

Bases de Datos 2023

Recuperatorio Parcial II: No-SQL

Juan Cabral, Tadeo Cocucci,
Ramiro Demasi, Guido Ivetta

Detalle Importante

El alumno puede usar sus notas personales e internet, pero queda prohibido utilizar modelos generativos (Chat-GPT, LLAMA, etc) y comunicarse con otras personas oral o digitalmente.

Romper alguna de estas dos restricciones invalida el examen

Contexto

La base de datos **mflix** contiene datos sobre películas y salas de cine. Adicionalmente incluye datos de usuarios y comentarios sobre las películas. A continuación se describen todas las colecciones de la base de datos mflix.

Colección	Descripción
movies	Contiene información de las películas, incluyendo título, año de estreno, y director.
theaters	Contiene ubicaciones de salas de cine.
users	Contiene información del usuario.
comments	Contiene comentarios asociados a películas específicas
sessions	Campo de metadatos. Contiene JSON Web Token de los usuarios.

Preparación

- I. Descargar el archivo mflix.tar.gz que contiene el dump de la base datos mflix
- II. Cargar los datos desde el dump

- Si va a trabajar en la computadora del laboratorio debe cargar los datos con el siguiente comando, donde <ID> es tu identificador y <PASSWORD> es tu contraseña de mongo,

```
$ tar xzvf mflix.tar.gz
$ mongorestore --host 172.18.7.140 --drop --gzip --db recu_u<ID>
mflix/ -uumongo_g<ID> -p<PASSWORD>
```

Luego puede acceder al shell de mongo con

```
$ mongo --host 172.18.7.140 --authenticationDatabase recu_u<ID>
-uumongo_g<ID> -p<PASSWORD>
```

- Si va a trabajar desde su computadora personal

```
$ tar xzvf mflix.tar.gz
$ mongorestore --host <host> --drop --gzip --db mflix mflix/
```

Consignas

1. Identificar los tres géneros con el tiempo promedio de duración más alto de sus películas. Incluir el nombre del género y el tiempo promedio, y ordenar el resultado según el tiempo promedio en orden descendente. Si una película pertenece a varios géneros, contar esa película para cada uno de ellos.
2. Calcular el número promedio de caracteres en los títulos de las películas para cada género. Incluye el nombre del género y el número promedio de caracteres, y ordena el resultado en orden descendente según este número. Si una película pertenece a varios géneros, contar esa película para cada uno de ellos.
3. Identificar a los usuarios que tienen un gusto diverso en películas. Considerar la diversidad como la cantidad de películas de años diferentes a las que un usuario ha comentado. Es decir, si un usuario tiene comentarios en Titanic (1997), Forrest Gump (1994) y Pulp Fiction (1994), la diversidad es 2 porque hay exactamente dos años distintos entre todas las películas. Incluir el correo electrónico y la cantidad de años únicos que ha comentado. Ordena el resultado por la cantidad de años únicos descendentemente y por email alfabéticamente.
4. Especificar reglas de validación en la colección theaters utilizando JSON Schema.
 - a. Las reglas se deben aplicar sobre los campos: theaterId y location, (y todos sus campos anidados). Inferir los tipos y otras restricciones que considere

adecuados para especificar las reglas a partir de los documentos de la colección.

- b. Para testear las reglas de validación crear un caso de falla en la regla de validación y un caso de éxito (Indicar si es caso de falla o éxito)

Puntos a tener en cuenta

- Algunas consultas pueden resolverse con `find` y `aggregate`. En estos casos se puede resolver eligiendo alguno de estos métodos.
- Resolver las consultas sin vistas salvo que se lo pida explícitamente.
- Mostrar únicamente los campos pedidos en la consigna.
- Se piden que los campos que se devuelven sean valores escalares a menos que se pida los valores de los campos devueltos podrán ser documentos anidados, arreglos de escalares o arreglos de documentos.
- Buscar hacer la consulta de la forma más sencilla y eficiente posible.
- Se evaluará el correcto formato de las soluciones:
 - El código entregado debe ser legible.
 - Utilizar indentación de espacios de manera uniforme.

Entrega

- Se entregará un archivo comprimido ``soluciones.js.gz`` o ``soluciones.zip`` (con ``soluciones.js`` adentro) con las soluciones de los 5 ejercicios. Separar las soluciones mediante comentarios.
- La entrega se hará mediante el Aula Virtual en el correspondiente apartado.
 - Tendrán hasta las 17:00 para que se considere una entrega completa. La recomendación es empezar a subir el archivo a las 16:55 para evitar cualquier eventualidad.