

Base de Datos (75.15 / 75.28 / 95.05)

Evaluación Integradora - 21 de diciembre de 2022

TEMA 20222C2						Padrón: _____
CRT		Proc.		SQL		Apellido: _____
NoSQL		CyT		Rec.		Nombre: _____
Nota:						Cantidad de hojas: _____
						<input type="checkbox"/> Aprobado <input type="checkbox"/> Insuficiente

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. (*Cálculo Relacional de Tuplas*) Las siguientes tablas indican la cantidad de ejemplares de distintas especies presentes en cada una de las reservas naturales del país:

- Reservas(nombre_reserva, superficie, año_creación)
- Ejemplares(nombre_reserva, nombre_especie, cantidad_ejemplares)

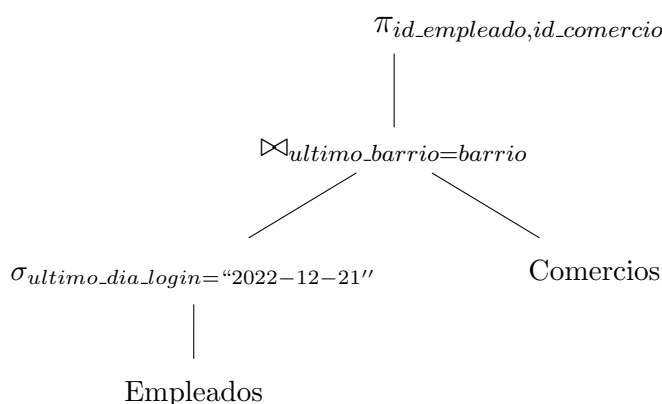
Escriba una consulta en Cálculo Relacional de Tuplas que encuentre el nombre y la cantidad de ejemplares de la especie “tatú carreta” presentes en aquellas reservas que tienen ejemplares de “tatú carreta” pero que no tienen ningún ejemplar de “zorro gris”.

Nota: Asuma que el atributo `cantidad_ejemplares` sólo puede tomar valores mayores a cero.

2. (*Procesamiento de consultas*) La empresa de envío de pedidos *Lentti* analiza constantemente en qué barrios se encuentran sus empleados, para decidir a qué comercios enviarlos. En la tabla **Empleados** registra, entre otros datos de sus empleados, el último barrio en el que estuvieron y el último día en que activaron la aplicación. También mantiene una tabla con la información de los comercios que venden productos a través de su aplicación:

- **Empleados**(id_empleado, DNI, nombre, último_barrio, último_día_login, calif)
- **Comercios**(id_comercio, nombre, dirección, barrio, calificación, tipo_contrato)

El gerente de despachos quiere ver en la cercanía de qué comercios se encuentra cada empleado activo. Para ello, considera activo a un empleado que se logueó en el día y considera que un empleado está cerca de un comercio si está en el mismo barrio. De esta forma, ejecuta la siguiente consulta sobre la base de datos:



La base de datos dispone de un índice de tipo árbol por **id_empleado** en **Empleados** y de un índice por **id_comercio** en **Comercios**. Además, puede considerar para sus cálculos la siguiente información de catálogo:

EMPLEADOS	COMERCIOS
$n(\text{Empleados}) = 150.000$	$n(\text{Comercios}) = 30.000$
$B(\text{Empleados}) = 30.000$	$B(\text{Comercios}) = 3.000$
$H(I(\text{id_empleado}, \text{Empleados})) = 3$	$H(I(\text{id_comercio}, \text{Comercios})) = 3$

También se sabe que cada día están activos alrededor de un 10% de los empleados.

Se pide:

- Teniendo la anterior información, y asumiendo que no hay problemas de memoria, estime el costo de la consulta eligiendo los métodos óptimos.
 - Indique qué ocurriría si hubiera únicamente 502 bloques de memoria disponibles.
3. (*SQL*) Para las mismas tablas del *Ejercicio 2*, escriba una consulta en SQL que encuentre los nombres de los barrios en los cuales no hay ningún comercio con calificación mayor o igual a 8.

4. (NoSQL) La nueva red social *Rhinos* publicó una nueva API que permite consultar su base de datos en MongoDB, en la que cada posteo tiene la siguiente estructura en formato JSON:

```
1 {
2   post_id: "1928500576325_104344819840",
3   user_id: 2401740,
4   user_name: "Lionel Messi",
5   picture: "3f2103ab01a20299ee0.jpg",
6   texto: "Buen día!",
7   tags: ["futbol", "copa", "qatar2022"],
8   likes: ["jorge", "celia"]
9 }
```

Lionel quiere revisar a cuántos de sus posts etiquetados como “futbol” les dieron likes su padre “jorge” y su madre “celia”. Quiere la respuesta en el siguiente formato:

```
1 { nombre : "jorge",
2   cantidad : 18 },
3 { nombre : "celia",
4   cantidad : 20 }
```

Escriba la consulta en MongoDB que permita a Lionel obtener dicho resultado.

5. (Concurrencia y transacciones) Para cada una de las siguientes afirmaciones, indique si la misma es verdadera ó falsa, justificando su respuesta:
- (a) Si se utiliza el protocolo de lock de dos fases (2PL) para el control de concurrencia, no puede ocurrir la anomalía del fantasma.
 - (b) Si se utiliza el protocolo de lock de dos fases (2PL) para el control de concurrencia, no es posible que ocurra la anomalía de la lectura sucia.
 - (c) Bajo el mecanismo de control de concurrencia *Snapshot Isolation* no es posible que ocurra la anomalía del fantasma.
 - (d) Bajo el mecanismo de control de concurrencia *Snapshot Isolation* nunca resulta necesario abortar una transacción para asegurar la serializabilidad.
6. (Recuperación) Un SGBD implementa el algoritmo de recuperación UNDO/REDO con checkpoint activo. Indique si las siguientes afirmaciones sobre el funcionamiento del algoritmo son verdaderas ó falsas, justificando su respuesta:
- (a) Los ítems de datos modificados por una transacción T_i deben ser *flushados* a disco antes de escribir $(COMMIT, T_i)$ en el archivo de *log*.
 - (b) Cuando se modifica un ítem de datos X es necesario registrar en el *log* tanto su valor anterior como su nuevo valor.
 - (c) Cuando se modifica un ítem de datos X es obligatorio *flushear* el nuevo valor de X a disco antes de *flushear* el registro de *log* correspondiente a disco.
 - (d) Si el sistema se reinicia y el algoritmo detecta que una transacción T_i no había llegado a *commit*ear, se debe escribir $(ABORT, T_i)$ en el archivo de *log* y *flushearlo* a disco antes de deshacer las modificaciones realizadas por T_i en disco.