

Notacion: Se va a usar el signo “[]” para indicar que se redondea en exceso y “{ }” para indicar que se redondea por defecto

Cliente (nro_cliente (8), nombre (100), cod_categoria(4), localidad (100))

Categoria(cod_categoria(4) , des_categoria(100))

Linea(nro_linea(15) , nro_cliente(8), fecha_alta (10))

$L_{cliente} = 8 \text{ bytes} + 100 \text{ bytes} + 4 \text{ bytes} + 100 \text{ bytes} = 212 \text{ bytes}$

$FB_{cliente} = \{4096/212 \text{ bytes}\} = \{19,3 \text{ tuplas/ bloques}\} = 19 \text{ tuplas / bloques}$

$B_{cliente} = [500.000 \text{ tuplas} / 19 \text{ tuplas / bloque}] = [26.315,78] \text{ bloques} = 26.316$

$L_{categoria} = 4 \text{ bytes} + 100 \text{ bytes} = 104 \text{ bytes}$

$FB_{categoria} = \{4096/104\} = \{39,38\} = 39 \text{ tuplas/ bloques}$

$B_{categoria} = [15/39] = 1 \text{ bloque}$

$L_{linea} = 15 \text{ bytes} + 8 \text{ bytes} + 10 \text{ bytes} = 33 \text{ bytes}$

$FB_{linea} = \{4096/ 33\} = \{124,12\} = 124$

$B_{linea} = [3.000.000/ 124] = 24.194$

SELECT nombre, fecha_alta

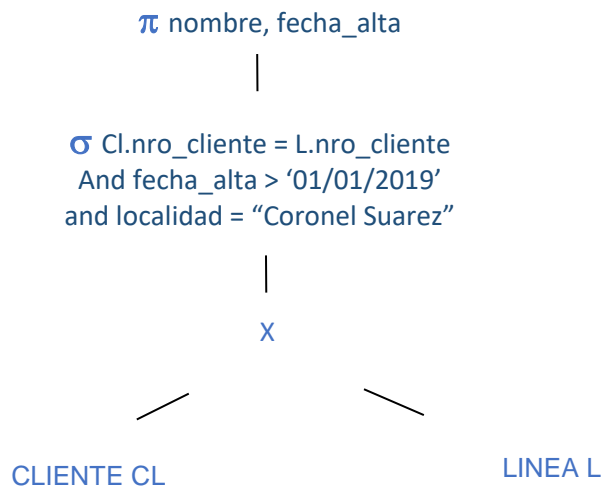
FROM Cliente Cl, Linea L

WHERE Cl.nro_cliente = L.nro_cliente

And fecha_alta > '01/01/2019'

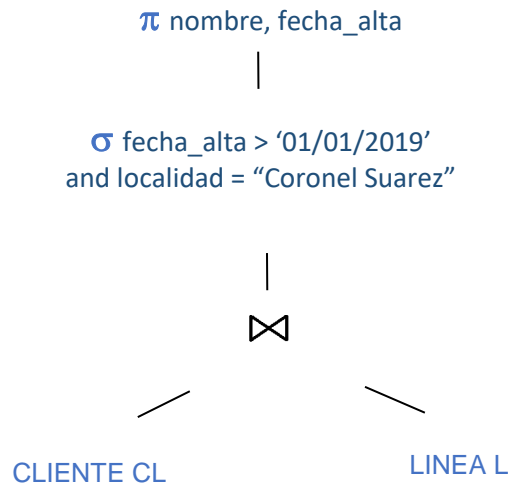
and localidad = "Coronel Suarez"

a) Árbol Canónico

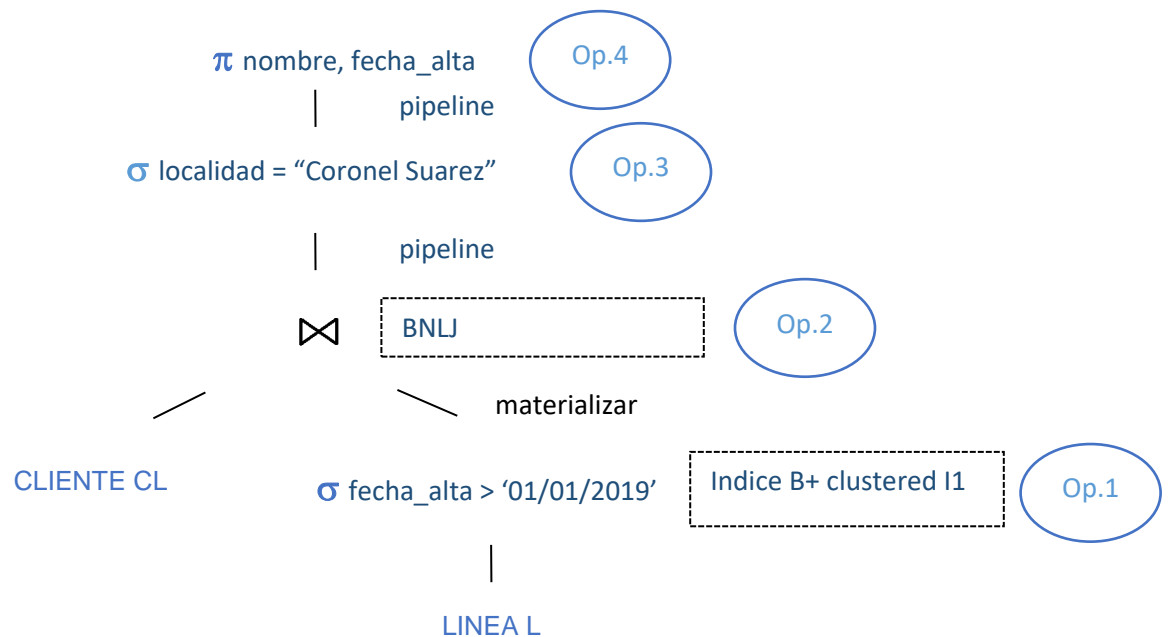


B) Arboles optimizados

De base transformamos el producto cartesiano en un join



B.1) Opción 1



Vamos a calcular el costo de cada operacion.

Op.1

Costo Input = $x + [T'/FBLinea]$ (costo de acceder a un indice clustered)

T' es la cantidad de registros que cumplen con "fecha de alta $> '01/01/2019'$ ", en este caso según el enunciado el 1% de las líneas cumple con esta condicion. Por lo tanto

$T' = 1\%$ de $3.000.000 = 30.000$ registros

Costo Input= $3 + [30.000/124] = [241,93] = 242 = 245$ bloques

Costo Ouput = costo de escribir 150.000 registros, que coincide con el costo de leerlos
 = 242 bloques

Op.2

Costo Input = $B_{cl} + [B_{cl}/(B-2)] * B_{Op1}$
 = $26.316 + \{26.316/(5-2)\} * 242 =$
 = $26.316 + 8.772 * 242 = 2.149.140$ bloques

Costo Output = 0 bloques (se pasan los datos pipeline)

De aca en mas todos los costos de input y de ouput son 0 , porque los datos se pasan por "pipeline"

Op. 3

Costo Input = 0 bloques

Costo Output = 0 bloques

Op. 4

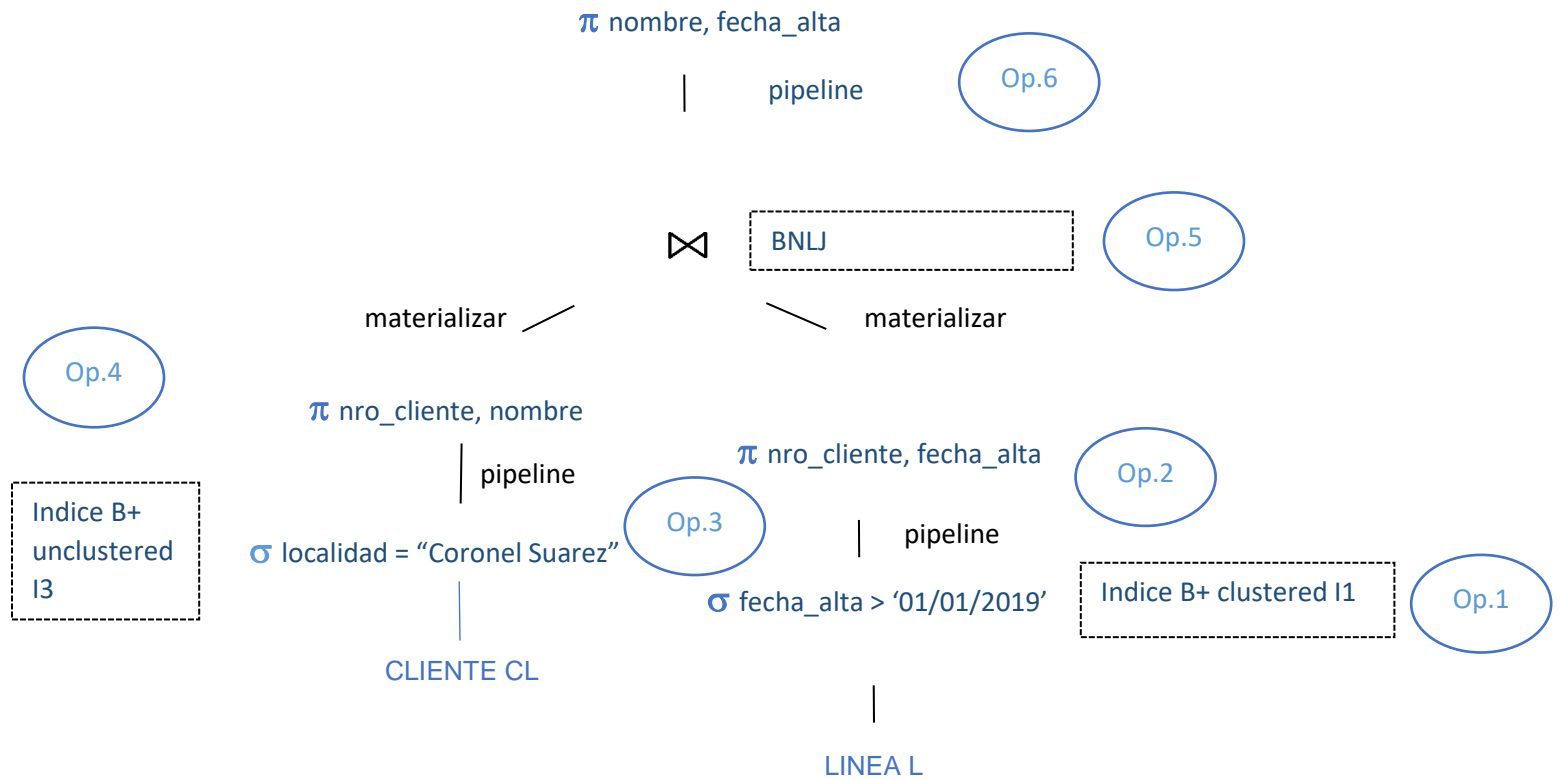
Costo Input = 0 bloques

Costo Output = 0 bloques

Para calcular el costo total de la opcion 1 es necesario sumar todos los numeros que estan en "rojo"

Costo Total Opcion 1= 245 bloques + 242 bloques + $2.149.140$ bloques + 0 bloques + 0 bloques + 0 bloques
 + 0 bloques + 0 bloques = $2.149.627$ bloques

B.2) Opción 2, llevar las selecciones y las proyecciones lo mas cerca posible de las hojas



El costo de input de la operacion 1 es igual al de la opcion anterior

Op.1

Costo Input = $x + [T'/FB_{\text{Linea}}]$ (costo de acceder a un indice clustered)

T' es la cantidad de registros que cumplen con "fecha de alta > '01/01/2019'", en este caso según el enunciado el 1% de las líneas cumple con esta condicion. Por lo tanto

$T' = 1\% \text{ de } 3.000.000 = 30.000$ registros

Costo Input = $3 + [30.000/124] = [241,93] = 242 = 245$ bloques

Costo Output = 0 bloques (es pipeline)

Op. 2

Costo Input = 0 bloques (es pipeline)

Costo Output, hay que escribir los 30.000 registros, pero es necesario calcular nuevamente el FB dado que ahora solo se escriben nro_cliente, fecha_alta,

$$L_{op2} = 10 + 8 = 18 \text{ bytes}$$

$$FB_{op2} = \lceil 4096/18 \rceil = \lceil 227.56 \rceil = 228 \text{ tuplas por bloque}$$

$$\text{Costo Output} = \lceil 30.000/228 \rceil = 132 \text{ bloques}$$

Op.3

$$\text{Costo Input} = X + T' = 3 + 1.000 = 1.003 \text{ bloques}$$

Como no tengo ningún otro dato asumo distribución uniforme

$$T' = \{T_{\text{Cliente}}/I_{\text{Cliente.localidad}}\} = 1.000$$

$$\text{Costo Output} = 0 \text{ bloques (es pipeline)}$$

Op. 4

$$\text{Costo Input} = 0 \text{ bloques (es pipeline)}$$

Costo Output, hay que escribir los 1.000 registros, pero es necesario calcular nuevamente el FB dado que ahora solo se escriben nro_cliente, nombre

$$L_{op3} = 100 + 8 = 108 \text{ bytes}$$

$$FB_{op3} = \lceil 4096/108 \rceil = \lceil 37,92 \rceil = 37 \text{ tuplas por bloque}$$

$$\text{Costo Output} = \lceil 1.000/37 \rceil = \lceil 27,02 \rceil = 28 \text{ bloques}$$

Op.5

$$\text{Costo Input} = B_{op4} + \{B_{op4}/(B-2)\} * B_{op2} = 28 + \{28/3\} * 132 = 28 + 10 * 132 = 1.348 \text{ bloques}$$

$$\text{Costo Output} = 0 \text{ bloques (es pipeline)}$$

Op. 6

$$\text{Costo Input} = 0 \text{ bloques (es pipeline)}$$

$$\text{Costo Total Opcion 2} = 245 \text{ bloques} + 0 \text{ bloques} + 0 \text{ bloques} + 132 \text{ bloques} + 1.003 \text{ bloques} + 0 \text{ bloques} + 0 \text{ bloques} + 28 \text{ bloques} + 1.348 \text{ bloques} + 0 \text{ bloques} + 0 \text{ bloques} = 2.756 \text{ bloques}$$

