

1. INTRODUCCIÓN

Hoy en día el mayor activo de las organizaciones son los datos y su gestión eficaz y segura. Por ello, si analizamos la mayoría de los ámbitos de actividad, nos encontramos que la utilización de las bases de datos está ampliamente extendida (al registrarse en una web, al acudir a la consulta médica, al consultar el catálogo de productos de una tienda online, etc.). Las bases de datos y los datos contenidos en ellas, son imprescindibles para llevar a cabo multitud de acciones. La necesidad de almacenar datos de forma masiva dio paso a la creación de los sistemas de bases de datos.

Las bases de datos originalmente almacenan la información de manera centralizada, pero con el paso del tiempo las necesidades aumentaron y esto produjo ciertos inconvenientes que no era posible solucionarlos o volverlos eficientes de la forma centralizada. Estos problemas impulsaron la creación de sistemas de base de datos de almacenamiento distribuido, los cuales hoy en día proveen características indispensables en el manejo de información.

Lo expuesto anteriormente justifica la importancia del tema donde vamos a empezar definiendo una base de datos y un sistema gestor de base de datos para posteriormente centrarnos en los sistemas de bases de datos distribuidos analizando sus características más importantes.

2. CONCEPTOS BÁSICOS

Podemos definir una **base de datos** como un conjunto, colección o depósito de datos almacenados en un soporte informático no volátil. Los datos están interrelacionados y estructurados.

Un **sistema gestor de base de datos o SGBD** es una colección de programas de aplicación que permite a los usuarios la creación y el mantenimiento de una base de datos, facilitando la definición, construcción y manipulación de la información contenida en ésta.

3. SISTEMAS DE BASE DE DATOS DISTRIBUIDOS. FUNCIONALIDADES. TIPOS.

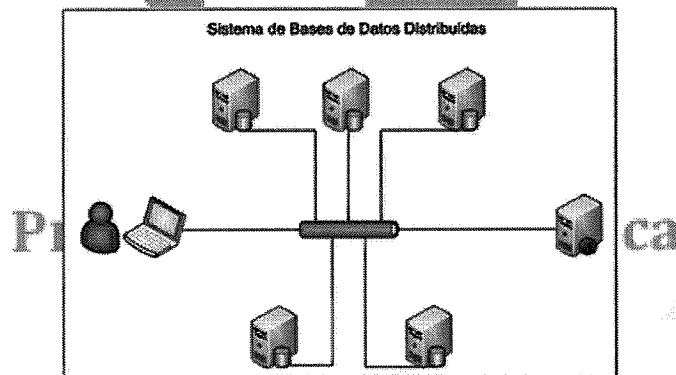
Una **base de datos distribuida (BDD)** es un conjunto de múltiples bases de datos lógicamente relacionadas las cuales se encuentran distribuidas en

diferentes espacios lógicos y geográficos e interconectados por una red de comunicaciones. Dichas BDD tienen la capacidad de realizar procesamiento autónomos, estos permiten realizar operaciones locales o distribuidas.

Un **Sistema de Bases de Datos Distribuida (SBDD)** es un sistema en el cual múltiples sitios de bases de datos están ligados entre sí 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 forma local.

Un sistema distribuido de bases de datos se almacena en varias computadoras. Los principales factores que distinguen un SBDD de un sistema centralizado son los siguientes:

- Hay múltiples computadores, llamados sitios o nodos.
- Estos nodos deben de estar comunicados por medio de algún tipo de red de comunicaciones para transmitir datos y órdenes entre los sitios.



3.1. FUNCIONALIDADES

- Accede a sitios remotos y transmite consultas y datos a través de varios sitios mediante una red de comunicación.
- Almacena el esquema de distribución y replicación de los datos en el catálogo del sistema.
- Establece las estrategias de ejecución de las consultas y las transacciones que acceden a los datos en más de un sitio.
- Decide sobre cual copia de los datos replicados acceder.
- Mantiene la consistencia de las copias de los datos replicados.
- Realiza la recuperación ante los fallos.

3.2. TIPOS

En un sistema de bases de datos distribuidas, existen varios factores que deben tomar en consideración que definen la arquitectura del sistema:

- **Distribución:** Los componentes del sistema están localizados en la misma computadora o no.
- **Heterogeneidad:** Un sistema es heterogéneo cuando existen en él componentes que se ejecutan en diversos sistemas operativos, de diferentes fuentes, etc.
- **Autonomía:** Se puede presentar en diferentes niveles, los cuales se describen a continuación:
 - **Autonomía de diseño:** Habilidad de un componente del sistema para decidir cuestiones relacionadas a su propio diseño.
 - **Autonomía de comunicación:** Habilidad de un componente del sistema para decidir cómo y cuándo comunicarse con otros SGBD.
 - **Autonomía de ejecución:** Habilidad de un componente del sistema para ejecutar operaciones locales como quiera.

4. COMPONENTES

Respecto al **hardware** utilizado debemos decir que no difiere mucho del hardware utilizado en un servidor normal. El hardware que compone una base de datos distribuida se reduce a servidores y la red.

Respecto a los componentes **software** debemos distinguir los siguientes:

- **Sistema manejador de base de datos distribuida (DDBMS):** Este sistema está formado por las transacciones y los administradores de la base de datos distribuidos. Un DDBMS implica un conjunto de programas que operan en diversas computadoras, estos programas pueden ser subsistemas de un único DDBMS de un fabricante o podría consistir de una colección de programas de diferentes fuentes.

- **Administrador de transacciones distribuidas (DTM):** Este es un programa que recibe las solicitudes de procesamiento de los programas de consulta o transacciones y las traduce en acciones para los administradores de la base de datos. Los DTM se encargan de coordinar y controlar estas acciones.

El manejador de transacciones es el encargado de definir la estructura de las transacciones, mantener la consistencia en la base de datos cuando se ejecuta una transacción o se cancela la ejecución de una, mantener protocolos de fiabilidad, implementar algoritmos para el control de la concurrencia y sincronizar las transacciones que se ejecutan simultáneamente.

- **Sistema manejador de base de datos (DBMS):** Es un programa que procesa cierta porción de la base de datos distribuida. Se encarga de recuperar y actualizar datos del usuario y generales de acuerdo con los comandos recibidos de los DTM.

- **Nodo:** Un nodo es una computadora que ejecuta un DTM o un DBM o ambos. Un nodo de transacción ejecuta un DTM y un nodo de base de datos ejecuta un DBM.

5. VENTAJAS E INCONVENIENTES

Ventajas:

- Refleja una estructura organizacional: los fragmentos de la base de datos se ubican en los departamentos a los que tienen relación.
- Autonomía local: un departamento 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.
- Modularidad: se pueden modificar, agregar o quitar sistemas de la base de datos distribuida sin afectar a los demás sistemas (módulos).

Inconvenientes:

- Complejidad: Se debe asegurar que la base de datos sea transparente, se debe lidiar con varios sistemas diferentes que pueden presentar dificultades únicas. El diseño de la base de datos se tiene que trabajar tomando en cuenta



su naturaleza distribuida, por lo cual no podemos pensar en hacer joins que afecten varios sistemas.

- **Economía:** la complejidad y la infraestructura necesaria implica que se necesitará una mayor mano de obra.
- **Seguridad:** se debe trabajar en la seguridad de la infraestructura, así como cada uno de los sistemas.
- **Integridad:** Se vuelve difícil mantener la integridad, aplicar las reglas de integridad a través de la red puede ser muy caro en términos de transmisión de datos.
- **Falta de experiencia:** las bases de datos distribuidas son un campo poco común por lo cual no existe mucho personal con experiencia o conocimientos adecuados.
- **Carencia de estándares:** aún no existen herramientas o metodologías que ayuden a los usuarios a convertir un DBMS centralizado en un DBMS distribuido.

6. DISEÑO DE SISTEMAS DE BASE DE DATOS DISTRIBUIDOS

El diseño de sistemas de bases de datos distribuidos se refiere, en general, a tomar decisiones acerca de la ubicación de datos y programas a través de los diferentes sitios de una red de computadoras. Para el diseño de las bases de datos distribuidas se tienen que considerar los dos problemas siguientes:

- 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

Cuando realizamos el diseño queremos cumplir varios objetivos importantes:

- **Procesamiento local:** La distribución de los datos, para maximizar el procesamiento local corresponde al principio simple de colocar los datos tan cerca como sea posible de las aplicaciones que los utilizan.
- **Distribución de la carga de trabajo:** Esta distribución de la carga se realiza para aprovechar mejor las diferentes características (potenciales) o

utilizaciones de las computadoras de cada sitio, y maximizar el grado de ejecución de paralelismo de las aplicaciones.

- **Costo de almacenamiento y disponibilidad:** La distribución de la base de datos refleja el costo y disponibilidad del almacenamiento en diferentes sitios. Para esto, es posible tener sitios especializados en la red para el almacenamiento de datos. Sin embargo, el costo de almacenamiento de datos no es tan relevante si éste se compara con el del CPU, I/O y costos de transmisión de las aplicaciones.

6.1. DISTINTOS ENFOQUES PARA EL DISEÑO

Existen dos estrategias generales para abordar el problema de diseño de bases de datos distribuidas:

- **El enfoque de arriba hacia abajo (top-down):** Este enfoque es más apropiado para aplicaciones nuevas y para sistemas homogéneos. Consiste en partir desde el análisis de requerimientos para definir el diseño conceptual y las vistas de usuario. A partir de ellas se define un esquema conceptual global y los esquemas externos necesarios. Se prosigue con el diseño de la fragmentación de la base de datos, y de aquí se continúa con la localización de los fragmentos en los sitios, creando las imágenes físicas. Esta aproximación se completa ejecutando en cada sitio, "el diseño físico" de los datos, que se localizan en éste.
- **El diseño de abajo hacia arriba (bottom-up):** Se utiliza particularmente a partir de bases de datos existentes, generando con esto bases de datos distribuidas. En forma resumida, el diseño bottom-up de una base de datos distribuida requiere de la selección de un modelo de bases de datos común para describir el esquema global de la base de datos. Después se hace la traducción de cada esquema local en el modelo de datos común y finalmente se hace la integración del esquema local en un esquema global común

El diseño de las bases de datos distribuidas introduce dos nuevos problemas: cómo particionar la base de datos en fragmentos que puedan ser replicados en diferentes sitios y dónde ubicar esos fragmentos.

La fragmentación de datos y la replicación de datos atacan el primer problema y la ubicación de los datos ataca el segundo problema.



6.2. FRAGMENTACIÓN DE DATOS

La fragmentación de los datos nos permite romper un objeto en uno o más segmentos o fragmentos. El objeto puede ser almacenado en cualquier sitio en la red de computadoras. La información de la fragmentación de datos es almacenada en el catálogo de datos distribuidos (DDC), al cual tendrá acceso el procesador de transacciones (TP) para procesar las peticiones de usuarios.

Existen diversos tipos de fragmentación:

- **Fragmentación Horizontal:** La fragmentación horizontal se realiza sobre las tuplas de la relación. Cada fragmento será un subconjunto de las tuplas de la relación.
- **Fragmentación Vertical:** La fragmentación vertical produce fragmentos los cuales contienen un subconjunto de los atributos, así como la clave principal de la relación. El objetivo de este tipo de fragmentaciones es particionar una relación en un conjunto de relaciones más pequeñas.
- **Fragmentación Mixta:** En la mayoría de los casos una simple fragmentación horizontal o vertical de un esquema de base de datos no será suficiente para satisfacer los requerimientos de una aplicación de usuario. En este caso una fragmentación vertical debería ir seguida por una horizontal, o viceversa, produciendo una partición con estructura de árbol. La fragmentación mixta requiere particionamiento de las tablas tanto horizontal como vertical. Supongamos que una compañía opera en tres diferentes lugares. En cada uno de estos lugares los departamentos de servicio y distribución operan en diferentes edificios. Cada uno de los departamentos requiere tener acceso a información acerca de sus clientes locales. La estructura de la compañía requiere entonces que los datos sean fragmentados horizontalmente para contener la información de los diferentes lugares y dentro de cada lugar, los datos deben estar fragmentados verticalmente para contener los diferentes departamentos. En este caso, se requiere fragmentación mixta.
- **Fragmentación Simultánea:** Consiste en aplicar sobre una relación, de forma simultánea y no secuencial, la fragmentación horizontal y la fragmentación vertical.

6.3. REPLICACIÓN DE DATOS

El sistema conserva varias copias o réplicas idénticas de una tabla. Cada réplica se almacena en un nodo diferente. La replicación de datos se refiere al almacenamiento de varias copias de datos en múltiples sitios conectados por una red de computadoras. Las copias de los fragmentos pueden estar almacenadas en diferentes sitios dependiendo de ciertos requerimientos de información específicos. Dado que existen copias de los fragmentos la disponibilidad de los datos y el tiempo de respuesta pueden ser mejores, reduciendo tiempo de red y costos de consultas

Existen tres condiciones de replicación:

a) Una base de datos **completamente replicada**: La cual almacena múltiples copias de cada fragmento de base de datos en múltiples sitios. En este caso todos los fragmentos de base de datos están replicados. Las base de datos completamente replicadas tienen a ser poco prácticas por la cantidad de carga que representa para el sistema.

b) Una base de datos **parcialmente replicada**: La cual almacena múltiples copias de algunos fragmentos de base de datos en múltiples sitios. La mayoría de los DDBMSs son capaces de administrar y manipular la base de datos parcialmente replicada bien.

c) Una base de datos que **no es replicada** almacena cada fragmento de base de datos en un solo sitio. Por tanto, no existen fragmentos duplicados.

Existen dos factores que permiten decidir si usar datos replicados o no:

a) El tamaño de la base de datos.

b) Frecuencia de utilización.

Cuando la frecuencia de utilización de datos ubicados remotamente es alta y la base de datos es grande, la replicación de datos puede ayudar a reducir los costos de acceso a datos. La información de los datos replicados se encuentra en el catálogo de los datos distribuidos DDC, cuyo contenido es usado por el procesador de transacciones TP para decidir cuál copia de fragmento de base de datos acceder.



6.4. UBICACIÓN DE LOS DATOS

Para la ubicación de los datos se deben observar tres estrategias de alojamiento:

- a) Centralizado: Toda la base de datos esta almacenada en un solo sitio.
- b) Particionada: La base de datos está dividida en varios fragmentos y almacenados en diversos sitios.
- c) Replicada: Copias de uno o más fragmentos de base de datos están almacenados en diversos sitios.

Respecto a los algoritmos de colocación de datos debemos decir que toman en consideración una gran variedad de factores como:

- a) Requerimientos de rendimiento y disponibilidad de datos.
- b) Tamaño, número de registros y el número de relaciones que una entidad mantiene con otras entidades.
- c) Tipos de transacciones a ser aplicados a la base de datos, los atributos accedidos por cada una de estas transacciones, etc.

7. CONCLUSIÓN

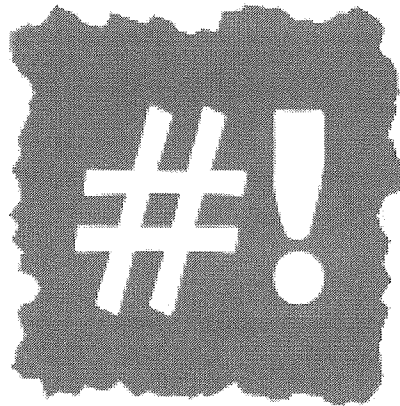
Hay varios factores que han hecho que las bases de datos evolucionen a bases de datos distribuidas. En el mundo de los negocios se ha dado una globalización y a la vez las operaciones de las empresas son cada vez más descentralizadas geográficamente. Además, la necesidad de compartir datos ha hecho que crezca el mercado de las bases de datos distribuidas.

En la actualidad las organizaciones que tienen numerosas oficinas en distintas ubicaciones geográficas suelen utilizar bases de datos distribuidas. Normalmente, una sucursal individual interactúa principalmente con los datos que pertenecen a sus propias operaciones, con una necesidad mucho menos frecuente de datos generales de la empresa.

En este tema se ha presentado una visión global sobre los sistemas de bases de datos distribuidos, con el objetivo de conocer sus características y funcionamiento debido a que han adquirido bastante importancia en el mundo de las bases de datos.

8. BIBLIOGRAFÍA

- Date D.J.: **Introducción a los sistemas de bases de datos**. Editorial Addison-Wesley
- De Miguel A,y Piattini M:**Fundamentos y modelos de BBDD**. Edit. Ra-Ma
- Korth H. y Silberschatz: **Fundamentos de bases de datos**. Editorial McGraw-Hill
- <https://modelosbd2012t1.wordpress.com/2012/03/08/bases-de-datos-distribuidas/>
- <http://profesores.fi-b.unam.mx/pilarang/docencia/Notas-BDDistribuidas.pdf>



Preparador Informática

