

Base de Datos (75.15 / 75.28 / 95.05)

Evaluación Integradora - 21 de febrero de 2018

TEMA 20172C4						Padrón: _____
Rel.		SQL		Proc.		Apellido: _____
DR		CyT		NoSQL		Nombre: _____
Corrigió:						Cantidad de hojas: _____
Nota:						<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. (*Modelo relacional*) Explique en qué consisten las restricciones de unicidad, de integridad de entidad y de integridad referencial del modelo relacional.
2. (*SQL*) Una base de datos guarda una matriz cuadrada rala A de grandes dimensiones a través de una tabla $A(i, j, v)$, en donde los atributos i y j identifican al elemento (i, j) de la matriz, mientras que v representa su valor, a_{ij} . La tabla sólo almacena aquellas posiciones para las cuales el valor a_{ij} es no nulo.

Escriba una consulta en lenguaje SQL que devuelva como resultado una tabla $B(i, j, v)$ que represente, con el mismo formato, a la matriz B que resulta de elevar A al cuadrado (es decir, $B = A^2 = A \cdot A$). Sólo almacene aquellas posiciones para las cuales el valor b_{ij} es no nulo. En el Cuadro 1 encontrará un pequeño ejemplo ilustrativo.

Nota: Recuerde que los elementos de la matriz $B = A^2$ se calculan como $b_{ij} = \sum_k a_{ik} \cdot a_{kj}$.

$$A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 4 \\ 0 & 1 & 0 \end{bmatrix} \quad B = A^2 = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 5 & 2 \\ 1 & 0 & 4 \end{bmatrix}$$

A			B		
i	j	v	i	j	v
1	2	1	1	1	1
1	3	2	1	2	2
2	1	1	1	3	4
2	3	4	2	2	5
3	2	1	2	3	2
			3	1	1
			3	3	4

Cuadro 1: Ejemplo correspondiente al Ejercicio 2. Matrices A y B (*arriba*) y su representación como tabla (*abajo*).

3. (*Procesamiento de Consultas*) En este ejercicio compararemos en una situación concreta el método de *sort-merge* para el cálculo de la junta natural con el método de iteración por bloques.

(*Parte I*) Suponga que dispone de 4GB de memoria RAM, y debe ordenar una tabla con 4 millones de bloques de 4KB, que ocupa 16GB. Se pide:

- Considere que se divide la tabla en cuatro partes de 4GB cada una, que denominaremos P1, P2, P3 y P4. El primer paso (*sort*) consistirá en leer cada una de las cuatro partes aprovechando toda la memoria disponible, ordenarlas y guardarlas en disco. Calcule el costo de este paso en términos de cantidad de accesos a disco.
- El segundo paso (*merge*) consiste en combinar las 4 partes ya ordenadas, guardando el resultado ordenado en disco. Calcule el costo de este paso en términos de cantidad de accesos a disco. Indique cuál es la mínima cantidad de memoria RAM que debe tener disponible para poder mantener ese costo en este paso.
- Calcule el costo total que le insumió ordenar la tabla y guardarla en términos de cantidad de accesos a disco.

(*Parte II*) Considere ahora los esquemas de relación $R(A, B, C)$ y $S(C, D, E)$, con 4 millones de bloques de 4KB (16GB) y 2 millones de bloques de 4KB (8GB) respectivamente. Se pide:

- Calcule el costo total de la junta natural $R * S$ utilizando el método de *sort-merge* disponiendo de 4GB de RAM, en base a la fórmula calculada en la *Parte I*. Recuerde que además de ordenar cada tabla, luego debe juntarlas.
- Calcule en forma aproximada el costo de la junta natural $R * S$ utilizando el método de iteración por bloques, suponiendo que emplea todos los bloques de memoria menos uno ($\approx 4GB$) para la lectura de la tabla S , y el bloque restante para recorrer R .
- ¿Qué método resultó más conveniente? Si aumenta el tamaño de las tablas manteniendo la cantidad de memoria RAM, ¿cree que esta relación se reafirma o se revierte? Justifique su respuesta.

Muestre todos los cálculos involucrados.

Nota: No considere el costo de guardado del resultado final en disco.

4. (*Diseño Relacional*) Considere el esquema de relación $R(A_1, A_2, \dots, A_n)$ con un cubrimiento minimal F_M y una descomposición $\{R_1(X_1, X_2, \dots, X_{x_M}), R_2(Y_1, Y_2, \dots, Y_{y_M})\}$ del mismo, en donde $\bar{X} = \{X_1, X_2, \dots, X_{x_M}\}$ e $\bar{Y} = \{Y_1, Y_2, \dots, Y_{y_M}\}$ forman un cubrimiento del conjunto de atributos¹ y tienen al menos un atributo en común ($\bar{X} \cap \bar{Y} \neq \emptyset$). Considere también que todas las dependencias funcionales del cubrimiento minimal F_M involucran exclusivamente a atributos de \bar{Y} ².

Para cada una de las siguientes afirmaciones, indique si la misma *es necesariamente verdadera*, si *puede ser verdadera* ó si *es necesariamente falsa*, justificando su respuesta.

- Para toda instancia de relación $r \in R$: $\pi_{\bar{X}}(r) * \pi_{\bar{Y}}(r) \subset r$.³
- Para toda instancia de relación $r \in R$: $n(\pi_{\bar{X}}(r)) = n(\pi_{\bar{Y}}(r))$.
- La descomposición es sin pérdida de información.
- La descomposición preserva todas las dependencias funcionales.

¹ Nota: Es decir, $\bar{X} \subset \bar{A}$, $\bar{Y} \subset \bar{A}$ y $\bar{X} \cup \bar{Y} = \bar{A}$, siendo $\bar{A} = \{A_1, A_2, \dots, A_n\}$.

² Nota: Ésto es, $\forall (C \rightarrow D) \in F_M : C \subset \bar{Y} \wedge D \subset \bar{Y}$.

³ Nota: Ésto es, la junta de las proyecciones r_1 y r_2 está contenida en r o es igual a r .

5. (*Concurrencia y Transacciones*) ¿Es posible que en un solapamiento de transacciones que respeta el *Protocolo de Lock de Dos Fases (2PL)* el aborto de una transacción produzca *rollbacks* en cascada de otras transacciones? Justifique su respuesta.
6. (*NoSQL*) Una clínica utiliza una base de datos Neo4J para realizar consultas sobre los medicamentos a recetar a sus pacientes. Esta base de datos almacena a los Medicamentos, Acciones Terapéuticas y Drogas como nodos del grafo:

```

1 (med1:Medicamento {nombre: 'Lotrial'})
2 (med2:Medicamento {nombre: 'Alloboxal'})
3 (med3:Medicamento {nombre: 'Benzetacil'})
4 (dr1:Droga {nombre: 'Enalapril'})
5 (dr2:Droga {nombre: 'Alopurinol'})
6 (dr3:Droga {nombre: 'Penicilina'})
7 (at1:AccionTerapeutica {nombre: 'Antihipertensivo'})
8 (at2:AccionTerapeutica {nombre: 'Antigotoso'})
9 (at3:AccionTerapeutica {nombre: 'Antibiotico'})

```

Además, se almacenan las siguientes interrelaciones:

■ (Medicamento)-[:CONTIENE]->(Droga)

Indica que un Medicamento contiene en su composición una determinada Droga. Cada medicamento puede contener una o más drogas.

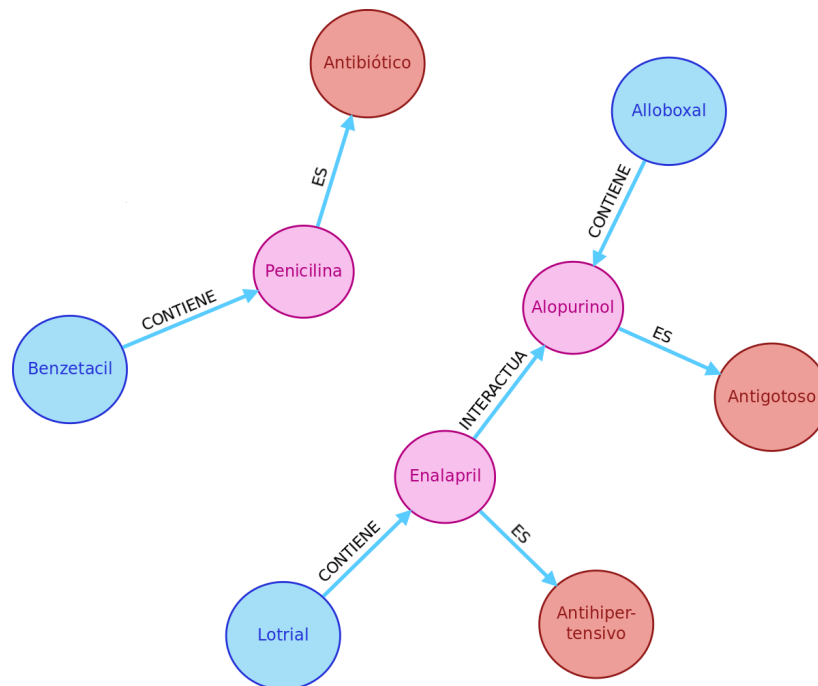
■ (Droga)-[:ES]->(AccionTerapeutica)

Indica que una Droga tiene una determinada Acción Terapéutica. Una droga puede tener una o más acciones terapéuticas.

■ (Droga)-[:INTERACTUA]-(Droga)

Indica que dos Drogas tienen una interacción medicamentosa entre sí. Esto quiere decir que, suministradas simultáneamente a un paciente, pueden tener efectos colaterales.

A continuación se muestra como ejemplo un subgrafo de la base de datos con algunos nodos e interrelaciones:



Para recetar un antibiótico a un paciente alérgico a la penicilina, el doctor Gregorio realizó ayer la siguiente consulta a la base de datos:

```

1  MATCH (m:Medicamento)-[:CONTIENE]->(:Droga)-[:ES]->(a:AccionTerapeutica)
2  WHERE NOT (m)-[:CONTIENE]->(:Droga {nombre: 'Penicilina'})
3  AND a.nombre = 'Antibiotico'
4  RETURN DISTINCT m.nombre

```

Acaba de llegar un paciente hipertenso con una infección, y el doctor Gregorio quiere recetarle dos medicamentos distintos: uno que contenga un antihipertensivo y otro que contenga un antibiótico. Además quiere asegurarse de que las drogas de un medicamento no tengan interacciones medicamentosas con las drogas del otro medicamento.

Escriba la consulta en Cypher que encuentre los posibles pares de medicamentos en la base de datos. El formato del resultado debería ser *(nombre_medicamento_1, nombre_medicamento_2)*.