

## Cuarto Trabajo de Bases de Datos 2 (12.5%)

**Punto 1 (60%)** Una de las consultas clásicas más interesantes en SQL es la implementación de la división del álgebra. El ejemplo convencional es: dada una tabla de productos (**P**), de proveedores (**S**) y de proveedorxproducto (**SP**), obtener los datos de los proveedores que están combinados con todos los productos. Veamos a continuación tres soluciones clásicas para este problema.

```
CREATE TABLE s(  
  sn VARCHAR (10) PRIMARY KEY,  
  snombre VARCHAR (20) NOT NULL,  
  ciudad VARCHAR (20) NOT NULL  
);
```

```
CREATE TABLE p(  
  pn VARCHAR (10) PRIMARY KEY,  
  pnombre VARCHAR (20) NOT NULL,  
  color VARCHAR (15) NOT NULL  
);
```

```
CREATE TABLE sp(  
  sn VARCHAR (10) REFERENCES s,  
  pn VARCHAR (10) REFERENCES p,  
  PRIMARY KEY (sn,pn)  
);
```

### Consulta 1

```
SELECT *  
FROM s  
WHERE NOT EXISTS(SELECT *  
                  FROM p  
                  WHERE NOT EXISTS (SELECT *  
                                     FROM sp  
                                     WHERE sp.pn = p.pn  
                                     AND  
                                     sp.sn = s.sn  
                                     )  
                  );
```

### Consulta 2.

```
SELECT *  
FROM s  
WHERE (SELECT COUNT(*)  
       FROM sp  
       WHERE s.sn = sp.sn) = (SELECT COUNT(*) FROM p);
```

### Consulta 3.

```
SELECT *  
FROM s  
WHERE sn IN (SELECT sn  
              FROM sp  
              GROUP BY sn  
              HAVING COUNT(*) = (SELECT COUNT(*)  
                                  FROM p)  
              );
```

Verifique que las tres consultas anteriores solucionan el problema planteado.

### Consulta 4.

Obtenga **mediante una herramienta de inteligencia artificial (como ChatGPT) una cuarta consulta** que solucione el mismo problema y que sea distinta de las consultas anteriores (la consulta debe ser distinta en el sentido que no sea una simple variación “estética” de alguna de las tres anteriores).

**Para cada una** de las consultas anteriores (1 a la 4), haga pruebas con **diversas muestras de datos** (desarrolle un programa PL/SQL para cargar las tablas con datos aleatorios). Aumente **gradualmente** el tamaño de las muestras **hasta llegar a volúmenes de datos que permitan observar el comportamiento del rendimiento de cada consulta**.

Presente el *explain plan* para cada consulta y una tabla resumen donde **incluya los tamaños de las muestras que usó** y los **costos (cost)** y tiempos (*time*) obtenidos en cada caso.

Aplique también el TKPROF y presente para cada consulta, una tabla resumen donde incluya los tamaños de las muestras que usó y los resultados obtenidos para las siguientes tasas:

- i) LIOs (f+g) sobre filas procesadas (h).
- ii) Filas retornadas (i) sobre traídas (*fetches*) (j).
- iii) Lecturas de disco (k) sobre LIOs (f+g).

**Presente conclusiones (análisis de resultados) sobre los resultados obtenidos.**

**Punto 2 (40%)** Considerando las mismas tablas del punto anterior, considere las siguientes dos consultas:

### Consulta 1.

```
SELECT s.*, (SELECT COUNT(*)  
              FROM sp
```

```
WHERE sp.sn = s.sn) AS cuantos  
FROM s;
```

### Consulta 2.

```
SELECT *  
FROM s  
NATURAL JOIN (SELECT sn, COUNT(*) AS cuantos  
               FROM sp  
               GROUP BY sn)  
UNION  
SELECT s.*, 0 AS cuantos  
FROM s WHERE sn NOT IN (SELECT sn FROM sp);
```

**Verifique si estas dos consultas son equivalentes (solucionan un mismo problema).**

### Consulta 3.

Obtenga **mediante una herramienta de inteligencia artificial (como ChatGPT) una tercera consulta** que solucione el mismo problema y que sea distinta de las dos consultas anteriores (la consulta debe ser distinta en el sentido que no sea una simple variación “estética” de alguna de las dos anteriores). Por ejemplo, hay una solución interesante basada en un *natural left outer join*, es probable que ChatGPT llegue a esa o a otra...

Realice para estas tres consultas los mismos experimentos y análisis solicitados en el punto 1.

### Notas:

- **No modifique la estructura de las tablas.**
- Fecha de entrega: **Jueves 26 de septiembre hasta las 11 am.** Impreso en papel físico (**le sugiero imprimir por ambos lados para ahorrar papel y por favor imprima en un tamaño de letra que sea legible, es decir, que no requiera lupa para leerse**). Entregar en la oficina M8A-312 (si no estoy en el momento en que usted entrega, por favor introduzca el trabajo por debajo de la puerta). También enviar todo el trabajo (informe y programa de carga) en un archivo comprimido a [fjmoreno@unal.edu.co](mailto:fjmoreno@unal.edu.co). Para facilitar la identificación del trabajo, por favor, nombrar el archivo con los apellidos de los integrantes, ejemplo: `Electra_Knightley_Dennis.zip`<sup>1</sup>
- No se reciben trabajos en hora posterior. Si no entrega el trabajo impreso físicamente en papel, así lo envíe por correo, no se le calificará. Si lo entrega físico, pero no lo envía al correo, tampoco se le calificará.
- No se reciben versiones “mejoradas”.
- Grupos máximo de tres personas.
- Los trabajos deben ser independientes entre los grupos. Trabajos copiados **así sea en un punto** se califican con 0 (cero) para todos los integrantes.

---

<sup>1</sup> **Carmen Electra** (¡qué extraño siempre aparece!), **Keira Knightley** (a Knightley siempre la recuerdo por “*The Jacket*”, película basada en la novela “*The Star Rover*” (“El Peregrino de las Estrellas”) de Jack London, aunque la traducción debería ser más bien “El Viajero Astral”) y a super **Cathy Dennis** (de quien estoy revisitando su discografía).

- Cualquier duda consultarla personalmente o por correo electrónico con el profesor.

## **Pregunta frecuente**

**¿Con cuántos datos se deben hacer las pruebas?**

**Rta.** Usted decide con cuantos datos iniciar las pruebas y de a cuantos datos va aumentando gradualmente, hasta llegar a volúmenes de datos que permitan observar el comportamiento del rendimiento de cada consulta. En el trabajo deben quedar claro los tamaños de muestras usados.

F. Moreno

Septiembre 12 de 2024.

**Éxitos en su vida profesional.**