

Certamen 3

Nicolás Gómez Morgado

Administración y programación de base de datos

24 de julio de 2024

1. Ejercicios guía 1

- Supongamos que estamos utilizando un sistema de disco donde el tiempo para mover la cabeza de lectura/escritura a un bloque de 15ms, y el tiempo de transferencia de un bloque es de 0.4ms. Supongamos que queremos calcular el join de R con S, y tenemos que $B(R)=1000$, $B(S)=500$ y $M=101$. Para acelerar el join, queremos leer y escribir tantos bloques como podamos en posiciones consecutivas del disco, y usar buffers que puedan ser múltiplos de un bloque. Responda las siguientes preguntas.

Tiempo en mover cabeza a un bloque: 15ms.

Tiempo transferencia de un bloque: 0.4ms

$R \bowtie S$

$B(R) = 1000$

$B(S) = 500$

$M = 101 \rightarrow 100$ (Se deja uno para la salida)

- ¿Cuántas E/S de disco se requieren para realizar esta operación join?

1 ^{er} pasada	Leer M bloques de R en MP[Memoria principal] (ordenar, escribir contenido ordenando)	$2B(R)$ crear sublistas ordenadas
2 ^{da} pasada	El atributo de ordenación y unión	$B(R)$ leer cada sublista
	Total	$3B(R)$

Que para R son $3B(R)$, por lo tanto para S, es lo mismo $3B(S)$.

Se requiere $B(R) + B(S)$.

Total:

Lo mismo para S $\rightarrow 3B(S)$ App. requerimientos $\lfloor \sqrt{B(R) + B(S)} \rfloor$

$$\begin{aligned}
 \text{Disco E/S} &= 3(B(R) + B(S)) \\
 &= 3(1000 + 500) \\
 &= 4500
 \end{aligned}$$

Por lo tanto son necesarias 4500 E/S de disco para realizar la union.

- b) ¿Cuanto tiempo toma un join basado en ordenamiento (sort-merge join), suponiendo que escribimos sublistas ordenadas en bloques consecutivos del disco?

Datos:

$$SB(R) = \text{Sublista de } R = \frac{B(R)}{M} = \frac{1000}{100} = 10$$

$$SB(S) = \text{Sublista de } S = \frac{B(S)}{M} = \frac{500}{100} = 5$$

$$tt = \text{Tiempo de transferencia} = 0.4ms.$$

$$M = 101 \rightarrow 100 \text{ (Se deja uno para la salida)}$$

$$tm = \text{Tiempo de mover la cabeza} = 15ms.$$

1 ^{er} pasada	R (Lectura y escritura) y S (Lectura y escritura): misma pista de lectura/escritura secuenciales	
	$tm + (SB(R) \cdot 2((M-1) \cdot tt) + tm + (SB(S) \cdot 2((M-1) \cdot tt)) + S(B(S))$ $[15 + (10) \cdot [2 \cdot (100 \cdot 0.4)]] + [15 + (5 \cdot [2 \cdot (100 \cdot 0.4)])]$ $[15 + (10) \cdot [80]] + [15 + (5 \cdot [80])]$ $[15 + 800] + [15 + 400]$ $815 + 415$ 1230 ms.	
2 ^{da} pasada	Original 15 buffers c/u con 1 bloque	Nueva: 15 buffers c/u con 6 bloques
	$M/\text{bloques} \cdot \text{buffers} \cdot tm$ $= 100 \text{ veces} \cdot 15 \text{ buffers} \cdot 15 \text{ ms}$ $= 25500 \text{ ms} = 22.5 \text{ seg}$	$M/\text{bloques} \cdot \text{buffers} \cdot tm$ $17 \text{ veces} \cdot 15 \text{ buffers} \cdot 15 \text{ ms}$ $\sim 3825 \text{ ms} \sim 3.825 \text{ seg}$

$$1 \text{ buffer} + 1 \text{ bloque} = 1230ms + 22500ms = 23730ms \approx 23,7seg.$$

$$1 \text{ buffer} + 6 \text{ bloques} = 1230ms + 3825ms = 5055ms \approx 5,05seg.$$

2. Supongamos que tenemos las relaciones $R(a, b)$, $S(b, c)$, $T(c, d)$, $U(d, e)$ con las siguientes características:

$$\begin{array}{ll}
 T(R) = 100 & T(T) = 100 \\
 V(R, b) = 100 & V(T, c) = 10 \\
 T(S) = 100 & V(T, d) = 100 \\
 V(S, b) = 100 & T(U) = 100 \\
 V(S, c) = 10 & V(U, d) = 100
 \end{array}$$

Computar un orden de Join de R sobre S Sobre T Sobre U $[R \bowtie S \bowtie T \bowtie U]$, utilizando:

- a) Dinámica (Dynamic Programming method)

R	S	T	U
100	100	100	100
0	0	0	0
R	S	T	U

Considerando los pares:

- 1) $R \bowtie S = 100$
- 2) $R \bowtie T = 10000$
- 3) $R \bowtie U = 10000$
- 4) $S \bowtie T = 1000$
- 5) $S \bowtie U = 10000$
- 6) $T \bowtie U = 100$

	R,S	R,T	R,U	S,T	S,U	T,U
Size	100	10000	10000	1000	10000	100
Cost.						
Best plan						

Ahora considerar la unión de 3 de las 4 relaciones. Elegir 2 para unir primero.

R,S,TR,S,UR,T,US,T,U

- 1) (R,S,T) =
 $R \bowtie S = 100 \leftarrow$
 $R \bowtie T = 10000$
 $S \bowtie T = 1000$

$$T((R \bowtie S) \bowtie T) = \frac{T(R \bowtie S) \cdot T(T)}{\max\{V(R \bowtie S, c), V(T, c)\}} = \frac{100 \cdot 100}{\max\{10, 10\}} \\ = \frac{10000}{10} = 1000$$

- 2) (R,S,U) =
 $R \bowtie S = 100 \leftarrow$
 $R \bowtie U = 10000$
 $S \bowtie U = 10000$

$$T((R \bowtie S) \bowtie U) = \frac{T(R \bowtie S) \cdot T(U)}{\max\{V(R \bowtie S, c), V(U, c)\}} = \frac{1000 \cdot 100}{\max\{10, 10\}} \\ = \frac{100000}{10} = 10000$$

- 3) (R,T,U) =
 $R \bowtie T = 10000$
 $R \bowtie U = 10000$
 $T \bowtie U = 100 \leftarrow$

$$T((R \bowtie T) \bowtie U) = \frac{T(R \bowtie T) \cdot T(U)}{\max\{V(R \bowtie T, d), V(U, d)\}} = \frac{10000 \cdot 100}{\max\{100, 100\}} \\ = \frac{1000000}{100} = 10000$$

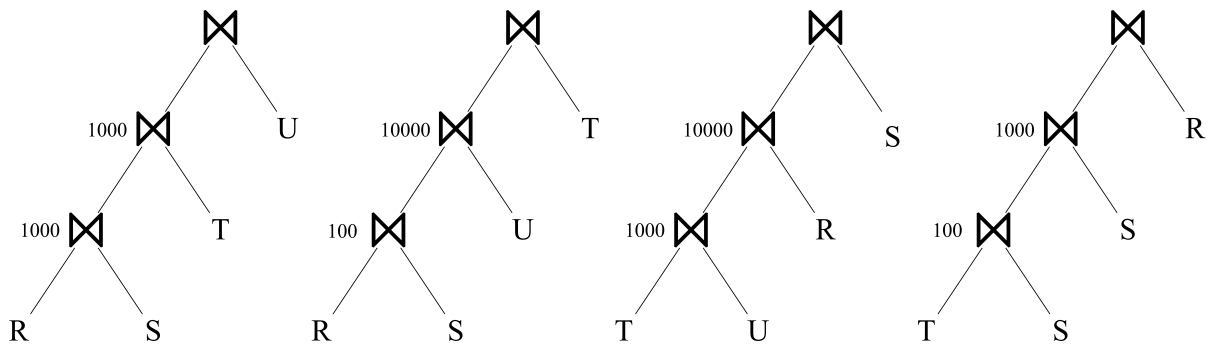
$$\begin{aligned}
 4) (S, T, U) = \\
 S \bowtie T &= 1000 \\
 S \bowtie U &= 10000 \\
 T \bowtie U &= 100 \leftarrow
 \end{aligned}$$

$$\begin{aligned}
 T((S \bowtie T) \bowtie U) &= \frac{T(S \bowtie T) \cdot T(U)}{\max\{V(S \bowtie T, d), V(U, d)\}} = \frac{1000 \cdot 100}{\max\{100, 100\}} \\
 &= \frac{100000}{100} = 1000
 \end{aligned}$$

	RST	RSU	RTU	STU
S	1000	10000	10000	1000
C	100	100	100	100
P	$(R \bowtie S) \bowtie T$	$(R \bowtie S) \bowtie U$	$(T \bowtie U) \bowtie R$	$(T \bowtie U) \bowtie S$

Tripletes.

Considerar arboles.



Agrupando:

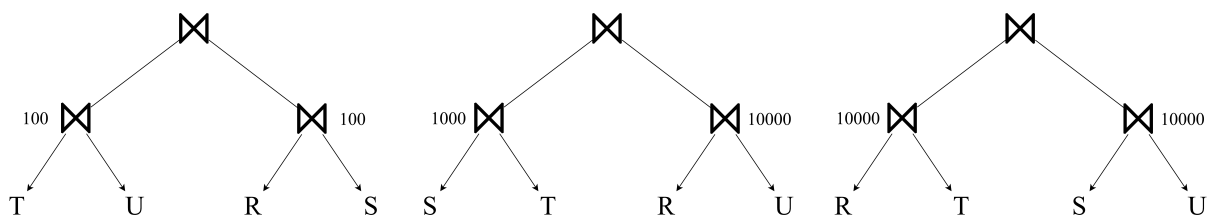
Costo mas tamaño de cada triplet

$$(((R \bowtie S) \bowtie T) \bowtie U) = 1000 + 100 = 1100$$

$$(((R \bowtie S) \bowtie U) \bowtie T) = 10000 + 100 = 10100$$

$$(((R \bowtie T) \bowtie U) \bowtie S) = 10000 + 100 = 10100$$

$$(((S \bowtie T) \bowtie U) \bowtie R) = 1000 + 100 = 1100$$



Agrupando:

$$(T \bowtie U) \bowtie (R \bowtie S) = 100 + 100 = 200 \leftarrow$$

$$(S \bowtie U) \bowtie (R \bowtie T) = 10000 + 1000 = 11000$$

$$(R \bowtie T) \bowtie (S \bowtie U) = 10000 + 10000 = 20000$$

b) Greedy method

Se toma una decision sin retroceder.

Base: Pares de relaciones cuyo tamaño estimado es el mas pequeño (árbol actual).

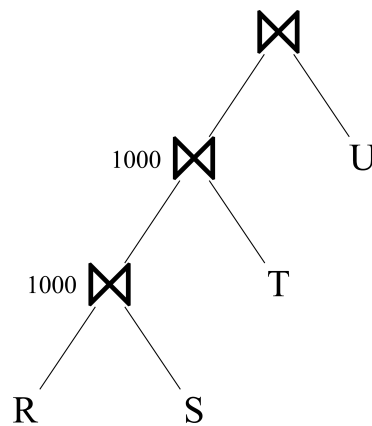
$$\begin{array}{ll} R, S = 100 & (1) \quad S, T = 10000 \\ R, T = 10000 & S, U = 10000 \\ R, U = 10000 & T, U = 100 \quad (2) \end{array}$$

Inducción: Encontrar todas las relaciones no incluidas, en este caso, T y U.

$$\begin{array}{l} R \bowtie S - ((R \bowtie S) \bowtie U) = 10000 \\ R \bowtie S - ((R \bowtie S) \bowtie T) = 1000 \end{array}$$

Se escoge T. Luego hay que unirse a U, no hay mas opciones.

$$(((R \bowtie S) \bowtie T) \bowtie U) = 1100$$

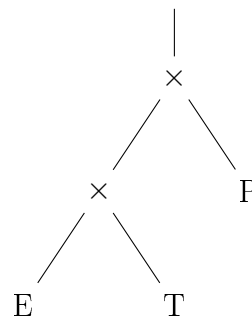


2. Ejercicios guía 2

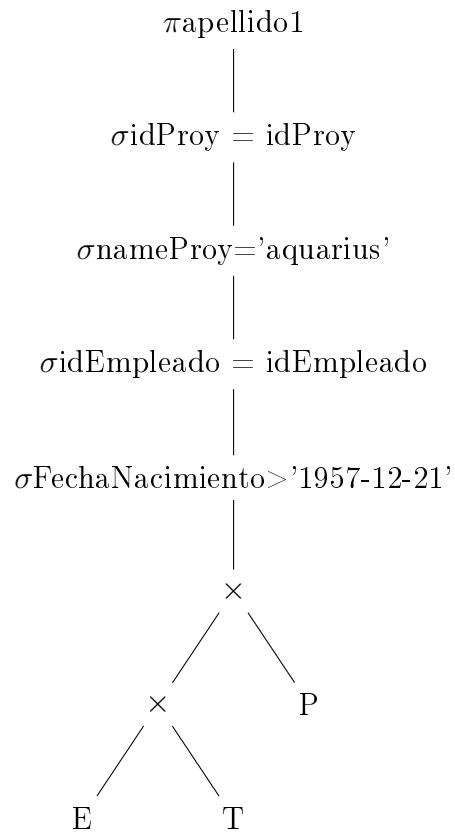
1. Traspasar la consulta SQL a un árbol de consulta.

- SELECT apellido1
- FROM Empleado E, trabajaEn T, proyecto P
- WHERE nameProy='aquarius'
- AND T.idProy = P.idProy
- AND E.idEmpleado = T.idEmpleado
- AND E.FechaNacimiento > '1957-12-21'

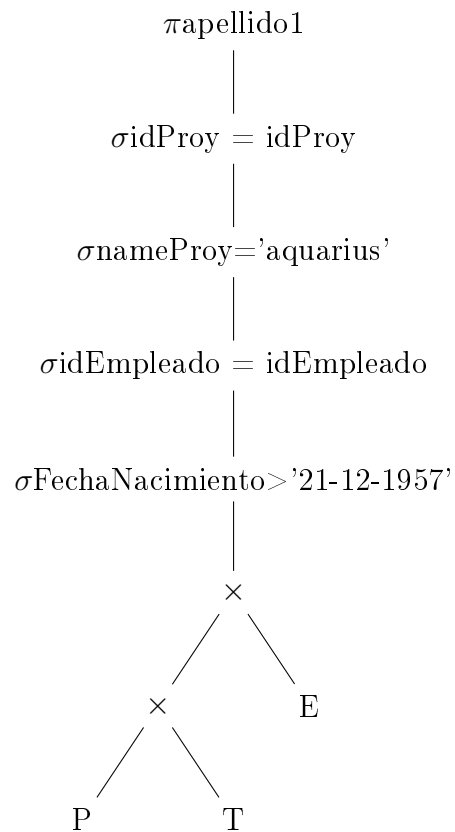
- Obtener el árbol inicial (canónico) de la consulta

$$\pi_{\text{apellido1}} \wedge \sigma_{\text{idProy} = \text{P.idProy} \wedge \text{nameProy} = \text{'aquarius'} \wedge \text{idEmpleado} = \text{T.idEmpleado} \wedge \text{E.FechaNacimiento} > \text{'1957-12-21'}}$$


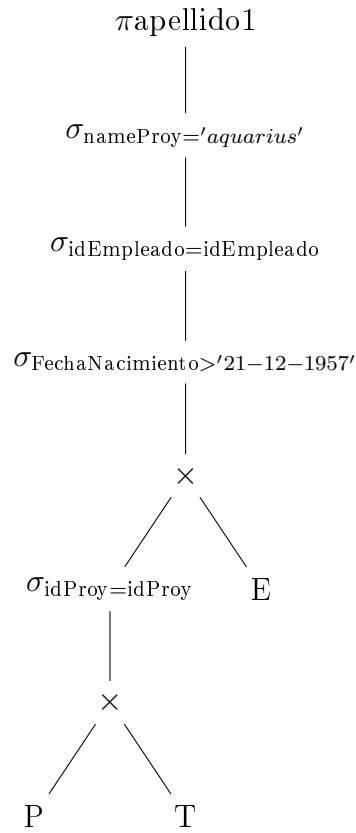
- Explique como se optimiza el árbol de consulta mediante la optimización vista en clases
 - a) Separar selección y proyección por tablas.



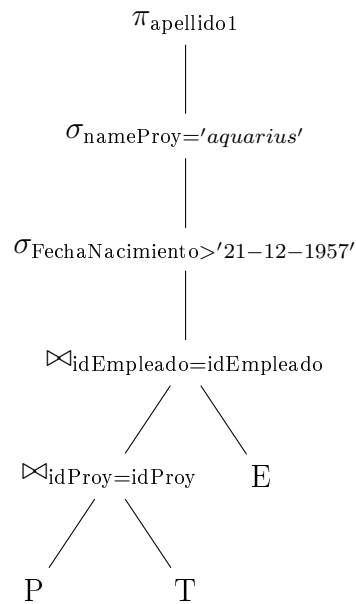
b) Reorganizar las tablas buscando la forma mas optima de unir las tablas.



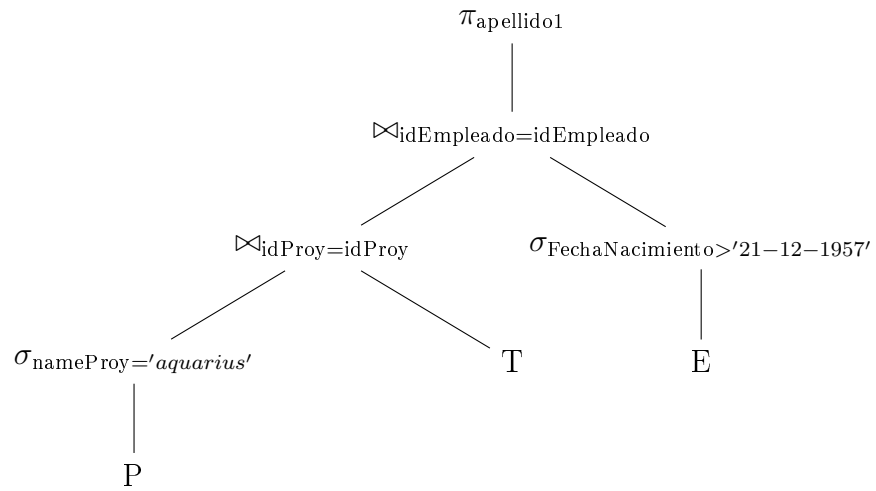
c) Se bajan las selecciones hasta su producto cartesiano.



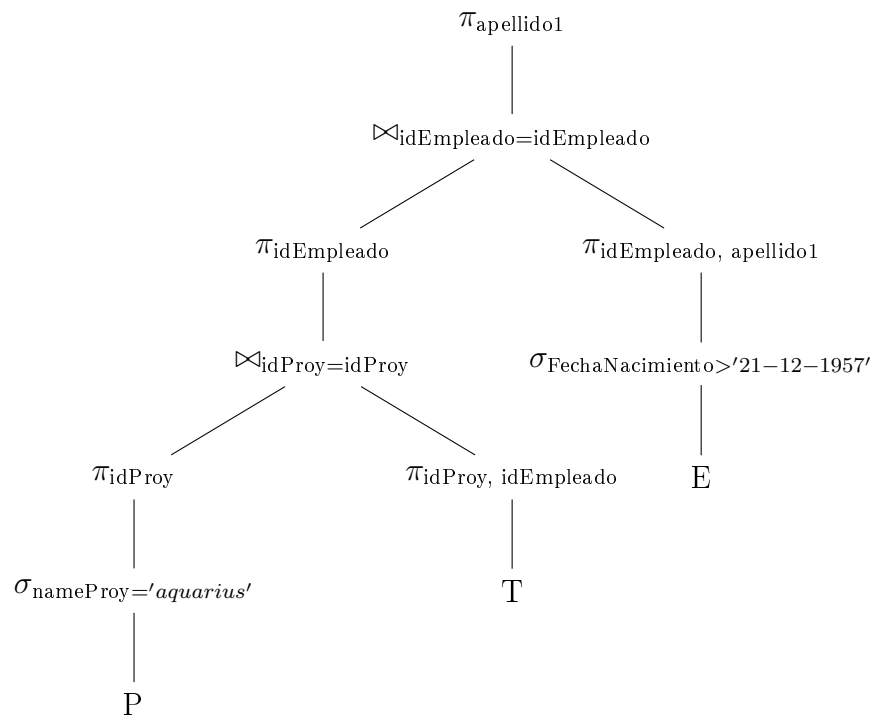
d) Cambia la selección y producto cartesiano por un join.



e) Se bajan las selecciones a sus respectivas tablas.



f) Se proyecta solo lo necesario en subtablas.



2. Considere las siguientes relaciones:

Variedades(IdVar, Nombre, Prog2, Prog1)

Predios(IdPredio, NombrePredio, Comuna, Superficie)

Siembra(IdPredio, IdVar, HaSem, Rdto, añoA)

Sea la siguiente consulta: “Listar los nombres de las variedades sembradas en el predio idPredio = 10 y que el año 2015 tuvieron un rendimiento mayor a 60 qq/ha”.

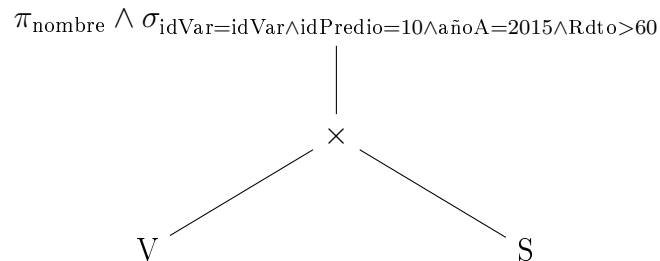
a) Escriba la consulta en SQL para la consulta anterior

- **SELECT** V.NOMBRE
- **FROM** VARIEDADES V, SIEMBRA S
- **WHERE** S.IDPREDIO = 10
AND S.AÑO A = 2015
AND S.RD TO > 60
AND S.IDVAR = V.IDVAR

b) Escriba la consulta en A.Relacional para la consulta anterior

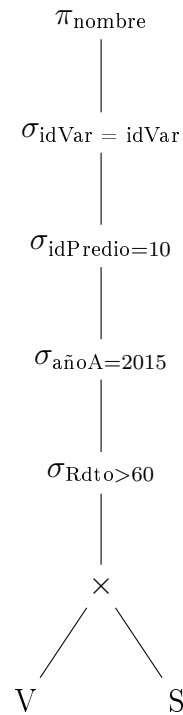
$$\pi_{v.nombre} \left(\sigma_{S.IdPredio=10 \wedge S.añoA=2015 \wedge S.Rdto>60} (VARIEDAD \bowtie SIEMBRA) \right)$$

c) Obtener el árbol inicial (canónico) de la consulta

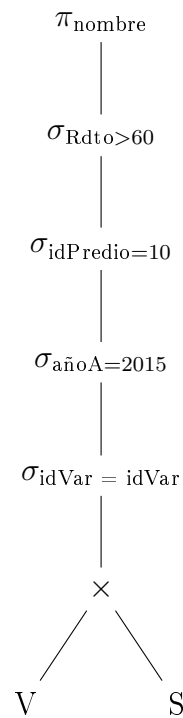


d) Explique como se optimiza el árbol de consulta mediante el algoritmo de optimización algebraica (visto en clases)

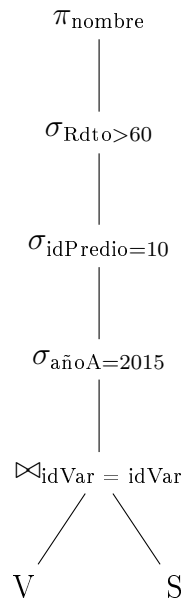
- 1) Separar selección y proyección por tablas.



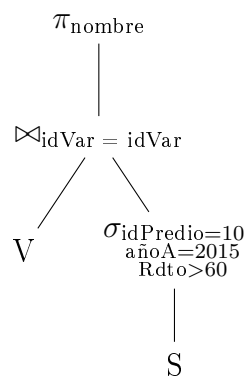
- 2) Reorganizar las tablas buscando la forma mas optima de unir las tablas.
Como solo son 2 tablas no es necesario reorganizar.
- 3) Se bajan las selecciones hasta su producto cartesiano.



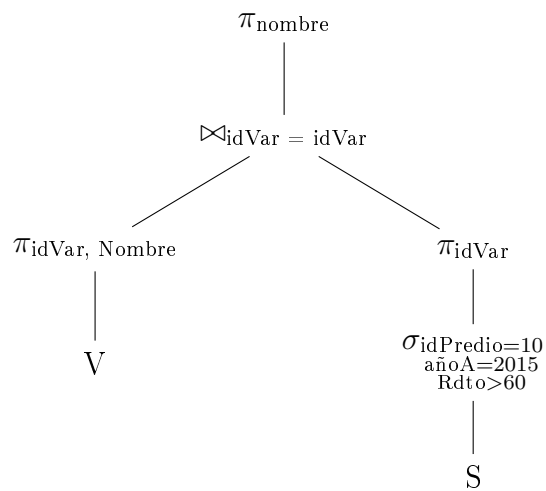
- 4) Se cambian las selecciones y producto cartesiano por un join.



5) Se baja la selección a su respectiva tabla.



6) Se proyectan solo lo necesario en subtablas.



3. Suponer que tenemos las relaciones $R(a, b)$, $S(b, c)$, $T(c, d)$, y $U(d, e)$ con las siguientes características:

$$T(R) = 300$$

$$V(R, b) = 100$$

$$T(S) = 200$$

$$V(S, b) = 100$$

$$T(T) = 150$$

$$T(U) = 500$$

$$V(S, c) = 20$$

$$V(T, c) = 20$$

$$V(T, d) = 300$$

$$V(U, d) = 300$$

Costos simples:

	R	S	T	U
Size	300	200	150	500
Cost.	0	0	0	0
Best plan	R	S	T	U

Cálculos simples:

- $T(R) = 300$
- $T(S) = 200$
- $T(T) = 150$
- $T(U) = 500$

Costos pares:

	R,S	R,T	R,U	S,T	S,U	T,U
Size	600	45000	150000	1500	100000	250
Cost.	0	0	0	0	0	0
Best plan	$R \bowtie S$	$R \bowtie T$	$R \bowtie U$	$S \bowtie T$	$S \bowtie U$	$T \bowtie U$

Cálculos pares:

$$\begin{aligned}
T(R \bowtie S) &= \frac{T(R) \cdot T(S)}{\max\{V(R, b), V(S, b)\}} = \frac{300 \cdot 200}{\max\{100, 100\}} \\
&= \frac{60000}{100} = 600 \\
T(R \bowtie T) &= \frac{T(R) \cdot T(T)}{\max\{V(R, b), V(T, c)\}} = \frac{300 \cdot 150}{\max\{0, 0\}} \\
&= 45000 \\
T(R \bowtie U) &= \frac{T(R) \cdot T(U)}{\max\{V(R, b), V(U, d)\}} = \frac{300 \cdot 500}{\max\{0, 0\}} \\
&= 150000 \\
T(S \bowtie T) &= \frac{T(S) \cdot T(T)}{\max\{V(S, c), V(T, c)\}} = \frac{200 \cdot 150}{\max\{20, 20\}} \\
&= \frac{30000}{20} = 1500 \\
T(S \bowtie U) &= \frac{T(S) \cdot T(U)}{\max\{V(S, c), V(U, d)\}} = \frac{200 \cdot 500}{\max\{0, 0\}} \\
&= 100000 \\
T(T \bowtie U) &= \frac{T(T) \cdot T(U)}{\max\{V(T, d), V(U, d)\}} = \frac{150 \cdot 500}{\max\{300, 300\}} \\
&= \frac{75000}{300} = 250
\end{aligned}$$

Costos triples:

	R,S,T	R,S,U	R,T,U	S,T,U
Size	4500	150000	1500	100000
Cost.	600	600	250	250
Best plan	$R \bowtie S \bowtie T$	$R \bowtie S \bowtie U$	$R \bowtie T \bowtie U$	$S \bowtie T \bowtie U$

Cálculos triples:

$$\begin{aligned}
1) (R, S, T) &= \\
R \bowtie S &= 600 \leftarrow \\
R \bowtie T &= 45000 \\
S \bowtie T &= 15000
\end{aligned}$$

$$\begin{aligned}
T((R \bowtie S) \bowtie T) &= \frac{T(R \bowtie S) \cdot T(T)}{\max\{V((R \bowtie S), c), V(T, c)\}} = \frac{600 \cdot 150}{\max\{\max[0, 20], 20\}} \\
&= \frac{180000}{20} = 4500
\end{aligned}$$

$$\begin{aligned}
2) (R, S, U) &= \\
R \bowtie S &= 600 \leftarrow \\
R \bowtie U &= 150000 \\
S \bowtie U &= 100000
\end{aligned}$$

$$T((R \bowtie S) \bowtie U) = \frac{T(R \bowtie S) \cdot T(U)}{\max\{V(R \bowtie S, c), V(U, c)\}} = \frac{600 \cdot 500}{\max\{0, 0\}} \\ = 300000$$

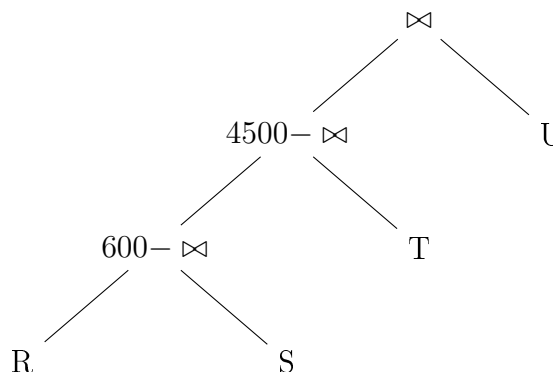
$$\begin{aligned} 3) (R, T, U) = \\ R \bowtie T &= 45000 \\ R \bowtie U &= 150000 \\ T \bowtie U &= 250 \leftarrow \end{aligned}$$

$$T((T \bowtie U) \bowtie R) = \frac{T(T \bowtie U) \cdot T(R)}{\max\{V(T \bowtie U, b), V(R, b)\}} = \frac{250 \cdot 300}{\max\{0, 0\}} \\ = 75000$$

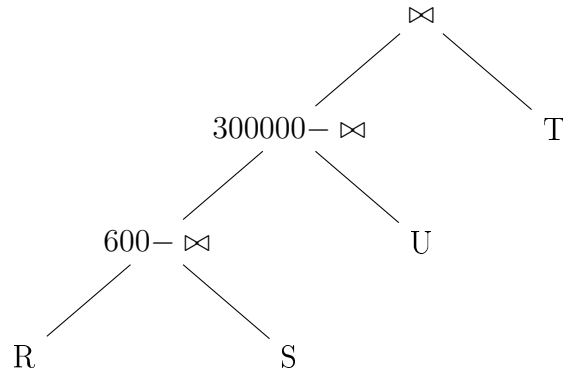
$$\begin{aligned} 4) (S, T, U) = \\ S \bowtie T &= 1500 \\ S \bowtie U &= 100000 \\ T \bowtie U &= 250 \leftarrow \end{aligned}$$

$$T((T \bowtie U) \bowtie S) = \frac{T(T \bowtie U) \cdot T(S)}{\max\{V(T \bowtie U, c), V(S, c)\}} = \frac{250 \cdot 200}{\max\{\max[20, 0], 20\}} \\ = \frac{50000}{20} = 2500$$

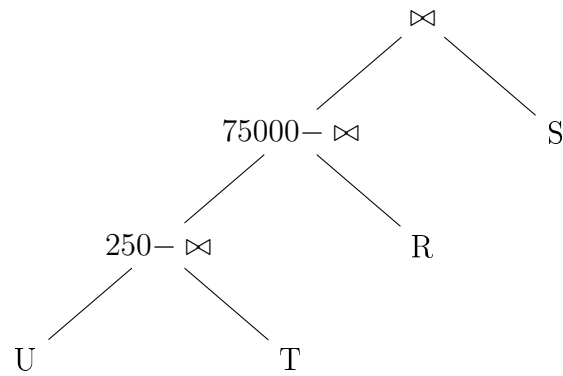
Árboles tripletas:
RST:



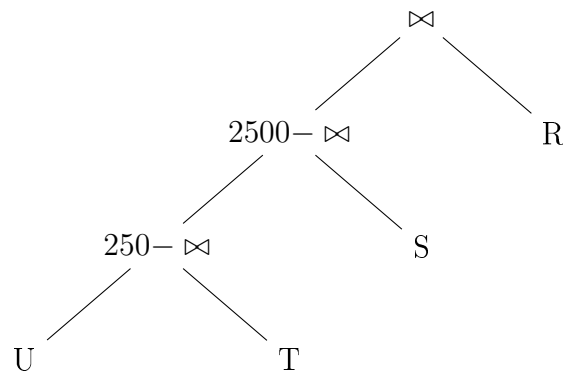
RSU:



RTU:



STU:



Agrupando(Costo mas tamaño de cada tripleta):

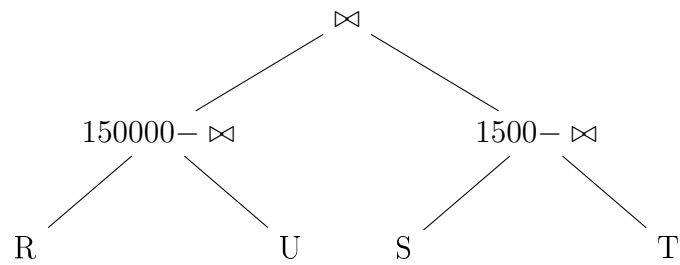
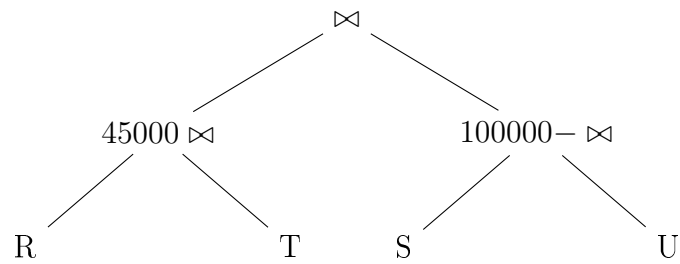
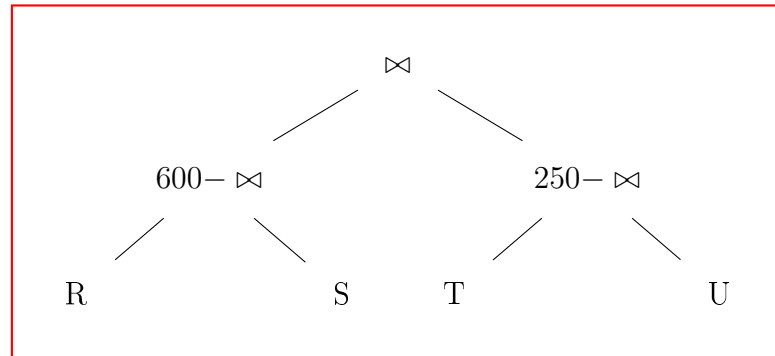
$$(((R \bowtie S) \bowtie T) \bowtie U) = 4500 + 600 = 5100$$

$$(((R \bowtie S) \bowtie U) \bowtie T) = 300000 + 600 = 300600$$

$$(((R \bowtie T) \bowtie U) \bowtie S) = 75000 + 250 = 75250$$

$$(((S \bowtie T) \bowtie U) \bowtie R) = 2500 + 250 = 2750$$

Árboles equilibrados:



Agrupando:

$$(R \bowtie S) \bowtie (T \bowtie U) = 600 + 250 = 850$$

$$(R \bowtie T) \bowtie (S \bowtie U) = 45000 + 100000 = 145000$$

$$(R \bowtie U) \bowtie (S \bowtie T) = 150000 + 1500 = 151500$$

Conclusion: Por lo tanto el árbol mas optimo es el de la primera agrupación, ya que es el que tiene el menor costo.

4. Suponer que tenemos las relaciones $R(a, b)$, $S(b, c)$, $T(c, d)$, y $U(d, e)$ con las siguientes características:

$$T(R) = 50$$

$$V(R, b) = 250$$

$$T(S) = 55$$

$$V(S, b) = 500$$

$$V(S, c) = 5$$

$$V(U, d) = 50$$

$$T(T) = 50$$

$$T(U) = 45$$

$$V(T, c) = 15$$

$$V(T, d) = 500$$

a) Según el método Greedy:

Pares de relaciones:

$$\blacksquare R \bowtie S = \frac{50 \cdot 55}{\max\{250, 500\}} = 5.5$$

$$\blacksquare R \bowtie T = \frac{50 \cdot 50}{\max\{0, 0\}} = 2500$$

$$\blacksquare R \bowtie U = \frac{50 \cdot 45}{\max\{0, 0\}} = 2250$$

$$\blacksquare S \bowtie T = \frac{55 \cdot 50}{\max\{5, 15\}} = 183$$

$$\blacksquare S \bowtie U = \frac{55 \cdot 45}{\max\{0, 0\}} = 2475$$

$$\blacksquare T \bowtie U = \frac{50 \cdot 45}{\max\{500, 50\}} = 4.5$$

Se escoge el menor:

$$((T \bowtie U) \bowtie R)$$

$$T \bowtie U = 4.5 \leftarrow$$

$$T \bowtie R = 2500$$

$$U \bowtie R = 2250$$

$$T((T \bowtie U) \bowtie R) = \frac{4.5 \cdot 50}{\max\{0, 0\}} = 225$$

$$((T \bowtie U) \bowtie S)$$

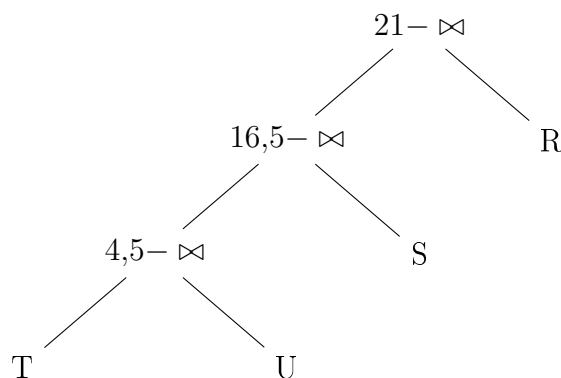
$$T \bowtie U = 4.5 \leftarrow$$

$$T \bowtie S = 183$$

$$U \bowtie S = 2475$$

$$T((T \bowtie U) \bowtie S) = \frac{4.5 \cdot 55}{\max\{\max[0, 15], 5\}} = 16.5 \leftarrow$$

Se agrega el que falta: $((T \bowtie U) \bowtie S) \bowtie R = 16.5 + 4.5 = 21$





Para 3 y 4, calcular un orden/JOINS para R, S, T, U, usando Programación dinámica (como visto en clases). Mostrar tabla inicial de costos, los cálculos de cada etapa y árboles.