
INSTITUTO POLITÉCNICO NACIONAL
UNIDAD PROFESIONAL INTERDISCIPLINARIA EN INGENIERÍA Y
TECNOLOGÍAS AVANZADAS

BASES DE DATOS DISTRIBUIDAS

Tarea Escrita 1
Investigación de Bases de Datos Distribuidas

Realizado por:
De Jesús Lucio Oscar Daniel
Pérez Iturbe Carolina

Profesor:
Carlos De La Cruz Sosa

Fecha de entrega: 14 de marzo de 2025
Grupo: 3TM3

Contents

1	Introducción	3
2	Conceptos de una Base de Datos Distribuida	3
3	Diferencias entre BD / BDD / BD-nube	4
4	Arquitecturas de una base de datos	4
5	Estrategias para distribuir una base de datos	5
6	Gestores y funciones para bases de datos distribuidas	6
6.1	Gestores	6
6.2	Funciones de un gestor de base de datos	6
7	Bibliografía	7

1 Introducción

BASE: Fundamento o apoyo principal de algo.

DATO: Información sobre algo concreto que permite su conocimiento exacto o sirve para deducir las consecuencias derivadas de un hecho.

BASE DE DATOS: Memoria informática controlada por un sistema de gestión de bases de datos (DBMS). El término base de datos surgió en 1963. En informática, una base de datos consiste en una colección de datos interrelacionados y un conjunto de programas para acceder a dichos datos. En otras palabras, una base de datos no es más que un conjunto de información (conjunto de datos) relacionada que se encuentra agrupada o estructurada. "Es una colección organizada de datos" (Daitel y Deitel, 2008)

(1) (2)

Una base de datos es una herramienta para recopilar y organizar información. Las bases de datos pueden almacenar información sobre personas, productos, pedidos u otras cosas. Muchas bases de datos comienzan como una lista en una hoja de cálculo o en un programa de procesamiento de texto. (3) Las bases de datos distribuidas se diferencian de las bases de datos centralizadas en que los datos se almacenan en varios nodos o ubicaciones geográficas.

En el presente trabajo, se explorará en profundidad el mundo de las bases de datos, con un enfoque particular en las bases de datos distribuidas. Se analizarán las diferencias clave entre las bases de datos centralizadas y distribuidas, así como los conceptos fundamentales que sustentan este tipo de sistemas. Además, se examinarán las diversas arquitecturas utilizadas en las bases de datos distribuidas, las estrategias para migrar de un sistema a otro y los gestores de bases de datos más eficaces para su implementación y manejo.

2 Conceptos de una Base de Datos Distribuida

Conceptos generales de BDD

- **Definición:** Una base de datos distribuida es un sistema de gestión de bases de datos en el que las bases de datos no están ubicadas en un sistema de servidor común. Las bases de datos pueden almacenarse en varios servidores de datos ubicados en la misma ubicación física, o pueden distribuirse a través de una red de área local o amplia.

Es una base de datos almacenada en una o más computadoras (4) (2)

- **Autonomía local** Los sitios distribuidos deben ser autónomos, es decir, que todas las operaciones en un sitio dado se controlan en ese sitio, pues cuenta con su propio SGBD (5).
- **Transparencia de ubicación** Permite a los usuarios tener acceso a los datos sin que tengan conocimiento de la ubicación de éstos.
- **Transparencia de duplicación** Los administradores de transacciones deben traducir las solicitudes de procesamiento de transacción en acciones para el administrador de datos.
- **Transparencia de concurrencia:** Cuando varias transacciones se ejecuten al mismo tiempo, los resultados de las transacciones no deberán afectarse.
- **Particionamiento de la Base de Datos:** La base de datos se distribuye de modo que no haya solapamiento o duplicación de los datos mantenidos en las diferentes localidades.
- **Fragmentación de datos:** Consiste en subdividir las relaciones y distribuirlas entre los sitios de la red. Tiene como objetivo buscar formas alternativas de dividir una de las instancias (tablas) de relaciones en otras más pequeñas.
- **Replicación de datos:** Se refiere al almacenamiento de copias de datos en sitios múltiples servidos por una red de computadoras.

- **Base de datos totalmente replicada:** Guarda varias copias de cada fragmento de la base de datos en varios sitios. No es práctica debido a la cantidad de carga impuesta al sistema.
 - **Base de datos parcialmente replicada:** Guarda múltiples copias de algunos fragmentos de la base de datos en múltiples sitios. Se tiene un buen manejo.
 - **Base de datos no replicada:** Guarda cada fragmento de base de datos en un solo sitio.
- ...(6)

3 Diferencias entre BD / BDD / BD-nube

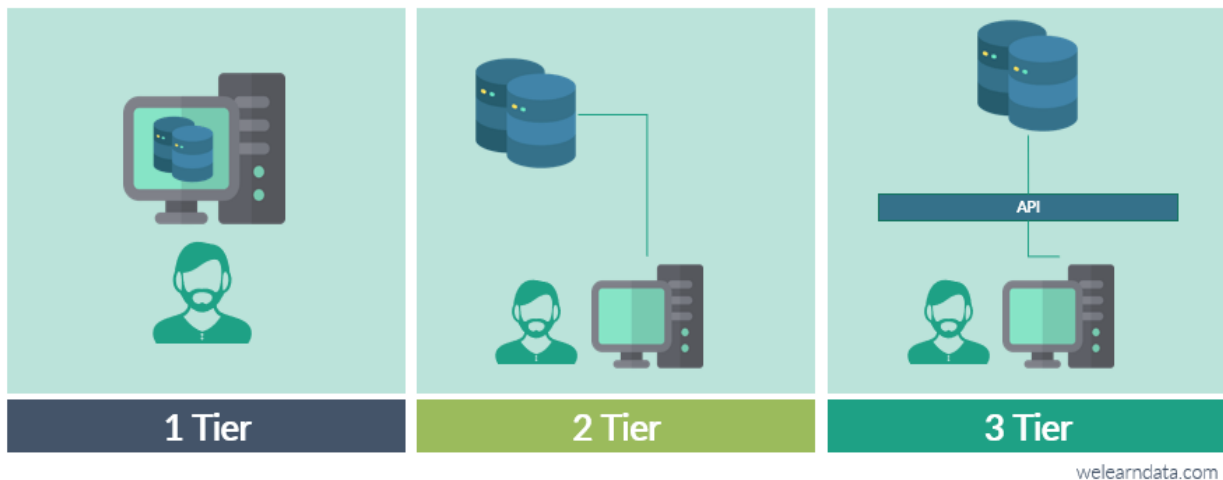
Principales diferencias entre bases de datos			
Aspecto	BD (Base de Datos)	BDD (Base de Datos Distribuida)	BD-Nube (Base de Datos en la Nube)
Ubicación de los datos	En una sola máquina o servidor	Distribuidos entre múltiples nodos	En servidores remotos en la nube
Escalabilidad	Limitada, generalmente escalabilidad vertical	Escalabilidad horizontal (agregar más nodos)	Escalabilidad automática y flexible
Redundancia	No siempre, depende de la configuración	Alta, ya que los datos se replican entre nodos	Alta, con copias automáticas en la nube
Mantenimiento	Requiere gestión manual de hardware y software	Gestión compleja, especialmente en redes grandes	Proveedor de la nube gestiona la infraestructura
Ejemplos	MySQL, PostgreSQL, SQLite	Cassandra, MongoDB (en modo distribuido)	Amazon RDS, Google Cloud SQL, MongoDB Atlas
Acceso y disponibilidad	Limitada a la ubicación física del servidor	Alta disponibilidad, pero depende de la configuración	Alta, acceso desde cualquier lugar con Internet

(7)

4 Arquitecturas de una base de datos

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. Autonomía de diseño (habilidad de un componente para decidir cuestiones relacionadas a su propio diseño). Autonomía de comunicación (habilidad de un componente para decidir cómo y cuándo comunicarse con otros SGBD). Autonomía de ejecución (habilidad de un componente para ejecutar operaciones locales como quiera). (6)



- **Arquitecturas de memoria compartida [Tier 1]** Se trata de una arquitectura en la que el servidor, el cliente y la base de datos residen en el mismo equipo. (Puede ser útil para entornos de prueba, pero no es común en producción).
- **Arquitectura de disco compartido [Tier 2]** La típica aplicación cliente-servidor. La capa de presentación de datos se encuentra en otra máquina o máquinas, independientemente del servidor donde reside la base de datos. (La base de datos no está expuesta a los servicios de cliente).
- **Arquitecturas de nada compartido [Tier 3]** Este es el más común. Separamos la capa de presentación (los usuarios), la capa de aplicación, que actúa como intermediaria comunicando las peticiones de los usuarios con la base de datos, y la capa con la base de datos en otro servicio donde se definirían las relaciones entre las tablas de la base de datos. (La capa de aplicación es donde se define la lógica de comunicación entre el usuario y la base de datos).

Las bases de datos distribuidas utilizan una arquitectura cliente/servidor para procesar las solicitudes de información.

Arquitectura de base de datos cliente/servidor Un servidor de base de datos es el software que administra una base de datos, y un cliente es una aplicación que solicita información a un servidor. Cada computadora en una red es un nodo que puede alojar una o más bases de datos. Cada nodo en un sistema de base de datos distribuida puede actuar como cliente, servidor o ambos, según la situación. Un cliente puede conectarse directa o indirectamente a un servidor de base de datos.

Una conexión directa se produce cuando un cliente se conecta a un servidor y accede a la información de una base de datos contenida en ese servidor. Por el contrario, una conexión indirecta se produce cuando un cliente se conecta a un servidor y luego accede a la información contenida en una base de datos en un servidor diferente.

5 Estrategias para distribuir una base de datos

- **Partición horizontal:** A menudo denominada fragmentación, esta estrategia implica dividir una base de datos en fragmentos y distribuirlos entre varios servidores. Cada fragmento funciona de forma independiente, lo que puede mejorar significativamente el rendimiento.
- **Partición vertical:** esta técnica divide una base de datos en diferentes tablas según la frecuencia de acceso o la correlación entre diferentes columnas. Es particularmente útil para separar los datos a los que se accede con frecuencia de la información menos crítica.

-
- **Partición basada en directorios:** al utilizar un directorio para realizar un seguimiento de dónde se almacena cada partición de datos, este método permite una distribución de datos flexible y dinámica.
 - **Partición Round-Robin:** un método de equilibrio de carga simple pero eficaz en el que cada nuevo registro de datos se asigna al siguiente servidor de la secuencia. Esto garantiza una distribución uniforme de la carga
 - **Partición basada en hash:** los datos se asignan a las particiones en función del resultado de una función hash aplicada a una determinada clave. Esto se puede visualizar a través del proceso de asignación de estudiantes a dormitorios en función del hash de sus números de identificación de estudiante, asegurando una distribución equitativa.

6 Gestores y funciones para bases de datos distribuidas

6.1 Gestores

Un sistema gestor de base de datos (SGBD) o Database Management System (DBMS) es un conjunto de programas invisibles para el usuario final con el que se administra y gestiona la información que incluye una base de datos. Según el modo en que se organizan los datos almacenados para algún proyecto, disponemos de diferentes gestores que se adaptan mejor a cada tipo (tipología).

- Con base en modelos de datos; Sistema jerárquico, Sistema en red, Sistema relacional, Sistema multidimensional, Sistema de objetos
- Con base en la cantidad de usuarios: el sistema de gestión de base de datos puede usarse para que solo un usuario acceda a la vez o para que admita varios.
- Con base en su distribución; Sistema de gestión de datos centralizados y Sistema de gestión de datos distribuidos.

(8)

6.2 Funciones de un gestor de base de datos

las principales funciones a considerar son

- Distribución y Fragmentación de Datos; Los datos se dividen y almacenan en múltiples nodos mediante fragmentación horizontal (por filas), vertical (por columnas) o una combinación de ambas.
- Replicación y Consistencia; Se crean copias de los datos en diferentes nodos para mejorar la disponibilidad y tolerancia a fallos. La consistencia se mantiene mediante modelos como la consistencia fuerte o eventual, asegurando que las réplicas reflejen los cambios de manera adecuada.
- Procesamiento y Optimización de Consultas; Las consultas se descomponen en subconsultas que se ejecutan en paralelo en distintos nodos. La optimización considera factores como la ubicación de los datos y el uso de índices para mejorar el rendimiento.
- Gestión de Transacciones Distribuidas; Se garantiza que las transacciones cumplan con las propiedades ACID (Atomicidad, Consistencia, Aislamiento y Durabilidad) incluso cuando involucran múltiples nodos, utilizando protocolos de compromiso y control de concurrencia distribuidos.
- Seguridad y Control de Acceso; Se implementan mecanismos de autenticación, autorización, cifrado y auditoría para proteger la integridad y confidencialidad de los datos en un entorno distribuido.
- Tolerancia a Fallos y Recuperación; El sistema está diseñado para continuar operando incluso si algunos nodos fallan, utilizando técnicas como la replicación y la recuperación automática para minimizar la pérdida de datos y el tiempo de inactividad.
- Escalabilidad y Flexibilidad; Permite agregar o eliminar nodos fácilmente, adaptándose a las necesidades cambiantes de almacenamiento y procesamiento sin afectar la disponibilidad del sistema.

(9)

7 Bibliografía

- 1.- ¿Qué es una base de datos? (2020, noviembre 24). Oracle.com; Oracle. <https://www.oracle.com/mx/database/what-is-database/>
- 2.- Notas del curso Bases de Datos. Gómez Fuentes, M. (2013). Bases de Datos. Uam.mx. <http://ilitia.cua.uam.mx:8080/jspui/bitstream/123456789/177/1/X190.pdf>
- 3.- Conceptos básicos sobre bases de datos. (s/f). Microsoft.com. Recuperado de <https://support.microsoft.com/es-es/topic/conceptos-b%C3%A1sicos-sobre-bases-de-datos-a849ac16-07c7-4a31-9948-3c8>
- 4.- Db2 11.1. (2022, noviembre 28). Ibm.com. <https://www.ibm.com/docs/es/db2/11.1?topic=managers-designing-distributed-databases>
- 5.- 1.1. Conceptos de base de datos distribuidas. (2016, diciembre 1). Tópicos de base de datos.<https://karenkurapikadb.wordpress.com/2016/12/01/1-1-conceptos-de-base-de-datos-distribuidas/>
- 6.-I UNIDAD BASES DE DATOS DISTRIBUIDAS. (s/f). Academia.edu. Recuperado de https://www.academia.edu/15492427/I_UNIDAD_BASES_DE_DATOS_DISTRIBUIDAS
- 7.- <https://www.oracle.com/mx/database/what-is-database/><https://www.datacamp.com/es/blog/types-of-databases-overview>
- 8.-<https://intelequia.com/es/blog/post/gestor-de-base-de-datos-qu-es-funcionalidades-y-ejemplos#:~:text=Qu%20es%20un%20sistema%20gestor,incluye%20una%20base%20de%20datos.>
- 9.-https://www-geeksforgeeks-org.translate.goog/functions-of-distributed-database-system/?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=tc