

# Bases de Datos

## Práctico 5: Optimización de Consultas

**Ejercicio 1:** Suponga que hay un árbol B+ en (dept\_name, building) en la tabla department. ¿Cuál es la mejor manera de manejar la siguiente selección?

$$\sigma_{(building < "Watson") \wedge (budget < 55000) \wedge (dept\_name = "Music")}(department)$$

**Ejercicio 2:** considerar las tablas r1(A, B, C) y r2(C, D, E) y r3(E, F) con claves primarias A, C y E respectivamente. Asumir que r1 tiene 1000 tuplas, r2 1500 tuplas y r3 750 tuplas. Estimar el tamaño de  $r1 \bowtie r2 \bowtie r3$  y dar una estrategia eficiente para computar la reunión.

**Ejercicio 3:** considerar las tablas r1(A, B, C) y r2(C, D, E) y r3(E, F); asumir que no hay claves primarias, excepto el esquema entero. Sea  $V(C, r1) = 900$ ,  $V(C, r2) = 1100$ ,  $V(E, r2) = 50$  y  $V(E, r3) = 100$ . Asumir que r1 tiene 1000 tuplas, r2 1500 tuplas y r3 750 tuplas. Estimar el tamaño de  $r1 \bowtie r2 \bowtie r3$  y dar una estrategia eficiente para computar la reunión.

**Ejercicio 4:** Usando reglas de equivalencia demostrar las siguientes transformaciones:

- $\sigma_{nombre="Optimización de Consultas"}(\sigma_{apellido="Selinger"}(profe \bowtie curso)) = \sigma_{apellido="Selinger"}(profe) \bowtie \sigma_{nombre="Optimización de Consultas"}(curso)$
- Con telefono(legajo, numero),  
 $\Pi_{legajo, numero}(\sigma_{nombre="Optimización de Consultas"}(curso \bowtie (profe \bowtie telefono))) = \Pi_{legajo}(\sigma_{nombre="Optimización de Consultas"}(curso)) \bowtie \Pi_{legajo, nombre}(profe \bowtie telefono)$

**Ejercicio 5:** sugerir al menos 2 reglas de equivalencia que involucren agregación. Intentar probarlas usando álgebra de tablas.

**Ejercicio 6:** contestar las siguientes preguntas. Si la respuesta es falso, dar un contraejemplo:

- Eliminación de duplicados puede ser empujada adentro de una proyección
- Eliminación de duplicados puede ser empujada adentro de unión
- Proyección puede ser empujada adentro de unión

**Ejercicio 7:** Considerar las tablas con la siguiente información estadística:

$W(a, b)$	$X(b, c)$	$Y(c, d)$	$Z(d, e)$
$T(W) = 400$	$T(X) = 300$	$T(Y) = 200$	$T(Z) = 100$
$V(W, a) = 50$	$V(X, b) = 60$	$V(Y, c) = 50$	$V(Z, d) = 10$
$V(W, b) = 40$	$V(X, c) = 100$	$V(Y, d) = 20$	$V(Z, e) = 50$

T(R) significa número de tuplas de la tabla R.

Dar las entradas de la tabla de programación dinámica que evalúa todos los ordenes posibles de natural join permitiendo:

- Árboles de reunión profunda a la izquierda solamente
- Todos los árboles.

¿Cuál es la mejor elección en cada caso?

**Ejercicio 8:** considerar la optimización heurística siguiente:

1. Realizar selección tempranamente
2. Realizar proyección tempranamente
3. Hacer selección más restrictiva
4. Hacer operaciones de reunión más restrictivas
5. Ciertas selecciones pueden ser combinadas con producto cartesiano para tornar las operaciones en una reunión (natural o selectiva)

Ver detalles de esta optimización heurística en filmina 47.

Aplicar esta optimización heurística a las siguientes consultas:

1.  $\sigma_{instructor.ID = teaches.ID}(\sigma_{dept\_name = "Physics"}(instructor \times teaches))$
2.  $\Pi_{name, course\_id} (\sigma_{instructor.ID = teaches.ID}(\sigma_{dept\_name = "Physics"}(instructor \times teaches)))$
3.  $\Pi_{name, title} (\sigma_{dept\_name = "Comp. Sci."} (instructor \bowtie teaches \bowtie course))$

Donde *name* es nombre de instructor, *title* es título del curso. Instructor tiene el atributo *dept\_name*.

Para cada expresión de consulta transformarla paso a paso indicando heurística usada y para cada heurística usada indicar reglas de equivalencia usadas y en qué orden se aplicaron (mostrar aplicación de heurística paso a paso).