

1. (Cálculo Relacional de Tuplas) La AASA (Asociación Argentina de Salto en Alto) está contactando a deportistas de salto en alto de todo el país para organizar un torneo nacional. La idea de la Asociación es que el torneo sea lo más federal y representativo posible, por eso se buscará que haya al menos un/a saltador/a de cada departamento del país, y que sea el mejor de su departamento.

A partir de los siguientes datos sobre los departamentos del país y sus saltadores, escriba una consulta en Cálculo Relacional de Tuplas que encuentre a aquellos/as saltadores/as que tengan la marca más alta de su departamento, devolviendo el nombre del saltador y el nombre del departamento. Si varios saltadores empatan dentro de su departamento, la Asociación quiere invitarlos a todos.

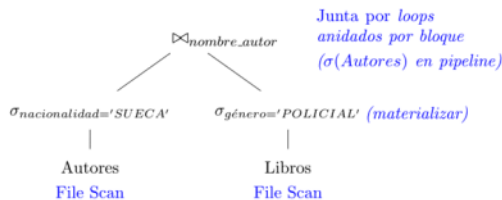
- Departamentos(nombre_dpto, población, superficie)
- Saltadores(nombre_saltador, nombre_dpto, marca)

$$\{ s.\text{nombre_saltador}, s.\text{nombre_dpto} \mid \text{Saltadores}(s) \wedge (\forall \theta) (\neg \text{Saltadores}(\theta) \vee \theta.\text{marca} \leq s.\text{marca}) \}$$

2. (Procesamiento de Consultas) Cándido trabaja en una biblioteca, y utiliza frecuentemente una base de datos relacional que él mismo creó para poder asistir a los lectores. Esta base contiene, entre otras, las siguientes dos tablas:

- Autores(nombre_autor, nacionalidad, fecha_nacimiento)
- Libros(ISBN_libro, nro_ejemplar, género, estante, nombre_autor)

Federica es una ávida lectora que acaba de consultarle a Cándido sobre alguna novela policial de origen sueco, y Cándido escribió una consulta sobre su base de datos que generó el siguiente plan de ejecución:



Dado que no se dispone de índices, el plan construido utiliza *file scans* como método de acceso, y debe materializar el resultado de la selección sobre **Libros** antes de la junta.

Considerando que la computadora de Cándido sólo dispone de $M=30$ bloques de memoria, y utilizando la información de catálogo que se brinda más abajo, se pide:

- Estime el costo del plan de ejecución generado, en términos de cantidad de bloques.
- Explique por qué la siguiente estimación $n(\bowtie) = \frac{500 \cdot 20000}{\max(500, 20000)} = 500$ para la cardinalidad del resultado no es correcta.
- Si tuviera que crear un único índice a efectos de mejorar significativamente el desempeño de esta consulta, indique qué índice crearía.
- ¿Cree que una mayor cantidad de memoria podría mejorar significativamente el costo de esta consulta? Justifique su respuesta.

AUTORES	LIBROS
$n(\text{Autores}) = 50.000$	$n(\text{Libros}) = 1.000.000$
$B(\text{Autores}) = 5.000$	$B(\text{Libros}) = 100.000$
$V(\text{nacionalidad}, \text{Autores}) = 100$	$V(\text{nombre_autor}, \text{Libros}) = 50.000$
	$V(\text{género}, \text{Libros}) = 50$

Si, hasta poder usar pipeline en libros

$$a) \quad c(\sigma_1) = B(L) \Rightarrow \text{materializar}$$

$$B(L) = 10K$$

$$\text{cluster } H(I) = \frac{10000}{50} = 500$$

$$n(\sigma_1) = \frac{n(L)}{V(g, L)} = 20.000$$

$$F(L) = 10$$

$$B(\sigma_1) = \frac{n(\sigma_1)}{F(L)} = 2000$$

$$\frac{50.000}{50} = 1000$$

$$c(\sigma_2) = B(A) \rightarrow 5000 \rightarrow H(I) + \frac{5000}{100} = 50 \rightarrow \text{Aprovechar el pipelining para utilizar idx}$$

$$c(\bowtie) = B(\sigma_1) + \left\lceil \frac{B(\sigma_1)}{M-2} \right\rceil \cdot B(A)$$

Ambos $B(A)$ es la misma lectura

$$60 \cdot \frac{2000}{30} \approx 334K$$

$$c(Q) = B(L) + 2B(\sigma_1) + \left\lceil \frac{B(\sigma_1)}{M-2} \right\rceil \cdot B(A)$$

$$2) n(\sigma_2) = \frac{n(A)}{V(nA)} = \frac{50000}{100} = 500$$

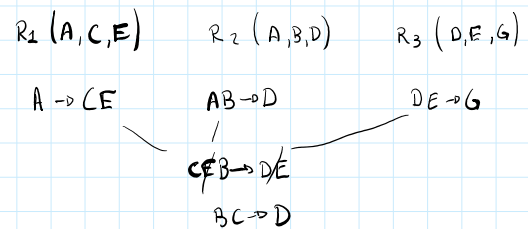
$$n(\bowtie) = \frac{n(\sigma_1) \cdot n(\sigma_2)}{\max(V(n_2, \sigma_2), V(n_2, \sigma_1))} = \frac{n(\sigma_1) \cdot n(\sigma_2)}{\max(n(\sigma_2), \frac{V(n_2, L)}{V(g, L)})} = \frac{500 \cdot 20000}{\max(500, 10000)} = 10000$$

3. (Diseño relacional) Considere una relación $R(A, B, C, D, E, G)$ con el siguiente conjunto de dependencias funcionales asociado $F = \{A \rightarrow CE, BC \rightarrow D, DE \rightarrow G\}$. Para cada una de las siguientes afirmaciones, indique si la misma es verdadera o falsa, justificando su respuesta.

- a) $\{A, C, D, E\}$ es superclave de R . ✓
- b) $\{A, B, E\}$ es clave candidata de R . F
- c) R se encuentra en Segunda Forma Normal (2FN). ✓
- d) La descomposición de R en $\{R_1(A, C, E), R_2(B, C, D), R_3(D, E, G)\}$ preserva la información. ✓
- e) La descomposición de R en $\{R_1(A, C, E), R_2(A, B, D), R_3(D, E, G)\}$ preserva todas las dependencias funcionales. ✓

$A^+ = \{A, C, E\}$
 $AB^+ = \{A, B, C, D, E, G\} \rightarrow$ clave candidata
 cualquier cosa extra es superclave

Clave primaria



Preserva

5. (NoSQL) La diosa griega *Afrodita* tuvo 16 descendientes, algunos de ellos mortales y otros inmortales (dioses), fruto de su relación con distintos dioses y mortales. Alejandro, que es fanático de la mitología, volcó esta información como parte de una base de datos en Neo4j para construir una base de conocimiento (*knowledge base*) de la mitología griega.

En esta base, los dioses y los mortales se representan como:

```
1 (afrodita:Dios:Ser { nombre: 'Afrodita', género: 'femenino'})
2 (eneas:Mortal:Ser { nombre: 'Eneas', género: 'masculino'})
3 ...
```

Mientras que las relaciones de filiación se describen de la siguiente forma:

```
1 (eneas)-[:HIJO_DE]->(afrodita)
2 (eneas)-[:HIJO_DE]->(anquises)
```

Escriba una consulta en Neo4j que encuentre a los dioses con los que Afrodita tuvo hijos, indicando para cada uno el nombre del dios y la cantidad de hijos que tuvo con él, y ordenando el resultado de manera de comenzar por los dioses con los que más hijos tuvo.

Ser

Dios Mortal

MATCH

```
(a:Dios {nombre:'Afrodita'})
  ↑
[:HIJO_DE]
  ↓
(h:Ser)
  ↓
[:HIJO_DE]
  ↓
(p:Ser)
```

WITH p, COUNT(h) AS cnt_hijos

RETURN p, cnt_hijos

ORDER BY cnt_hijos DESC