

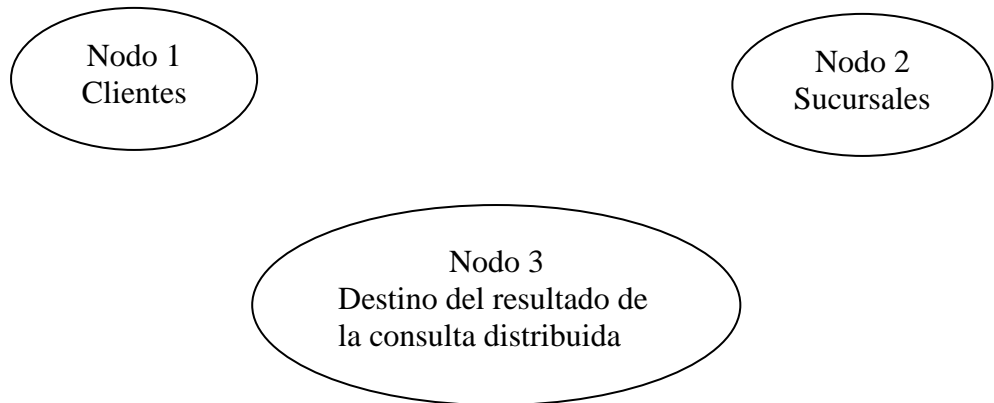
Bases de Datos II
Resoluciones Serie Ejercicios Prácticos 2
Bases de Datos Distribuidas

A continuación, detallamos los pasos a seguir para resolver estos ejercicios, agregando los comentarios necesarios para comprender lo que se realiza en cada ítem, sólo para el primer ejercicio.

Objetivo: es la reducción de la cantidad de datos transferidos a través de una red de computadoras.

1)

Nuestra red de bases de datos distribuidas está compuesta por tres nodos (sedes o sitios): Nodo 1, Nodo 2 y Nodo 3, en el Nodo 1 reside la tabla de Clientes, en el Nodo 2 se almacena la tabla de Sucursales y el Nodo 3 oficia como el sitio donde se va a enviar los resultados de la consulta solicitada, e iremos evaluando cuales son los costos de transferencia de los datos de procesamiento a través de esta red.



Para los nodos 1 y 2, se proporciona la cantidad de tuplas que contienen y cuál es su longitud en bytes.

- a) Para determinar el tamaño de la relación de cada nodo, debemos de multiplicar la cantidad de tuplas por la longitud de la tupla, únicamente para los Nodos 1 y 2, ya que el Nodo 3 oficia como sede donde se destinará el resultado de la consulta solicitada, por lo tanto, tenemos:

El tamaño de la relación de Clientes (Nodo 1): 5000 tuplas * 123 bytes = **615000 bytes**

El tamaño de la relación de Sucursales (Nodo 2): 100 tuplas * 71 bytes = **7100 bytes**

- b) La consulta requerida para este caso práctico, consiste en obtener para *cada cliente*, su *Apellido y Nombres* y el *Nombre de la Sucursal en la cual adquiere los libros*.

Vemos que la columna que contiene el Apellido y Nombres del cliente se halla en el Nodo 1 y la columna que contiene el Nombre de la sucursal se halla en el Nodo 2, por lo que la sentencia SQL involucrada en este caso, quedaría representada como la consulta resultante de la unión de ambas tablas:

```
Select Apellido_y_Nombre, Nombre_Sucursal
      from Clientes inner join Sucursales
      on Numero_Sucursal = Nro_Sucursal
```

En primer lugar, evaluaremos el costo de la transferencia de los datos resultantes al Nodo 3 y en segundo lugar optamos por considerar al Nodo 2 como la sede donde se desplegarán los resultados (I-II), y por último consideraremos la alternativa de semi unión con el Nodo 2 como el de resultado (III).

I. Nodo 3 como nodo resultado

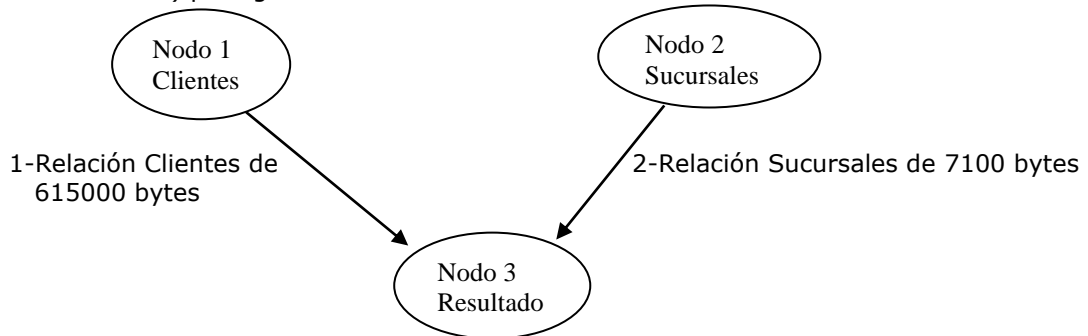
La consulta solicitada es enviada (viaja) al Nodo 3, denominado nodo resultado (o respuesta), es el lugar en el que se precisan los resultados, no tiene datos implicados en la consulta. Ni la relación Clientes ni Sucursales residen en este nodo.

Tenemos que el resultado de esta consulta constará de 5000 tuplas, que es la totalidad de los clientes existentes, y cada tupla resultante para la consulta solicitada será de una longitud de 60 bytes, ya que por cada cliente se selecciona su Apellido y Nombres de una longitud de 30 bytes y el nombre de la sucursal con longitud de 30 bytes, de allí que el **tamaño del resultado de la consulta** será: (30 bytes + 30 bytes) * 5000 tuplas o 60 bytes * 5000 tuplas = **300000 bytes**

Para ejecutar esta consulta distribuida con el Nodo 3 como el nodo resultante, existen tres tipos de estrategias u opciones a evaluar:

Opción 1:

Transferimos tanto la relación Clientes como la relación Sucursales al nodo resultado (Nodo 3) y realizar allí mismo la operación de **join** (reunión, concatenación de ambas tablas) para generar la consulta.

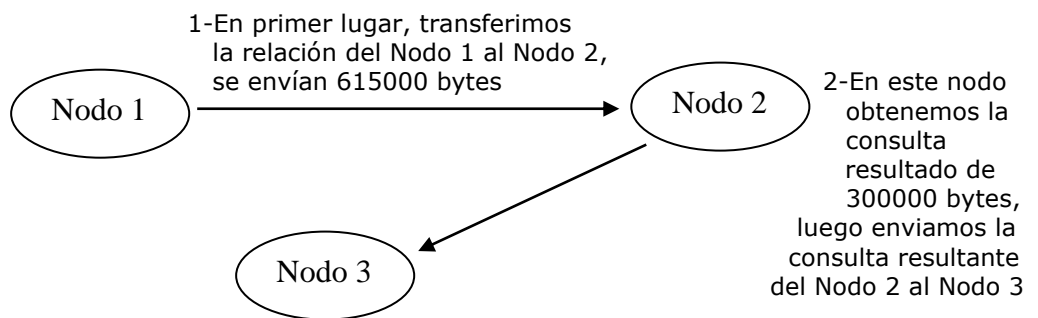


En este caso deben transferirse: **615000 bytes + 7100 bytes = 622100 bytes transferidos**

En esta opción sumamos los tamaños de las relaciones del Nodo 1 y Nodo 2.

Opción 2:

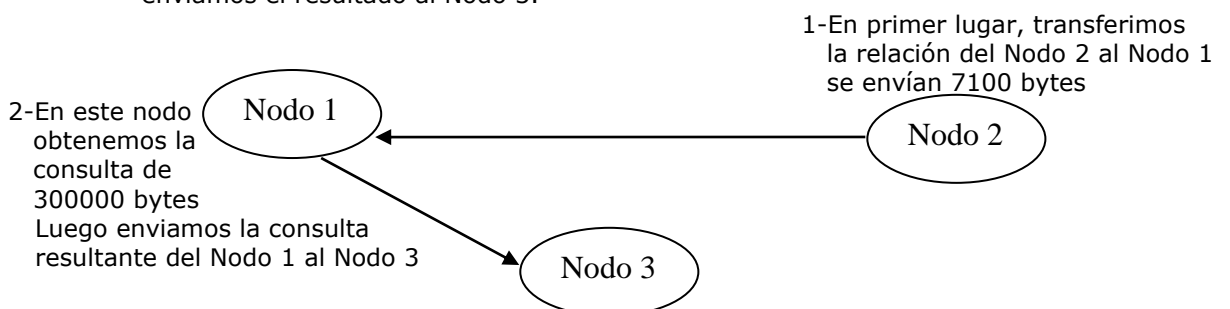
Transferimos la relación Clientes al Nodo 2, ejecutamos el join en este nodo y luego enviamos el resultado al Nodo 3.



Esto implicaría transferir **615000 bytes** de Clientes, **más 300000 bytes** del resultado de la consulta, es decir: **915000 bytes transferidos**

Opción 3:

Transferimos la relación Sucursales al Nodo 1, ejecutamos el join en este nodo y luego enviamos el resultado al Nodo 3.



En este caso, los bytes transferidos de la relación Sucursales serán **7100 bytes** más **300000 bytes** del resultado de la consulta, resultado **307100 bytes transferidos**.

De las tres opciones evaluadas, la Opción 3 es la estrategia más adecuada.

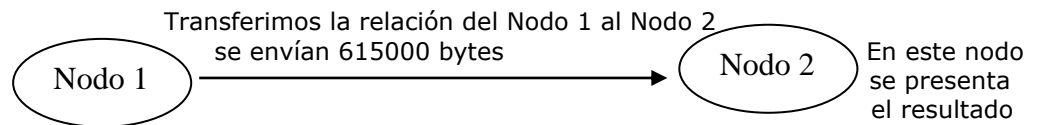
Como se observa, elegimos la tercera opción, ya que implica menos transferencia de datos por la red, siendo nuestro criterio de optimización aquel que minimiza la cantidad de transferencia de datos.

II. Nodo 2 como nodo resultado

Si ahora consideramos al **Nodo 2, como nodo resultado**, tendríamos en este caso dos opciones a evaluar:

Opción 1:

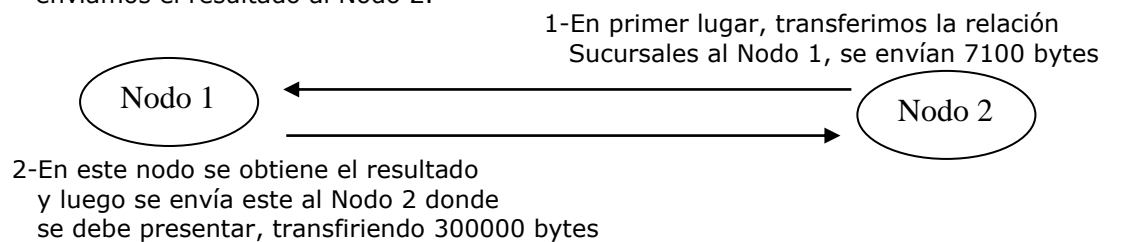
Transferimos la relación Clientes al nodo resultado (Nodo 2), realizar allí mismo la operación de join y presentamos el resultado al usuario en el Nodo 2.



En este caso deben transferirse: **5000 tuplas * 123 bytes = 615000 bytes transferidos** corresponde al tamaño de la relación Clientes.

Opción 2:

Transferimos la relación Sucursales al Nodo 1, ejecutamos el join en este nodo y luego enviamos el resultado al Nodo 2.



Esto implicaría transferir **7100 bytes** de la relación Sucursales al Nodo 1 más **300000 bytes** del resultado de la consulta del Nodo 1 al Nodo 2, dando **307100 bytes transferidos**.

De las dos opciones evaluadas, la Opción 2 es la estrategia más adecuada.

Como vimos, la segunda estrategia es la más óptima, en cuanto a la reducción en la cantidad de datos transferidos.

III. Consulta distribuida usando el tipo de semi reunión (semi join) respecto al ítem II

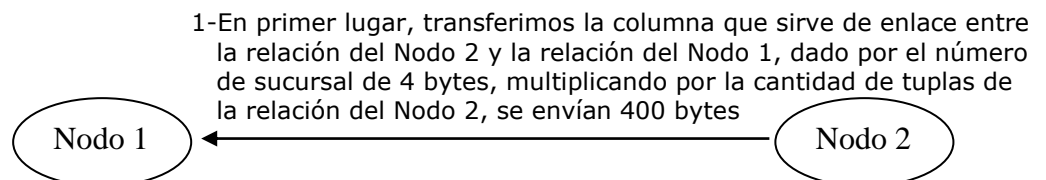
Ahora procesaremos la consulta distribuida de la estrategia anterior evaluada, usando el tipo de semi reunión (**semi join**), solicitada en el Nodo 2 (nodo resultado).

Con el fin de reducir el número de tuplas de una relación antes de ser transferidas a otro nodo, se envían únicamente las columnas (atributos) estrictamente necesarios para ejecutar la consulta.

Paso 1:

Para nuestra consulta, proyectamos los atributos de concatenación de la relación Sucursales al Nodo 1, que luego van a intervenir en la operación de join en el Nodo 2.

En el Nodo 2, la columna que sirve de enlace entre ambas relaciones es el atributo Numero de Sucursal, cuyo tamaño es de **4 bytes** por el número de **100** tuplas de Sucursales.

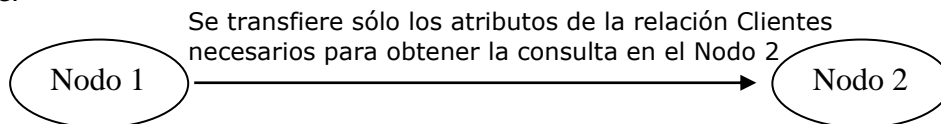


Numero de sucursal (longitud) = 4 bytes * 100 tuplas = **400 bytes transferidos**

Paso 2:

Realizamos el join de las columnas transferidas en el paso anterior con la relación Clientes en el Nodo 1, y luego devolvemos únicamente los atributos de reunión requeridos para luego obtener el resultado mediante el join final en el Nodo 2, en este caso, además del atributo de enlace debemos de transferir la columna de Apellido y Nombres de cada cliente del Nodo 1.

Tenemos que considerar a la columna Numero Sucursal de **4** bytes y a la columna de Apellido y Nombres de **30** bytes de la relación Clientes, por el número de **5000** tuplas de esta sede.



Numero de sucursal + Apellido y Nombres = (4 bytes + 30 bytes) * 5000 tuplas = 34 bytes * 5000 tuplas = **170000 bytes transferidos**

En este caso los atributos fueron proyectados y enviados desde el Nodo 1 al nodo original (Nodo 2).

Limitamos los atributos de Clientes transmitidos al Nodo 2 a sólo aquéllos que serán concatenados con las tuplas de Sucursales, para nuestra consulta, en esta reunión participan todas las tuplas de Clientes, con lo que se obtiene una sensible mejora.

Por último, para obtener el total transferido en ambos pasos, tenemos que:

400 bytes + 170000 bytes = 170400 bytes transferidos

Por lo tanto, la solución obtenida en el punto III es aún más óptima que la evaluada en ítem II.

Nos podemos preguntar, de que sirve enviar la columna de enlace de todas las tuplas desde el Nodo 2 al Nodo 1, así como lo planteamos anteriormente, si luego desde el nodo donde reside la relación Clientes, devolvemos para cada una de sus tuplas, esta columna además de la del Apellido y Nombres, por ello es que esta alternativa se visualiza de mejor manera, cuando por ejemplo en nuestra consulta distribuida, requerimos procesar un rango de sucursales, por ejemplo las comprendidas entre el rango 1521 al 2315, entonces sólo remitiremos este rango de sucursales desde el Nodo 2 al Nodo 1, y allí seleccionamos los apellidos y nombres de todos aquellos clientes que se encuentren dentro del rango de sucursales que fueron previamente enviadas desde la relación de Sucursales.

2)

- a) Tamaño de la relación Profesores (Nodo 1): 7000 tuplas * 88 bytes = **616000 bytes**
 Tamaño de la relación de Facultades (Nodo 2): 200 tuplas * 77 bytes = **15400 bytes**

b)

Resultado de la consulta requerida:

(Apellido y Nombres del profesor + Denominación de la facultad) * tuplas relación Profesores (30 bytes + 32 bytes) * 7000 tuplas = **434000 bytes**

I. Nodo 3 como nodo resultado

Se pueden presentar tres opciones:

Opción 1

Transferimos el Nodo 1 y el Nodo 2 al Nodo 3:

616000 bytes + 15400 bytes = **631400 bytes transferidos**

Opción 2

Se transfiere el Nodo 1 al Nodo 2 y del Nodo 2 transferimos el resultado al Nodo 3:

616000 bytes + 434000 bytes = **1050000 bytes transferidos**

Opción 3

Se transfiere el Nodo 2 al Nodo 1 y del Nodo 1 transferimos el resultado al Nodo 3:

15400 bytes + 434000 bytes = **449400 bytes transferidos**

De las opciones consideradas, la **Opción 3** es la mejor.

II. Nodo 2 como nodo resultado

Se pueden presentar dos opciones:

Opción 1

Se transfiere el Nodo 1 al Nodo 2: **616000 bytes transferidos**

Opción 2

Se transfiere el Nodo 2 al Nodo 1 y del Nodo 1 transferimos el resultado al Nodo 2:

15400 bytes + 434000 bytes = **449400 bytes transferidos**

De las opciones consideradas, la **Opción 2** es la mejor.

III. Consulta distribuida usando el tipo de semi reunión (semi join) respecto al ítem II

Paso 1:

Proyección de la columna de concatenación de la relación del Nodo 2 al Nodo 1, siendo:
Código de facultad (longitud) = 2 bytes * 200 tuplas = **400 bytes transferidos**

Paso 2:

Proyección de las columnas necesarias para la obtención de la consulta distribuida, desde la relación del Nodo 1 al Nodo 2: Código de facultad + Apellido y Nombres del profesor.
(2 bytes + 30 bytes) * 7000 tuplas = **224000 bytes transferidos**

El total transferido resulta: **400 bytes + 224000 bytes = 224400 bytes transferidos.**

Por lo tanto, la solución obtenida en el punto III es aún más óptima que la evaluada en ítem II.

3)

- a) Tamaño de la relación Remedios (Nodo 1): 100000 tuplas * 81 bytes = **8100000 bytes**
Tamaño de la relación de Laboratorios (Nodo 2): 63800 tuplas * 177 bytes = **11292600 bytes**
b)

Resultado de la consulta requerida:

(Nombre comercial del remedio + Nombre del laboratorio) * tuplas de la relación Remedios
(20 bytes + 40 bytes) * 100000 tuplas = **6000000 bytes**

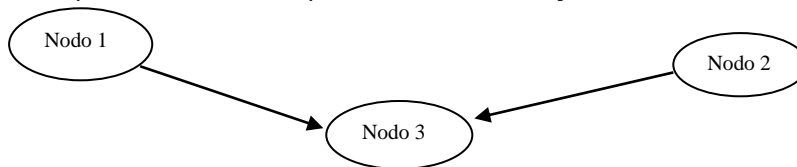
I. Nodo 3 como nodo resultado

Se pueden presentar tres opciones:

Opción 1

Transferimos el Nodo 1 y el Nodo 2 al Nodo 3:

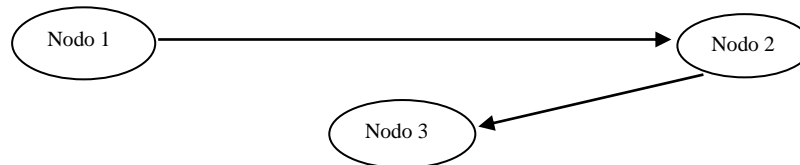
8100000 bytes + 11292600 bytes = **19392600 bytes transferidos**



Opción 2

Se transfiere el Nodo 1 al Nodo 2 y del Nodo 2 transferimos el resultado al Nodo 3:

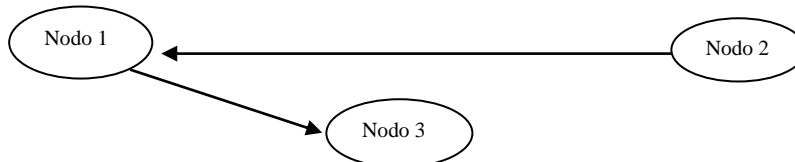
8100000 bytes + 6000000 bytes = **14100000 bytes transferidos**



Opción 3

Se transfiere el Nodo 2 al Nodo 1 y del Nodo 1 transferimos el resultado al Nodo 3:

11292600 bytes + 6000000 bytes = **17292600 bytes transferidos**



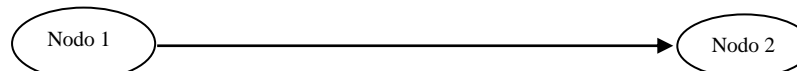
De las opciones consideradas, la **Opción 2** es la mejor.

II. Nodo 2 como nodo resultado

Se pueden presentar dos opciones:

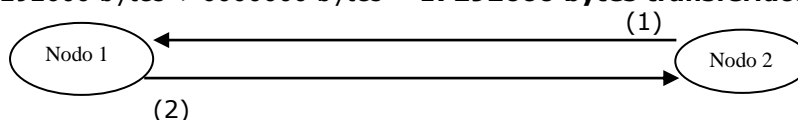
Opción 1

Se transfiere el Nodo 1 al Nodo 2: **8100000 bytes transferidos**



Opción 2

Se transfiere el Nodo 2 al Nodo 1 (1) y del Nodo 1 transferimos el resultado al Nodo 2 (2):
11292600 bytes + 6000000 bytes = **17292600 bytes transferidos**



De las opciones consideradas, la **Opción 1** es la mejor.

4)

- a) Tamaño de la relación Empleados (Nodo 1): 3200 tuplas * 56 bytes = **179200 bytes**
Tamaño de la relación de Departamentos (Nodo 2): 1500 tuplas * 130 bytes = **195000 bytes**

b) **Resultado de la consulta requerida:**

(Nombre y Apellido del empleado + Nombre del departamento) * tuplas relación Empleados
(25 bytes + 40 bytes) * 3200 tuplas = **208000 bytes**

I. Nodo 3 como nodo resultado

Se pueden presentar tres opciones:

Opción 1

Transferimos el Nodo 1 y el Nodo 2 al Nodo 3:

179200 bytes + 195000 bytes = **374200 bytes transferidos**

Opción 2

Se proyecta el Nodo 1 al Nodo 2 y del Nodo 2 transferimos el resultado al Nodo 3:

179200 bytes + 208000 bytes = **387200 bytes transferidos**

Opción 3

Se transfiere el Nodo 2 al Nodo 1 y del Nodo 1 proyectamos el resultado al Nodo 3:

195000 bytes + 208000 bytes = **403000 bytes transferidos**

De las opciones consideradas, la **Opción 1** es la mejor.

II. Nodo 2 como nodo resultado

Se pueden presentar dos opciones:

Opción 1

Se transfiere el Nodo 1 al Nodo 2: **179200 bytes transferidos**

Opción 2

Se transfiere el Nodo 2 al Nodo 1 y del Nodo 1 transferimos el resultado al Nodo 2:

195000 bytes + 208000 bytes = **403000 bytes transferidos**

De las opciones evaluadas, la **Opción 1** es la mejor.

III. Consulta distribuida usando el tipo de semi reunión (semi join) respecto al ítem II

Paso 1:

Proyección de la columna de concatenación de la relación del Nodo 2 al Nodo 1, siendo:

Código de departamento (longitud) = 3 bytes * 1500 tuplas = **4500 bytes transferidos**

Paso 2:

Proyección de las columnas necesarias para la obtención de la consulta distribuida, desde la relación del Nodo 1 al Nodo 2: Código de departamento + Nombre y Apellido del empleado.

(3 bytes + 25 bytes) * 3200 tuplas = **89600 bytes transferidos**

El total transferido resulta: **4500 bytes + 89600 bytes = 94100 bytes transferidos.**

Por lo tanto, la solución obtenida en el punto III es aún más óptima que la evaluada en ítem II.

Nota: en la solución planteada del ejercicio 1, se agregaron los comentarios y gráficos necesarios para entender la manera de resolverlo, la forma en que se proyectan los datos entre las diferentes sedes y las distintas alternativas que tenemos para evaluar cuál de ellas es la más óptima, aquella que minimice el transporte de bytes, entre los distintos nodos distribuidos.

En los demás ejercicios, solamente se especificaron todas las operaciones necesarias para calcular cada una de las opciones que se pueden plantear de acuerdo a las consignas solicitadas, y esta es la forma que aceptamos cuando los resuelvan en los exámenes parciales, a los cuáles se pueden agregar de manera optativa la representación de los gráficos que plasme la forma en que se proyectan los datos entre los sitios involucrados, como en el ejercicio 3).

- 5)** Pendiente para que agreguen su resolución.