

Bases de Datos No Relacionales

Instituto de Computación, Facultad de Ingeniería

Tarea 1: Bases de datos documentales

Marzo 2025

Objetivos

- analizar y discutir sobre el diseño de bases de datos documentales, identificar ventajas y desventajas de un ejemplo concreto.
- experimentar con un motor documental y su lenguaje de consulta. Deberá construir consultas sobre un esquema documental para resolver los requerimientos planteados.

Introducción

Las bases de datos documentales tienen una presencia muy fuerte en el mercado, ya sea en proyectos nuevos o migrando aplicaciones/datos desde motores relacionales existentes. En este trabajo se parte desde un conjunto de datos existente, con un diseño propio, y de allí se desencadenan una serie de pasos:

- seleccionar el ambiente de ejecución
- carga de los datos
- escribir consultas
- proponer un diseño alternativo para mejorar ciertas operaciones. Identificar decisiones de diseño basadas en patrones y mencionar sus implicancias.
- implementar operaciones sobre diseño mejorado

Datos a utilizar

Los datos a utilizar son datos de ejemplo oficiales de mongodb¹. Estos datos están disponibles para su descarga y uso², y se utilizará la base de datos “*Sample Mflix Dataset*” de la cual puede ver detalles de su definición aquí.

Ambiente de ejecución

El conjunto de datos a utilizar está disponible para su uso, ya sea utilizando mongodb como servicio en MongoDB Atlas³, o si lo prefiere puede realizar una instalación local de mongodb.

Para escribir las consultas y/o realizar programación se debe usar un Jupyter Notebook o similar que permita luego entregar el archivo con su solución en formato ipynb (que funcione por ejemplo en la plataforma Google Colaboratory⁴). Dentro del Notebook se sugiere utilizar python y pueden utilizar las bibliotecas que necesiten. Para mostrar los resultados dentro del notebook es sugerido el uso de DataFrames de Pandas.

Lo anterior es un conjunto mínimo de herramientas que se deben usar para la resolución del trabajo, pero puede utilizar cualquier otra herramienta que considere útil, siempre que respete el formato de los archivos que debe entregar.

¹<https://www.mongodb.com/developer/products/atlas/atlas-sample-datasets/>

²<https://www.mongodb.com/developer/products/atlas/atlas-sample-datasets/#sql-atlas-sample-data-local-installation>

³<https://www.mongodb.com/products/platform/atlas-database>

⁴<https://colab.research.google.com/>

Trabajo a realizar

La aplicación web <https://www.atlassearchmovies.com/> es un frontend que por detrás resuelve consultas contra un motor mongodb, el cual tiene cargado el conjunto de datos “*Sample Mflix Dataset*” que fue mencionado antes.

Responda y resuelva las siguientes consultas, describiendo y justificando adecuadamente cada solución propuesta:

1. ¿Cuál es la consulta que ejecuta esta aplicación web y sobre qué colección o colecciones se realiza la búsqueda?
2. Suponga que se le pide mejorar la aplicación con los siguientes requerimientos:
 1. Dado un string que debe ser parte del nombre de la película, muestre como resultado el nombre, descripción completa, rating y año de la película. Los resultados deben estar ordenados por rating descendente y luego por año en forma descendente.
 2. Dado el nombre de una película, mostrar todos los comentarios asociados a esa película, ordenados por fecha en forma descendente.
 3. Se decide quitar el campo `email` de la colección `comments`, ¿hay algún tipo de pérdida de información? Escriba la consulta para eliminar este campo.
 4. Se pretende poder sacar estadísticas sobre los comentarios basadas en el país de origen del usuario que la comenta. Agregue un campo `country` a los objetos de la colección `users`.
 5. Dado un país, mostrar todas las películas que fueron vistas (comentadas) por usuarios de ese país.
 6. Dado el género de una película y una fecha, dar el reporte de todas las películas que pertenecen a ese género y fueron posteriores a dicha fecha. El reporte debe incluir el nombre de la película, la descripción completa, los directores, escritores, premios y nominaciones. Los resultados deben ser ordenados por premios y nominaciones en forma descendente.

Evalúe las siguientes situaciones desde el punto de vista del diseño y la aplicación de patrones, justificando de forma clara su conclusión.

1. ¿Mantendría los campos de género y países de las películas modelado como está actualmente o lo cambiaría?
2. ¿Cuál es el objetivo del campo `num_mflix_comments` de la colección `movies`? ¿puede asociarlo a algún patrón de diseño? ¿cuál sería un requerimiento que se resuelve eficientemente gracias a éste campo, y como sería esa consulta si no existiera?

Formato y plazo de entrega:

Las entregas deben realizarse en grupos de dos estudiantes. Se deberá entregar a través de EVA un reporte en formato PDF, y un Jupyter Notebook (que funcione por ejemplo en la plataforma Google Colaboratory⁵) en el que se carguen los datos y se ejecuten las consultas. El plazo de entrega vence el viernes 11 de abril a las 23:59.

⁵<https://colab.research.google.com/>