

Base de Datos (75.15 / 75.28 / 95.05)

Evaluación Integradora - 27 de julio de 2022

TEMA 20221C3						Padrón: _____ Apellido: _____ Nombre: _____ Cantidad de hojas: _____ <input type="checkbox"/> Aprobado <input type="checkbox"/> Insuficiente
CyT		Proc.		SQL		
NoSQL		NoSQL		NoSQL		
Nota:						

Criterio de aprobación: El examen está compuesto por 6 ítems, cada uno de los cuales se corrige como B/B-/Reg/Reg-/M. Se aprueba con nota mayor o igual a 4(cuatro), equivalente a desarrollar el 60% del examen correctamente.

1. (*Concurrencia y transacciones*) A partir del siguiente conjunto de 3 transacciones:

$$T_1 : b_{T_1}; R_{T_1}(X); R_{T_1}(Y); W_{T_1}(Z); c_{T_1};$$

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

$$T_3 : b_{T_3}; R_{T_3}(Z); W_{T_3}(Y); c_{T_3};$$

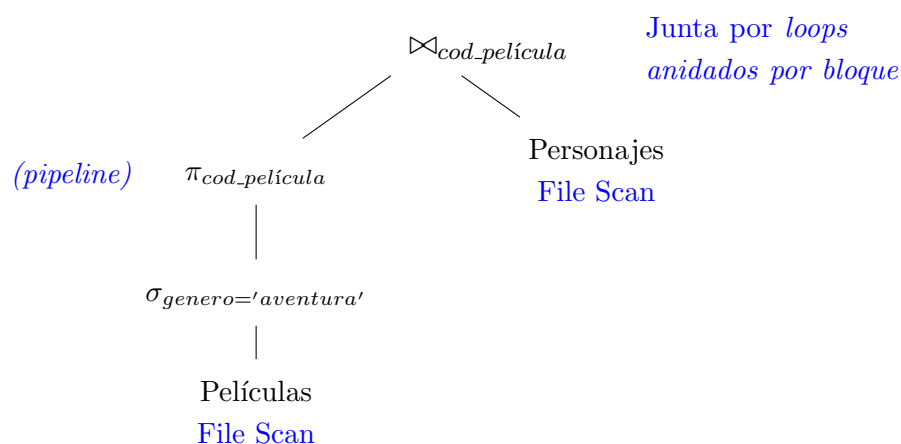
Se pide:

- Reescriba cada una de las 3 transacciones colocando locks de acuerdo con el *Protocolo de Lock de 2 Fases (2PL)*.
- Proponga un solapamiento de las transacciones que respete los locks que colocó en el punto (a), pero que no sea recuperable.

2. (*Procesamiento de Consultas*) Marcela está organizando una fiesta de disfraces en su quinta, con la consigna de que sólo serán admitidos personajes de películas de aventura. Se espera encontrar a Lara Croft, Indiana Jones, Buzz Lightyear o Alicia deambulando por los pasillos. Para que no queden dudas, Marcela quiere armar un listado de los personajes admisibles a la fiesta en base a la información que posee en las siguientes dos tablas:

- Películas(cod_pelicula, nombre, año, género)
- Personajes(cod_pelicula, nombre_personaje, actor)

Dado que Marcela no dispone de índices en la base de datos para ninguna de las dos tablas, su plan de ejecución para encontrar el listado de personajes de películas de aventura es el siguiente:



Se pide:

- Estime el costo del plan propuesto en términos de cantidad de bloques, utilizando la información de catálogo indicada debajo y suponiendo que se dispone de $M=100$ bloques de memoria. Nota: Considere que el atributo `cod_pelicula` representa en promedio el 10% de la longitud de una tupla de `Películas`.
- Estime la cardinalidad del resultado obtenido, en términos de cantidad de tuplas.
- ¿Cree que podría mejorarse el costo de este plan de ejecución consiguiendo más memoria? En caso afirmativo, indique hasta qué cantidad de memoria considera que convendría adquirir.
- ¿Cree que podría mejorarse el costo de este plan de ejecución construyendo uno o más índices sobre las tablas? En caso afirmativo, indique qué índice/s crearía.

PELICULAS	PERSONAJES
$n(\text{Películas}) = 100.000$	$n(\text{Personajes}) = 500.000$
$B(\text{Películas}) = 10.000$	$B(\text{Personajes}) = 50.000$
$V(\text{género}, \text{Películas}) = 10$	$V(\text{cod_pelicula}, \text{Personajes}) = 100.000$
Histograma para género :	
<i>comedia</i> : 30.000	
<i>romántica</i> : 15.000	
<i>acción</i> : 20.000	
<i>otros</i> : 35.000	

3. (SQL) El *Ente Regulador de la Electricidad* está analizando los cortes de suministro de energía eléctrica padecidos durante el año 2021 por clientes de distintas empresas distribuidoras de electricidad en el país. La ley estipula que por cada corte la empresa deberá abonar una sanción de \$100 por cada día entero sin servicio. Por ejemplo, para un corte de 23 horas la empresa no deberá abonar ninguna sanción, pero para un corte de 28 horas la empresa deberá abonar \$100.

En la siguiente tabla se registran todos los reclamos realizados durante el 2021, y los distintos estados por los que los mismos pasaron (obviaremos los reclamos que habían iniciado antes del 2021 y que no habían sido aún resueltos al comenzar el año).

- `Reclamos2021(cod_reclamo, num_cliente, nombre_empresa_distribuidora)`
- `EstadoReclamos2021(cod_reclamo, fecha, tipo_reclamo, estado)`

Los reclamos por falta de suministro se identifican con el `tipo_reclamo='FALTA_SUMINISTRO'`, y pueden pasar por los siguientes estados: `{'ABIERTO', 'DESESTIMADO', 'RESUELTO'}`. Se considera que la duración de un corte es la diferencia de tiempo entre el instante que se abrió el reclamo y el instante en que se cerró. El atributo `fecha` es de tipo `TIMESTAMP` en la tabla `EstadoReclamos2021`.

Nos interesa calcular la sanción total a pagar por cada empresa distribuidora. Para esto, se pide:

- (a) En primer lugar verificaremos la consistencia de los datos. Escriba una consulta SQL que verifique que ningún reclamo pasó por un mismo estado más de una vez. ¿Hay alguna forma de forzar esta restricción desde la definición de datos?
- (b) Continuando con las verificaciones de consistencia, escriba una consulta SQL que verifique que no existen reclamos en estado `DESESTIMADO` o `RESUELTO` que no habían sido previamente abiertos.
- (c) Ahora sí, escriba una consulta SQL que calcule para cada empresa distribuidora el monto de sanción total a pagar.

Nota: Puede calcular la cantidad entera de días contenida en un dato de tipo `INTERVAL` utilizando `EXTRACT(DAY FROM variable)`. Por ejemplo, la consulta `SELECT EXTRACT(DAY FROM INTERVAL '40 days 1 minute')` devuelve 40.

4. (NoSQL) En una familia de columnas en Cassandra con un determinado conjunto de atributos (PK, (CK)) como clave de partición y de clustering, existen ciertas limitaciones respecto a las consultas CQL que pueden ejecutarse bajo un despliegue estándar de la base. En particular, es obligatorio que la clave de partición sea comparada por igual contra un valor en toda consulta. Explique el por qué de esta decisión de diseño del gestor, y por qué en cambio no es obligatorio que la clave de clustering sea comparada por igual.

5. (*NoSQL*) Un sistema de gestión de bases de datos (SGBD) distribuido posee sus datos fragmentados en 10 partes, cada una de las cuales se encuentra replicada en 5 nodos de almacenamiento, haciendo un total de 50 nodos de almacenamiento. Además de ello, posee 5 nodos ruteadores para orquestar el procesamiento de las consultas y el armado de las respuestas. En este contexto, indique si las siguientes afirmaciones son verdaderas (V) o falsas (F), justificando su respuesta.

- Si este SGBD distribuido ofrece consistencia secuencial, entonces cualquier solapamiento de transacciones que se ejecute sobre él será serializable.
- Para ofrecer consistencia eventual sin mantener un orden global de todos los eventos, el SGBD distribuido deberá contar con algún mecanismo de anti-entropía que permita sincronizar a los nodos.
- Si este SGBD distribuido debe ofrecer consistencia secuencial, la latencia percibida por el usuario será mayor que si debe ofrecer consistencia eventual, aún cuando ningún nodo de la base se encuentre caído ni haya particiones en la red.
- Si este SGBD distribuido debe ofrecer consistencia secuencial, la disponibilidad de la base será menor, aún cuando ningún nodo de la base se encuentre caído ni haya particiones en la red.

6. (*NoSQL*) La facultad guarda las notas finales de los alumnos en una colección en *MongoDB* con la siguiente estructura:

```

1 {   padron: 70000,
2     nombre: "Deshawn Watson",
3     notas: [
4         {depto: 71, numero: 14, nombre: "Modelos y Optimización I", nota: 7},
5         {depto: 75, numero: 1, nombre: "Computación", nota: 4},
6         {depto: 75, numero: 15, nombre: "Base de datos", nota: 10}
7     ]
8 },
9 {   padron: 81000,
10    nombre: "Patrick Mahomes",
11    notas: [
12        {depto: 71, numero: 14, nombre: "Modelos y Optimización I", nota: 9},
13        {depto: 75, numero: 1, nombre: "Computación", nota: 8},
14    ]
15 }
```

Debido a un pedido de la Secretaría Académica, se desea evaluar el desempeño de cada departamento de la facultad, por lo que se solicitó un reporte que tenga el promedio de notas finales para cada departamento.

Escriba una consulta que resuelva dicho pedido, con el siguiente formato de respuesta:

```

1 {   "departamento" : 75,
2     "promedio" : 7.333333333333333
3 },
4 {   "departamento" : 71,
5     "promedio" : 8
6 }
```