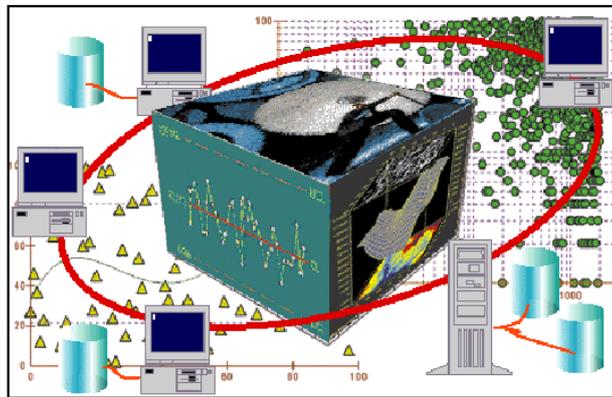


1

BASE DE DATOS DISTRIBUIDOS



Contenido:

Base de Datos Distribuidos

- 1.1. Introducción
 - 1.1.1 Almacenamiento Distribuido
- 1.2. Sistemas de gestión de bases de datos distribuidos
 - 1.2.1 Funciones y Arquitectura de un SGBDD
- 1.3. Las Doce Reglas
- 1.4. Ventajas e Inconvenientes
- 1.5. Base de Datos Distribuidas Vs Centralizada
 - 1.5.1 Base de Datos Centralizada.
 - 1.5.2 Ventajas e Inconvenientes.



CAPITULO I

1. BASE DE DATOS DISTRIBUIDA

1.1. INTRODUCCIÓN

Las Bases de Datos en la actualidad son muy importantes debido a que en ellas se guarda gran información de suma importancia.

Este tema trata de simplificar las consultas de una base de datos, para poder acceder desde cualquier sitio, en cualquier punto de la red tal como si todos los datos estuvieran almacenados en el sitio mismo.

Para esto se debe construir sistemas distribuidos de bases de datos, ya que se están utilizando cada vez más al igual que se usan las arquitecturas de cliente-servidor. Los principales problemas que se generan por el uso de la tecnología de bases de datos distribuidas son en lo referente a duplicidad de datos y su integridad al momento de realizar actualizaciones a los mismos.



1.1.1 ALMACENAMIENTO DISTRIBUIDO

BASES DE DATOS DISTRIBUIDAS.

Es una colección de datos que pertenecen lógicamente a un sólo sistema, pero se encuentra físicamente esparcido en varios "sitios" de la red. Un sistema de base de datos distribuidos se compone de un conjunto de sitios, conectados entre sí mediante algún tipo de red de comunicaciones, en el cual:

- Cada sitio es un sistema de base de datos en sí mismo.
- Los sitios trabajan en conjunto si es necesario con el fin de que un usuario de cualquier sitio pueda obtener acceso a los datos de cualquier punto de la red tal como si todos los datos estuvieran almacenados en el sitio propio del usuario.

OBJETIVOS DE LAS BASES DE DATOS DISTRIBUIDAS

Es una especie de objeto virtual, cuyos componentes se almacenan físicamente en varias bases de datos “reales” distintas, ubicadas en diferentes sitios.

Cada sitio tiene sus propias bases de datos “reales” locales, sus propios usuarios locales, sus propios DBMS y programas para la administración de transacciones y su propio administrador local de comunicación de datos.



Un usuario puede realizar operaciones sobre los datos en su propio sitio local exactamente como si ese sitio no participara en absoluto en el sistema distribuido.

1.2. SISTEMA DE GESTIÓN DE BASE DE DATOS DISTRIBUIDAS (SGBDD)

Es necesaria una aplicación que actúe de interfaz entre el usuario, los modelos y el sistema físico. Esta es la función que desempeñan los SGBD. Una base de datos permite al usuario acceder a datos almacenados sin que él deba saber cómo están organizados esos datos en el sistema, garantiza que ningún usuario sin autorización pueda ver o manipular los datos.

Las principales funciones que debe cumplir un SGBD se relacionan con la creación y mantenimiento de la base de datos, el control de accesos, la manipulación de datos de acuerdo con las necesidades del usuario, el cumplimiento de las normas de tratamiento de datos, evitar redundancias e inconsistencias y mantener la integridad.



1.2.1. FUNCIONES Y ARQUITECTURA DE UN SGBDD

OBJETIVOS DE LOS SGBDD

- **Permitir la definición de todos los datos.** Es decir, incorporar a las estructuras todos los objetos necesarios para completarlas e incluir los atributos necesarios para definir a los objetos.
- **Permitir la manipulación de los datos:** Operaciones de intercambio de datos entre las tablas que pueden ser de consulta o de puesta al día (inserción, modificación, supresión)
- **Establecer controles de seguridad:** Garantizando que sólo los usuarios autorizados puedan efectuar operaciones correctas bien sobre toda la base de datos o sobre algunas tablas.
- **Permitir los accesos concurrentes:** El principal objetivo de la implantación de una base de datos es poner a disposición de un gran número de usuarios un conjunto integrado de datos y que estos datos puedan ser manipulados por los diferentes usuarios. El SGBD debe garantizar que esos datos seguirán siendo coherentes después de las diversas manipulaciones.

Una mala concepción a la hora de diseñar una base de datos puede dar origen a:

- ✓ La redundancia de datos
- ✓ Incoherencia en los datos
- ✓ Pérdida de datos



✓ Estado de la tabla

COMPONENTES DE UN SISTEMA GESTOR DISTRIBUIDO

Hardware

- Los nodos de procesamiento pueden estar formados por múltiples y distintos tipos de hardware.

Software

- El programa principal que necesitamos considerar en un sistema de base de datos distribuida es el DDBMS.
- En un SGBD distribuido, la base de datos se almacena en varios computadores que se pueden comunicar a su vez por distintos medios de comunicación. Se denomina con el término de emplazamientos o nodos a todos aquellos computadores que pertenecen a un sistema distribuido.

Datos

- Una base de datos distribuida puede no estar duplicada, parcialmente duplicada o totalmente duplicada. Una base de datos parcialmente duplicada contiene ciertos elementos de datos que están duplicados y algunos que no lo están.

Procedimientos

- Los sistemas de base de datos distribuidos contienen una multitud de componentes de procesamiento. En los sistemas distribuidos más sencillos, los datos no están duplicados, y sólo el nodo que almacena los datos puede actualizarlos.



Personal

- Los sistemas con un DDBMS complejo y poderoso imponen pocas demandas especiales sobre los usuarios. De hecho, los usuarios no saben que están procesando datos distribuidos.

1.3. DOCE REGLAS DE LAS BASES DE DATOS DISTRIBUIDAS

Principio fundamental: Para el usuario un sistema distribuido debe ser igual que uno centralizado.

1. **Autonomía Local:** Los sitios distribuidos deben ser autónomos, es decir que todas las operaciones en un sitio dado se controlan en ese sitio.
2. **No dependencia de un sitio central:** No debe de haber dependencia de un sitio central para obtener un servicio. La dependencia de un sitio sería indeseable por las siguientes razones: ese sitio podría ser un cuello de botella. El sistema sería vulnerable; si el sitio sufriera un desperfecto, todo el sistema dejaría de funcionar.
3. **Operación Continua:** Nunca debería haber necesidad de apagarse a propósito para que se pueda realizar alguna función, como añadir un nuevo sitio, o instalar una versión mejorada
4. **Independencia con respecto a la localización:** No debe de ser necesario que los usuarios sepan dónde están almacenados



físicamente los datos, el usuario debe de ver como si solo existiera un sitio local.

- 5. Independencia con respecto a la fragmentación:** La fragmentación es deseable por razones de desempeño, los datos, pueden almacenarse en la localidad donde se utilizan con mayor frecuencia de manera que la mayor parte de las operaciones sean sólo locales y se reduzca el tráfico en la red.
- 6. Independencia de réplica:** Si una relación dada, es decir, un fragmento dado de una relación se puede presentar en el nivel físico mediante varias copias almacenadas o réplicas, en muchos sitios distintos. La réplica es viable por dos razones: las aplicaciones pueden operar sobre copias locales en vez de tener que comunicarse con sitios remotos; una mejor disponibilidad. La desventaja principal de las réplicas es, cuando se pone al día un cierto objeto copiado, deben ponerse al día todas las réplicas de ese objeto.
- 7. Procesamiento Distribuido de Consultas:** El objetivo es convertir transacciones de usuario en instrucciones para manipulación de datos, y así reducir el tráfico en la red, esto implica que el proceso mismo de optimización de consultas debe ser distribuido.
- 8. Manejo Distribuido de Transacciones:** Tiene dos aspectos principales, el control de recuperación y el control de



conurrencia, cada uno de los cuales requiere un tratamiento más amplio en el ambiente distribuido.

9. Independencia con respecto al equipo: El SGBDD (Sistema de Gestión de Base de Datos Distribuidos) debe ser ejecutable en diferentes plataformas hardware.

10. Independencia con respecto al Sistema Operativo: El sistema debe ser ejecutable en diferentes SO.

11. Independencia con respecto a la red: El sistema debe poder ejecutarse en diferentes redes.

12. Independencia con respecto al DBMS: No se requiere sino que los DBMS en los diferentes sitios manejen todos la misma interfaz; no necesitan ser por fuerza copias del mismo sistema.

1.4. VENTAJAS E INCONVENIENTES DE LAS BASES DE DATOS DISTRIBUIDAS

VENTAJAS DE LAS BASES DE DATOS DISTRIBUIDAS

- **Compartimiento de datos.** Los usuarios de un nodo son capaces de acceder a los datos de otro nodo.
- **Autonomía.** Cada nodo tiene cierto grado de control sobre sus datos, en un sistema centralizado, hay un administrador del sistema responsable de los datos a nivel global. Cada administrador local puede tener un nivel de autonomía local diferente.



- **Disponibilidad.** Si en un sistema distribuido falla un nodo, los nodos restantes pueden seguir funcionando. Si se duplican los datos en varios nodos, la transacción que necesite un determinado dato puede encontrarlo en cualquiera de los diferentes nodos.

INCONVENIENTES DE LAS BASES DE DATOS DISTRIBUIDAS

- **Coste 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.
- **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.
- **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.



1.5. BASES DE DATOS DISTRIBUIDAS VRS. CENTRALIZADA

Los sistemas de bases de datos centralizados, residen en una sola localidad.

Los sistemas de bases de datos distribuidos se encuentran en varias localidades.

1.5.1. BASE DE DATOS CENTRALIZADA (BDC).

Es una base de datos almacenada en su totalidad en un solo lugar físico, es decir, es una base de datos almacenada en una sola máquina y una sola CPU, en donde los usuarios trabajan en terminales que sólo muestran resultados.

Son aquellos que se ejecutan en un único sistema informático sin interaccionar con ninguna otra computadora.

CARACTERÍSTICAS DE LAS BASES DE DATOS CENTRALIZADAS.

- Se almacena completamente en una localidad central, es decir, todos los componentes del sistema residen en un solo computador o sitio.



- No posee múltiples elementos de procesamiento ni mecanismos de intercomunicación como las Bases de Datos Distribuidas.
- Los componentes de la Bases de Datos Centralizadas son los datos, el software del SGBD (Sistema de Gestión de Base de Datos) y los dispositivos de almacenamiento secundario asociados, como disco de almacenamiento en línea de la base de datos y cintas para las copias de seguridad.
- El problema de seguridad es inherentemente fácil en estos sistemas de bases de datos.

1.5.2. VENTAJAS E INCONVENIENTES

VENTAJAS DE LAS BASES DE DATOS CENTRALIZADAS.

- 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.
- Pueden aplicarse restricciones de seguridad. El Administrador de Base de Datos (DBA) puede:
 - ✓ 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
- Puede conservarse la integridad. Que los datos se encuentren en una sola máquina ayuda a evitar la inconsistencia de los datos, por el mismo hecho de encontrarse en una sola máquina.
- El procesamiento de los datos ofrece un mejor rendimiento y resulta más confiable que en los sistemas distribuidos.

DESVENTAJAS DE LAS BASES DE DATOS CENTRALIZADAS.

- Los mainframes (computadora central) no ofrecen mejor proporción precio/rendimiento que los microprocesadores de los sistemas distribuidos.



- 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.
- En caso de un desastre o catástrofe, la recuperación es difícil de sincronizar.
- Las cargas de trabajo no se pueden difundir entre diferentes computadoras, ya que los trabajos siempre se ejecutarán en la misma máquina.
- No se puede añadir poder de cómputo en pequeños incrementos, debido a lo complicado de esta operación.

COMPARACIÓN

CENTRALIZADO	DISTRIBUIDO
Control centralizado: Un solo DBA	Control jerárquico: DBA global y DBAs locales
Independencia de Datos: Organización de los datos es transparente para el programador	Transparencia en la Distribución: Localización de los datos es un aspecto adicional de independencia de datos
Reducción de redundancia: Una sola copia de datos que se comparta	Replicación de Datos: Copias múltiples de datos que incrementa la localidad y la disponibilidad de datos
Estructuras físicas complejas para accesos eficientes	No hay estructuras intersitios. Uso de optimización global para reducir transferencia de datos
Seguridad	Problemas de seguridad Intrínsecos

Comparación de Base de Datos Distribuida vr Centralizada