

MÓDULO XI

PROCESAMIENTO DE CONSULTA DISTRIBUIDA
PROCESAMIENTO DE TRANSACCIÓN DISTRIBUIDA

Marcela Russo
Laboratorio IV

PROCESAMIENTO DE CONSULTA DISTRIBUIDA IMPLICA

- Optimización local (intra-sitio) y optimización global (inter-sitio - decisiones de movimiento de datos y de selección de sitio). Por tal motivo existen muchos más planes de acceso posibles para una consulta distribuida que para su correspondiente consulta centralizada.
- Las complejidades también existen en las bases de datos paralelas con arquitecturas nada compartidas y la diferencia principal es la comunicación en redes mucho más rápida y más confiable que las que son usadas en el procesamiento de bases de datos paralelas.
- Persigue múltiples objetivos de optimización
- En un entorno centralizado, minimizar el uso de recursos (entrada-salida y procesamiento) es consistente con la reducción del tiempo de respuesta.
- En un entorno distribuido, minimizar los recursos puede entrar en conflicto con la reducción del tiempo de respuesta, debido a las oportunidades de procesamiento paralelo.
- El procesamiento paralelo puede reducir el tiempo de respuesta para aumentar la cantidad global de recursos consumidos (entrada-salida, procesamiento y comunicación).
- Además, la ponderación de los costos de comunicación contra los costos locales (entrada-salida y procesamiento) depende de las características de la red.

PROCESAMIENTO DE CONSULTA DISTRIBUIDA IMPLICA

- Para redes de área amplia, los costos de comunicación pueden dominar sobre los costos locales.
- Para redes de área local, los costos de comunicación están más equitativamente ponderados con los costos locales.
- La elección de un mal plan de acceso puede conducir a un rendimiento extremadamente pobre.
- Los planes de acceso distribuidos muchas veces se ajustan para las condiciones del sitio.
- Si un sitio no está disponible o está sobrecargado, un plan de acceso distribuido deberá elegir dinámicamente otro sitio. Por tal motivo, parte del proceso de optimización puede requerir que se ejecute dinámicamente (durante el tiempo de ejecución) en vez de estáticamente (durante el tiempo de compilación).
- Para determinar el mejor plan de acceso se muchas veces es necesario realizar análisis adicionales de los costos de procesamiento locales.

PROCESAMIENTO DE TRANSACCIÓN DISTRIBUIDA

- El procesamiento de transacción distribuida se ajusta a los principios de procesamiento de cualquier transacción.
- Las transacciones obedecen las propiedades ACID.
- Los DBMS distribuidos proporcionan concurrencia y transparencia de recuperación.
- Los sitios que operan independientemente deben estar coordinados.
- El entorno distribuido hace más difícil la implementación de los principios de optimización, debido a que en la red de comunicación se presentan nuevos tipos de falla y es necesario aplicar nuevos protocolos

PROCESAMIENTO DE TRANSACCIÓN DISTRIBUIDA

- **CONTROL DE CONCURRENCIA DISTRIBUIDA:**

- ✓ Los sitios locales deben coordinarse a través de mensajes sobre una red de comunicación, derivando en más gastos generales que el control de concurrencia centralizada.
- ✓ El esquema más simple implica coordinación centralizada.
- ✓ Al comienzo de una transacción, se elige el sitio de coordinación y la transacción se divide en subtransacciones realizadas en otros sitios.
- ✓ Cada sitio que alberga una transacción emite solicitudes de candado y liberación al sitio coordinador mediante el uso de reglas normales de candado en dos fases.
- ✓ La coordinación centralizada implica la menor cantidad de mensajes y la detección más simple de bloqueo.
- ✓ la confiabilidad en un coordinador centralizado puede hacer que el procesamiento de transacción sea menos confiable.
- ✓ Para aliviar la confianza en un sitio centralizado, el administrador de candados puede distribuirse entre sitios.
- ✓ El precio por mayor confiabilidad es más gastos generales por mensaje y detección de bloqueo más compleja.

PROCESAMIENTO DE TRANSACCIÓN DISTRIBUIDA

- **CONTROL DE CONCURRENCIA DISTRIBUIDA:**

- ✓ El número de mensajes puede duplicarse en el esquema de coordinación distribuida en comparación con el esquema de coordinación centralizada.
- ✓ Los datos copiados implican un problema tanto en la coordinación centralizada como en la distribuida.
- ✓ Actualizar los datos copiados implica gastos generales adicionales debido a que se debe obtener un candado de escritura en todas las copias antes de que cualquier copia se actualice.
- ✓ Obtener candados de escritura en múltiples copias puede causar retardos e incluso regresiones si una copia no está disponible.
- ✓ Para reducir la saturación con múltiples copias de candado puede usarse el protocolo de copia primaria.
- ✓ En el protocolo de copia primaria, una copia de cada fragmento copiado se designa como la copia primaria, mientras que las otras copias son secundarias.
- ✓ Los candados de escritura son necesarios sólo para la copia primaria.

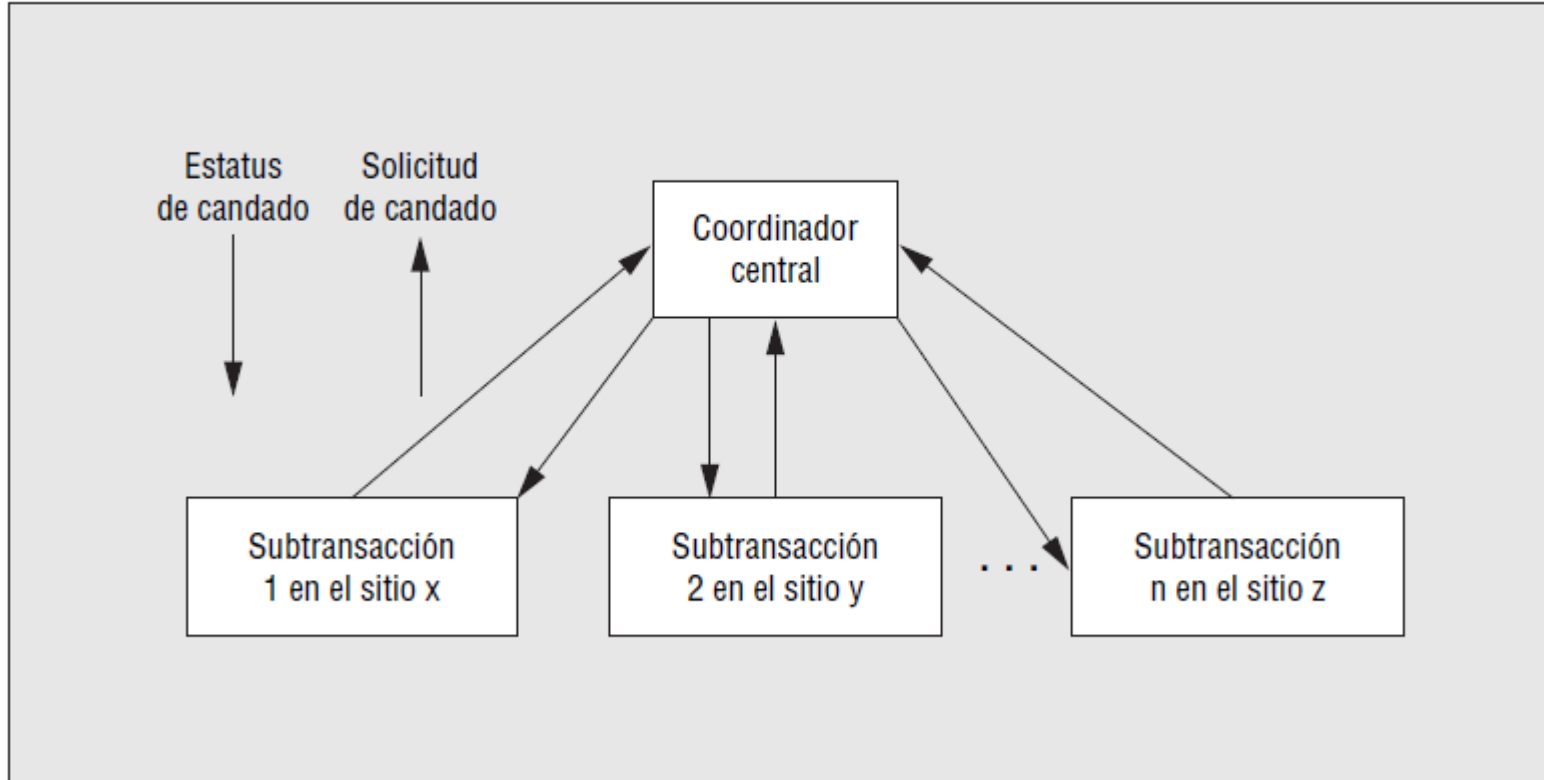
PROCESAMIENTO DE TRANSACCIÓN DISTRIBUIDA

- **CONTROL DE CONCURRENCIA DISTRIBUIDA:**

- ✓ Después de que una transacción actualiza la copia primaria, las actualizaciones se propagan a las copias secundarias.
- ✓ Sin embargo, las copias secundarias pueden no actualizarse sino hasta después de comprometer la transacción.
- ✓ El protocolo de copia primaria proporciona rendimiento mejorado pero al costo de copias secundarias no concurrentes.
- ✓ Debido a que la reducción de gastos generales con frecuencia es más importante que las copias secundarias actuales, muchos DBMS distribuidos usan el protocolo de copia primaria.

PROCESAMIENTO DE TRANSACCIÓN DISTRIBUIDA

Control de concurrencia centralizada



Mannino tabla 17.17 página 633

PROCESAMIENTO DE TRANSACCIÓN DISTRIBUIDA

- **PROTOCOLO DE COPIA PRIMARIA:**

- ✓ Es un protocolo para el control de concurrencia de transacciones distribuidas.
- ✓ Cada fragmento copiado está designado como la copia primaria o como la copia secundaria.
- ✓ Durante el procesamiento de transacción distribuida sólo la copia primaria tiene garantía de ser la actual al final de una transacción.
- ✓ Las actualizaciones pueden propagarse a copias secundarias después del fin de la transacción.

PROCESAMIENTO DE TRANSACCIÓN DISTRIBUIDA

- **PROCESAMIENTO DE COMPROMISO DISTRIBUIDO:**

- ✓ Los DBMS distribuidos deben enfrentar fallas de comunicación en vínculos y sitios.
- ✓ La detección de fallas implica coordinación entre sitios.
- ✓ Si un vínculo o sitio falla, debe abortarse cualquier transacción que involucre al sitio y ese sitio debe evitarse para futuras transacciones hasta que la falla se resuelva.
- ✓ Las fallas pueden ser más complejas que sólo un sitio sencillo o un vínculo de comunicación.
- ✓ Muchos sitios y vínculos pueden fallar simultáneamente dejando una red particionada.
- ✓ En una red particionada las diferentes particiones (colecciones de sitios) no pueden comunicarse, aunque se comuniquen los sitios en la misma partición.
- ✓ El administrador de transacción debe asegurar que diferentes partes de una red particionada actúan al unísono.
- ✓ No es posible que los sitios en una partición decidan comprometer una transacción, sino que los sitios en otra partición deben decidir no comprometer una transacción.
- ✓ Todos los sitios deben comprometer o abortar.

PROCESAMIENTO DE TRANSACCIÓN DISTRIBUIDA

- **PROTOCOLO DE COMPROMISO EN DOS FASES:**

- ✓ **Es el protocolo más utilizado para procesamiento de compromiso distribuido.**
- ✓ **Para cada transacción, se elige un sitio como el coordinador y la transacción se divide en subtransacciones realizadas en otros sitios participantes.**
- ✓ **El coordinador y los sitios participantes interactúan en una fase de votación y una fase de decisión.**
- ✓ **Al final de ambas fases, cada sitio participante actúa al unísono para comprometer o abortar su subtransacción.**
- ✓ **Las fases de votación y decisión requieren acciones tanto del coordinador como de los sitios participantes.**
- ✓ **En la fase de decisión, el coordinador envía un mensaje a cada participante preguntando si está listo para comprometer.**
- ✓ **Antes de responder, cada participante obliga todas las actualizaciones a disco cuando termina el trabajo de la transacción local.**
- ✓ **Si no ocurre una falla, el participante escribe un registro de READY-COMMIT y envía un voto READY al coordinador.**

PROCESAMIENTO DE TRANSACCIÓN DISTRIBUIDA

- **PROTOCOLO DE COMPROMISO EN DOS FASES:**

- ✓ En este punto, el participante tiene un estatus incierto porque el coordinador puede solicitar más tarde al participante abortar.
- ✓ La fase de decisión comienza cuando el coordinador recibe votos de cada participante u ocurre un vencimiento.
- ✓ Si ocurre un vencimiento o al menos un participante envía un voto de ABORT, el coordinador aborta toda la transacción mediante el envío de mensajes ABORT a cada participante, cada participante realiza una regresión a sus cambios.
- ✓ Si todos los participantes votan READY, el coordinador escribe el registro de compromiso global y pide a cada participante comprometer sus subtransacciones.
- ✓ Cada participante escribe un registro de compromiso, libera los cierres y envía un reconocimiento al coordinador.
- ✓ Cuando el coordinador recibe el reconocimiento de todos los participantes, el coordinador escribe el registro de fin de transacción global.
- ✓ Si ocurre una falla en la fase de votación o en la de decisión, el coordinador envía un mensaje ABORT a todos los sitios participantes.

PROCESAMIENTO DE TRANSACCIÓN DISTRIBUIDA

- **PROTOCOLO DE COMPROMISO EN DOS FASES:**

- ✓ La falla durante la recuperación y vencimientos, pueden complicar el protocolo.
- ✓ Las modificaciones al protocolo básico pueden reducir el número de mensajes necesarios para forzar al protocolo.
- ✓ El protocolo de compromiso en dos fases puede usar un coordinador centralizado o uno distribuido.
- ✓ Las negociaciones son similares a la coordinación centralizada frente a la distribuida para el control de concurrencias.
- ✓ La coordinación centralizada es más simple que la coordinación distribuida, pero puede ser menos confiable.
- ✓ El protocolo de compromiso en dos fases no maneja ningún tipo concebible de falla.
- ✓ El protocolo de compromiso en dos fases puede tener un funcionamiento inadecuado si los registros de log se pierden.

PROCESAMIENTO DE TRANSACCIÓN DISTRIBUIDA

- **PROTOCOLO DE COMPROMISO EN DOS FASES:**
 - ✓ **No hay protocolo que garantice que todos los sitios actúan al unísono para comprometer o abortar al presentarse fallas arbitrarias.**
 - ✓ **Dado que el protocolo de compromiso en dos fases maneja eficientemente tipos comunes de fallas, se usa ampliamente en procesamiento de transacción distribuida.**

PROCESAMIENTO DE TRANSACCIÓN DISTRIBUIDA

Procesamiento de compromiso en dos fases para coordinador y participantes

