

4. Consultas originales a mantener en el nuevo diseño

La aplicación utiliza algunas de las vistas definidas en Sakila³, de manera que debe asegurarse de que sea posible reproducir las que se enumeran a continuación en el diseño de la nueva base documental:

- (a) customer_list
- (b) film_list
- (c) sales_by_store
- (d) sales_by_film_category

5. Estimación de costos

En este apartado se enumerarán algunos costos referentes a las operaciones involucradas en las consultas sobre el motor documental, para poder cuantificar el costo de ejecutarlas y en base a eso tomar decisiones para mejorar el desempeño y la carga general del sistema.

1. El costo por espacio de almacenamiento es cero, es decir, no hay penalización por el uso extra de espacio en disco.
2. Asuma que siempre habrá memoria y capacidad de procesamiento suficiente para ejecutar las consultas.
3. La distribución de las consultas en un día tipo es de esta forma:
 - (1) customer_list=100 veces/día
 - (2) film_list=80 veces/día
 - (3) sales_by_store=400 veces/día
 - (4) sales_by_film_category=500 veces/día
4. El costo por acceder a un documento por su `_id` o clave primaria/índice único es 1 unidad.
5. El costo por acceder a un documento en una colección sin ningún índice se puede asumir que es en promedio la mitad del tamaño de la colección.
6. El costo por acceder a cada atributo simple dentro de un documento es 0.01 (por ejemplo se accede al momento de proyectar/imprimir)
7. El costo por recorrer "claves foráneas" es decir seguir referencias que llevan de un documento a otro tiene un costo que está relacionado a los puntos anteriores

6. Trabajo a realizar

Esta tarea intenta replicar conceptualmente lo que se debería hacer para un trabajo bien planificado de migración desde un motor relacional a uno documental. El estado del arte en el área está trabajando en el diseño de estrategias automáticas de conversión, teniendo como entrada el diseño relacional, la carga de trabajo (*workload*) representativa y una función de costos asociados, y dando como resultado diseños documentales equivalentes a los relacionales, que optimizan una función de costo que depende de los datos de entrada y de las cargas de trabajo, entre otros.

En concreto, este trabajo tomará la base relacional Sakila y dos modelos/representaciones documentales resultantes de transformar Sakila (aunque no de forma automática sino manual). Ambas bases documentales serán evaluadas mediante los costos de las consultas pedidas y el

³ La definición de las vistas está incluida en el archivo sakila-schema.sql dentro de <https://downloads.mysql.com/docs/sakila-db.tar.gz>

workload detallado, para finalizar concluyendo sobre la conveniencia de utilizar uno de los dos modelos/representación. El plan de trabajo a seguir es:

1. La primer base documental DD1 corresponde a un mapeo directo y está disponible en MongoDB Atlas para utilizarla. La base migrada tendrá una colección para cada entidad, y todas las relaciones que no son 1:1 se representan usando referencias entre documentos (no hay documentos embebidos). Para acceder a la base desde MongoDB Compass utilizar `mongodb+srv://alumno_2023:bdnr.2023@bdnr.afutv.mongodb.net/sakilaDirectMap` y si accede desde python `mongodb+srv://alumno_2023:bdnr.2023@bdnr.afutv.mongodb.net/?retryWrites=true&w=majority` y el nombre de la base es `sakilaDirectMap`.
2. Generar sobre DD1 las consultas MongoDB que se traducen de las consultas sobre el modelo relacional presentadas en la sección 4.
3. Realizar el cálculo total de una carga típica de trabajo para DD1 indicado en la sección 5.
4. Generar un nuevo diseño documental para una nueva base de datos, que llamaremos DD2, en el que puede aplicar todos los patrones de diseño⁴ y decisiones que usted considere necesarios. Puede generar todos los índices (o similar) que considere necesario. El objetivo es mejorar los costos de las consultas de la sección 4.
5. Repetir los pasos 2 y 3 ahora sobre DD2.

¿Qué se debe entregar?

Elaborar un reporte que contenga :

- i. Descripción/diagrama del diseño DD2. Debe justificar las decisiones tomadas y los patrones utilizados. Agregue información adicional si es que la considera necesaria.
- ii. Detalle de los cálculos asociados a la evaluación de carga de trabajo para DD1 y DD2.
- iii. Compare y concluya sobre la conveniencia de utilizar DD1 o DD2.

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) que cargue las bases y ejecute en la nube las consultas. Se publicará una notebook de base que uds deberán modificar para entregar. El plazo de entrega vence miércoles 10 de mayo a las 23:59.

4 Catálogo de patrones disponible en <https://www.mongodb.com/blog/post/building-with-patterns-a-summary>

5 <https://colab.research.google.com/>