date": "2001-07-13T19:25:09.000+00:00"
}

```

Luego, se ha añadido otro stage con \$count sobre el field email para realizar el recuento:

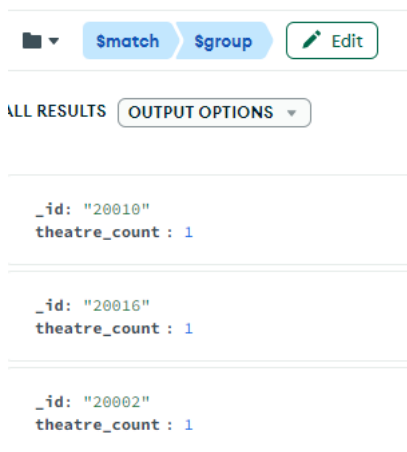


EJERCICIO 2

Enunciado:

Quants cinemes hi ha en cada codi postal situats dins de l'estat Washington D. C. (DC)?

Respuesta: Hay un cine para cada código postal situado en el estado DC. En total en ese estado se han encontrado 3 cines:



Para crear la query se han utilizado dos stages en la pestaña de Aggregations. En el primer stage se ha utilizado el método \$match para filtrar aquellos cines ubicados en DC.

```
{
  "location.address.state": "DC"
}
```

En el segundo stage se ha utilizado el método \$group para agruparlos por zipcode:

```
{
  "_id": "$location.address.zipcode",
  "theatre_count": {
    $sum:1
  }
}
```

Stage 1

\$match

```

1 /**
2  * query: The query in MQL.
3  */
4 {
5   "location.address.state": "DC"
6 }

```

Output after \$match stage (Sample of 3 documents)

```

_id: ObjectId('59a47286cfa9a3a73e51e785')
theaterId: 1092
location: Object
  address: Object
    street1: "3100 14th St Nw"
    city: "Washington"
    state: "DC"
    zipcode: "20010"
  geo: Object

```

```

_id: ObjectId('59a47287cfa9a3a73e51e786')
theaterId: 801
location: Object

```

Stage 2

\$group

```

1 /**
2  * _id: The id of the group.
3  * fieldN: The first field name.
4  */
5 {
6   "_id": "$location.address.zipcode",
7   "theatre_count": {
8     $sum: 1
9   }
10 }

```

Output after \$group stage (Sample of 3 documents)

```

_id: "20002"
theatre_count: 1

```

```

_id: "20010"
theatre_count: 1

```

NIVEL 3

EJERCICIO 1

Enunciado:

Troba totes les pel·lícules dirigides per John Landis amb una puntuació IMDb (Internet Movie Database) d'entre 7,5 i 8.

Respuesta:

Se han encontrado 4 películas.

La query utilizada es la siguiente:

```
{ $and: [{ directors: "John Landis" }, { "imdb.rating": { $gte: 7.5, $lte: 8 } } ] }
```

localhost:27017 > entertainment > movies

Documents 23.5K

Aggregations

Schema

Indexes 1

Validation

\$and: [{ directors: "John Landis" }, { "imdb.rating": { \$gte: 7.5, \$lte: 8 } }]

Generate query

Explain

Reset

Find

ADD DATA

EXPORT DATA

UPDATE

DELETE

1 - 4 of 4

```

_id: ObjectId('573a1397f29313caabce6d94')
fullplot: "Faber College has one frat house so disreputable it will take anyone. ..."
imdb: Object
  year: 1978
  plot: "At a 1962 college, Dean Vernon Wormer is determined to expel the entire..."
  genres: Array (1)
  rated: "R"

```


EJERCICIO 2

Enunciado:

Mostra en un mapa la ubicació de tots els teatres de la base de dades.

Respuesta: _