

INSTITUTO POLITECNICO NACIONAL

UNIDAD PROFESIONAL INTERDISCIPLINARIA EN INGENIERIA
Y TECNOLOGIAS AVANZADAS - IPN

MATERIA

Bases de Datos Distribuidas

ALUMNOS

Fernández Guerrero Keb Sebastián

Ramírez Orozco Juan Carlos

Sánchez Herrera Armando Eduardo

PROFESOR

Carlos De la Cruz Sosa

Tarea 1 Conceptos de Bases de Datos Distribuidas

Grupo 3TM3 Tarea No. 01

Fecha de asignación:

07/03/25

Fecha de entrega:

14/03/25

1. Conceptos de Bases de Datos Distribuidas

Una base de datos distribuida (BDD) es un sistema en el que los datos están almacenados en múltiples ubicaciones físicas, ya sea en diferentes servidores o nodos, que están interconectados a través de una red. Estos sistemas permiten que los datos sean accesibles y gestionados de manera transparente, como si estuvieran en una sola ubicación, aunque físicamente estén dispersos (Özsu & Valduriez, 2020). Las BDDs son especialmente útiles en entornos donde se requiere alta disponibilidad, escalabilidad y tolerancia a fallos. Además, permiten la distribución de cargas de trabajo, lo que mejora el rendimiento en aplicaciones con grandes volúmenes de datos (Garcia-Molina et al., 2008).

Un aspecto clave de las BDDs es la **transparencia**, que permite a los usuarios interactuar con el sistema sin necesidad de conocer la ubicación física de los datos. Esto incluye transparencia de ubicación, fragmentación y replicación (Elmasri & Navathe, 2016).

2. Diferencia entre Base de Datos Centralizada y una Base de Datos Distribuida

- **Base de Datos Centralizada:**
 - Todos los datos se almacenan en un único servidor o ubicación física.
 - Los usuarios acceden a los datos a través de una red, pero la gestión y el almacenamiento son centralizados.
 - Este enfoque es más simple de administrar, pero tiene limitaciones en escalabilidad y tolerancia a fallos (Elmasri & Navathe, 2016).
 - Es más vulnerable a fallos, ya que un problema en el servidor central puede dejar todo el sistema inoperativo (Garcia-Molina et al., 2008).
- **Base de Datos Distribuida:**
 - Los datos se almacenan en múltiples servidores o nodos, que pueden estar geográficamente dispersos.
 - Cada nodo puede operar de manera independiente, pero el sistema global permite el acceso transparente a los datos.
 - Este enfoque ofrece mayor escalabilidad, disponibilidad y tolerancia a fallos, pero es más complejo de gestionar (Özsu & Valduriez, 2020).
 - En el contexto de las bases de datos en la nube, las BDDs son comunes debido a su capacidad para distribuir cargas de trabajo y almacenamiento en múltiples centros de datos, lo que mejora el rendimiento y la resiliencia (Kaur & Kaur, 2019).

3. Arquitectura de Bases de Datos Distribuidas

- **Cliente-Servidor:**
 - En esta arquitectura, los clientes solicitan datos a uno o más servidores que almacenan y gestionan la información.
 - Los servidores pueden estar distribuidos geográficamente, pero los clientes interactúan con ellos como si fueran una única entidad (Elmasri & Navathe, 2016).
 - Es un modelo jerárquico donde los servidores centralizan la gestión de los datos, mientras que los clientes solo realizan solicitudes (Garcia-Molina et al., 2008).
- **Peer-to-Peer (P2P):**
 - En este modelo, todos los nodos tienen la misma capacidad y pueden actuar tanto como clientes como servidores.
 - No hay una jerarquía centralizada, lo que permite una mayor descentralización y resiliencia (Kaur & Kaur, 2019).
 - Este enfoque es común en sistemas distribuidos modernos como blockchain, donde la confianza y la responsabilidad se distribuyen entre todos los participantes (Wang et al., 2021).

4. Estrategias para Distribuir una Base de Datos

- **Replicación:**
 - Consiste en copiar los datos en múltiples nodos para mejorar la disponibilidad y el rendimiento.
 - Puede ser síncrona (los datos se replican en tiempo real) o asíncrona (los datos se replican con cierto retraso) (Özsu & Valduriez, 2020).
 - La replicación también facilita la recuperación ante desastres, ya que los datos están disponibles en múltiples ubicaciones (Garcia-Molina et al., 2008).
- **Fragmentación:**
 - Divide los datos en partes más pequeñas que se almacenan en diferentes nodos.
 - Puede ser:
 - **Horizontal:** Se dividen las filas de una tabla en diferentes nodos. Por ejemplo, los clientes de diferentes regiones se almacenan en servidores locales (Elmasri & Navathe, 2016).
 - **Vertical:** Se dividen las columnas de una tabla en diferentes nodos. Por ejemplo, los datos personales y financieros de los clientes se almacenan en servidores separados (Kaur & Kaur, 2019).
 - La fragmentación reduce la carga en cada nodo y mejora el rendimiento de las consultas locales (Wang et al., 2021).
- **Combinación de Replicación y Fragmentación:**

- En muchos sistemas distribuidos, se utilizan ambas estrategias para optimizar el rendimiento y la disponibilidad.
- Por ejemplo, un sistema puede fragmentar los datos por región y replicar cada fragmento en múltiples nodos para garantizar redundancia (Garcia-Molina et al., 2008).

5. Gestores y Funciones para Bases de Datos Distribuidas

- **Relacionales:**

- Sistemas como PostgreSQL, MySQL y Oracle ofrecen funcionalidades para la distribución de datos, incluyendo replicación y fragmentación.
- Estos sistemas mantienen la consistencia de los datos a través de mecanismos como el protocolo de dos fases (2PC) (Özsu & Valduriez, 2020).
- También soportan transacciones distribuidas, lo que permite operaciones atómicas en múltiples nodos (Elmasri & Navathe, 2016).

- **NoSQL:**

- Bases de datos como MongoDB, Cassandra y Redis están diseñadas para ser distribuidas desde su concepción.
- Ofrecen alta escalabilidad y disponibilidad, aunque a veces sacrifican consistencia fuerte en favor de la consistencia eventual (Kaur & Kaur, 2019).
- Estas bases de datos son ideales para aplicaciones que requieren alto rendimiento y tolerancia a fallos, como redes sociales y sistemas de recomendación (Wang et al., 2021).

- **Funciones de Replicación:**

- La replicación es una función clave en las BDDs, ya que permite la redundancia de datos, lo que mejora la disponibilidad y la tolerancia a fallos.
- Además, facilita la recuperación ante desastres y la distribución de cargas de trabajo (Garcia-Molina et al., 2008).

Conclusiones

Las bases de datos distribuidas ofrecen ventajas significativas en términos de escalabilidad, disponibilidad y tolerancia a fallos, especialmente en entornos de nube. Las estrategias de replicación y fragmentación son fundamentales para optimizar el rendimiento y la gestión de los datos en estos sistemas. Sin embargo, la complejidad de su administración y la necesidad de mantener la consistencia de los datos son desafíos que deben ser abordados cuidadosamente.

En el contexto actual, donde el volumen de datos y las demandas de disponibilidad son cada vez mayores, las BDDs se han convertido en una solución indispensable para muchas organizaciones. La elección entre sistemas

relacionales y NoSQL dependerá de los requisitos específicos de cada aplicación, como la necesidad de consistencia fuerte o la priorización del rendimiento y la escalabilidad.

Referencias

Elmasri, R., & Navathe, S. B. (2016). *Fundamentals of Database Systems* (7th ed.). Pearson.

Garcia-Molina, H., Ullman, J. D., & Widom, J. (2008). *Database Systems: The Complete Book* (2nd ed.). Pearson.

Kaur, R., & Kaur, A. (2019). A Review on Distributed Database Systems. *International Journal of Computer Applications*, 178(14), 1-5. <https://doi.org/10.5120/ijca2019918731>

Özsu, M. T., & Valduriez, P. (2020). *Principles of Distributed Database Systems* (4th ed.). Springer.

Wang, L., Zhang, J., & Chen, Y. (2021). Distributed Database Systems: Challenges and Opportunities. *Journal of Computer Science and Technology*, 36(2), 245-260. <https://doi.org/10.1007/s11390-021-9815-9>