



INSTITUTO POLITÉCNICO NACIONAL
UNIDAD PROFESIONAL INTERDISCIPLINARIA EN
INGENIERÍA Y TECNOLOGÍAS AVANZADAS



Bases de Datos Distribuidas

Tarea 1. Investigación de BDD

Integrantes del Equipo:

Legorreta Rodríguez María Fernanda

Macías Galván Arturo Daniel

Palacios Reyes Leslie Noemi

Profesor:

Carlos De la Cruz Sosa

3TM3

Viernes 14 de Marzo del 2025

Conceptos de base de datos distribuidos

Se trata de un conjunto de datos que pertenece a un sistema y que está distribuido entre varios ordenadores, 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. Cada uno de estos ordenadores opera de forma autónoma y tiene la capacidad de realizar operaciones de manera local. Además, es necesario que en cada uno de ellos se esté ejecutando una aplicación a nivel global que permita consultar todos los datos como si fueran uno solo.

Diferencias entre una base de datos, distribuida y en la nube

Base de datos	Base de datos distribuida	Base de datos en la nube
<ul style="list-style-type: none">• Se almacena, ubica y mantiene en un solo lugar.• Integridad de los datos• Seguridad• Fácil acceso a toda la información• Los datos son fácilmente transportables• Facilidad de mantenimiento porque todos los datos e información están disponibles en un solo lugar• Mucho menos costosa a comparación de las otras	<ul style="list-style-type: none">• Múltiples bases de datos que están conectadas entre sí y distribuidas en diferentes ubicaciones físicas.• Alto rendimiento debido a la división de la carga de trabajo.• Alta disponibilidad debido a la preparación de los nodos disponibles para realizar el trabajo.• Nodos independientes y mejor control sobre los recursos	<ul style="list-style-type: none">• Se implementa, entrega y accede a la nube.• Tienen la velocidad, la escalabilidad, la agilidad y los costos reducidos que son parte de la computación en la nube• Reducción de la sobrecarga operativa• Mayor agilidad y escalabilidad• Menor costo total de propiedad (TCO)• Opciones flexibles de base de datos• Datos seguros y protegidos• Mejor confiabilidad

Arquitectura de bases de datos relacionales distribuidas

La estructura de las bases de datos distribuidas no obedece a un único estándar, lo que genera la necesidad de establecer un modelo de referencia como un **marco conceptual**. Este modelo tiene como objetivo *descomponer el proceso de estandarización* en partes manejables, mostrando la forma en que los diferentes elementos del sistema se interconectan.

Los métodos de desarrollo de bases de datos distribuidas se agrupan principalmente en tres tipos:

1. **Enfoque Basado en Componentes:** Este método organiza la base de datos en función de sus componentes clave, como el procesador de consultas, la gestión de recuperación ante fallos y el apoyo a la ejecución distribuida.
2. **Enfoque Basado en Funciones:** En este esquema, la organización de la base de datos se fundamenta en las funciones que cumple cada componente, como el procesador de datos que se encarga del almacenamiento y la recuperación.
3. **Enfoque Basado en Datos:** Este modelo se centra en cómo los datos son distribuidos y gestionados a través de los nodos del sistema de base de datos.

Clasificación Funcional de los Sistemas de Bases de Datos Distribuidas

Los sistemas de bases de datos distribuidas pueden ser categorizados según la uniformidad o diversidad de sus componentes:

Sistemas de Bases de Datos Homogéneos: En este diseño, todos los nodos comparten una interfaz común. También existe un procesador de datos que permite una interacción uniforme con el sistema, independientemente de la ubicación física de los datos.

Software de Gestión de Datos:

1. Consultas: Habilita la ejecución de consultas en un sistema distribuido.
2. Recuperación de Fallas: Supervisión de la consistencia y disponibilidad de los datos en situaciones de fallo.
3. Soporte de Tiempo de Ejecución: Gestión y control de las operaciones en tiempo real.

Sistemas de Bases de Datos Distribuidas Heterogéneas: En este caso, los nodos pueden ser distintos entre sí, necesitando un procesador de datos que facilite la coordinación entre los sistemas diversos.

Software de Gestión de Datos:

Sistemas de Gestión Multi-Base de Datos: Este modelo incorpora un diccionario global y permite la interacción entre bases de datos locales y globales. Se segmenta en cuatro partes:

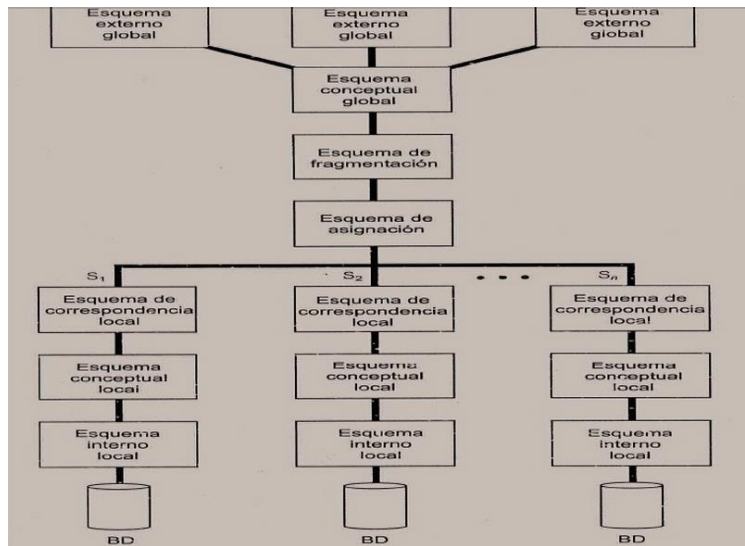
1. Interfaz Global al Usuario: Gestiona la comunicación entre el usuario y el sistema de bases de datos global.
2. Controlador de Datos Semánticos: Garantiza la consistencia de los datos en distintas bases de datos distribuidas.
3. Optimizador de Consultas: Aumenta la eficacia en la ejecución de consultas distribuidas.
4. Supervisor de Ejecución Global: Coordina la ejecución general de las consultas y asegura la integridad de los datos.

Esquemas de Fragmentación y Asignación

Para facilitar la distribución de datos, es esencial ampliar la arquitectura estándar ANSI-SPARC al agregar dos niveles adicionales:

1. **Esquema de Fragmentación:** Explica cómo se separan las conexiones y la información a través de las bases de datos locales. La fragmentación facilita la distribución lógica y efectiva de los datos entre los diferentes nodos.
2. **Esquema de Asignación:** Establece el sitio de almacenamiento de los fragmentos de datos, detallando el lugar físico de cada fragmento en los nodos de la red distribuida.

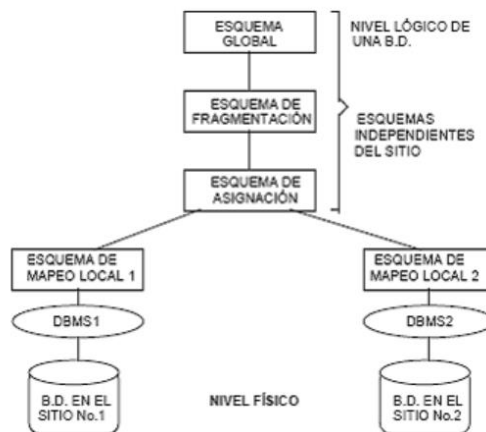
Por otra parte, tenemos el esquema de **Mapeo local**: el cual establece la correspondencia entre los fragmentos del esquema de asignación y los objetos externos de la base de datos local. Si nos enfocamos en el ambiente heterogéneo hay diferentes tipos de mapeos locales para los sitios, como se muestra en la siguiente imagen.



De esta manera se genera una arquitectura de referencia, como se muestra a continuación.

se genera una

Arquitectura de una b.d. distribuida



Estrategias para distribuir una base de datos

Replicación de datos

Esta estrategia se base en conservar replicas idénticas de una tabla. Cada replica se almacena en un nodo diferente.

Las ventajas que tenemos con las replicas son las siguientes:

Disponibilidad

Aumento del paralelismo

Fragmentación de datos

Consiste en dividir los datos en fragmentos, que se van a distribuir en diferentes nodos (servidores), tenemos tres tipos:

1. Fragmentación horizontal
2. Fragmentación vertical
3. Fragmentación mixta

La mayor ventaja es reducir la carga y el tamaño de sus datos. Sin embargo, la fragmentación también puede complicar sus consultas, uniones y transacciones, así como aumentar el riesgo de pérdida o corrupción de datos.

Sharding

Divide sus datos en subconjuntos más pequeños e independientes para asignarlos a diferentes servidores o ubicaciones, esto con el fin de que se pueda reducir la carga y el tamaño de sus datos.

Gestores y funciones para bases de datos distribuidas relacionales (NoSQL) que soportan la replicación.

Las bases de datos NoSQL, conocidas como bases de datos no relacionadas, están creadas para ajustarse a tipos de datos concretos y almacenan información en estructuras flexibles que pueden expandirse con facilidad para aplicaciones actuales.

Datos	Replicación	Descripción	Adicionales
MongoDB	Replica Set	Modelo de replicación donde el nodo es primario (máster) y los demás secundarios (slaves).	Alta disponibilidad y tolerancia a fallos, con promoción automática de secundarios a primarios.
Cassandra	Modelo de anillos	Replicación síncrona y asíncrona.	Factor de replicación configurable. Permite distribuir los datos eficientemente entre los nodos.
DynamoDB	Replicación síncrona	Base de datos NoSQL, clave-valor, con replicación en múltiples zonas de disponibilidad dentro de AWS.	Garantiza alta disponibilidad y resiliencia en la nube de AWS.

Redis	Replicación maestro-esclavo	Base de datos clave-valor en memoria.	Principalmente usado para cache y gestión de sesiones. Soporta persistencia para garantizar la durabilidad.
-------	-----------------------------	---------------------------------------	---

Referencias

1. - YouTube. (s/f). Youtu.Be. Recuperado el 14 de marzo de 2025, de <https://youtu.be/aHfvo1RGXTA>
2. Arquitectura de Base de Datos Distribuidas. (s/f). Blogspot.com. Recuperado el 14 de marzo de 2025, de <https://base-datos-ruiz-salas-fl.blogspot.com/2015/02/arquitectura-de-base-de-datos.html>
3. Marín, R. (2024, abril 8). Los gestores de bases de datos más usados en la actualidad. Canal Informática y TICS. <https://www.inesem.es/revistadigital/informatica-y-tics/los-gestores-de-bases-de-datos-mas-usados/>
4. Gestores de bases de datos NoSQL más recomendados. (s/f). E-dea.co. Recuperado el 14 de marzo de 2025, de <https://www.e-dea.co/blog/gestores-bases-de-datos-nosql>
5. Miralles, V. T. —. (s/f). *Bases de Datos Distribuidas*. lessanvicente.com. Recuperado el 14 de marzo de 2025, de <https://iessanvicente.com/colaboraciones/BBDDdistribuidas.pdf>
6. (S/f). Amazon.com. Recuperado el 14 de marzo de 2025, de <https://aws.amazon.com/es/nosql/>
7. ¿Qué es una base de datos NoSQL? (2025, enero 23). lbm.com. <https://www.ibm.com/mx-es/think/topics/nosql-databases>
8. Tablado, F. (2020, septiembre 10). *Base de datos distribuida. ¿Qué es? Características*. Ayuda Ley Protección Datos; AyudaLeyProteccionDatos. <https://ayudaleyprotecciondatos.es/bases-de-datos/distribuida/>
9. ¿Cómo se escala y distribuye la base de datos en varios servidores o ubicaciones? (2023, agosto 17). LinkedIn.com; www.linkedin.com. <https://es.linkedin.com/advice/0/how-do-you-scale-distribute-your-database?lang=es&lang=es>
10. *Difference between centralized database and distributed database*. (2020, junio 11). GeeksforGeeks. <https://www.geeksforgeeks.org/difference-between-centralized-database-and-distributed-database/>
11. ¿Qué es una base de datos en la nube? (s/f). Google Cloud. Recuperado el 14 de marzo de 2025, de <https://cloud.google.com/learn/what-is-a-cloud-database?hl=es-419>