



<> Code

Issues

Pull requests

Actions

Projects

Wiki

Security



main

ud2-practica-4-mongo-db-Dansarasix-DML / Readme.md



Dansarasix-DML update changes 4

f2ef221 · 7 months ago



764 lines (574 loc) · 27.3 KB

Preview

Code

Blame



Raw



Big Data Aplicado

UD2 - Procesado y Presentación de Datos Almacenados

Práctica 4 MongoDB

Recuerda que para hacer la prácticas puedes optar por cualquiera de la las 3 opciones.

- Instalarla en tu máquina local
- Usar [Atlas MongoDB](#)
- Crear un contenedor docker.

*El objetivo de esta tarea es poner en práctica lo aprendido en la UD2 sobre **bases de datos documentales noSQL** a través de **MongoDB**. Esta debe cumplir con los siguientes objetivos:*

1. La fuente de datos a almacenar son de libre temática.
 2. Usa cualquiera de los recursos libres para la carga de datos
- <https://www.kaggle.com>
 - <https://www.ine.es>
 - <https://ec.europa.eu/eurostat>
 - <https://www.who.int/es/data/gho/publications/world-health-statistics>
 - <https://archive.ics.uci.edu/ml/index.php>
 - Cualquier otra fuente (con volumen de datos de mínimo más de 1 MB)

3. La Base de Datos debe tener **al menos 10.000 de documentos (mientras más mejor) o al menos, que tenga suficientes datos que faciliten la extracción de Conocimiento.**
4. Cada uno de los Documentos deben tener al menos **1 array de más de 3 elementos.**
5. Cada uno de los Documentos deben tener al menos un objeto que contenga **al menos 3 elementos clave-valor.** Se valorará positivamente cuantos más elementos anidados tengan tus documentos.
6. Si tenéis dificultades para obtener un dataset con dichos requisitos, podéis hacer uso de uno de las bases de datos que carga MongoDB de prueba en su servicio Cloud. Podéis elegir entre (algunos son mejores que otros para obtener conocimiento, piensa bien antes cuál es tu elección):
 - i. sample_airbnb
 - ii. sample_mflix
 - iii. sample_training
 - iv. sample_weather
7. La elección de un propio dataset será tomada en cuenta en la calificación.
8. Crea una base de datos llamada "db_pract4"
9. Realiza operaciones CRUD sobre los datos cargados.
 - Importa la Collection o Collections, dependiendo de la fuente de datos a cargar.
 - En **todas las operaciones** hay que trabajar con los **array y/o objetos** que tengan los documentos.
 - Realiza operaciones READ. Estas se pueden combinar (puede haber alguna sentencia que contenga varias operaciones al mismo tiempo, por ejemplo, lógica, comparación y de elemento):
 - Al menos 2 con operaciones de comparación.
 - Al menos 2 con operaciones lógicas.
 - Al menos 2 con operaciones sobre elementos y/o evaluación.
 - Al menos 3 con operaciones de agrupación (contar, ordenar,...).
 - Al menos 4 que tengan más de 3 condiciones clave-valor.
 - Al menos 1 con tipo de dato fecha.
 - Al menos 2 con operaciones de geoposición (Opcional)
 - Operaciones UPDATE
 - Al menos 2 con operaciones de actualización de campos (\$min , \$max , \$currentDate).

- Al menos 2 con operaciones de actualización de un array del documento (añadir o eliminar un elemento, cambiar la posición,...)
 - Al menos 2 con operaciones de modificación de actualización (*Modifiers update operators*) de un array del documento.
 - Al menos 2 que actualicen más de un documento simultáneamente
 - Comprueba el resultado en cada una de ellas. Es decir, muestra el estado antes y después de la modificación.
 - Operaciones DELETE
 - Al menos 2 que borren un documento utilizando al menos 2 campos como condición
 - Al menos 2 que borren un documento por algún elemento de un array
 - Al menos 1 que borren un documento por algún elemento de un objeto
 - Al menos 2 que borren más de un documento simultáneamente
 - Comprueba el resultado en cada una de ellas
 - Al menos 5 consultas AGGREGATE con al menos 3 condiciones diferentes. Se valorará cuanto mayor sea el número diferente de operadores de agregación usados.
10. Mientras mayor dificultad tengas las operaciones sobre la Base de Datos, más calificación llevará.
11. Es Imprescindible que todas las operaciones sobre la Base de Datos tenga la intención de Obtener CONOCIMIENTO a partir de la INFORMACIÓN que tienen las colecciones elegidas
12. Demuestra y justifica el punto anterior en cada una de las operaciones READ
13. Estamos en un *Curso de Especialización* de Big Data. No debe ser tomada como una práctica cualquiera. La práctica debe ser lo suficientemente completa como para el nivel de formación en el que estamos. Además, esta práctica esta pensada para evitar hacer un examen de Mongo de la UD2

Creación de la base de datos

Creamos nuestra base de datos como hecho anteriormente.

Create Database

Database Name

db_pract4

Collection Name

db_pract4

☐ Time-Series

Time-series collections efficiently store sequences of measurements over a period of time. [Learn More](#)

> Additional preferences (e.g. Custom collation, Capped, Clustered collections)

Cancel Create Database

La colección escogida es la de **movies** de **sample_mflix** la cual usaremos para extraer conocimiento acerca de las películas y como pueden los estudios cinematográficos pueden tomar mejores decisiones al respecto de como fue la recepción de la audiencia a lo largo del mundo.

Esta colección cuenta con **21.349** documentos a nuestra disposición.

Import

To collection db_pract4.db_pract4

Import file: sample_mflix.movies.json

Options

☐ Stop on errors

Cancel Import

Operaciones READ

Podemos empezar por lo siguiente:

Vamos a identificar películas largas y con alta valoración, lo cual podría sugerir que la duración no afecta negativamente la percepción del público.

Para ello, buscaremos películas con un duración superior a los 100 minutos y con una calificación en IMDB superior a 8.

```
db.db_pract4.find({  
  "runtime": { $gt: 100 },  
  "imdb.rating": { $gt: 8.0 }  
})
```



```
> db.db_pract4.find({  
  "runtime": { $gt: 100 },  
  "imdb.rating": { $gt: 8.0 }  
}).count()  
< 576
```

Hemos obtenido **576** películas, si comparamos con el total estas solo equivalen a 2,70% del total. Esto significa que el público no está dispuesto a ver películas que superen los 100 minutos de duración y que tengan una calificación alta en IMDB.

Queremos saber si el reconocimiento de premios se ha incrementado en películas de los últimos años.

Esto podría darnos indicios de si las películas de la década de los 2000 hasta ahora han sido premiadas con muchos premios (como mínimo 5 premios).

```
db.db_pract4.find({  
  "year": { $gt: 2000 },  
  "awards.wins": { $gt: 5 }  
})
```



```
> db.db_pract4.find({
  "year": { $gt: 2000 },
  "awards.wins": { $gt: 5 }
}).count()
< 2419
```

Hemos obtenido **2419** películas. Esto solo es un 11,33%. Por lo que podemos deducir que las películas de los 2000 no han tenido mucho éxito en las ceremonias de premios, obteniendo solamente menos de 5 premios o ninguno.

Vamos a estudiar si las películas de un género particular tienden a tener mayor aceptación, comparando con el puntaje de IMDb.

De esta forma podemos averiguar que géneros aún no siendo exitosos superan la cantidad de películas que si fueron exitosas y además pertenecen a otros géneros.

```
db.db_pract4.find({
  $or: [
    { "genres": "Western" },
    { "imdb.rating": { $gt: 8.0 } }
  ]
})
```



//Extracción de aquellas películas con un género concreto

```
db.db_pract4.find({
  "genres": "Western"
})
```

```
> db.db_pract4.find({
  $or: [
    { "genres": "Western" },
    { "imdb.rating": { $gt: 8.0 } }
  ]
}).count()
< 1197
```

La consulta vuelca **1197** películas, si realizamos una consulta para sacar sólo el género Western, está vuelca **242** resultados. Eso solo es un 20% del total de la consulta anterior.

Esto lleva a deducir que el género Western no ha tenido mucha relevancia histórica, y también nos indica que las películas con alta calificación IMDB no tienen por qué pertenecer a un género concreto sino que pueden tener varios géneros.

Busquemos aquellas películas que son populares entre el público, pero que no fueron reconocidas por instituciones de premios.

Esto nos indicaría que películas fueron ignoradas por las ceremonias de premios pero no por la audiencia, lo que nos indicaría que una película no necesita de premios para ser un éxito.

```
db.db_pract4.find({
  $and: [
    { "imdb.votes": { $gt: 5000 } },
    { "awards.nominations": 0 }
  ]
})
```



```
> db.db_pract4.find({
  $and: [
    { "imdb.votes": { $gt: 5000 } },
    { "awards.nominations": 0 }
  ]
}).count()
< 562
```

Hemos recuperado **562** películas, esto solo es 2,63%. Por lo que podemos concluir que una película con muchos votos en IMDB es raro que no tenga alguna premiación.

Conocer qué películas se categorizan bajo múltiples géneros, lo que puede indicar una trama compleja o un cruce de estilos.

Podemos saber si las películas suelen tener más de un género o si en verdad solo se especializan en un género.

```
db.db_pract4.find({
  "genres": { $size: 3 }
```



```
}}
```

```
> db.db_pract4.find({
  "genres": { $size: 3 }
}).count()
< 8696
```

Hemos obtenido **8696** películas, esto es 41% del total. Una cantidad bastante grande sin superar el 50%.

Por lo que podemos concluir que casi la mitad de las películas tienen más de un género.

Investigar la recurrencia de actores con un apellido común, que puede indicar una representación frecuente en ciertas películas.

Saber la recurrencia de los actores nos puede servir para saber que actores han participado más veces, lo cual puede influir en la audiencia que les puede gustar como actua, su carisma, etc.

```
db.db_pract4.find({
  "cast": { $elemMatch: { $regex: /Smith/i } }
})
```



```
> db.db_pract4.find({
  "cast": { $elemMatch: { $regex: /Smith/i } }
}).count()
< 226
```

Con solo 226 películas, solo es un 1%. Por lo que actores que se llamen Smith no han tenido mucha participación o nuestro dataset no tienen suficientes películas que lo incluyan para ser un valor significativo.

Observar patrones de lanzamientos de películas por año, para identificar periodos de alta o baja producción.

Sabiendo el nº de películas lanzadas por año podemos identificar si hubo momentos de baja productividad e identificar el por qué de esa baja productividad (huelgas, accidentes, regrabaciones, etc.)

Haremos una agregación en donde primero haremos un `$group` con el año y el nº de películas usando `$sum`. Y luego un `$sort` sobre el año.

```
db.getCollection('db_pract4').aggregate(  
  [  
    {  
      $group: {  
        _id: '$year',  
        n_peliculas: { $sum: 1 }  
      }  
    },  
    { $sort: { _id: 1 } }  
  ],  
  { maxTimeMS: 60000, allowDiskUse: true }  
);
```

**ALL RESULTS****OUTPUT OPTIONS** ▼

_id: 1896
n_peliculas : 2

_id: 1903
n_peliculas : 1

_id: 1909
n_peliculas : 1

_id: 1911
n_peliculas : 2

_id: 1913
n_peliculas : 1

Los resultados muestran que los primeros años, el cine en su nacimiento tenía muy poca producción debido a que era una nueva forma de entretenimiento y la gente todavía no estaba especializada en el campo.

Pero el pasar del tiempo hizo que esta productividad incrementara gracias a las compañías que salieron adelante y se especializaron en esta materia.

ALL RESULTS**OUTPUT OPTIONS** ▼

_id: 2010
n_peliculas : 866

_id: 2011
n_peliculas : 893

_id: 2012
n_peliculas : 955

_id: 2013
n_peliculas : 1105

_id: 2014
n_peliculas : 1073

Veamos las películas agrupadas por el género principal y ordenadas por el rating promedio

Podemos intentar ver que géneros son más aceptados por la audiencia. Saber los gustos de la audiencia nos permitirá la posibilidad de atraer público potencial en próximas producciones.

Para ello haremos la siguiente agregación:

```
db.getCollection('db_pract4').aggregate(  
  [  
    { $unwind: { path: '$genres' } },  
    {  
      $group: {  
        _id: '$genres',  
        avgRating: { $avg: '$imdb.rating' }  
      }  
    }  
  ]  
)
```



```
    }  
  },  
  { $sort: { avgRating: -1 } },  
  {  
    $project: {  
      genre: '$_id',  
      _id: 0,  
      avgRating: { $round: ['$avgRating', 2] }  
    }  
  }  
],  
{ maxTimeMS: 60000, allowDiskUse: true }  
);
```

ALL RESULTS**OUTPUT OPTIONS** ▼

genre : "Film-Noir"
avgRating : 7.4

genre : "Short"
avgRating : 7.38

genre : "Documentary"
avgRating : 7.37

genre : "News"
avgRating : 7.25

genre : "History"
avgRating : 7.17

Aquí estamos viendo los 5 primeros géneros con mayor calificación en IMDB. Siendo el ranking el siguiente:

1. Cine Negro

2. Corto
3. Documental
4. Noticias (o Nuevas)
5. Historia

Esta información podría dar pistas de lo que más le gusta a la audiencia y hacer producciones nuevas con estos géneros que pueden atraer nuevas audiencias.

Identificar los directores más prolíficos en la colección y sus contribuciones al cine.

Encontrar a los mejores directores es crucial para que una producción sea exitosa. Es por ello que si los identificamos a tiempo, podremos entablar negociaciones con ellos para que trabajen en nuestro equipo.

Si además tenemos en cuenta los premios que han ganado, tendremos más posibilidades de tener un éxito.

Hacemos la siguiente agregación:

```
db.getCollection('db_pract4').aggregate(  
  [  
    { $unwind: { path: '$directors' } },  
    {  
      $group: {  
        _id: '$directors',  
        peliculasDirigidas: { $sum: 1 },  
        premiosGanados: { $sum: '$awards.wins' }  
      }  
    },  
    { $sort: { premiosGanados: -1 } },  
    {  
      $project: {  
        directo: '$_id',  
        _id: 0,  
        peliculasDirigidas: '$peliculasDirigidas',  
        premiosGanados: '$premiosGanados'  
      }  
    }  
  ],  
  { maxTimeMS: 60000, allowDiskUse: true }  
);
```



```
directo : "Steven Spielberg"  
peliculasDirigidas : 29  
premiosGanados : 747
```

```
directo : "Martin Scorsese"  
peliculasDirigidas : 30  
premiosGanados : 619
```

```
directo : "Alfonso Cuarèn"  
peliculasDirigidas : 8  
premiosGanados : 586
```

```
directo : "Peter Jackson"  
peliculasDirigidas : 13  
premiosGanados : 546
```

```
directo : "Joel Coen"  
peliculasDirigidas : 18  
premiosGanados : 520
```

La lista resultante es la siguiente:

1. Steven Spielberg
2. Martin Scorsese
3. Alfonso Cuarèn
4. Peter Jackson
5. Joel Coen

Tenemos que tener en cuenta también si los directores han fallecido o se han jubilado. Pero en nuestro dataset no tenemos dicha información, habría que consultar con los datasets necesarios para averiguar dicha información.

Examinar películas en un idioma que hayan tenido una buena recepción a pesar de no ser nominadas en gran medida.

Podemos buscar esas películas que fueron éxitosas o incluso nominadas a pesar de solo estar en un solo idioma. De esa forma se pueden doblar en otros idiomas. En este caso, buscaremos solo aquellas que están en inglés.

```
db.db_pract4.find({
  "languages": "English",
  "imdb.rating": { $gt: 7.0 },
  "awards.nominations": { $lt: 2 },
  "awards.wins": { $gte: 1 }
})
```



```
> db.db_pract4.find({
  "languages": "English",
  "imdb.rating": { $gt: 7.0 },
  "awards.nominations": { $lt: 2 },
  "awards.wins": { $gte: 1 }
}).count()
< 1397
```

Vemos que **1397** películas con solo el idioma inglés con buenas calificaciones en IMDB y con nominaciones y premios. Las plataformas podrían usar esta información para doblar estas películas y acercar audiencias de otros países.

Podemos analizar el impacto de un elenco reducido en películas bien recibidas por la audiencia.

Películas bien valoradas con un elenco pequeño nos puede indicar que no necesitas muchos actores en la mayoría de los casos para tener un éxito. También nos puede servir para buscar actores que pueden ser nuevos en el medio y despegar su carrera.

En este caso, vamos a usar la puntuación de la audiencia de Rotten Tomatoes en vez de IMDB.

```
db.db_pract4.find({
  "type": "movie",
  "tomatoes.viewer.meter": { $gt: 70 },
  "year": { $gt: 1990 },
  "cast": { $exists: true, $type: "array" },
  "$expr": { $lt: [{ $size: "$cast" }, 3] }
})
```



```
> db.db_pract4.find({
  "type": "movie",
  "tomatoes.viewer.meter": { $gt: 70 },
  "year": { $gt: 1990 },
  "cast": { $exists: true, $type: "array" },
  "$expr": { $lt: [{ $size: "$cast" }, 3] }
}).count()
< 128
```

Hemos obtenido 128 películas, esta cantidad es bastante pequeña. Lo cual indica que es más probable tener películas con un cast muy amplio que son muy exitosas.

Evaluar si las películas largas y de una clasificación obtienen una valoración favorable del público.

Podemos intentar buscar las películas largas de una clasificación concreta en una región concreta si tienen mucha aceptación por parte de la audiencia.

```
db.db_pract4.find({
  "runtime": { $gt: 120 },
  "imdb.rating": { $gt: 8 },
  "countries": "USA",
  "rated": "PG-13"
})
```



```
> db.db_pract4.find({
  "runtime": { $gt: 120 },
  "imdb.rating": { $gt: 8 },
  "countries": "USA",
  "rated": "PG-13"
}).count()
< 30
> db.db_pract4.find({
  "runtime": { $gt: 120 },
  "countries": "USA",
  "rated": "PG-13"
}).count()
< 393
```

De 393 películas estadounidenses con una duración mayor a 2 horas y con una calificación de PG-13, solo 30 tienen una alta puntuación en IMDB.

Lo cual nos puede indicar que el público no está en muchas ocasiones a ver una película extremadamente larga y que aspectos como que la trama esté bien escrita pueden influir lo suficiente en la audiencia para que les guste.

Podemos examinar cómo se perciben las películas de modernas de un género concreto que son populares entre el público y la crítica.

Investigar que géneros son más populares en la actualidad puede darnos una pista de aquellos géneros en los que podemos trabajar para una próxima producción.

Usaremos de ejemplo el género de drama.

```
db.db_pract4.find({
  "genres": "Drama",
  "year": { $gt: 2000 },
  "imdb.votes": { $gt: 5000 },
  "tomatoes.viewer.meter": { $gte: 80 }
})
```




```
> db.db_pract4.find({
  "genres": "Drama",
  "year": { $gt: 2000 },
  "imdb.votes": { $gt: 5000 },
  "tomatoes.viewer.meter": { $gte: 80 }
}).count()
< 494
```

Hemos obtenido **494** películas, un 2,3% del total. Esto nos indica que el género Drama no ha tenido mucho éxito en la década de los 2000 hasta nuestros días.

Obtener las películas con información reciente para identificar datos que pueden haber cambiado, como nuevas premiaciones o valoraciones.

Es importante saber si se han actualizado los datos de nuestro dataset para comprobar que la nueva información añadida sea veráz y de calidad.

```
db.db_pract4.find({
  "lastupdated": { $gt: new Date("2015-01-01T00:00:00Z") }
})
```



```
> db.db_pract4.find({
  "lastupdated": { $gt: new Date("2015-01-01T00:00:00Z") }
}).count()
< 0
```

No hemos obtenido nada, por lo que no hay actualizaciones ni nueva información que debamos destacar en nuestro dataset.

Obtener el promedio de calificación IMDb, Rotten Tomatoes y duración por país para películas con un rating IMDb

Examinar la producción cinematográfica de distintos países en términos de duración, calidad (según IMDb y Rotten Tomatoes) y número de películas, para ver si existen diferencias regionales en la producción y la recepción de películas.

```
db.getCollection('db_pract4').aggregate(
[
  { $match: { 'imdb.rating': { $gt: 7 } } },
  { $unwind: { path: '$countries' } },
  {
```



```
$group: {
  _id: '$countries',
  promedioIMDB: { $avg: '$imdb.rating' },
  promedioTomatoes: {
    $avg: '$tomatoes.viewer.rating'
  },
  promedioDuracion: { $avg: '$runtime' },
  totalPelículas: { $sum: 1 }
},
{ $match: { totalPelículas: { $gt: 10 } } },
{ $sort: { promedioIMDB: -1 } },
{
  $project: {
    _id: 0,
    pais: '$_id',
    promedioIMDB: {
      $round: ['$promedioIMDB', 2]
    },
    promedioTomatoes: {
      $round: ['$promedioTomatoes', 2]
    },
    promedioDuracion: {
      $round: ['$promedioDuracion', 2]
    },
    totalPelículas: '$totalPelículas'
  }
},
{ maxTimeMS: 60000, allowDiskUse: true }
);
```

```
pais : "India"  
promedioIMDB : 7.77  
promedioTomatoes : 3.63  
promedioDuracion : 141.56  
totalPeliculas : 320
```

```
pais : "Turkey"  
promedioIMDB : 7.73  
promedioTomatoes : 3.74  
promedioDuracion : 106.51  
totalPeliculas : 70
```

```
pais : "Soviet Union"  
promedioIMDB : 7.72  
promedioTomatoes : 3.5  
promedioDuracion : 106.8  
totalPeliculas : 133
```

```
pais : "Yugoslavia"  
promedioIMDB : 7.72  
promedioTomatoes : 4  
promedioDuracion : 110.15  
totalPeliculas : 26
```

```
pais : "Indonesia"  
promedioIMDB : 7.71  
promedioTomatoes : 3.65  
promedioDuracion : 105.75  
totalPeliculas : 12
```

Vemos que los países que mayor producción de películas son:

1. India
2. Turquía
3. Unión Soviética (Rusia)
4. Yugoslavia (Desapareció)
5. Indonesia

Lo cuál nos indica que las películas mejor valoradas por IMDB son de India y Turquía, esto nos puede indicar que realizar una producción india o turca puede tener mucho éxito (según IMDB).

Top 5 actores más frecuentes, el promedio de rating de las películas en las que participan, y el número de premios de esas películas

Analizar la carrera de los actores más prolíficos para identificar si su popularidad y alta participación están relacionadas con el éxito y los premios obtenidos en las películas donde actúan.

```
db.getCollection('db_pract4').aggregate(  
  [  
    { $unwind: { path: '$cast' } },  
    {  
      $group: {  
        _id: '$cast',  
        promedioRating: { $avg: '$imdb.rating' },  
        totalPeliculas: { $sum: 1 },  
        premiosTotales: { $sum: '$awards.wins' }  
      }  
    },  
    {  
      $match: {  
        totalPeliculas: { $gt: 5 },  
        promedioRating: { $gte: 6.5 }  
      }  
    },  
    { $sort: { totalPeliculas: -1 } }, //Hemos ordenado según el nº de  
    películas realizadas  
    { $limit: 5 },  
    {  
      $project: {  
        _id: 0,  
        actor: '$_id',  
        promedioRating: {  
          $round: ['$promedioRating', 2]  
        },  
        totalPeliculas: '$totalPeliculas',  
        premiosTotales: '$premiosTotales'  
      }  
    }  
  ],  
  { maxTimeMS: 60000, allowDiskUse: true }  
);
```



ALL RESULTS

OUTPUT OPTIONS ▾

```
actor : "G rard Depardieu"
promedioRating : 6.69
totalPeliculas : 67
premiosTotales : 224
```

```
actor : "Robert De Niro"
promedioRating : 6.96
totalPeliculas : 58
premiosTotales : 391
```

```
actor : "Michael Caine"
promedioRating : 6.71
totalPeliculas : 51
premiosTotales : 403
```

```
actor : "Morgan Freeman"
promedioRating : 7.16
totalPeliculas : 48
premiosTotales : 483
```

```
actor : "Christopher Plummer"
promedioRating : 6.99
totalPeliculas : 47
premiosTotales : 471
```

Los 5 actores m s frecuentes son:

1. G rard Depardieu
2. Rober De Niro
3. Michael Caine
4. Morgan Freeman
5. Christopher Plummer

Vemos que aunque G rard haya producido m s pel culas, Morgan Freeman es el actor con m s reconocimiento a n siendo el cuarto de la lista.

Buscar el promedio de duraci n y rating de IMDb y Rotten Tomatoes de pel culas por g nero.

Nos puede interesar qué géneros tienden a tener películas de mayor duración y aceptación en plataformas de calificación. Puede ser de ayuda saber los géneros donde una mayor duración no reduce la valoración y, de hecho, puede estar correlacionada con un mayor éxito de la audiencia y la crítica.

```
db.getCollection('db_pract4').aggregate(  
  [  
    {  
      $match: {  
        type: 'movie',  
        'imdb.votes': { $gt: 3000 },  
        'tomatoes.viewer.meter': { $gt: 80 }  
      }  
    },  
    { $unwind: { path: '$genres' } },  
    {  
      $group: {  
        _id: '$genres',  
        avgDuration: { $avg: '$runtime' },  
        avgIMDB: { $avg: '$imdb.rating' },  
        avgTomatoes: {  
          $avg: '$tomatoes.viewer.rating'  
        },  
        totalPelículas: { $sum: 1 }  
      }  
    },  
    { $sort: { avgIMDB: -1 } },  
    {  
      $project: {  
        _id: 0,  
        genre: '$_id',  
        avgDuration: {  
          $round: ['$avgDuration', 2]  
        },  
        avgIMDB: { $round: ['$avgIMDB', 2] },  
        avgTomatoes: {  
          $round: ['$avgTomatoes', 2]  
        },  
        totalPelículas: '$totalPelículas'  
      }  
    }  
  ],  
  { maxTimeMS: 60000, allowDiskUse: true }  
);
```



```
genre : "Short"  
avgDuration : 17.35  
avgIMDB : 8.08  
avgTomatoes : 4.12  
totalPelículas : 23
```

```
genre : "Film-Noir"  
avgDuration : 95.94  
avgIMDB : 7.92  
avgTomatoes : 3.98  
totalPelículas : 18
```

```
genre : "Documentary"  
avgDuration : 99.99  
avgIMDB : 7.89  
avgTomatoes : 4.01  
totalPelículas : 121
```

```
genre : "Animation"  
avgDuration : 79.18  
avgIMDB : 7.87  
avgTomatoes : 3.97  
totalPelículas : 120
```

```
genre : "War"  
avgDuration : 127.02  
avgIMDB : 7.86  
avgTomatoes : 3.99  
totalPelículas : 122
```

Con la información obtenida, vemos que los cortos tienen una puntuación mucho más alta que el resto de películas. También aparecen otros géneros como la animación o las que tratan de guerras que no habían aparecido con anterioridad.

Lo más curioso es la puntuación de Rotten Tomatoes (audiencia), la cual no supera el 5. Esta discrepancia de opiniones puede ser crucial para entender a nuestra audiencia y llegar a puntos comunes para atraer público de ambos sitios.

Operaciones UPDATE

Para las operaciones UPDATE usaremos la siguiente tabla:

Usamos `$min` para cambiar el rating de IMDB.

```
db.db_pract4.updateOne(
  { "_id": ObjectId("573a1390f29313caabcd42e8") },
  { $min: { "imdb.rating": 7.0 } }
);
```



Dato Anterior	Dato Actualizado
<pre> _id: ObjectId('573a1390f29313caabcd42e8') plot: "A group of bandits stage a brazen train hold-up, only to find a determ..." genres: Array (2) runtime: 11 cast: Array (4) poster: "https://m.media-amazon.com/images/M/MV5BMTU3NjESNzYtYTYyNS08MDVmlWIwYj..." title: "The Great Train Robbery" fullplot: "Among the earliest existing films in American cinema - notable as the ..." languages: Array (1) released: 1903-12-01T00:00:00.000+00:00 directors: Array (1) rated: "TV-G" awards: Object lastupdated: "2015-08-13 00:27:59.177000000" year: 1903 imdb: Object rating: 7.4 votes: 9847 id: 439 </pre>	<pre> _id: ObjectId('573a1390f29313caabcd42e8') plot: "A group of bandits stage a brazen train hold-up, only to find a determ..." genres: Array (2) runtime: 11 cast: Array (4) poster: "https://m.media-amazon.com/images/M/MV5BMTU3NjESNzYtYTYyNS08MDVmlWIwYj..." title: "The Great Train Robbery" fullplot: "Among the earliest existing films in American cinema - notable as the ..." languages: Array (1) released: 1903-12-01T00:00:00.000+00:00 directors: Array (1) rated: "TV-G" awards: Object lastupdated: "2015-08-13 00:27:59.177000000" year: 1903 imdb: Object rating: 7 votes: 9847 id: 439 </pre>

Actualizamos el documento con \$currentDate .

```
db.db_pract4.updateOne(
  { "_id": ObjectId("573a1390f29313caabcd446f") },
  { $currentDate: { "lastupdated": true } }
);
```



Dato Anterior	Dato Actualizado
<pre> _id: ObjectId('573a1390f29313caabcd446f') plot: "A greedy tycoon decides, on a whim, to corner the world market in whea..." genres: Array (2) runtime: 14 cast: Array (4) num_mflix_comments: 1 title: "A Corner in Wheat" fullplot: "A greedy tycoon decides, on a whim, to corner the world market in whea..." languages: Array (1) released: 1909-12-13T00:00:00.000+00:00 directors: Array (1) rated: "G" awards: Object lastupdated: "2015-08-13 00:46:30.660000000" year: 1909 imdb: Object countries: Array (1) </pre>	<pre> _id: ObjectId('573a1390f29313caabcd446f') plot: "A greedy tycoon decides, on a whim, to corner the world market in whea..." genres: Array (2) runtime: 14 cast: Array (4) num_mflix_comments: 1 title: "A Corner in Wheat" fullplot: "A greedy tycoon decides, on a whim, to corner the world market in whea..." languages: Array (1) released: 1909-12-13T00:00:00.000+00:00 directors: Array (1) rated: "G" awards: Object lastupdated: 2024-11-09T15:09:30.713+00:00 year: 1909 imdb: Object countries: Array (1) type: "movie" </pre>

Añadir un nuevo género a la lista de géneros (\$push)

```
db.db_pract4.updateOne(
  { "_id": ObjectId("573a1390f29313caabcd4803") },
  { $push: { "genres": "Adventure" } }
);
```



Dato Anterior	Dato Actualizado
<pre> _id: ObjectId('573a1390f29313caabcd4803') plot: "Cartoon figures announce, via comic strip balloons, that they will mov..." genres: Array (3) 0: "Animation" 1: "Short" 2: "Comedy" runtime: 7 cast: Array (1) 0: num_mflix_comments: 0 poster: "https://m.media-amazon.com/images/M/MV5BYzg2NjNHNtctMjUxMj00ZWU4LWI3Zj..." 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 (1) </pre>	<pre> _id: ObjectId('573a1390f29313caabcd4803') plot: "Cartoon figures announce, via comic strip balloons, that they will mov..." genres: Array (4) 0: "Animation" 1: "Short" 2: "Comedy" 3: "Adventure" runtime: 7 cast: Array (1) 0: num_mflix_comments: 0 poster: "https://m.media-amazon.com/images/M/MV5BYzg2NjNHNtctMjUxMj00ZWU4LWI3Zj..." 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 (1) </pre>

Eliminar el último actor de la lista de `cast` (`$pop`)

```

db.db_pract4.updateOne(
  { "_id": ObjectId("573a1390f29313caabcd4eaf") },
  { $pop: { "cast": 1 } }
);

```



Dato Anterior	Dato Actualizado
<pre> _id: ObjectId('573a1390f29313caabcd4eaf') plot: "A woman, with the aid of her police officer sweetheart, endeavors to u..." genres: Array (2) runtime: 88 cast: Array (4) 0: "Jane Gail" 1: "Ethel Grandin" 2: "William H. Turner" 3: "Matt Moore" num_mflix_comments: 1 poster: "https://m.media-amazon.com/images/M/MV5BYzk6YWQzMGVtYTM5M000NjM2LWESYz..." title: "Traffic in Souls" lastupdated: "2015-09-15 02:07:14.247000000" languages: Array (1) released: 1913-11-24T00:00:00.000+00:00 </pre>	<pre> _id: ObjectId('573a1390f29313caabcd4eaf') plot: "A woman, with the aid of her police officer sweetheart, endeavors to u..." genres: Array (2) runtime: 88 cast: Array (3) 0: "Jane Gail" 1: "Ethel Grandin" 2: "William H. Turner" num_mflix_comments: 1 poster: "https://m.media-amazon.com/images/M/MV5BYzk6YWQzMGVtYTM5M000NjM2LWESYz..." title: "Traffic in Souls" lastupdated: "2015-09-15 02:07:14.247000000" languages: Array (1) released: 1913-11-24T00:00:00.000+00:00 directors: Array (1) rated: "TV-PG" </pre>

Actualizar el nombre de un actor específico dentro de `cast` (`$set` con índice de array)

```

db.db_pract4.updateOne(
  { "_id": ObjectId("573a1390f29313caabcd50e5") },
  { $set: { "cast.0": "A.C. Abadie (Updated)" } }
);

```



Dato Anterior	Dato Actualizado
<pre> _id: ObjectId('573a1390f29313caabcd50e5') plot: "The cartoonist, Winsor McCay, brings the Dinosaur back to life in th..." genres: Array (3) runtime: 12 cast: Array (3) 0: "Winsor McCay" 1: "George McManus" 2: "Roy L. McCardell" num_mflix_comments: 0 poster: "https://m.media-amazon.com/images/M/MV5BMTQxNzI4ODQ3NF5BM15BanBnXkFtZT..." title: "Gertie the Dinosaur" fullplot: "Winsor Z. McCay bets another cartoonist that he can animate a dinosaur..." languages: Array (1) released: 1914-09-15T00:00:00.000+00:00 directors: Array (1) writers: Array (1) </pre>	<pre> _id: ObjectId('573a1390f29313caabcd50e5') plot: "The cartoonist, Winsor McCay, brings the Dinosaur back to life in th..." genres: Array (3) runtime: 12 cast: Array (3) 0: "A.C. Abadie (Updated)" 1: "George McManus" 2: "Roy L. McCardell" num_mflix_comments: 0 poster: "https://m.media-amazon.com/images/M/MV5BMTQxNzI4ODQ3NF5BM15BanBnXkFtZT..." title: "Gertie the Dinosaur" fullplot: "Winsor Z. McCay bets another cartoonist that he can animate a dinosaur..." languages: Array (1) released: 1914-09-15T00:00:00.000+00:00 directors: Array (1) </pre>

Agregar múltiples elementos al array `genres` solo si no existen (`$addToSet`)

```

db.db_pract4.updateOne(
  { "_id": ObjectId("573a1390f29313caabcd516c") },
  { $addToSet: { "genres": { $each: ["Comedy", "Action"] } } }
);

```



Dato Anterior	Dato Actualizado
<pre><code>_id: ObjectId('573a1390f29313caabcd516c') plot: "Original advertising for the film describes it as a drama of primitive..." genres: Array (2) 0: "Drama" 1: "History" runtime: 65 cast: Array (4) num_mflix_comments: 1 poster: "https://m.media-amazon.com/images/M/MV5BMjE3MjAyNzMSNV5BMl5BanBxKftZT..." title: "In the Land of the Head Hunters" lastupdated: "2015-09-16 12:11:37.770000000" languages: Array (1) released: 1914-12-07T00:00:00.000+00:00 directors: Array (1)</code></pre>	<pre><code>_id: ObjectId('573a1390f29313caabcd516c') plot: "Original advertising for the film describes it as a drama of primitive..." genres: Array (4) 0: "Drama" 1: "History" 2: "Comedy" 3: "Action" runtime: 65 cast: Array (4) num_mflix_comments: 1 poster: "https://m.media-amazon.com/images/M/MV5BMjE3MjAyNzMSNV5BMl5BanBxKftZT..." title: "In the Land of the Head Hunters" lastupdated: "2015-09-16 12:11:37.770000000" languages: Array (1) released: 1914-12-07T00:00:00.000+00:00</code></pre>

Incrementar el número de premios en 1 para todas las películas con más de 5 nominaciones.

```
db.db_pract4.updateMany(
  { "awards.nominations": { $gt: 5 } },
  { $inc: { "awards.wins": 1 } }
);
```



Dato Anterior	Dato Actualizado
<pre><code>_id: ObjectId('573a1393f29313caabddad5') plot: "A reporter pretends to be Jewish in order to cover a story on anti-Semitism..." genres: Array (1) runtime: 118 rated: "APPROVED" cast: Array (4) num_mflix_comments: 0 poster: "https://m.media-amazon.com/images/M/MV5BMTZhYmMzMjYzE2OS00YzljLWI2YT..." title: "Gentleman's Agreement" fullplot: "Philip Green is a highly respected writer who is recruited by a nation..." languages: Array (1) released: 1948-02-01T00:00:00.000+00:00 directors: Array (1) writers: Array (2) awards: Object wins: 14 nominations: 6 text: "Won 3 Oscars. Another 10 wins & 6 nominations."</code></pre>	<pre><code>_id: ObjectId('573a1393f29313caabddad5') plot: "A reporter pretends to be Jewish in order to cover a story on anti-Semitism..." genres: Array (1) runtime: 118 rated: "APPROVED" cast: Array (4) num_mflix_comments: 0 poster: "https://m.media-amazon.com/images/M/MV5BMTZhYmMzMjYzE2OS00YzljLWI2YT..." title: "Gentleman's Agreement" fullplot: "Philip Green is a highly respected writer who is recruited by a nation..." languages: Array (1) released: 1948-02-01T00:00:00.000+00:00 directors: Array (1) writers: Array (2) awards: Object wins: 13 nominations: 6 text: "Won 3 Oscars. Another 10 wins & 6 nominations."</code></pre>

Añadir la fecha actual en lastUpdated para todas las películas de un director específico.

```
db.db_pract4.updateMany(
  { "directors": "Edwin S. Porter" },
  { $currentDate: { "lastUpdated": true } }
);
```



Dato Anterior	Dato Actualizado
<pre><code>cast: Array (4) poster: "https://m.media-amazon.com/images/M/MV5BMTU3NjE5NzYyYyNS00MDVmLWlwYj..." title: "The Great Train Robbery" fullplot: "Among the earliest existing films in American cinema - notable as the ..." languages: Array (1) released: 1903-12-01T00:00:00.000+00:00 directors: Array (1) rated: "TV-G" awards: Object lastupdated: "2015-08-13 00:27:59.177000000" year: 1903 imdb: Object countries: Array (1) type: "movie" tomatoes: Object num_mflix_comments: 0</code></pre>	<pre><code>poster: "https://m.media-amazon.com/images/M/MV5BMTU3NjE5NzYyYyNS00MDVmLWlwYj..." title: "The Great Train Robbery" fullplot: "Among the earliest existing films in American cinema - notable as the ..." languages: Array (1) released: 1903-12-01T00:00:00.000+00:00 directors: Array (1) rated: "TV-G" awards: Object lastupdated: "2015-08-13 00:27:59.177000000" year: 1903 imdb: Object countries: Array (1) type: "movie" tomatoes: Object num_mflix_comments: 0 lastUpdated: 2024-11-09T17:33:09.211+00:00</code></pre>

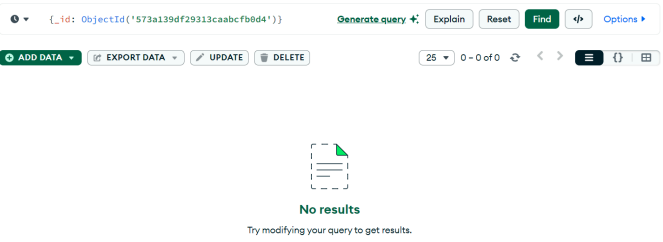
Operaciones DELETE

Para las operaciones DELETE haremos lo mismo que con las operaciones UPDATE.

Eliminar una película con duración menor a 50 minutos y rating de IMDB inferior a 5.0

```
db.db_pract4.deleteOne({
  "runtime": { $lt: 50 },
  "imdb.rating": { $lt: 5.0 }
})
```



Dato Anterior	Dato Actualizado
<pre>{ "_id": "573a139df29313caabcfb0d4", "plot": "Ally Hayden is a teenager who shares her father's interest in dinosaur...", "genres": Array (3) runtime: 45 cast: Array (4) poster: "https://m.media-amazon.com/images/M/MV5BMTHYNDI3NTA3OF5BM158anBkxkFCZT..." title: "T-Rex: Back to the Cretaceous" fullplot: "Ally Hayden is a teenager who shares her father's interest in dinosaur..." languages: Array (1) released: 1998-10-23T08:08:08.000+08:00 directors: Array (1) writers: Array (4) awards: Object lastupdated: "2015-09-05 01:00:06.930000000" year: 1998 imdb: Object countries: Array (1) }</pre>	

Eliminar una película lanzada antes de 1950 y del género "Western"

```
db.db_pract4.deleteOne({
  "year": { $lt: 1950 },
  "genres": "Western"
})
```



Dato Anterior	Dato Actualizado
<pre>{ "_id": "573a1390f29313caabcf42e8", "plot": "A group of bandits stage a brazen train hold-up, only to find a determ...", "genres": Array (2) runtime: 11 cast: Array (4) poster: "https://m.media-amazon.com/images/M/MV5BMTU3NjESNzYkYTYNS08MDVNLWZlWj..." title: "The Great Train Robbery" fullplot: "Among the earliest existing films in American cinema - notable as the ..." languages: Array (1) released: 1903-12-01T00:00:00.000+08:00 directors: Array (1) rated: "TV-G" awards: Object lastupdated: "2015-08-13 00:27:59.177000000" year: 1903 imdb: Object }</pre>	

Eliminar una película que tenga el actor "John Doe" en el reparto

```
db.db_pract4.deleteOne({
  "cast": "John Doe"
})
```



Dato Anterior	Dato Actualizado
<pre>{ "_id": "573a1398f29313caabce6714", "plot": "The life and career of the wildly controversial rock 'n' roll star, 3e...", "genres": Array (3) runtime: 188 rated: "PG-13" cast: Array (4) 0: "Dennis Quaid" 1: "Winona Ryder" 2: "John Doe" 3: "Stephen Tobolowsky" num_mflix_comments: 2 poster: "https://m.media-amazon.com/images/M/MV5BYWU1NDI2MTctOWMyby90MjAzZWFlMT..." title: "Great Balls of Fire!" fullplot: "The story of Jerry Lee Lewis, arguably the greatest and certainly one ..." languages: Array (1) }</pre>	

Eliminar una película del género "Documentary" que esté dentro del array de géneros

```
db.db_pract4.deleteOne({
  "genres": "Documentary"
})
```



Dato Anterior	Dato Actualizado
<pre>{ "_id": "573a1391f29313caabce7f11", "plot": "The struggles of a hardscrabble Iranian nomadic tribe as they journey ...", "genres": Array (1) 0: "Documentary" runtime: 71 rated: "PASSEp" cast: Array (4) poster: "https://m.media-amazon.com/images/M/MV5BMTE6MzA5OTk2M15BM15BanBnXkFtZT..." title: "Grass: A Nation's Battle for Life" fullplot: "The real-life struggles of the Bakhtiari tribe of Southern Persia (Ira..." countries: Array (1) released: 1925-03-28T00:00:00.000+00:00 directors: Array (2) }</pre>	

Eliminar una película con rating en IMDB inferior a 3.0 (en el objeto `imdb`)

```
db.db_pract4.deleteOne({
  "imdb.rating": { $lt: 3.0 }
})
```



Dato Anterior	Dato Actualizado
<pre>{ "_id": "573a1395f29313caabce188e", "plot": "With the aid of Merlin, Santa Claus must defeat the evil machinations ...", "genres": Array (2) runtime: 94 cast: Array (4) poster: "https://m.media-amazon.com/images/M/MV5BNmJmODcyMTctOWU5ZC80B0VhLTk1ZT..." title: "Santa Claus" fullplot: "Santa Claus, high above the North Pole in his cloud-borne castle equip..." languages: Array (1) released: 1960-10-01T00:00:00.000+00:00 directors: Array (1) writers: Array (4) awards: Object lastupdated: "2015-06-05 00:40:34.993000000" year: 1959 imdb: Object countries: Array (1) }</pre>	

Eliminar todas las películas con más de 10 años de antigüedad y menos de 1000 votos en IMDB

```
db.db_pract4.deleteMany({
  "year": { $lt: (new Date()).getFullYear() - 10) },
})
```



```
"imdb.votes": { $lt: 1000 }  
}))
```

Dato Anterior	Dato Actualizado
<pre>> db.db_pract4.find({ "year": { \$lt: (new Date().getFullYear() - 10) }, "imdb.votes": { \$lt: 1000 } }).count() < 7033</pre>	<pre>> db.db_pract4.find({ "year": { \$lt: (new Date().getFullYear() - 10) }, "imdb.votes": { \$lt: 1000 } }).count() < 0</pre>

Eliminar todas las películas de tipo "documentary" con rating en IMDB inferior a 6.0

```
db.db_pract4.deleteMany({  
  "genres": "Documentary",  
  "imdb.rating": { $lt: 6.0 }  
}))
```



Dato Anterior	Dato Actualizado
<pre>> db.db_pract4.find({ "genres": "Documentary", "imdb.rating": { \$lt: 6.0 } }).count() < 23</pre>	<pre>> db.db_pract4.find({ "genres": "Documentary", "imdb.rating": { \$lt: 6.0 } }).count() < 0</pre>