

19.4.2. Compromiso de tres fases

El protocolo de compromiso de tres fases (C3F) es una extensión del protocolo de compromiso de dos fases que evita el problema del bloqueo con determinadas suposiciones. En concreto, se supone que no se produce ninguna fragmentación de la red y que no fallan más de k sitios, donde k es un número predeterminado. Con estas suposiciones el protocolo evita el bloqueo introduciendo una tercera fase adicional en que se implican varios sitios en la decisión de comprometer. En lugar de anotar directamente la decisión de comprometer en su almacenamiento persistente, el coordinador se asegura antes de que al menos otros k sitios sepan que pretende comprometer la transacción. Si el coordinador falla, los sitios restantes seleccionan primero un nuevo coordinador. Este nuevo coordinador verifica el estado del protocolo a partir de los demás sitios; si el coordinador había decidido comprometer, al menos uno de los otros k sitios a los que informó estará funcionando y asegurará que se respete la decisión de comprometer. El nuevo coordinador vuelve a iniciar la tercera fase del protocolo si algún sitio sabía que el antiguo coordinador pretendía comprometer la transacción. En caso contrario, el nuevo coordinador aborta la transacción.

Aunque el protocolo C3F tiene la propiedad deseable de no bloquearse a menos que fallen k sitios, tiene el inconveniente de que una división de la red parece lo mismo que el fallo de más de k sitios, lo que puede producir un bloqueo. El protocolo también tiene que implementarse con mucho cuidado para asegurarse de que la división de la red (o el fallo de más de k sitios) no provoque inconsistencias, de modo que una transacción se comprometa en una de las particiones y se aborte en otra. Debido a la sobrecarga que implica el protocolo C3F, no se utiliza mucho. Véanse las notas bibliográficas para hallar referencias que den más detalles del protocolo C3F.

Par1a2016sem2

Tema 6: Determine si las siguientes planificaciones son o no secuenciables en cuanto a conflictos e indique cuál es el orden de las transacciones en caso afirmativo

Tema 6 – 5p – Determine si las siguientes planificaciones son o no secuenciables en cuanto a conflictos e indique cual es el orden de las transacciones en caso afirmativo.

P1	
T1	T2
Leer(A)	Escribir(B)
Escribir(A)	Escribir(A)

P2	
T1	T2
Leer(A)	Escribir(B)
Leer(B)	Escribir(A)

P3		
T1	T2	T3
Escribir(A)	Escribir(B)	Escribir(C)
	Leer(C)	

Si

P1: NO (si es escribir A) y SI (SI es escribir B);

P2: NO

P3: SI => T1, T3, T2

Tema 7: Determine si las siguientes planificaciones son o no secuenciables en cuanto a vistas e indique cuál es el orden de las transacciones en caso afirmativo

Tema 7 – 5p – Determine si las siguientes planificaciones son o no secuenciables en cuanto a vistas e indique cual es el orden de las transacciones en caso afirmativo.

P1	
T1	T2
Leer(A)	Escribir (A)
Escribir(A)	Escribir(A)

P2	
T1	T2
Leer(A)	Escribir (B)
Leer(B)	Escribir (A)

P3		
T1	T2	T3
Leer(A)	Leer(B)	Escribir(B)
Escribir(A)	Escribir(B)	Leer(A)
		Escribir(B)

P1: Si , primero leer(A) y escribir(a) de T1 luego escribir(a), escribir(A) de T2
 P2: NO
 P3: SI => T1, T2, T3

A – Determine si las siguientes planificaciones son o no secuenciables en cuanto a conflictos, e indique cual es el orden de las transacciones en caso afirmativo.

P1	
T1	T2
Leer(A)	
Escribir(A)	Escribir (A)
	Escribir(A)

P2	
T1	T2
Leer(A)	
	Escribir (B)
Leer(B)	Escribir (A)

P3		
T1	T2	T3
Escribir(A)	Escribir(B)	
	Leer(C)	Escribir (C)

B – Determine si las siguientes planificaciones son o no secuenciables en cuanto a vistas, e indique cual es el orden de las transacciones en caso afirmativo.

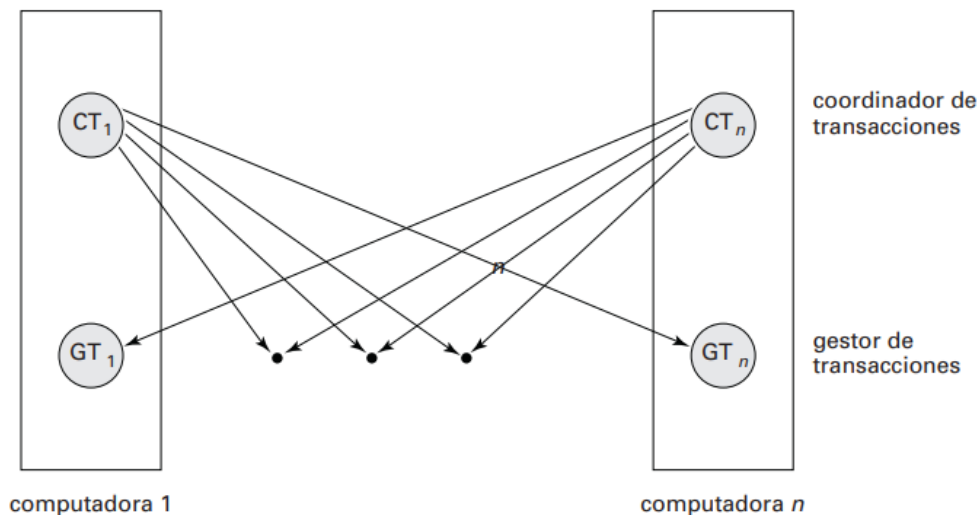
P1	
T1	T2
Leer(A)	
Escribir(A)	Escribir (A)
	Escribir(A)

P2	
T1	T2
Leer(A)	
	Escribir (B)
Leer(B)	Escribir (A)

P3		
T1	T2	T3
Leer(A)	Leer(B)	Escribir(B)
Escribir(A)	Escribir(B)	Leer(A)
		Escribir(B)

Fin1a2017Sem2

Tema 3: Explique detalladamente el Modelo abstracto del sistema de transacción en un sistema de bases de datos distribuidas orientados a asegurar las propiedades ACID



Gestor de transacciones (GT): Administra la ejecución de transacciones del sitio.

Coordinador de transacciones (CT): Coordina la ejecución de diferentes transacciones iniciadas en el sitio.

Cada GT es responsable de mantener un registro histórico y coordinar la ejecución concurrente de transacciones.

Cada CT es responsable de iniciar una transacción, dividir la transacción en varias sub-transacciones y distribuirlo a los diferentes sitios, coordinar la terminación de la transacción.

Tema 4: Explique detalladamente las propiedades de las transacciones que deben ser aseguradas por un SGBD (Sistema gestor de bases de datos) conforme el estándar SQL.

Para preservar la integridad de los datos las siguientes propiedades de las transacciones deben ser aseguradas por el SGBD:

- **Atomicidad** (Atomicity): Una transacción es una unidad atómica de procesamiento; es realizada enteramente o no es realizada en nada.
- **Consistencia** (Consistency): La ejecución de la transacción debe pasar la BD desde un estado consistente a otro también consistente.
- **Aislamiento** (Isolation): Una transacción no deberá hacer visible sus modificaciones a otras transacciones hasta que esté confirmada.
- **Persistencia** (Durability): Cuando una transacción cambia la BD y los cambios son confirmados, estos cambios no deben perderse por fallos posteriores.

Tema 5: Explique qué entiende por un gestor de bloqueos en un SGBD que implementa un Control de Concurrency basado en Bloqueos. Indique además como él mismo podría estar implementado y que pasos debería de tener en cuenta para evitar la inanición de las transacciones.

Un control de concurrency basado en bloqueos consiste en un mecanismo que permite reservar un elemento de la base de datos a una transacción dada y evitar que otras transacciones tengan acceso al mismo y pueden causar inconsistencias.

- Estos bloqueos están implementados para ejecutarse en dos modos: compartido y exclusivo. Los bloqueos compartidos permiten a un elemento bloqueado en una

transacción T1 ser leído por una transacción T2, pero no permite su modificación hasta que haya cesado el bloqueo. Los bloqueos exclusivos prohíben tanto la lectura como la escritura de un elemento bloqueado por una transacción a cualquier otra transacción que lo solicite.

- Se puede evitar la inanición de las transacciones al conceder los bloqueos de la siguiente manera: cuando una transacción T_i solicita un bloqueo sobre un elemento de datos Q en un modo particular M , el gestor de control de concurrencia concede el bloqueo siempre que:
 - No exista otra transacción que posea un bloqueo sobre Q en un modo que esté en conflicto con M .
 - No exista otra transacción que esté esperando un bloqueo sobre Q y que lo haya solicitado antes que T_i .

Tema 6: Explique en el contexto de los protocolos de control de concurrencia que se entiende por: (1) planificación, (2) planificación secuenciable (3) Secuencialidad en cuanto a conflicto y (4) Secuencialidad en cuanto a vistas. Proporcione ejemplos positivos y negativos para los dos últimos casos

- Una **planificación** para un conjunto de transacciones consiste en un ordenamiento de todas las operaciones de dichas transacciones, pero conservando el orden interno en cada transacción individual.
- Un plan S de n transacciones es **secuenciable** si es equivalente a algún plan secuencial de las mismas n transacciones. Un plan secuenciable es siempre correcto. Las planificaciones pueden ser secuenciales en cuanto a vistas o en cuanto a conflictos.
- Se dice que existe secuencialidad en cuanto a conflictos cuando dos transacciones realizan operaciones sobre un mismo dato, tal que al menos una de ellas es escribir, entonces, el orden de ambas operaciones es importante. Se dice que una planificación es secuenciable en cuanto a conflictos cuando puede convertirse en una planificación secuencial

Positivo:

T_1	T_2
leer(A) escribir(A)	
	leer(A) escribir(A)
leer(B) escribir(B)	
	leer(B) escribir(B)

Negativo:

Plan E

T_1	T_2
Leer_item(X)	
Escribir_item(X)	Escribir_item(X)

Está en conflicto

- Se dice que hay secuencialidad en cuanto a vistas cuando es posible convertir una transacción T concurrente en una secuencial, cumpliendo los criterios: todas las operaciones "leer" deben leer el mismo dato inicial o producido por una operación

escribir previa a la misma, y toda operación “escribir” que haya ocurrido al final de una transacción, debe mantenerse al final de la misma.

Positivo:

Negativo:

Plan S₁		Plan S₂	
T ₁	T ₂	T ₁	T ₂
Leer_item(X) X := X-N Escribir_item(X)	Leer_item(X) X := X+M Escribir_item(X)	Leer_item(X) X := X-N Escribir_item(X) Leer_item(Y) Y := Y+N Escribir_item(Y)	Leer_item(X) X := X+M Escribir_item(X)

Cada operación de lectura en S2 debe leer el mismo valor que en S1. Además, el último valor escrito en la BD debe ser el mismo en ambos planes.

Fin1a2016sem2

Tema 1: Explique concisa y ordenadamente

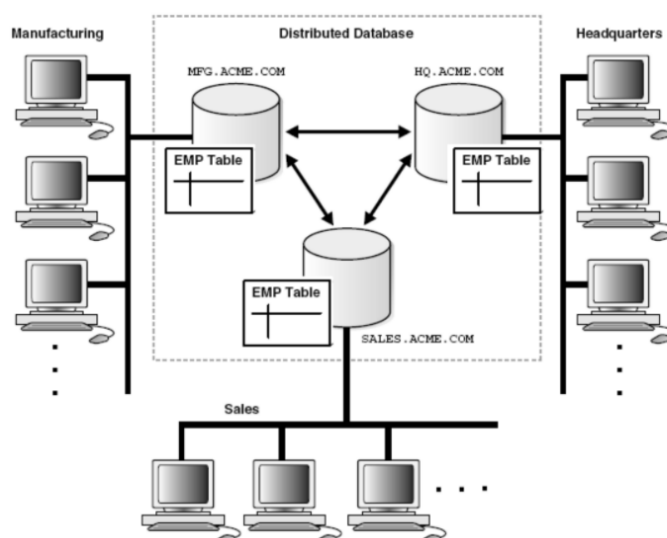
b- las formas de almacenamiento distribuido en un SGBD. Proporcione ejemplos.

Considérese una relación r que hay que almacenar en la base de datos. Hay dos enfoques del almacenamiento de esta relación en la base de datos distribuida:

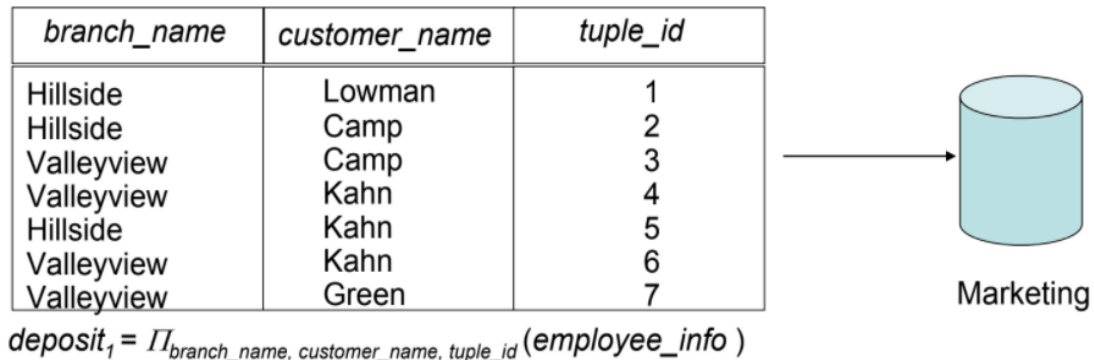
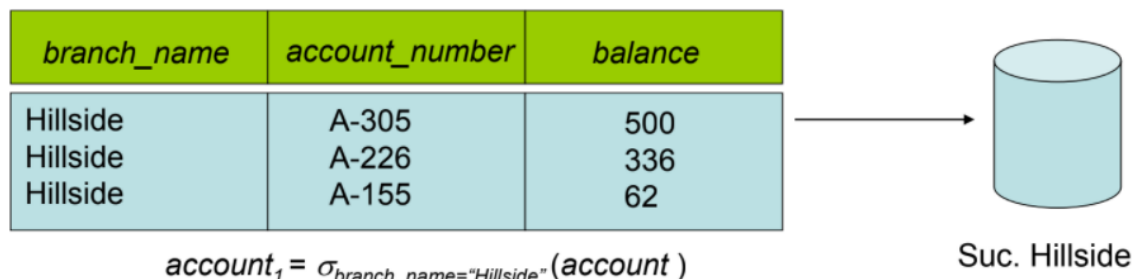
- Réplica. El sistema conserva réplicas (copias) idénticas de la relación y guarda cada réplica en un sitio diferente. La alternativa a las réplicas es almacenar sólo una copia de la relación r.

Ejemplo:

Replicación de Datos



- Fragmentación. El sistema divide la relación en varios fragmentos y guarda cada fragmento en un sitio diferente.



Tema 5: cite y explique brevemente 4 características del Dataware

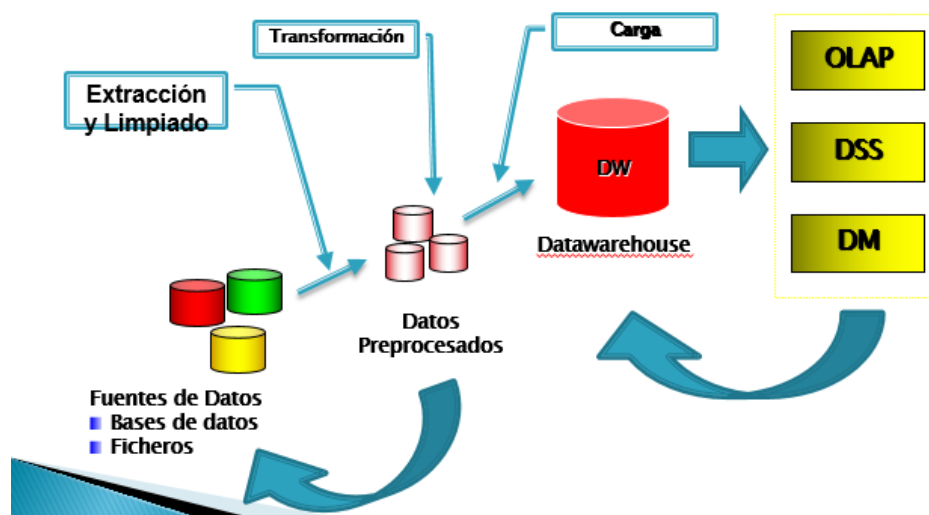
- **Orientado hacia la información relevante de la organización:** se diseña para consultar eficientemente información relativa a las actividades (ventas, compras, producción, ...) básicas de la organización, no para soportar los procesos que se realizan en ella (gestión de pedidos, facturación, etc).
- **Integrado:** integra datos recogidos de diferentes sistemas operacionales de la organización (y/o fuentes externas).
- **Variable en el tiempo:** los datos son relativos a un periodo de tiempo y deben ser incrementados periódicamente.
- **No volátil:** los datos almacenados no son actualizados, sólo son incrementados.

Richard:

- **Orientado a un Tema.**
 - Colección de información relacionada organizada alrededor de un tema central.
- **Integrado.**
 - Datos de múltiples orígenes; consistencia de datos.
- **Histórico - Variable en el tiempo.**
 - 'Fotos' en el tiempo.
 - Basado en fechas/periodos.
- **No-volátil.**
 - Sólo lectura para usuarios finales.
- **Menos frecuencia de cambios/actualizaciones.**
 - Usado para el Soporte a Decisiones y Análisis de Negocio.

Tema 6: Ilustra el proceso de Extracción, Transformación y carga (ETL) de datos en DW

- **Extracción:** este es el primer paso de obtener la información hacia el ambiente del Data Warehouse.
- **Transformación:** una vez que la información es extraída hacia el área de trafico de datos, hay posibles paso de transformación como; limpieza de la información, tirar la basura que no nos sirve, seleccionar únicamente los campos necesarios para el Data Warehouse, combinar fuentes de datos, haciéndolas coincidir por los valores de las llaves, creando nuevas llaves para cada registro de una dimensión.
- **Carga:** al final del proceso de transformación, los datos están en forma para ser cargados.



Fin2a2016sem2A

Tema 1: Explique concisa y ordenadamente:

c- Las formas de almacenamiento distribuido en un SGBD y su definición en álgebra relacional

Definición ya está. Álgebra lineal:

-Fragmentación:

*horizontal: $R_i = \langle \text{condición } i \rangle (R)$ tal que $R = R_1 \cup R_2 \cup \dots \cup R_n$

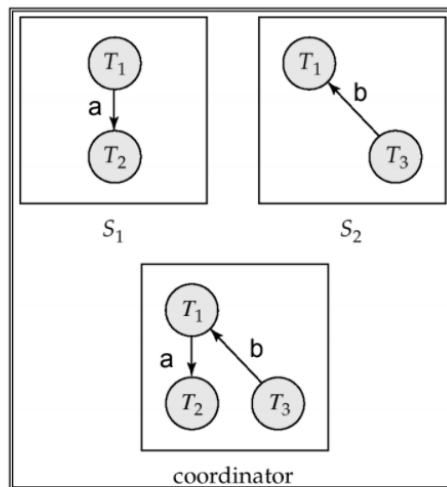
*vertical: $R_i = \langle \text{subesquema } i \rangle (R)$ tal que $R = R_1 \text{ join } R_2 \text{ join } \dots \text{ join } R_n$

- Replicación: No tiene

Fin1a2019sem2

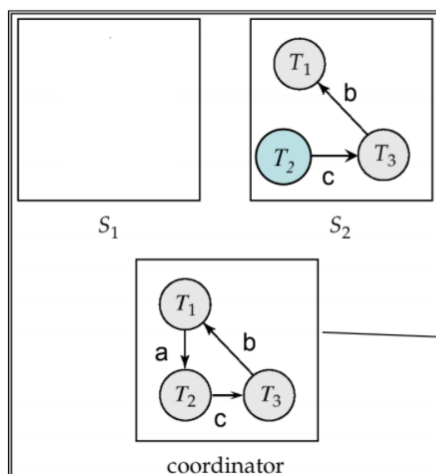
Tema 1: Explique en que consisten los problemas que deben evitarse en la detección de interbloqueos en un Sistema de base de datos distribuida.

- Detección de ciclos falsos:



Luego ocurre en el sgte. orden que:

1. T2 libera el dato a
2. S1 envía el mensaje de eliminar el arco $T1 \rightarrow T2$
3. T2 solicita un bloqueo sobre c en el sitio S2, pero T3 tiene bloqueado el dato c
4. S2 envía el mensaje de añadir el arco $T2 \rightarrow T3$

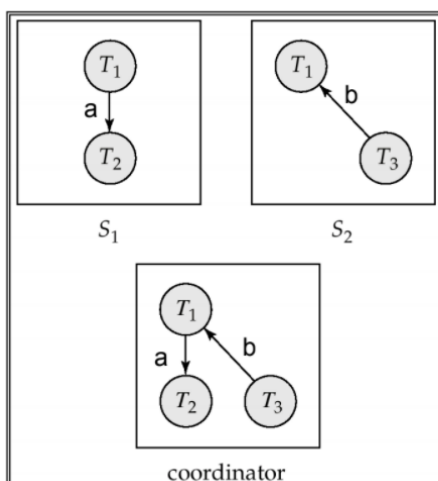


Se detecta un falso ciclo si:

1. El mensaje de S1 llega luego del mensaje de S2
2. El mensaje de S1 se pierde.

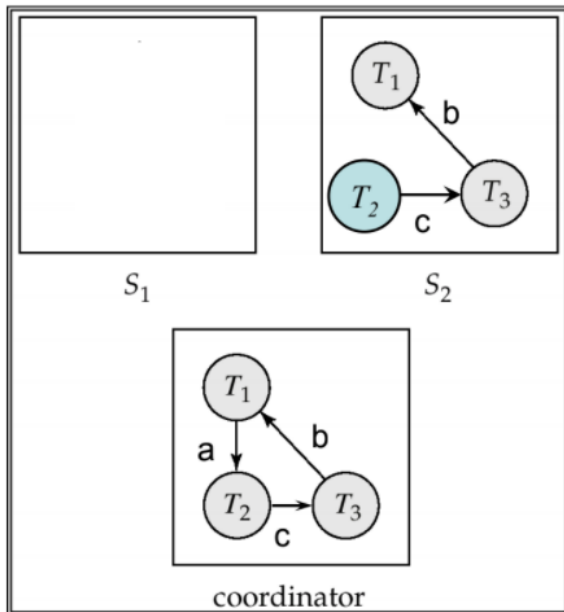
No Real

- Retroceso no relacionado:



Luego ocurre en el sgte. orden que:

- T2 solicita un bloqueo sobre c en el sitio S2, pero T3 tiene bloqueado el dato c
- S2 envía el mensaje de añadir el arco $T2 \rightarrow T3$
-



Se detecta un falso ciclo si:

1. En S_1 se aborta T_1 por algún otro problema no relacionado a la detección de ciclos.

Ocorre que:

1. El coordinador pudo haber elegido como víctima a T_2 o a T_3
2. Se Retrocede innecesariamente a T_2 o a T_3 .

Final

Tema X: fundamentar al menos dos ventajas del enfoque de semi reuniones que no tengan que ver con la reducción de costos de transmisión de datos

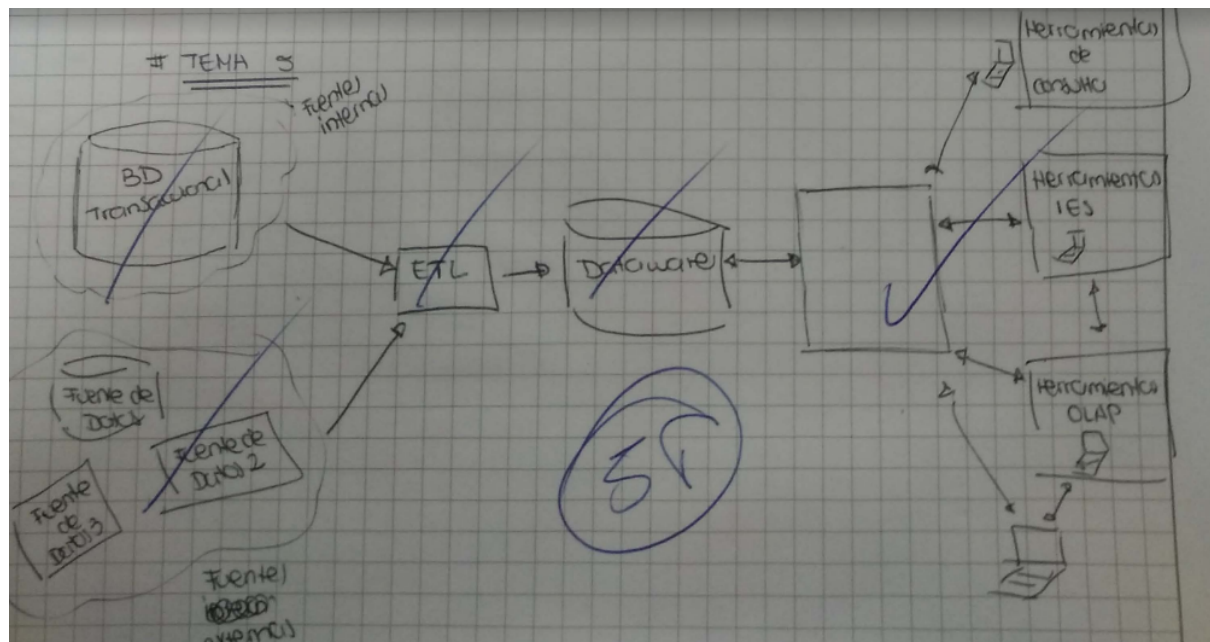
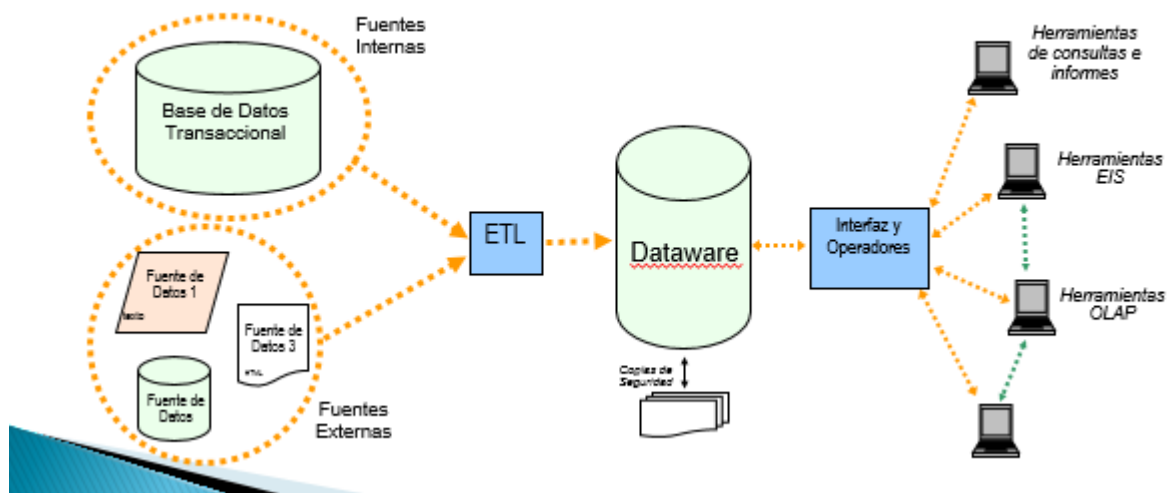
- Costo de Disco.
- Costo de Memoria.
- Costo CPU.

Esta estrategia resulta especialmente ventajosa cuando relativamente pocas tuplas de r_2 contribuyen a la reunión. Esta situación es probable que se produzca si r_1 es resultado de una expresión de álgebra relacional que implique a una selección. En esos casos puede que r_2 tenga significativamente menos tuplas que r_1 . Los ahorros de costes de la estrategia proceden de tener que enviar a S_1 sólo r_2 , en vez de toda r_1 . El envío de r_1 a S_2 supone un coste adicional. Si una fracción de tuplas de r_2 lo bastante pequeña contribuye a la reunión, la sobrecarga del envío de r_1 queda dominada por el ahorro de tener que enviar sólo una parte de las tuplas de r_1 . Esta estrategia se denomina estrategia de semi-reunión, del operador de semirreunión del álgebra relacional, denotado por \bowtie . La semirreunión de r_1 con r_2 , denotada por $r_1 \bowtie r_2$, es $W R_1 (r_1 \bowtie r_2)$

Por tanto, $r_1 \bowtie r_2$ selecciona las tuplas de r_1 que han contribuido a $r_1 \bowtie r_2$. En el paso 3 $r_2 = r_2 \bowtie r_1$. Para las reuniones de varias relaciones esta estrategia puede ampliarse a una serie de pasos de semirreunión. Se ha desarrollado un importante corpus teórico en relación con el empleo de la semirreunión para la optimización de consultas. A parte de esta teoría se hace referencia en las notas bibliográficas.

Tema Y : Esquematizar la arquitectura de un DW

del taa



✓ **Tema 3 – 10 p.** – Explique: (i) el protocolo de control de bloqueo distribuido de Quórum de Consenso, (ii) que implica el que se cumpla la condición del protocolo y (iii) como pueden emularse los protocolos de Mayoría y Sesgado con el mismo.

Tema 4 – 10 p. – Esquematice el protocolo de compromiso de 2 fases e indique como se procede: (a) en caso de falla del coordinador y (b) en caso de falla de un sitio participante.

✓ **Tema 5 – 5 p.** – Ilustre y describa la arquitectura de implementación de un Dataware.

Tema 6 – 5 p. – Explique las 3 V's que caracterizan a las iniciativas de Big Data.

✓ **Tema 2 – 10 p.** – Muestre los pasos para la reunión de las siguientes relaciones mediante la estrategia de la semireunión. Luego, plantee el ahorro que conlleva el mismo en transmisión respecto a la estrategia más simple que realiza el menor movimiento de datos entre los sitios. Suponga que la consulta ha sido recibida en el Sitio 1 y que el resultado debe producirse en este mismo sitio.

$R \text{ join } S \text{ on } (R.A2 = S.A2)$

R@S1		S@S2		
A1	A2	A2	A3	A4
1	3	3	13	16
1	4	3	14	16
1	6	7	13	17
2	3	10	14	16
2	6	10	15	17
3	7	11	15	16
3	8	11	15	16
3	9	12	15	16

Tema 2 (10 p.).

- Desarrolle los pasos correspondientes para la semi-reunión de las relaciones ilustradas. Suponga que la consulta ha sido recibida en el Sitio 2.
- Explique o fundamente al menos dos (2) ventajas de este enfoque.

$R \text{ join } S \text{ on } (R.A2 = S.A2)$

R@S1		S@S2		
A1	A2	A2	A3	A4
1	3	3	13	16
1	4	3	14	16
1	6	7	13	17
2	3	10	14	16
2	6	10	15	17
3	7	11	15	16
3	8	11	15	16
3	9	12	15	16

Sitio 1 - R		Sitio 2 - S			
A1	A2	A2	A3	A4	
1	3	3	13	16	
1	4	3	14	16	
1	6	7	13	17	
2	3	10	14	16	
2	6	10	15	17	
3	7	11	15	16	
3	8	11	15	16	
3	9	12	15	16	



Paso 1

Original
Costo 1: 16
Costo 2: 11
Surrender

Paso 1 = $\begin{bmatrix} 3 \\ 7 \\ 11 \\ 12 \end{bmatrix}$ SMI
Enviar a S1

$\begin{bmatrix} 1 & 3 \\ 2 & 3 \\ 3 & 7 \end{bmatrix}$ SMI
→ Enviar a S2

$S_2A_1 \times S_1A_2$ $\begin{bmatrix} 3 & 1 & 7 & 13 & 17 \\ 1 & 3 & 3 & 13 & 16 \\ 1 & 3 & 3 & 14 & 16 \\ 2 & 3 & 3 & 13 & 16 \\ 2 & 3 & 3 & 14 & 16 \end{bmatrix}$

Control de Concurrencia en BDD

- Se utilizan los esquemas de control de concurrencia de sistemas centralizados modificados para BDD
- Se asume que todos los sitios participan del protocolo de compromiso para asegurar así la atomicidad global de las transacciones.
- Se asume inicialmente que todas la réplicas son actualizadas.

Concurrencia: Gestor único de bloqueos

- El sistema mantiene un único gestor de bloqueos que reside en un sitio S_i
- Todas las peticiones de bloqueo y desbloqueo se realizan en S_i
 - Cuando una transacción necesita bloquear un dato, envía una solicitud a S_i
- Si el bloqueo puede concederse se envía un respuesta al sitio.
- Si el bloqueo no puede concederse, se retrasa la solicitud hasta que se pueda.
- Una transacción que posea el bloqueo de un dato puede leerlo de cualquier sitio.
- Las operaciones de escrituras deben realizarse en todos los sitios.
- Para liberar un bloqueo se envía un mensaje de desbloqueo a S_i

Concurrencia: Gestor único de bloqueos

- Este esquema tiene las siguientes ventajas
 - Implementación sencilla.
 - Tratamiento simple de interbloqueos
- Los inconvenientes son:
 - Si puede ser un potencial cuello de botella
 - El protocolo es vulnerable a fallas de Si
- Se debe detener el procesamiento de transacciones
- Se debe tener un gestor de bloqueos de respaldo

Concurrencia: Control de bloqueo distribuido

- Cada sitio posee un administrador de bloqueos
 - Controlan el bloqueo de los datos locales
 - Se necesitan protocolos especiales para tratar con datos replicados.

- Ventaja
 - El sistema es más tolerable a fallos.
- Desventaja
 - La detección de interbloqueos es más complicada

- Varias variantes
 - Copia Primaria
 - Protocolo de Mayoría
 - Protocolo Sesgado
 - Quórum de consenso.

Bloqueos distribuidos: Copia Primaria

- Una replica es seleccionada como Copia Primaria
 - El sitio que contiene la Copia Primaria es llamado Sitio Primario
 - Diferentes datos pueden tener diferentes Sitios Primarios
- Cuando una transacción necesita bloquear un dato Q, esta pide el bloqueo al sitio primario de Q.
 - Si el bloqueo está disponible se concede el bloqueo, sino se retrasa.
 - Un bloqueo es concedido sobre todas las replicas.
- La solicitud de desbloqueo se envía al sitio primario de Q

- Ventaja:
 - El control de concurrencia para los datos replicados es similar al de datos no replicados – Implementación Sencilla.

- Desventaja:
 - Si el sitio primario de Q falla, Q es inaccesible, aun cuando posea copias accesibles.

Bloqueos distribuidos: Protocolo de Mayoría

- Los administradores de bloqueos de cada sitio administran los bloqueos y desbloqueos de todos los datos en su sitio.

- Si el dato Q está replicado en N sitios, la solicitud de bloqueo de Q debe ser enviada a más de la mitad de los N sitios.
- Una transacción no puede operar sobre Q hasta que no haya obtenido un bloqueo en la mayoría de las replicas.

- Ventaja

- Protocolo válido aun ante falla de los sitios.

- Desventaja

- Requiere de muchos mensajes, $2(N/2 + 1)$ para bloqueos y $(N/2 + 1)$ para desbloques.
- Problema de interbloqueo de solicitudes de bloqueo.

Bloqueos distribuidos: Protocolo Sesgado

- Cada sitio posee un administrador de bloqueos para los datos en el mismo.
- Las peticiones de bloqueos compartidos y exclusivos se manejan de forma diferente.
 - Bloqueo Compartido
 - Para bloquear un ítem Q en modo compartido solo realiza la solicitud en cualquier sitio donde haya una replica de Q
 - Bloqueo Exclusivo
 - Para bloquear Q en modo exclusivo se hace la solicitud a todos los sitios en donde este replicado Q.

- Ventaja

- Menos sobrecarga en las operaciones de lectura.

- Desventaja

- Mayor sobrecarga en las operaciones de escritura.
- Necesidad de detección de interbloqueos distribuida

Bloqueos distribuidos: Quórum de Consenso

- Es una generalización de los protocolos de Mayoría y Sesgado.
- Cada Sitio tiene asignado un peso
 - Siendo el S la suma de todos los pesos de los sitios.
- Para cada dato Q se escogen dos valores
 - Q_r , que es el quórum de lectura
 - Q_w , que es el quórum de escritura
 - Se debe cumplir la siguiente condición:
- $Q_r + Q_w > S$ y $2 * Q_w > S$
- Si un sitio necesita realizar una lectura debe bloquear tantas replicas tal que la suma de sus pesos sea mayor o igual a Q_r
- Si un sitio necesita realizar una escritura debe bloquear tantas replicas tal que la suma de sus pesos sea mayor o igual a Q_w
- Se puede asumir que cualquier replica puede ser escrita.

FINALES

fin1a2017semB

Tema 9: Detalle apropiadamente como se implementa la organización física de bloques denominada estructura de páginas por ranuras para el almacenamiento de registros.

Existen diferentes técnicas para implementar los registros de longitud variable.

La **estructura de páginas con ranuras** se utiliza habitualmente para organizar los registros en bloques, y puede verse en la Figura 11.9. Hay una cabecera al principio de cada bloque, que contiene la información siguiente:

1. El número de elementos del registro de la cabecera.
2. El final del espacio vacío del bloque.
3. Un *array* cuyas entradas contienen la ubicación y el tamaño de cada registro.

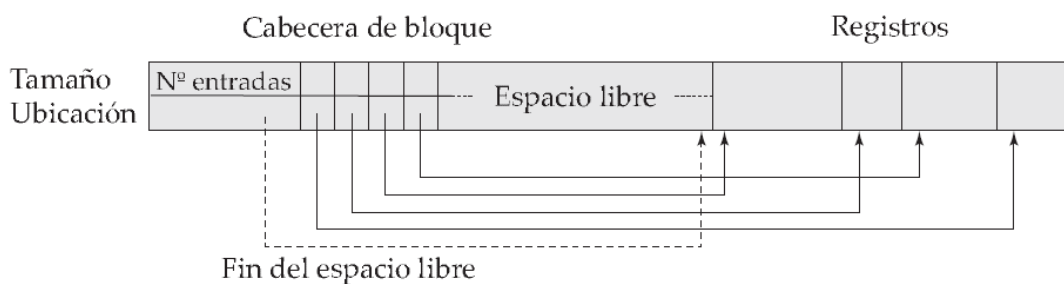


Figura 11.9 Estructura de páginas con ranuras.

El método de la página con ranuras se usa mucho para manejar los registros de longitud variable en los bloques de disco.

Cuestionario 2, 2do. Parcial, 2021

Explique detalladamente las formas de almacenamiento distribuido en Bases de Datos Distribuidas Relacionales.

- (a) Detalle los pasos de cada fase del protocolo C2F en Bases de Datos Distribuidas. (b) Indique como proceden los participantes en caso de falla del coordinador. (c) Indique como se procede un sitio en caso de falla del mismo.

Conforme la figura de abajo y la consulta "select * from R join S". Indique los pasos y el costo total de transmisión de realizar la Estrategia de la Semireunión si la consulta fue recibida: (a) en el Sitio 1, (b) en el Sitio 2.

Sitio 1 – R		Sitio 2 – S		
A1	A2	A2	A3	A4
1	3	3	13	16
1	4	3	14	16
1	6	7	13	17
2	3	10	14	16
2	6	10	15	17
3	7	11	15	16
3	8	11	15	16
3	9	12	15	16

Cuestionario 3, 2do. Parcial, 2021

Explique en sus términos: ¿Cuál es la diferencia entre bases de datos OLTP y OLAP?. Así mismo, cite al menos dos ejemplos de aplicación para cada tipo.

Explique en sus términos en qué consisten las tablas de dimensiones, tablas de hechos y medidas en el modelado multidimensional. Así mismo, enumere al menos tres ejemplos para cada uno.

Resuelva el siguiente planteamiento

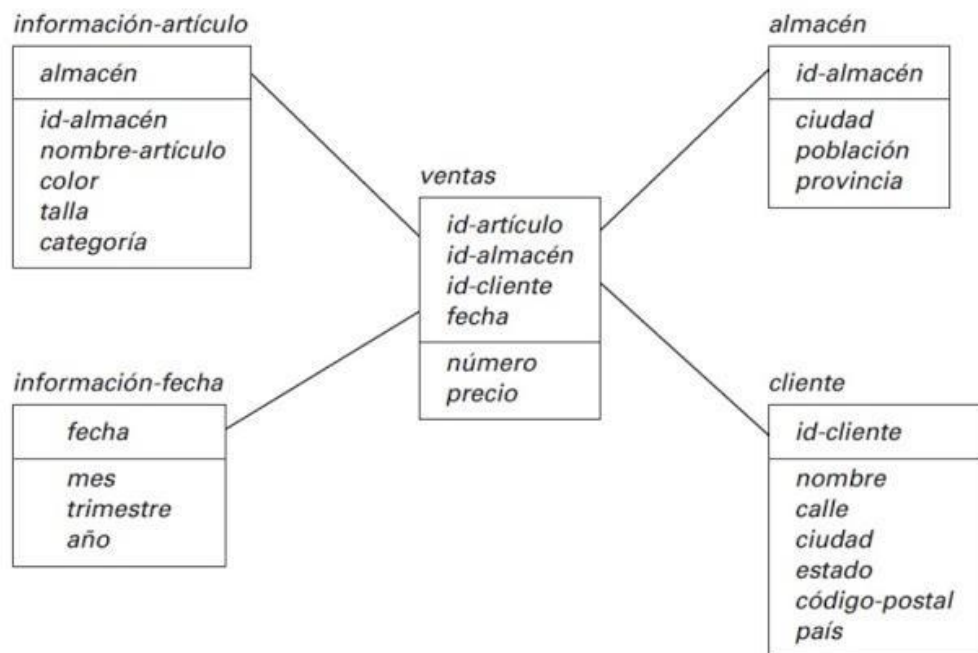
Una empresa te contrata como su nuevo Analista de Datos. Actualmente la empresa, tiene cierto grado de madurez y, cuenta con un Data Warehouse (DWH) sobre un PostgreSQL para el procesamiento analítico de sus datos.

Usted sabe que se cuenta con datos históricos de ventas de los últimos 10 años de operativa. El periodo de actualización del DWH es diario.

A continuación, el Departamento de Marketing te solicita la extracción de información del DWH para su campaña de promoción. Por lo tanto, se desea saber:

- Los clientes activos que no han realizado ninguna compra en los últimos 6 meses.
- Los 10 artículos más vendidos en los últimos 5 años, no se debe considerar las ventas del año en vigencia. Además, incluir el monto total de venta.

Dado el siguiente Esquema en Estrella. Escriba las sentencias SQL genéricas para la extracción de información y satisfacer los requerimientos de usuarios.



```
*select cliente.nombre , fecha.mes, fecha. año from cliente join ventas on ventas.id_cliente = cliente.id_cliente join información_fecha.fecha = ventas.fecha where cliente.estado = activo and (información_venta.mes < mes_actual - 6 and informacion_venta.año <= año_actual );
*select ventas.numero and articulo.nombre and información_fecha.año from información_artículo join ventas on ventas.id_articulo = información_artículo.id_artículo join información_fecha join ventas on información_fecha.fecha = ventas.fecha where año > año_actual - 5 order by ventas.numero desc limit 10;
```

Cuestionario 1, 2do. Parcial, 2021

(a) Defina el concepto de Transacción en SGBDs. (b) Indique y defina las fases del Ciclo de Vida de una Transacción. (b) Indique y defina las propiedades que un SGBD debe asegurar para las Transacciones.

(a) Explique que es una Planificación Secuencial. (b) Explique que es una Planificación Secuenciable. (c) Fundamente a su criterio la importancia de la Secuencialidad en SGBD.

(a) Describa el protocolo de Bloqueo de 2 Fases. (b) Indique cuales son las variantes del mismo.

(a) Describa la función del Componente de Gestión de Concurrencia en un SGDB.
(b) Describa la estructura de datos utilizada para la gestión y concesión de bloqueos.

Link utiles:

<https://www.cs.usfca.edu/~galles/visualization/BPlusTree.html>

Dada una operación de selección cualquiera basada en una condición de igualdad, explique en qué casos un SGBD utilizará los siguientes algoritmos e indique cual es el costo asintótico de los mismos: (a) Búsqueda Lineal. (b) Búsqueda Binaria. (c) Búsqueda en índice primario para un atributo clave. (d) Búsqueda en índice secundario para atributo no clave.

Dada una operación de selección cualquiera basada en una condición de igualdad simple, indique que condiciones físicas y/p lógicas obligarían a un SGBD utilizar cada uno de los siguientes algoritmos, indicando además el costo asintótico correspondiente en cada caso: (1) Búsqueda Lineal. (2) Búsqueda Binaria. (3) Búsqueda en índice primario. (4) Búsqueda en índice secundario.

a) Cuando no se tiene un índice de clave de búsqueda ni se ordena la tabla, necesariamente se revisan todos los bloques. (Costo = B_r) Aunque podría ser más lento que otros algoritmos para la implementación de la selección, el algoritmo de búsqueda lineal se puede aplicar a cualquier archivo, sin importar su ordenación, la presencia de índices o la naturaleza de la operación selección

b) Cuando no se tiene un índice de clave pero la tabla está ordenada por la clave de búsqueda. (Costo = $\log_2(B_r) + \lceil cs(A,r)/fr \rceil - 1$)

c) Cuando se tiene un índice de clave de búsqueda que además el atributo tiene clave (Costo = $AA_i + 1$)

d) Cuando no se tiene ningún índice de clave de búsqueda ni atributo clave. (Costo = $br + c(A,r) + 1$)

Explique detalladamente las formas de almacenamiento distribuido en Bases de Datos Distribuidas Relacionales.

Considérese una relación r que hay que almacenar en la base de datos. Hay dos enfoques del almacenamiento de esta relación en la base de datos distribuida:

Réplica. El sistema conserva copias idénticas de la relación y guarda cada réplica en un sitio diferente. La alternativa a las réplicas es almacenar sólo una copia de la relación r .

Fragmentación. El sistema divide la relación en varios fragmentos y guarda los fragmentos en cada sitio diferente.

Explique en sus términos las formas organización de archivos en posibles para la implementación de archivos de tablas/datos.

-Heap: los registros se guardan en cualquier lugar del disco donde hay espacio suficiente.

- Secuencial: los registros se guardan de forma secuencial en el disco por medio de una clave, la clave de búsqueda.
- Hash: los registros son almacenados en un bloque gracias a la función hash del valor de un campo del registro específico.
- Organización por agrupación: los datos de diferentes relaciones se guardan en el mismo archivo.

Dadas las relaciones "A (a1, a2, ..., a20)", "B (b1, b2, ..., b12)" y "C (c1, c2, ..., c15)", y la siguiente consulta: "select distinct A.a1, C.c1 from A join B on (A.a2 = B.b3) join C on (C.c2 = B.b4) where A.a1 > 10 and B.b1 = 50;". Muestre: (1) Su traducción directa al algebra relacional, (2) Los pasos para llegar a la expresión equivalente tal que se minimice la cantidad de datos procesados por cada operación. (3) El árbol de evaluación de la expresión final. Debe levantar una imagen de su desarrollo.