



INSTITUTO POLITÉCNICO NACIONAL



**UNIDAD PROFESIONAL INTERDISCIPLINARIA EN
INGENIERÍA Y TECNOLOGÍAS AVANZADAS**

UPIITA

PROFESOR: Carlos De La Cruz Sosa

ASIGNATURA: Bases de Datos Distribuidas

GRUPO: 3TM3

Tarea I

EQUIPO 5:

- Bernal Aguilar Yuvia Abigail
- Contreras Jimenez Mariana Montserrat
- Medina Gómez Jimena Zarahí

FECHA: 12/03/2025

Conceptos de bases de datos distribuidos

Una Base de Datos Distribuida (BDD) es un conjunto de bases de datos lógicamente relacionadas que están físicamente almacenadas en distintos nodos de una red. En otras palabras, es un sistema en el que los datos no residen en un solo lugar, sino que se distribuyen en múltiples servidores interconectados, permitiendo mayor escalabilidad, tolerancia a fallos y acceso optimizado según la ubicación de los usuarios.

Algunos conceptos generales que parten de ahí son:

- Arquitectura de BDD: Donde se define de manera general la manera de distribuir una base de datos, dependiendo de su estructura y nivel de independencia de cada nodo. (Más adelante se encontrará el concepto más específico y mejor explicado.)
- Fragmentación de datos: Optimiza el acceso y el almacenamiento para que se puedan dividir los datos.

Dentro de esta sección se encuentra el fragmento horizontal (Donde cada nodo almacena sólo un subconjunto de filas de una tabla), fragmento vertical (Cada nodo almacena solo algunas columnas de una tabla), fragmento mixto (Es la combinación de ambos fragmentos, es decir horizontal y vertical) y por último la replicación (Creación de copias de los datos en varios nodos).

- Replicación y Consistencia: Se refiere a que existen varias copias de los datos en distintos nodos.

Existen diferentes tipos de consistencia como son la consistencia fuerte (Todas las copias de los datos se actualizan de inmediato), consistencia eventual (Los cambios se tardan en reflejarse por lo que los nodos llegan a tener diferentes versiones) y por último la consistencia relajada (En la que se personalizan las reglas en relación a el rendimiento, así como los datos se sincronizan de manera periódica).

Características principales de las BDD:

- Autonomía local: Sitios distribuidos que son autónomos, donde todas las operaciones de un sitio se controlan solo en ese sitio.
- Réplicas: Se tienen copias de los datos y estos se almacenan en los sitios requeridos.
- Fragmentación: los datos pueden almacenarse en la localidad donde se utilizan con mayor frecuencia y las operaciones generalmente serán locales.
- Transparencia de localización de datos: No es necesario que los usuarios sepan dónde están almacenados físicamente los datos, sino que el usuario debe verlo como si solo existiera un sitio local.

- Independencia con respecto a la red: Se permitirá leer o escribir datos que se encuentren en diferentes nodos.
- Dos tipos de transacciones: Locales, cuando se accede a los datos del único sitio y globales que es cuando se accede a los datos de diferentes sitios.

Diferencias entre Bases de Datos y Base de Datos Distribuidas

De manera general su principal diferencia es la ubicación de los datos, ya que en una Base de Datos (BD) sus datos se encontraran en un solo lugar mientras que en una Base de Datos Distribuidas (BDD) los datos se encontraran en diferentes localidades.

Ya de manera más particular sobre estas diferencias están presentes en muchos aspectos, pero por mencionar algunos son los siguientes:

- **Fiabilidad y tolerancia a fallos:**
 - Centralizada: Si el servidor falla, la base de datos queda inaccesible.
 - Distribuida: Si un nodo falla, otros pueden seguir operando.
- **Procesamiento de Consultas:**
 - Centralizada: Todas las consultas se ejecutan en un solo servidor, lo que facilita la optimización y ejecución de las mismas.
 - Distribuida: Las consultas pueden requerir datos de múltiples nodos, lo que implica dividir las, ejecutarlas en diferentes sitios y luego combinar los resultados.
- **Gestión de Transacciones:**
 - Centralizada: Se ejecutan de manera secuencial o con bloqueo en un solo sistema, garantizando atomicidad y consistencia con facilidad.
 - Distribuida: Se usan protocolos como 2PC (Two-Phase Commit) o 3PC (Three-Phase Commit) para coordinar transacciones entre nodos y garantizar la atomicidad.
- **Rendimiento y escalabilidad:**
 - Centralizada: Puede volverse lenta con muchas solicitudes y es difícil de escalar.

- Distribuida: Mejora el rendimiento al repartir la carga y permite agregar más nodos fácilmente.
- Consistencia:
 - Centralizada: Siempre mantiene datos consistentes.
 - Distribuida: Puede usar diferentes niveles de consistencia (fuerte, eventual, relajada).
- Costo y mantenimiento:
 - Centralizada: Más barata y fácil de administrar.
 - Distribuida: Puede ser más costosa y compleja de gestionar

Base de Datos en la NUBE

Una Base de Datos en la Nube es aquella que se implementa, gestiona y accede a través de plataformas en la nube. En este entorno, se organizan y almacenan datos tanto estructurados (como en bases de datos relacionales) como no estructurados (como en bases de datos NoSQL). Las bases de datos en la nube están diseñadas para operar en entornos públicos o híbridos, dependiendo de la necesidad y el modelo de servicio elegido. Estas bases de datos pueden ejecutarse en plataformas como AWS, Google Cloud, Azure, IBM Cloud, entre otras.

Algunas de sus características son:

- Escalabilidad Automática: Es la BD diseñada para que los recursos se ajusten de manera eficaz.
- Alta Disponibilidad y Recuperación ante Desastres: Las plataformas en la nube hacen que tus datos se almacenen en diferentes puntos por si el servidor falla se pueda recuperar los datos por medio de estas copias.
- Seguridad: cifrado de datos en tránsito y en reposo, autenticación multifactor, auditoría de acceso y firewalls avanzados.
- Soporte para Diversos Tipos de Bases de Datos: Los proveedores de bases de datos en la nube ofrecen soporte para diferentes tipos de bases de datos, como relacionales (SQL), NoSQL (como MongoDB, Firebase), y bases de datos en memoria (como Redis).

Arquitectura de una base de datos

En una arquitectura de una base de datos distribuida se busca establecer una estructura de un sistema en donde se identifican los componentes junto con sus funciones y sus interacciones entre éstas.

Estos sistemas se encuentran la mayoría de veces regidos por la arquitectura ANSI/SPARC (American National Standards Institute/Standards Planning and Requirements Committee) el cual fue propuesto en 1975. Esta arquitectura se divide en tres niveles:

1. Externo o Nivel de vistas: Los usuarios visualizan una parte de la base de datos. Se permite crear vistas de datos excluyendo datos irrelevantes.
2. Conceptual o Lógico: Describe cómo están relacionadas las tablas de la base de datos para un sistema de información. Aquí se hace uso del administrador y es el único nivel donde existe una visión global de la base de datos. Es independiente del hardware.
3. Interno: Describe como los datos se almacenan de forma física y el hardware utilizado para este proceso

Es necesario establecer un esquema de estandarización y para hacer esto, se necesita un marco de trabajo en específico para de este modo dividir el trabajo en piezas más pequeñas y demostrar cómo se relacionan entre ellas. Existen tres marcos:

1. Basado en componentes: Se establecen los componentes que existen en el sistema, la relación que existen entre ellas y de igual forma su funcionalidad.
2. Basado en funciones: Se identifica al tipo de usuario y cuál funcionalidad será utilizada para las clases.
3. Basado en datos: Identificamos la descripción de datos y se especificará un marco de trabajo arquitectural

Los sistemas de base de datos distribuidos se dividen en las siguientes clases:

1. Sistema de manejo de base de datos distribuidos homogéneos: Este podría ser referido como un sistema centralizado en cual se ejecuta en un solo sistema, pero en este caso, los datos se distribuyen por distintos lados que se encuentran comunicados en la red. Aquí no existen los usuarios locales y todos acceden a través de una interfaz global.
2. Sistema de manejo de base de datos distribuido heterogéneos: Manejan distintos manejadores de sistemas en los nodos locales que existen. Aquí existen los usuarios globales y locales. En cada nodo hay un procesador de datos que ayudan a gestionar a los usuarios.

Estrategias de distribución de base de datos

1. Escalado horizontal: Se busca aumentar el número de servidores o también conocidos como nodos, que controlan una aplicación y de este modo trabajar de manera conjunta distribuyendo la carga de trabajo a través de un balanceador. Este escalado necesita un mayor nivel de escalado para poder realizar distintas funciones como lo son la creación de la red de servidores de un cluster.
2. Escalado vertical: En este se busca aumentar los recursos del servidor como lo son el almacenamiento y procesamiento. Una de sus ventajas es su fácil implementación y configuración ya que sólo gestiona un servidor, mejora el rendimiento y además de que no es necesario un diseño en específico para implementarlo.
3. Particionamiento: Proceso de almacenar una base de datos en distintas máquinas que se denominan nodos, dividiendo en particiones los datos los cuales contienen filas únicas de información y guardándolos en servidores de la base de datos. Esto ayuda a que, aunque las bases vayan adquiriendo más datos conforme para el tiempo y los usuarios requieran leer la información que contienen, no se ralentice el proceso y de este modo, permitiendo que exista un procesamiento en paralelo.

Existen diferentes tipos de partición:

- Partición por rangos: Se hacen particiones a través de un rango en específico de valores.
 - Partición con hash: Se realiza a través de una fórmula matemática y asigna una clave de partición a cada fila de la base.
 - Partición de directorios: Es necesario utilizar una tabla de búsqueda para establecer una relación en la información que se encuentra en la base de datos y la Partición física que corresponda.
 - Partición geográfica: Divide la información dependiendo la ubicación geográfica.
4. Replicación: Se basa en que los datos importantes se repliquen en varias ubicaciones y servidores para de este modo salvaguardarlos, mejorar el rendimiento de la base de datos y distribuir la carga de trabajo. Se necesita una Replicación continua para que los datos en todo momento se encuentren actualizados y facilita la recuperación por errores.

Gestores y Funciones para las BDD tanto Relacionales NoSQL

A continuación se presentarán algunos gestores para las Bases de Datos relacionales son:

- MySQL
- MariaDB
- SQLite
- PostgreSQL
- Microsoft SQL Server
- Oracle:

SQL el cual es un lenguaje para este tipo de gestión y se llegan a utilizar diferentes gestores para la relación de la nube como es:

- Amazon RDS
- Google Cloud SQL
- Azure SQL
- IBM Db2 para Cloud

Para las BDD en relación con NoSQL son más flexibles y se diseñan para el manejo de datos con gran cantidad de datos tanto estructurados como semiestructurados, algunos gestores que se utilizan son:

- Redis
- Cassandra
- ObjectDB
- Mongo DB

Por otro lado para los gestores de las BDD en NoSQL con relación a la nube se pueden gestionar en los siguientes gestores:

- Amazon DynamoDB
- Google Firestore
- Azure Cosmos DB
- MongoDB Atlas
- Cassandra (AWS, Azure, Google Cloud)

Referencias:

- Platzi. (s. f.). *Bases de datos distribuidas*. <https://platzi.com/tutoriales/2059-practico-sql/9025-bases-de-datos-distribuidas/>
- 1.1. *Conceptos de base de datos distribuidas*. (2016, 2 diciembre). *Tópicos de Base de Datos*. <https://karenkurapikadb.wordpress.com/2016/12/01/1-1-conceptos-de-base-de-datos-distribuidas/>
- Toledo, V., & Miralles, I. (n.d.). *Bases de Datos Distribuidas*. <https://iessanvicente.com/colaboraciones/BBDDdistribuidas.pdf>
- *¿Qué es una base de datos distribuida y en qué se diferencia de una base de datos?* (2024, 14 enero). [www.linkedin.com. https://www.linkedin.com/advice/3/what-distributed-database-how-does-differ-from-88mxf?lang=es&originalSubdomain=es](https://www.linkedin.com/advice/3/what-distributed-database-how-does-differ-from-88mxf?lang=es&originalSubdomain=es)
- *¿Qué es una base de datos en la nube? | Google Cloud*. (s. f.). Google Cloud. <https://cloud.google.com/learn/what-is-a-cloud-database?hl=es-419>
- *Bases de datos en la nube: ¿qué son y para qué sirven?* (2022, 2 noviembre). Personio. <https://www.personio.es/glosario/bases-de-datos-en-la-nube/>
- (S/f). Recuperado el 13 de marzo de 2025, de <https://repositorio.utn.edu.ec/bitstream/123456789/522/6/04%20ISC%20151%20CAPITULO%20IV.pdf>

- Alejandro, M., & Díaz, G. (n.d.). *BASES DE DATOS DISTRIBUIDAS Introducción BASES DE DATOS DISTRIBUIDAS (BDD)*. <https://cursos.aiu.edu/Base%20de%20Datos%20Distribuidas/pdf/Tema%201.pdf>
- Jose Atonaltzin. (2020, 25 junio). *Arquitectura de bases de datos distribuidas* [Vídeo]. YouTube. <https://www.youtube.com/watch?v=aHfvo1RGXTA>
- González, Ó., & Ruiz González, F. (1999). *UNIVERSIDAD DE CASTILLA LA MANCHA ESCUELA SUPERIOR DE INFORMÁTICA BASES DE DATOS ARQUITECTURAS DE SISTEMAS DE BASES DE DATOS*. <https://rudockblog.wordpress.com/wp-content/uploads/2016/03/cliente-servidor-y-distribuidas.pdf>
- *Ansi-sparc tutorial - ¿Qué es ANSI/SPARC? Es una arquitectura de SGBDD (Sistema de Gestión de Bases - Studocu.* (2022). Studocu. <https://www.studocu.com/es/document/universidad-de-malaga/fundamentos-de-informatica/ansi-sparc-tutorial/43106805>
- *Db2 11.1.* (2022, November 28). Ibm.com. <https://www.ibm.com/docs/es/db2/11.1?topic=partitions-database-partitioning-across-multiple-database>
- *¿Qué es la partición de bases de datos? - Explicación de la partición de bases de datos - AWS.* (2022). Amazon Web Services, Inc. <https://aws.amazon.com/es/what-is/database-sharding/>
- León, M. (2019, July 3). *Escalabilidad horizontal y vertical: modelos de escalado* | Arsys. Blog de Arsys.es. <https://www.arsys.es/blog/escalado-horizontal-vs-vertical>
- *Replicación de bases de datos 101: todo lo que necesita saber.* (2024, January 24). Astera. <https://www.astera.com/es/type/blog/database-replication-101/>
- *Gestores de Bases de Datos Más Usados.* <https://www.inesem.es/revistadigital/informatica-y-tics/los-gestores-de-bases-de-datos-mas-usados/>

- Gestores de Bases de Datos NoSQL Más Populares.
<https://www.e-dea.co/blog/gestores-bases-de-datos-nosql>
- Base de datos relacional.
<https://cloud.google.com/learn/what-is-a-relational-database?hl=es-419>