

## Práctico 9

### *Procesamiento y optimización de consultas*

**Ejercicio 1. (\*)**

Dadas dos relaciones R(A, B, C) y S(C, D, E), escribir 3 planes lógicos válidos para la siguiente consulta SQL:

```
SELECT R.B, R.C, S.D
FROM R, S
WHERE R.C = S.C AND R.A = 5;
```

**Ejercicio 2.**

Sea el siguiente esquema relacional que representa empleados, departamentos, proyectos y la asignación de empleados a proyectos, de una empresa familiar donde el 90% de los empleados tienen apellido "Silva".

**Empleado** (CI, nombreP, apellido, direccion, salario, CIsuper, numD)

En esta tabla CIsuper es la cédula de identidad del supervisor del empleado y numD es el número del departamento al cual está asignado dicho empleado.

**Departamento** (numeroD, nombreD, CIgte)

En esta tabla CIgte es la cédula de identidad del gerente del departamento.

**Proyecto** (numeroP, nombrePr, lugarP, numD)

En esta tabla se almacena información de los proyectos, número y nombre de proyecto, lugar donde se realizó el proyecto y número de departamento al cual pertenece el proyecto.

**TrabajaEn** (CIemp, numP, horas)

En esta tabla se representa la asignación de empleados a proyectos.

En este esquema se cumplen las siguientes dependencias de inclusión:

- $\prod_{CIsuper}(\text{Empleado}) \subseteq \prod_{CI}(\text{Empleado})$
- $\prod_{CIgte}(\text{Departamento}) \subseteq \prod_{CI}(\text{Empleado})$
- $\prod_{nd}(\text{Empleado}) \subseteq \prod_{numerod}(\text{Departamento})$
- $\prod_{numd}(\text{Proyecto}) = \prod_{numerod}(\text{Departamento})$
- $\prod_{CIemp}(\text{TrabajaEn}) \subseteq \prod_{CI}(\text{Empleado})$
- $\prod_{nump}(\text{TrabajaEn}) \subseteq \prod_{numerop}(\text{Proyecto})$

Además se conoce la siguiente relación entre la cardinalidad de las relaciones:

$$|\text{Departamento}| < |\text{Proyecto}| < |\text{Empleado}| < |\text{TrabajaEn}|$$

Basándose en las reglas de transformación de los operadores del álgebra relacional, obtenga el plan lógico para cada una de las siguientes consultas aplicando las heurísticas vistas en el curso.

- (a) 

```
SELECT    nombrep, apellido, direccion
FROM      Empleado, Departamento
WHERE     nombred = 'Investigacion' AND numerod = numD
```

- (b) SELECT E.nombrep, E.apellido, S.nombrep, S.apellido  
FROM Empleado E, Empleado S  
WHERE E.CIsuper = S.CI
- (c) SELECT nombrep, apellido, salario  
FROM Empleado, TrabajaEn, proyecto  
WHERE CI = CIemp AND nump = numerop AND nombrepr = 'ProductoX'
- (d) SELECT numerop  
FROM Proyecto, Departamento, Empleado  
WHERE numd = numerod AND CIgte = CI AND apellido = 'Silva'  
UNION  
SELECT numerop  
FROM Proyecto, TrabajaEn, empleado  
WHERE numerop = nump AND CIemp = CI AND apellido = 'Silva'

**Ejercicio 3. (\*)**

Sean dos relaciones R y S no ordenadas. Suponga además que existen dos índices sobre la relación R, uno por el atributo A y otro por el C (y ningún índice más). Sabiendo que se dispone de un algoritmo de Sort que puede ser utilizado, dar 2 planes físicos indicando el algoritmo a utilizar para cada operación, para el plan lógico de la Figura 1

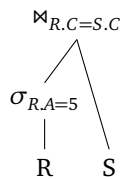


Figura 1: Plan lógico para el Ejercicio 3

**Ejercicio 4.**

Dadas las relaciones R(A,B,C) y S(B,C,D,E), dar la estimación de los tamaños en cantidad de tuplas de los resultados de las siguientes operaciones:

- (a) Sabiendo que  $V(D,S)=500$  (cantidad de valores distintos de D en S), que los valores de D en S se encuentran distribuidos en forma uniforme, y que  $|S| = 3400$ , estime el tamaño de  $\sigma_{D='Montevideo'}S$
- (b) Sabiendo que  $|S| = 3400$  y que el 60% de las tuplas de S tienen el valor 'Maldonado' en el atributo D y de estas el 20% tienen el valor 350 en el atributo E, estime el tamaño de  $\sigma_{D='Maldonado' \wedge E=350}S$
- (c) Sabiendo que  $|R| = 138$ ,  $|S| = 3400$  y que la selectividad del join entre dichas tablas es  $js = 0.30$ , estime el tamaño de  $R * S$
- (d) Sabiendo que  $|R| = 138$ ,  $|S| = 3400$ ,  $\prod_{B,C}(R) \subseteq \prod_{B,C}(S)$  y que (BC) es clave en S, estime el tamaño de  $R * S$

**Ejercicio 5.**

En la base de datos de una empresa de servicios informáticos se tienen las siguientes tablas relacionales:

**Abonados**(IdEquipo, vendedor, tipoAbono, vCuota, fecha,marca)

En esta tabla, para cada equipo se tiene el vendedor que realizó la suscripción, el valor actual de la cuota y la fecha en que se realizó la suscripción.

**Servicios**(IdEquipo, CodServicio, fecha,importe, status)

Esta tabla representa los servicios realizados a cada equipo. Para cada equipo, dado un código de servicio y una fecha, la tabla contiene el importe de ese servicio y el status. El atributo status indica si el importe fue saldado al contado, si es a crédito o si está con mora.

**Cientes**(IdEquipo,CICli)

Esta tabla contiene los equipos abonados junto con los clientes dueños de los mismos.

Dada la siguiente consulta SQL:

```
SELECT  CICli , vendedor
FROM    Clientes C, Servicios S, Abonados A
WHERE   A.IdEquipo = S.IdEquipo AND A.IdEquipo = C.IdEquipo AND
        A.marca = 'Dell' AND S.status = 'Mora'
```

- (a) Construir el árbol canónico para la consulta dada.
- (b) Dar un plan lógico para la consulta, optimizado mediante las heurísticas vistas en el curso **sin considerar los tamaños**. Mostrar los pasos aplicados.
- (c) Teniendo en cuenta la tabla e información indicados a continuación:

Tabla	Cant. Tuplas
Abonados	720
Clientes	720
Servicios	3500

- un equipo está asociado a un solo cliente
- hay 29 equipos abonados de marca “Dell”
- hay 1167 servicios en estado de mora
- solo 50 equipos de los que se encuentran en mora son marca “Dell”

¿ Considera necesario modificar el plan lógico definido en la parte b? Justifique.

- (d) Dar un plan físico cualquiera para el plan lógico de la parte c, utilizando (cuando sea posible) la información dada en la siguiente tabla:

Indice	Tabla/Atributo	Tipo
IndCli1	Clientes/IdEquipo	Primario
IndCli2	Clientes/CiCli	Secundario
IndAbo1	Abonados/IdEquipo	Primario
IndAbo2	Abonados/marca	Secundario

**Ejercicio 6.**

Una empresa de turismo tiene personal que cumple la función de Guías Turísticos. A estos guías, se les asigna algún tipo de vehículo para llevar pasajeros. De los vehículos se conoce la matrícula y el modelo. De cada modelo de vehículo se conoce la capacidad, si tiene dirección hidráulica o no y el tipo de combustible que usa. En cada fecha, a cada guía se le asigna un determinado vehículo. La base de datos que mantiene esta información contiene (entre otras) las siguientes tablas:

**Vehiculos**(mat, modelo)

**Modelos**(modelo, cap, DH, comb)

**Asignacion**(CIG, fecha, mat)

Considere la consulta “Capacidad y tipo de combustible de los vehículos asignados al guía con cédula de identidad 7.458.256 el día 23/1/2007”

- (a) Indique la cantidad de tuplas del resultado y justifique su respuesta.

- (b) La empresa está evaluando qué manejador utilizar para implementar su base de datos. Para eso comparan los árboles que generan dos posibles manejadores luego de la optimización heurística. Considerando los árboles que se presentan en la Figura 2, donde V, M y A representan las tablas Vehículos, Modelos y Asignación respectivamente, y sabiendo que en ambos casos existe un índice primario por cada clave primaria, el cual indexa la concatenación de los atributos en el orden en que están declarados, ¿cuál de los dos manejadores elegiría? y ¿por qué?

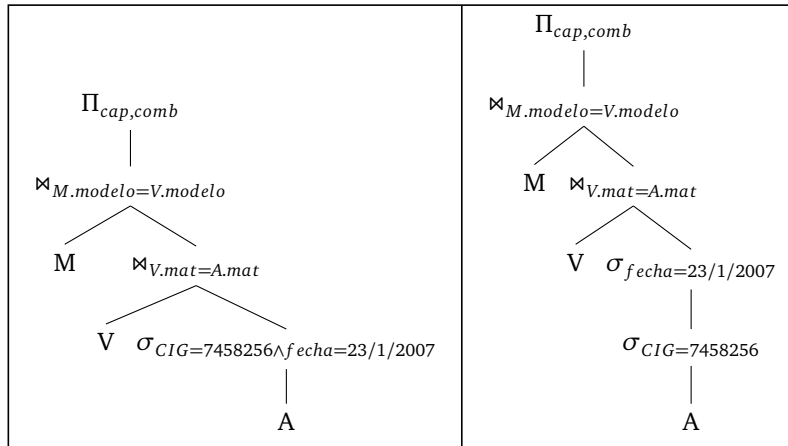


Figura 2: Planes lógicos para el RDBMS1 (izq) y RDBMS2 (der)

**Ejercicio 7.**

Considere las tablas representadas en la Figura 3, donde:

- **Tweets:** Representa *tweets* extraídos de la red social *Twitter* y contiene 16.000 tuplas.
- **Votos:** Representa votos de personas sobre los *tweets* y contiene 40.000 tuplas.
- **Features:** Representa características de los *tweets* y contiene 500.000 tuplas; Existen 10 valores distintos para el atributo *nombre\_features* y su distribución es uniforme.

Se sabe además que:

- Los atributos marcados con una llave en cada relación son clave primaria.
- *Votos.id\_tweet* es clave foránea de *Tweets.id\_tweet*.
- *Features.id\_tweet* es clave foránea de *Tweets.id\_tweet*.

Sobre el esquema de la Figura 3 se desea realizar la siguiente consulta:

```
SELECT T.id_tweet, texto_tweet, voto, nombre_feature
FROM Votos AS V, Tweets AS T, Features AS F
WHERE
    V.id_tweet = T.id_tweet AND
    T.id_tweet = F.id_tweet AND
    F.nombre_feature = 'Dialogo'
```

- (a) Construir el árbol canónico de la consulta dada.
- (b) Construir un plan lógico para la consulta, utilizando las heurísticas y teniendo en cuenta los tamaños.
- (c) Al ejecutar la consulta en *Postgres* se obtuvo el siguiente resultado utilizando la opción *Explain*
  1. Reconozca los algoritmos utilizados y explique su funcionamiento.

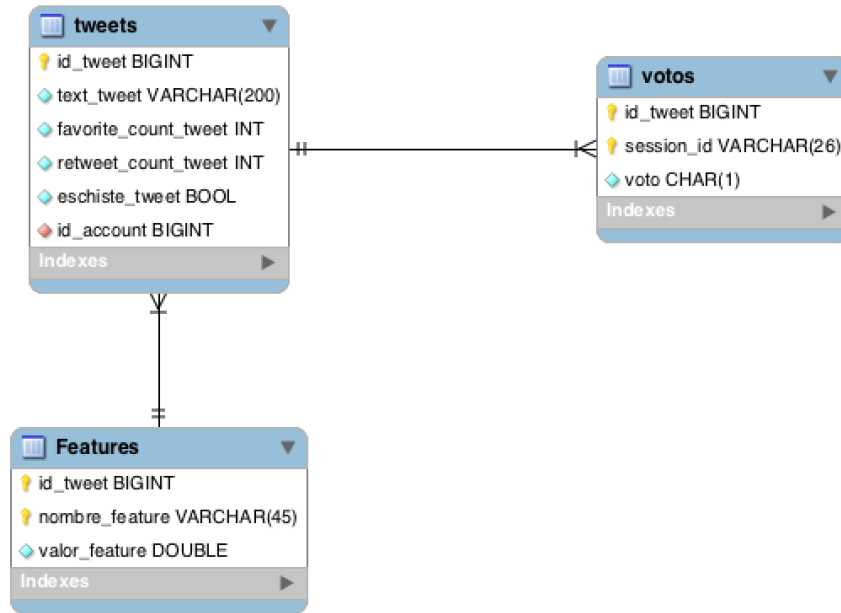


Figura 3: Esquema relacional para el ejercicio 7

- II. ¿Qué conclusiones puede sacar sobre la organización de los datos en las tablas?
- III. Se desea optimizar la consulta. ¿Qué estructura agregaría? Justifique.
- IV. ¿Qué algoritmo se ejecutaría al agregar la nueva estructura?

