

Diseño de Bases de Datos Distribuidas

M.C. Carlos Rojas Sánchez¹

¹Licenciatura en Informática
Universidad del Mar - Puerto Escodido

[Bases de Datos Distribuidas]

1 Consideraciones de diseño de bases de datos distribuidas

Consideraciones de diseño

El problema de diseño de bases de datos distribuidos se refiere a hacer decisiones acerca de la ubicación de datos y programas a través de los diferentes sitios de una red de computadoras. Este problema debería estar relacionado al diseño de la misma red.

Consideraciones de diseño

- Diseño de la fragmentación, este se determina por la forma en que las relaciones globales se subdividen en fragmentos horizontales, verticales o mixtos
- Diseño de la asignación de los fragmentos, esto se determina en la forma en que los fragmentos se mapean a las imágenes físicas, en esta forma, también se determina la solicitud de fragmentos

2 Diccionario de datos

Diccionario de Datos Global

El procesador de usuario consiste de cuatro partes:

- Un manejador de la interfaz con el usuario
- Un controlador semántico de datos
- Un optimizador global de consultas
- Un supervisor de la ejecución global

3 Niveles de transparencia

Niveles de Transparencia

La transparencia se puede entender como la separación de la semántica de alto nivel de un sistema de los aspectos de bajo nivel relacionados a la implementación del mismo. Un nivel de transparencia adecuado permite ocultar los detalles de implementación a las capas de alto nivel de un sistema y a otros usuarios.

Niveles de Transparencia

En sistemas de bases de datos distribuidos el propósito fundamental de la transparencia es proporcionar independencia de datos en el ambiente distribuido. Se pueden encontrar diferentes aspectos relacionados con la transparencia.

Niveles de Transparencia

La independencia de datos es la inmunidad de las aplicaciones de usuario a los cambios en la definición y/u organización de los datos y viceversa. La independencia de datos se puede dar en dos aspectos: lógica y física.

Niveles de Transparencia

Independencia Lógica

Se refiere a la inmunidad de las aplicaciones de usuario a los cambios en la estructura lógica de la base de datos. Esto permite que un cambio en la definición de un esquema no debe afectar a las aplicaciones de usuario.

Niveles de Transparencia

Independencia Física

Se refiere al ocultamiento de los detalles sobre las estructuras de almacenamiento a las aplicaciones de usuario. Esto es, la descripción física de datos puede cambiar sin afectar a las aplicaciones de usuario.

Transparencia al nivel de red

Se refiere a que los datos en un SBDD se accedan sobre una red de computadoras, sin embargo, las aplicaciones no deben notar su existencia.

Transparencia al nivel de red

Transparencia sobre la localización de datos

Esto es, el comando que se usa es independiente de la ubicación de los datos en la red y del lugar en donde la operación se lleve a cabo.

Transparencia al nivel de red

Transparencia sobre el esquema de nombramiento

Lo anterior se logra proporcionando un nombre único a cada objeto en el sistema distribuido. Así, no se debe mezclar la información de la localización con en el nombre de un objeto.

Transparencia sobre replicación

Se refiere a que si existen réplicas de objetos de la base de datos, su existencia debe ser controlada por el sistema no por el usuario. Se debe tener en cuenta que cuando el usuario se encarga de manejar las réplicas en un sistema, el trabajo de éste es mínimo por lo que se puede obtener una eficiencia mayor. Sin embargo, el usuario puede olvidarse de mantener la consistencia de las réplicas teniendo así datos diferentes.

4 Fragmentación de datos

Fragmentación horizontal

La división o fragmentación horizontal trabaja sobre las tuplas, dividiendo la relación en subrelaciones que contienen un subconjunto de las tuplas que alberga la primera.

Fragmentación vertical

Se basa en los atributos de la relación para efectuar la división.

Fragmentación mixta

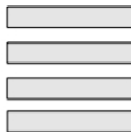
Cuando el proceso de partición hace uso de los dos tipos anteriores. La fragmentación mixta puede llevarse a cabo de tres formas diferentes:

- Desarrollando primero la fragmentación vertical y, posteriormente, aplicando la partición horizontal sobre los fragmentos verticales (denominada partición VH)
- Aplicando primero una división horizontal para luego, sobre los fragmentos generados, desarrollar una fragmentación vertical (llamada partición HV)
- Forma directa considerando la semántica de las transacciones.

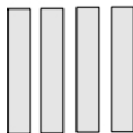
Fragmentación



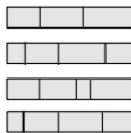
Relación



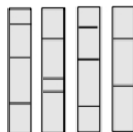
Frag. horizontal



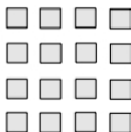
Frag. Vertical



Frag. HV



Frag. VH



Celdas

Fragmentación

- Encontrar la unidad de relación apropiada; ya que el acceso a la aplicación no se hace en la totalidad de las relaciones de la base pero si de un subconjunto de ellas.

Fragmentación

- Si tenemos las sub-relaciones como unidades de relación cuando tratemos de acceder a datos que estén en otros sitios será mas conveniente hacer la replicación de los datos que están en las subrelaciones en lugar de traer los datos de todas las relaciones.

Fragmentación

- Si tomamos las subrelaciones como unidades de relación se pueden ejecutar una cantidad de transacciones concurrentemente, además la fragmentación de relaciones típicamente resulta en la ejecución paralela de un único query dividiéndolo en sub-queries que operan sobre los fragmentos.



Grado de la fragmentación

El grado de fragmentación dependerá de las necesidades de nuestra aplicación la cual correrá en la base de datos.

Reglas para la fragmentación

- **Compleitud:** Esta referida a una descomposición sin pérdida de datos, es decir que se asegure que los datos en una relación global pueden ser mapeados en fragmentos si ningún tipo de perdida.
- **Reconstrucción:** la posibilidad de reconstrucción de una relación global a partir de fragmentos de relaciones asegura que las restricciones definidas sobre los datos en forma de dependencias sean preservadas.
- **Disyunción:** Dice que si particionamos la relación de forma horizontal un dato que tengamos en un fragmento no debe estar presente en algún otro de los fragmentos de la relación.

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.