

Base de Datos II

Unidad 4 – Bases de Datos Distribuidas

Objetivos

- Motivación de BD Distribuidas.
 - Comparación de BD Centralizada y BD Distribuida.
 - Administración de BD Distribuidas
 - Diseño de una Base de Datos Distribuida.
 - Procesamiento distribuido de consultas
-

Motivación BD Distribuidas

Las organizaciones necesitan acceder a los datos/información que se encuentra en diferentes sitios/localizaciones

Los avances las tecnologías de comunicación que permiten un flujo constante entre diferentes sitios.

La creciente globalización de las empresas y la necesidad de compartir información con clientes y proveedores.

BD Distribuidas - Ventajas

- Las empresas ya están distribuidas al menos físicamente y ven como algo natural el enfoque de BDD que refleja la estructura organizacional.
 - El uso mas frecuente de los datos almacenados localmente, brinda una mayor eficiencia en las transacciones.
 - Crecimiento proporcional, cuando se suman nuevas sucursales, sedes, etc de las organizaciones, las BDD proporcionan una solución con mínimo impacto.
-

BD Distribuidas - Ventajas

- Se reduce la sobrecarga de las comunicaciones, ya que mayormente, una alta proporción de las transacciones se resuelve de manera local.
- Mejora el rendimiento, teniendo un alto grado de paralelismo, la carga es distribuida entre los diferentes nodos/sitios.
- Mayor confiabilidad y disponibilidad, al tener redundancia de datos, lo que brinda una alta disponibilidad de los sistemas. Esto requiere además mayores controles,

Comparación - Definición

BD Centralizada	BD Distribuida
Los sistemas de base de datos centralizados son aquellos que se ejecutan en un único Software de Administración de BD. Son accedidos desde diferente terminales o sistemas , pero todas las transacciones se resuelven sobre ese único Software de administración.	Una base de datos distribuida (BDD) es un conjunto de múltiples bases de datos lógicamente relacionadas las cuales se encuentran distribuidas en diferentes espacios lógicos e interconectados por una red de comunicaciones. Tienen la capacidad de realizar procesamiento autónomo, esto permite realizar operaciones locales o distribuidas. Un SBDD es un sistema en el cual múltiples sitios de bases de datos están ligados por un sistema de comunicaciones de tal forma que, un usuario en cualquier sitio puede acceder los datos en cualquier parte de la red exactamente como si estos fueran accedidos de

Comparación - Ventajas	
BD Centralizada	BD Distribuida
<p>Se evita la redundancia. En sistemas que no usan Bases de Datos Centralizadas, cada aplicación tiene sus propios archivos privados o se encuentran en diferentes localidades. Esto a menudo origina enorme redundancia en los datos almacenados, así como desperdicio resultante del espacio de almacenamiento. Se evita la inconsistencia. Ya que si un hecho específico se representa por una sola entrada, la no concordancia de datos no puede ocurrir. Pueden hacerse cumplir las normas establecidas. Con un control central de la base de datos, el Administrador de Base de Datos (DBA) puede garantizar que se cumplan todas las formas aplicables a la representación de los datos.</p>	<p>Autonomía local: un sitio puede controlar los datos que le pertenecen. Disponibilidad: un fallo en una parte del sistema solo afectará a un fragmento, en lugar de a toda la base de datos. Rendimiento: los datos generalmente se ubican cerca del sitio con mayor demanda, también los sistemas trabajan en paralelo, lo cual permite balancear la carga en los servidores. Economía: es más barato crear una red de muchas computadoras pequeñas, que tener una sola computadora muy poderosa.</p>

Comparación - Ventajas	
BD Centralizada	BD Distribuida
<p>Pueden aplicarse restricciones de seguridad. El Administrador de Base de Datos (DBA) puede:</p> <ul style="list-style-type: none"> • Asegurar que el único medio de acceder la base de datos sea a través de los canales establecidos. • Definir controles de autorización para que se apliquen cada vez que se intente el acceso a datos sensibles 	<p>Modularidad: se pueden modificar, agregar o quitar sistemas de la base de datos distribuida sin afectar a los demás sistemas (módulos).</p>

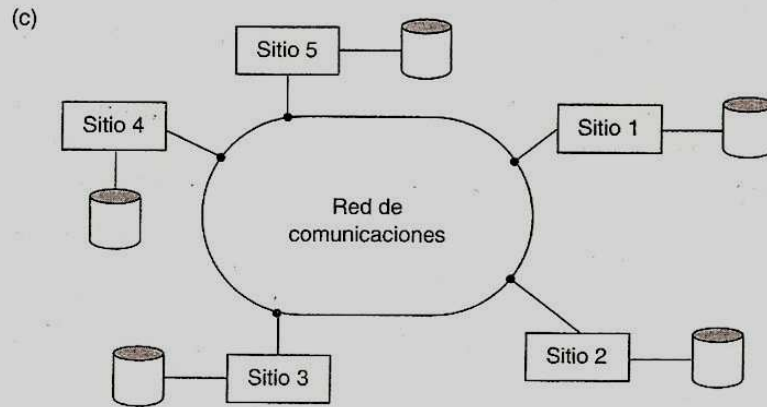
Comparación - Desventajas

BD Centralizada	BD Distribuida
<p>Los mainframes (computadora central) no ofrecen mejor proporción precio/rendimiento que los microprocesadores de los sistemas distribuidos.</p> <p>Cuando un sistema de Base de Datos Centralizada falla, se pierde toda la disponibilidad de procesamiento y sobre todo de la información confiada al sistema.</p> <p>Las cargas de trabajo no se pueden difundir entre diferentes computadoras, ya que los trabajos siempre se ejecutarán en la misma máquina.</p>	<p>Costo de desarrollo del software. La complejidad añadida que es necesaria para mantener la coordinación entre nodos hace que el desarrollo de software sea más costoso.</p> <p>Mayor probabilidad de errores. Como los nodos que constituyen el sistema funcionan en paralelo, es más difícil asegurar el funcionamiento correcto de los algoritmos, así como de los procedimientos de recuperación de fallos del sistema.</p> <p>Mayor sobrecarga de procesamiento. El intercambio de mensajes y ejecución de algoritmos para el mantenimiento de la coordinación entre nodos supone una sobrecarga que no se da en los sistemas centralizados.</p>

Tareas de Administración de una BDD

1. Actividades típicas de toda BD: Seguridad, Integridad, Optimización, etc.
2. Actividades específicas de una BDD:
 1. Descripción del esquema global.
 2. Descripción de la fragmentación/replicación.
 3. Descripción de las localizaciones.
 4. Estadísticas de la BD.

Arquitectura de una BDD



Pasos a seguir para diseñar una BDD

1. Diseño del **"esquema conceptual"** el cual describe la base de datos integrada (esto es, todos los datos que son utilizados por las aplicaciones que tienen acceso a las bases de datos).
2. Diseño **"físico" de la base de datos**, esto es, mapear el esquema conceptual a las áreas de almacenamiento y determinar los métodos de acceso.
3. Diseño de la **fragmentación**, este es, determinar la forma en que las relaciones globales se subdividen en fragmentos horizontales, verticales o mixtos.
4. Diseño de la **asignación de los fragmentos**, esto es, determinar la forma en que los fragmentos se mapean a las imágenes físicas.

Pasos a seguir para diseñar una BDD

1. Diseño del **"esquema conceptual"** el cual describe la base de datos integrada (esto es, todos los datos que son utilizados por las aplicaciones que tienen acceso a las bases de datos).
 2. Diseño **"físico" de la base de datos**, esto es, mapear el esquema conceptual a las áreas de almacenamiento y determinar los métodos de acceso.
 3. Diseño de la **fragmentación**, este es, determinar la forma en que las relaciones globales se subdividen en fragmentos horizontales, verticales o mixtos.
 4. Diseño de la **asignación de los fragmentos**, esto es, determinar la forma en que los fragmentos se mapean a las imágenes físicas.
-

Procesamiento distribuido de consultas

- **NODO1: EMPLEADO**
 - **NODO2: DEPARTAMENTO**
-

Procesamiento distribuido de consultas

*“Por cada empleado, obtener el **nombre del empleado** y el **nombre del departamento** al que pertenece”*

Q1(1): IINombre,Apellido,NombreDPto(EMPLEADO ⋈ DEPARTAMENTO)

- La consulta se lanza desde el **nodo3** (*nodo respuesta*) que no tiene datos implicados en la consulta.
 - El resultado de ésta consulta constará de 10.000 tuplas. Cada tupla resultante será de una longitud de 40 bytes. El tamaño del resultado será por tanto de 400.000 bytes.
 - Existen tres alternativas para resolver la consulta.
-

Procesamiento distribuido de consultas

Primera alternativa:

- Transferir, tanto la relación EMPLEADO, como la relación DEPARTAMENTO al nodo respuesta (**nodo3**) y realizar allí mismo la operación de join.
 - En éste caso se transfieren:
 - **1.000.000 + 3.500 = 1.003.500 bytes.**
-

Procesamiento distribuido de consultas

Segunda alternativa:

- Transferir la relación EMPLEADO al **nodo2**, ejecutar el join en este nodo y enviar el resultado al **nodo3**.
 - Esto implicaría transferir:
 - **$1.000.000 + 400.000 \text{ (resultado)} = 1.400.000 \text{ bytes}$**
-

Procesamiento distribuido de consultas

Tercera alternativa:

- Transferir la relación DEPARTAMENTO al **nodo1**, ejecutar el join en este nodo y enviar el resultado al **nodo3**.
 - En este caso, los bytes transferidos serán:
 - **$3.500 + 400.000 \text{ (resultado)} = 403.500 \text{ bytes}$** .
-

Procesamiento distribuido de consultas

- “Para cada departamento, obtener el nombre del departamento y el de su director”

La consulta se lanza desde el **nodo3**. El resultado de ésta consulta constará de 100 tuplas (4.000 bytes).

Opción 1: transferimos las relaciones **DEPARTAMENTO** y **EMPLEADO** al **nodo3**. Se transfieren:

$$3.500 + 1.000.000 = 1.003.500 \text{ bytes.}$$

Opción 2: transferimos la relación **EMPLEADO** al **nodo2** y enviamos el resultado del join al **nodo3**. Se transfieren:

$$1.000.000 + 4.000 = 1.004.000 \text{ bytes.}$$

Opción 3: transferimos la relación **DEPARTAMENTO** al **nodo1** y enviamos el resultado del join al **nodo3**. Se transfieren:

$$3.500 + 4.000 = 7.500 \text{ bytes.}$$

Procesamiento distribuido de consultas utilizando semijoin

Reducción del número de columnas antes de transferir a otro nodo.

- Se envía la **columna** con la que se va a realizar el join de una **relación R** al **nodo** donde se encuentra la otra relación, allí se realiza el join con la otra relación **S**
 - Se envían las **columnas implicadas en el resultado** al **nodo inicial** y se vuelve a realizar el join con **R**.
 - Sólo se transfieren las columnas de **R** que intervienen en la realización del join en una dirección y el subconjunto de columnas de **S** resultantes en la otra.
-

Objetivos

- ✓ Motivación de BD Distribuidas.
 - ✓ Comparación de BD Centralizada y BD Distribuida.
 - ✓ Administración de BD Distribuidas
 - ✓ Diseño de una Base de Datos Distribuida.
 - ✓ Procesamiento distribuido de consultas
-



¿Dudas,
consultas?

Unidad 4