

## Bases de Datos 2025 - Segundo Parcial de muestra

**Ejercicio 1:** Definir un operador que produce la unión de dos tablas sin repetición de tuplas en el resultado. Se pide:

- Hacer una definición recursiva del operador.
- ¿Como se puede definir el operador usando operaciones enseñadas?
- Sugerir algún operador físico para el operador y estimar su costo.

**Ejercicio 2:** Sea la siguiente base de datos de fotografías:

usuarios (uid, nombre)

fotos (fid, dueño, tamaño)

comentarios (cid, autor, foto, texto)

Observar que *dueño* y *autor* son usuarios (uid), *foto* es una fid.

Considerar la consulta:  $\prod_{fid, uid} (fotos \bowtie \sigma_{nombre='Luis Pérez'} (usuario))$ .

Asumir que hay índice primario en *nombre* para usuario, que usuario tiene 100.000 tuplas, que un bloque tiene 4096 bytes, que hay 10.000 nombres diferentes de usuarios, que un mismo nombre no se repite más de 15 veces en usuario.

1. Calcular la cantidad de claves de búsqueda por nodo del índice (árbol B+)
2. Calcular la altura del árbol B+.
3. Indicar el operador físico que conviene usar para la selección y estimar el costo en número de transferencias de bloques para este operador físico.

**Ejercicio 3:** Sea la BD con las siguientes tablas:

persona (DNI, nombre, edad)

bibliotecario (DNI, antigüedad)

trabajaEn (DNI, nombreBib, rol)

Sea la siguiente consulta:

$\prod_{nombre, DNI, nombreBib} ( \sigma_{antigüedad > 5 \text{ and } rol = 'bibliotecario'} (Persona \bowtie bibliotecario \bowtie trabajaEn) )$

persona	bibliotecario	trabajaEn
1000 tuplas	400 tuplas	1500 tuplas
$V(persona, edad) = 60$	$V(bibliotecario, antigüedad) = 40$	$V(trabajaEn, rol) = 5$
$V(persona, nombre) = 900$		$V(trabajaEn, DNI) = 750$
		$V(trabajaEn, nombreBib) = 200$

Se pide:

1. Aplicar primero optimización por costo usando programación dinámica para las reuniones naturales asumiendo la tabla de abajo.
2. Luego transformar la consulta obtenida en el paso anterior aplicando optimización heurística usando las reglas enseñadas.

No se pide elegir algoritmos más adecuados para operadores de la consulta obtenida.

**Ejercicio 4:** Recuperación de la información: Responder:

1. Supongamos que queremos hacer búsquedas ordenadas por relevancia de documentos en una colección de documentos en mi disco rígido; asumir que se usa modelo de espacio vectorial e índices invertidos. Las consultas son booleanas. Dar significado de las componentes de las tuplas que van en una lista de índice invertido asociada a un término.
2. Listar los pasos para procesar una consulta basada en relevancia de documentos; es una colección que está en un disco rígido, se usa modelo de espacio vectorial, similitud de vectores para la relevancia, y consultas booleanas. Evitar pasos innecesarios.