

Gestión de Transacciones Distribuidas

M.C. Carlos Rojas Sánchez¹

¹Licenciatura en Informática
Universidad del Mar - Puerto Escodido

[Bases de Datos Distribuidas]

1 Diseño de la BDD

En general

La ubicación de los datos y los programas, en los sitios de una red, así como el diseño de la red.

Sobre la BDD

- Ubicación del software de los sistemas administradores de las Bases de Datos Distribuidas.
- Ubicación de las aplicaciones que corren sobre la base de datos.

Diseño

- Descendente
 - Se inician de cero.
 - Sistemas homogéneos.
- Ascendente
 - Cuando las base de datos ya existen en los diferentes sitios.

Fragmentación

- ¿Cómo Fragmentar?
- ¿Cuánto Fragmentar?
- ¿Cómo probar la correcta fragmentación?
- ¿Cómo ubicar los fragmentos?

Distribución

- Los fragmentos no pueden tratarse como archivos individuales.
 - Se tienen que tomar en cuenta sus relaciones.
- El acceso a la base de datos es más complicado.
 - El modelo de acceso a archivos remotos no aplica.
 - Relación entre distribución y procesamiento de consultas.
- El costo de mantener la integridad debe considerarse.
- El costo de control de concurrencia debe considerarse.

Alternativas de Distribución

- No replicado.
 - Particionada: cada fragmento reside en un solo sitio.
- Replicado.
 - Réplica completa: cada fragmento en cada sitio.
 - Réplica parcial: cada fragmento en algunos de los sitios.

2 Procesamiento de consultas distribuidas

Componentes del Procesamiento de Consultas

- Lenguaje de la consulta
- Arquitectura de la red
- Distribución de datos en los nodos
- Costos de la consulta

Optimización de Consultas

Es el proceso de selección del plan de evaluación de las consultas más eficientes de entre las muchas estrategias generalmente disponibles para el procesamiento de una consulta dada, especialmente si la consulta es compleja.

Modelo

- Clientes
- Optimizador de consultas
- BDMS

Métricas

- Costo
 - Tamaño
 - Velocidad

Arquitectura del procesamiento

Cuatro capas principales están involucradas en mapear una consulta a una base de datos distribuida en una secuencia optimizada de operaciones locales, cada una de ellas actuando en una base de datos local.

- Descomposición de consultas
- Localización de datos
- Optimización global de consultas
- Optimización local de consultas

Las primeras tres se realizan en un nodo central usando información global. La cuarta capa se realiza en cada nodo local.

Descomposición de consultas

- Normalización
- Análisis. Se detecta y rechazan consultas semánticamente incorrectas.
- Simplificación. Elimina predicados redundantes.
- Reestructuración.

Localización de Datos

La entrada a esta capa es una consulta algebraica definida sobre relaciones distribuidas. El objetivo de esta capa es localizar los datos de la consulta usando la información sobre la distribución de datos. Esta capa determina cuales fragmentos están involucrados en la consulta y transforma la consulta distribuida en una consulta sobre fragmentos.



Optimización Global de Consultas

Dada una consulta algebraica sobre fragmentos, el objetivo de esta capa es hallar una estrategia de ejecución para la consulta cercana a la óptima. La estrategia de ejecución para una consulta distribuida puede ser descrita con los operadores del álgebra relacional y con primitivas de comunicación para transferir datos entre nodos.

Optimización Local de Consultas

El trabajo de la última capa se efectúa en todos los nodos con fragmentos involucrados en la consulta. Cada subconsulta que se ejecuta en un nodo, llamada consulta local, es optimizada usando el esquema local del nodo. La optimización local utiliza los algoritmos de sistemas centralizados

Bibliografía I

-  M. Tamer Özsu and Patrick Valduriez.
Principles of Distributed Database Systems.
Springer, 2011.
-  Abraham Silberschatz, Henry F. Korth and S. Sudarshan.
Fundamentos de Bases de Datos.
McGRAW-HILL, 2002.