

UNIVERSIDAD DE BUENOS AIRES

FIUBA - BASE DE DATOS

PARCIALITO COSTOS

Nombre: Molina, Taiel Alexis

Padrón: 109458

Fecha: 20/05/2025

Compromiso ético: Esta evaluación es domiciliaria, le permite evaluar su nivel de comprensión de la materia antes de la evaluación definitiva. Para que tenga validez usted debe garantizar que ha intentado resolverlo con su mejor esfuerzo; y que las respuestas sean el resultado únicamente de su trabajo y conocimiento individuales.

Instrucciones: Genere un archivo en formato pdf con la resolución del examen. Este archivo llámelo "Costos_(nro padrón)", es decir por ejemplo: Costos_76463. Suba este archivo al campus antes de las 24:00hs. del sábado 17/05/2025 para que sea corregido y devuelto el miércoles siguiente, de lo contrario tendrá tiempo hasta las 24:00 del martes 20/05/2025 y la devolución será el miércoles de la semana siguiente.

La famosa empresa AQD, creadora del ERP "AQD", consiguió una lista de empresas del rubro "Laboratorios" y quiere asignarles vendedores para que intenten venderle su maravilloso producto empresarial. Para vincular a las empresas con los vendedores de la misma ciudad, hizo la siguiente consulta a dos tablas:

- listadoempresas (id_empresa, nombre, nombre_contacto, ciudad)
- vendedores (legajo, nombre, celular, ciudad, fecha_ingreso)

```
SELECT
FROM listadoempresas
  INNER JOIN
    vendedores USING (ciudad)
```

Estime el costo de realizar esta junta y la cantidad de filas que serán devueltas. Se conocen los siguientes datos de las tablas, y, debido a que el potente ERP consume una cantidad importante de memoria, sólo se tienen 12 bloques de memoria disponibles para hacerla:

LISTADOEMPRESAS	VENDEDORES
$n(\text{listadoempresas}) = 10000$	$n(\text{vendedores}) = 50000$
$B(\text{listadoempresas}) = 1000$	$B(\text{vendedores}) = 2000$
$V(\text{ciudad, listadoempresas}) = 100$	$V(\text{ciudad, vendedores}) = 50$

Resolución

Como vimos con Lucas en clase, el join suele ser la operación más demandante a nivel costo, ya que implica recorrer dos tablas al completo. También vimos que existen 4 alternativas para resolver la operación *join*:

- Loops anidados por bloques.
- Loop con único índice.
- Sort-merge.
- Junta Hash grace.

A partir de ahora calculo el costo de realizar el join con cada una de estas cuatro opciones, llamando E a la tabla de empresas y V a la tabla de vendedores:

Loops anidados por bloques

Utilizando la fórmula

$$C(E \bowtie V) = B(E) + \frac{B(E)}{M-2} * B(V)$$

Reemplazando con nuestros valores:

$$C(E \bowtie V) = 1000 + \frac{1,000}{12-2} * 2000$$

$$C(E \bowtie V) = 201000$$

Loops con único índice

Como no hay un índice no podemos aplicar esta alternativa en este ejercicio.

Sort-merge

Como las tablas no están ordenadas tenemos que calcular también el costo de ordenarlas.

Primero calculamos si nos alcanzan los bloques de memoria:

$$M \geq \frac{B(E)}{V(C,E)} + \frac{B(V)}{V(C,V)}$$

$$\frac{1000}{100} + \frac{2000}{50} + 1 = 141 > 12$$

→ tengo que utilizar un sort externo.

$$C(E \bowtie V) = Cost(Ord_M(E)) + Cost(Ord_M(V)) + B(E) + B(V)$$

Calculo los costos de los ordenamientos primero. Inicio con el costo de ordenar las empresas

$$Cost(Ord_M(E)) = 2 \cdot B(E) \cdot \log_{12-1} B(E)$$

$$Cost(Ord_M(E)) = 2 \cdot 1000 \cdot \log_{11} 1000 = 2 \cdot 1000 \cdot 2,958$$

Redondeando el logaritmo a 3

$$Cost(Ord_M(E)) = 6000$$

Continuo con el costo de ordenar los vendedores:

$$Cost(Ord_M(V)) = 2 \cdot B(V) \cdot \log_{12-1} B(V)$$

$$Cost(Ord_M(V)) = 2 \cdot 2000 \cdot \log_{11} 2000 = 2 \cdot 2000 \cdot 3,16$$

Redondeando el logaritmo a 4

$$Cost(Ord_M(V)) = 16000$$

Finalmente calculo el costo total:

$$C(E \bowtie V) = Cost(Ord_M(E)) + Cost(Ord_M(V)) + B(E) + B(V)$$

$$C(E \bowtie V) = 6000 + 16000 + 1000 + 2000$$

$$C(E \bowtie V) = 25000$$

Junta Hash Grace

Antes que calcular nada tengo que verificar que se cumplan las condiciones para poder aplicar el método:

Sabemos que, siendo P la cantidad de particiones:

$$P \leq M - 1$$

$$P \leq 12 - 1$$

$$\min(\frac{B(E)}{P}; \frac{B(V)}{P}) \leq M - 2$$

$$\min(\frac{1000}{11}; \frac{2000}{11}) \leq 12 - 2$$

$$\min(90, 09; 181, 81) \leq 10$$

Como no se cumple, no se puede utilizar este método.

Calculo de filas devueltas

$$n(E \bowtie V) = \frac{n(E) \cdot n(V)}{\max(V(C,E), V(C,V))}$$

$$n(E \bowtie V) = \frac{10000 \cdot 50000}{100}$$

$$n(E \bowtie V) = 5000000$$

Conclusión

Finalmente sabemos que las filas obtenidas son 5.000.000 y que entre los 4 métodos posibles solo pudimos aplicar el método de sort-merge y el de loop anidados por bloques siendo el primero el que menos costo tuvo.

Dejo un cuadro comparativo expresando estos datos:

Costo	Tipos de resolución de operación join
Loops anidados por bloques	201.000
Loops con único índice	No se puede aplicar
Sort-merge	25.000
Junta hash grace	No se puede aplicar
Cantidad de filas devueltas	
5.000.000	