

Base de Datos (75.15 / 75.28 / 95.05)

Segundo Parcial Promocional

TEMA 2023241	Proc.			Fecha: 6 de diciembre de 2023 Padrón: _____ Apellido: _____ Nombre: _____ Cantidad de hojas: _____ <input type="checkbox"/> Aprobado <input type="checkbox"/> Insuficiente
	NoSQL			
	CRT			
	CyR			
Corrigió: Nota:				

Criterio de aprobación: El examen está compuesto por 7 ítems, cada uno de los cuales se corrige como B/B-/Reg/Reg-/M. El examen se aprueba con nota mayor o igual a 4(cuatro) y la condición de aprobación es desarrollar un ítem bien (B/B-) en al menos 3 de los 4 grupos de ejercicios (procesamiento de consultas, NoSQL, CRT, concurrencia/recuperación). Adicionalmente, no deberá haber más de dos ítems mal o no desarrollados.

1. (*Procesamiento de consultas*) Como recordará de exámenes previos, los siguientes esquemas de relación almacenan información sobre las multas de tránsito de la Ciudad de Buenos Aires:

- PERSONA(DNI, nombre, apellido, direccion, ciudad, celular)
// (21.454.201, 'Ramón', 'Mercury', 'Av. Rivadavia 500', 'Rosario', 5240-6544)
- VEHICULO(matricula, marca, modelo, fecha_VTV)
// ('AR 251 GH', 'RENAULT', 'DUSTER', '2022-12-05')
- MULTA(nro_multa, DNI_infractor, matricula, tipo, hora, fecha, lugar, importe)
// (1809, 21.454.201, 'AR 251 GH', 2, 12:23:21, '2022-01-01', 'ruta 205 KM 34.5', 350.000)
- PROPIETARIO(DNI,matricula)
// (21.454.201, 'AR 251 GH')

Tenga en cuenta que el infractor que cometió una multa con un vehículo no es necesariamente la misma persona registrada como propietaria de ese vehículo.

Esta base de datos registra distintos tipos de multa, que van desde el tipo 1 (menos severa) al tipo 4 (muy severa). Entre las multas de tipo 4 se incluyen infracciones como “ingresar al Paseo del Bajo en contramano” o “cruzar semáforo en rojo sonando una sirena falsa”, que determinan el quite de la licencia de por vida al propietario del vehículo.

Por lo tanto, cuando una persona se presenta para renovar su licencia de conducir, además de chequear que la persona no haya cometido ella misma infracciones, se chequea que jamás se haya cometido una infracción de tipo 4 con ninguno de los vehículos que son de su propiedad. De lo contrario, la licencia es denegada. La siguiente consulta, en particular, es realizada

cuando Marlon Siniestra (DNI 18.324.715) se presenta a renovar su licencia en las oficinas de la Dirección de Tránsito:

```
SELECT *
FROM Propietario p INNER JOIN Multa m USING(matricula)
WHERE p.DNI = 18324715 AND m.tipo = 4;
```

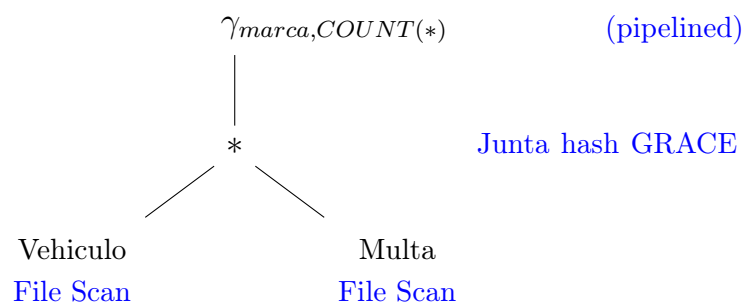
Se pide:

- Sugiera dos índices que puedan utilizarse para ejecutar más eficientemente esta consulta. Luego dibuje un plan de ejecución que haga uso de estos dos índices para la consulta. Específicamente, dibuje el árbol del plan de consulta y anote sobre el mismo los métodos de acceso o algoritmos que se utilizarán en cada paso.
- Estime el costo del plan de ejecución que armó en el punto anterior, en términos de cantidad de accesos a bloques de disco.

Para el ejercicio considere que los índices son de tipo árbol y tienen altura 4. Además, considere para sus cálculos la siguiente información de catálogo:

PROPIETARIO	MULTA	VEHICULO
$n(\text{Propietario}) = 600.000$	$n(\text{Multa}) = 300.000$	$n(\text{Vehiculo}) = 600.000$
$B(\text{Propietario}) = 100.000$	$B(\text{Multa}) = 30.000$	$B(\text{Vehículo}) = 200.000$
$V(\text{DNI, Propietario}) = 300.000$	$V(\text{matricula, Multa}) = 100.000$	
	$V(\text{tipo, Multa}) = 4$	

- (Procesamiento de consultas) Para las mismas tablas del ejercicio anterior y la misma información de catálogo, se quiere calcular cuántas multas se han cometido con cada marca de vehículo. Asumiendo que no se cuenta con ningún tipo de índice y que se dispone de $M=20.000$ bloques de memoria, se arma el siguiente plan de ejecución en el que la junta se realiza por el método de hash GRACE:



Se pide:

- Indique qué cantidad k de particiones intentaría generar para que la junta hash GRACE sea factible de ser realizada, justificando su respuesta. Estime qué tamaño promedio –en términos de cantidad de bloques– tendrían las particiones en ese caso.
- Indique cuál es el/los atributo/s a los que habrá que aplicar la función de hash en cada tabla, a efecto de construir las particiones.

- c) Estime cuál sería el costo de realizar la junta planteada, en términos de cantidad de accesos a bloques de disco. Considere que el agrupamiento posterior se realiza en pipeline utilizando un diccionario en memoria, y por lo tanto no incurre en costos de acceso a disco.

3. (NoSQL)

- a) (*MongoDB*) El *Club de Cinéfilos* quiere armar un ranking de actores que permita a sus miembros saber quiénes son los 100 actores a los que más vale la pena seguir. Para ello, cuentan con una base de datos de películas en MongoDB, que indica el puntaje en IMDB de cada película junto con el listado de actores de la película, tal como ejemplifica el siguiente documento:

```

1 {
2   "_id": 10910355903998401931,
3   "nombre_pelicula": "Interstellar",
4   "genero_principal": "Ciencia Ficción",
5   "puntaje_IMDB": 8.7,
6   "actores": ['Matthew McConaughey', 'Jessica Chastain', 'Anne Hathaway',
7               'Mackenzie Foy', 'Timothée Chalamet', 'Matt Damon', 'Michael Caine']
8 }
```

Mientras discutían qué métrica utilizar para rankear a los actores, algunos sugerían usar el puntaje promedio en IMDB de sus películas como un valor representativo. Otros en cambio consideraban que se debía tomar el mejor puntaje de entre todas las películas en las que el actor participó, ya que si un actor participó en una película muy buena, entonces valía la pena seguirlo aún cuando su promedio fuera malo.

Finalmente, se decidió por una estrategia híbrida en que se ordenará a los actores por su puntaje promedio en todas sus películas, pero se excluirá luego a aquellos actores cuya mejor película tenga un puntaje mayor o igual a 8.0.

- 1) Escriba una consulta en MongoDB que devuelva el listado de los 100 mejores actores ordenados por este criterio, indicando para cada actor su nombre y apellido, su puntaje promedio, y la máxima puntuación obtenida por sus películas.
 - 2) Explique si la consulta anterior puede ser ejecutada con la colección shardeada por el atributo `_id`. En caso afirmativo, explique brevemente cómo podría realizarse el cálculo anterior en forma distribuida entre los shards y los servidores de agregación. En caso negativo, explique cuál debería ser el/los atributo/s de sharding, y cómo se realizaría el cálculo en forma distribuida en ese caso.
- b) (*Neo4j*) La *Facultad de Ingeniería* quiere implementar un sistema de búsqueda de empleo en el que los estudiantes puedan cargar las asignaturas que cursaron y las empresas puedan cargar los conocimientos que requieren en cada búsqueda. El objetivo del sistema es recomendar a los estudiantes ofertas de empleo para las cuales estén potencialmente calificados, a partir de la información que se almacenará en una base de datos en Neo4j con los siguientes nodos y aristas:

```

1 (est:Estudiante {nombre: 'Guillermina', apellido: 'Fabbri',
2                  username: 'laguille99', padrón: 109329})
3 (asign:Asignatura {nombre: 'Base de Datos', código: '7515'})
4 (conoc:Conocimiento {nombre: 'BASES DE DATOS NOSQL'})
5 (conoc2:Conocimiento {nombre: 'POSTGRESQL'})
6 (busq:Busqueda {nombre_empresa: 'El Sauce S.R.L.',
```

```

7           fecha: '2023-10-27', contacto: 'rrhh@elsauce.com.ar' })
8   ...
9   ( est )-[ :APROBO ]->( asign )
10  ( asign )-[ :OFRECE ]->( conoc )
11  ( busq )-[ :DESEA ]->( conoc )

```

Así, cuando un estudiante se encuentre buscando empleo, el sistema le rankeará las búsquedas disponibles mostrando primero aquellas para las cuales el estudiante se encuentra mejor preparado, en el sentido de que cumpla con poseer la mayor cantidad de conocimientos que la misma exige.

Escriba una consulta en Neo4j que, dado un estudiante específico de padrón p liste las búsquedas ordenadas con dicho criterio, indicando los datos de la búsqueda y la cantidad de conocimientos deseados por la misma que el alumno posee.

4. (CRT) Considere las mismas relaciones del ejercicio 1, y escriba una expresión en *Cálculo Relacional de Tuplas* que encuentre el nombre, apellido y celular de aquellas personas propietarias de algún vehículo cuya última VTV (Verificación Técnica Vehicular) haya sido hace más de un año.

5. (Concurrencia y Recuperación)

a) (Concurrencia) Dado el siguiente solapamiento de transacciones:

$b_{T_1}; b_{T_2}; b_{T_3}; W_{T_3}(Y); W_{T_3}(X); W_{T_1}(Y); c_{T_3}; R_{T_2}(Y); R_{T_2}(X); c_{T_2}; W_{T_1}(X); c_{T_1}$

- 1) Dibuje el grafo de precedencias del solapamiento.
- 2) Indique si el solapamiento es serializable. Justifique su respuesta.
- 3) Indique si el solapamiento es recuperable. Justifique su respuesta.

b) (Recuperación) Un SGBD implementa el algoritmo de recuperación UNDO con checkpoint activo. Luego de una falla, el sistema encuentra el siguiente archivo de log:

01 (BEGIN, T1);	08 (WRITE T3, A, 8);
02 (BEGIN, T2);	09 (WRITE T2, B, 15);
03 (WRITE T1, A, 13);	10 (COMMIT, T2);
04 (WRITE T2, C, 7);	11 (END CKPT);
05 (COMMIT, T1);	12 (BEGIN, T4);
06 (BEGIN CKPT, T2);	13 (WRITE T3, B, 8);
07 (BEGIN, T3);	14 (WRITE T4, C, 9);

Explique cómo se llevará a cabo el procedimiento de recuperación, indicando hasta qué punto del archivo de log se deberá retroceder, y qué cambios deberán ser realizados en disco y en el archivo de log.

Padrón: _____

Apellido y nombre: _____