



INSTITUTO POLITÉCNICO NACIONAL

UNIDAD PROFESIONAL INTERDISCIPLINARIA
EN INGENIERIA Y TECNOLOGÍAS AVANZADAS

BASES DE DATOS DISTRIBUIDAS

GRUPO: 3TM3

TAREA 1

PROFESOR: DE LA CRUZ SOSA CARLOS

INTEGRANTES EQUIPO 2:

OCAMPO CABRERA ASael	2022640070
ROBERT ROA KARLA GUADALUPE	2022640710
SANCHEZ VILLAGRANA OSMAR ROBERTO	2022640514

Bases de Datos Distribuidas

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. Las actualizaciones a una base de datos distribuida pueden llevarse a cabo contra una sola base de datos como una unidad de trabajo remota o a través de varias bases de datos como una actualización multisitio. [1]

- *Diferencias entre Bases de Datos Centralizadas y Bases de Datos Distribuidas (Incluyendo Bases de Datos en la Nube).*

Las bases de datos centralizadas y distribuidas representan dos enfoques fundamentalmente distintos para la gestión de información, y la aparición de las bases de datos en la nube representa una nueva perspectiva en el panorama.

Una base de datos centralizada actúa como el único lugar para todos los datos, en una PC o sistema informático el cual funge como un servidor principal. Este modelo tradicional simplifica la administración y el mantenimiento, ya que todos los datos se encuentran en un mismo lugar, lo cual ayuda a la implementación de medidas de seguridad y la preservación de la integridad de los datos almacenados.

Elmasri y Navathe en *Fundamentals of Database Systems* (7ma edición) coinciden en que las bases de datos centralizadas son más fáciles de gestionar debido a su ubicación única, pero añaden que "la escalabilidad es un problema significativo, ya que el rendimiento se degrada a medida que aumenta el número de usuarios o transacciones" (p. 845). [3]

Por otro lado, una base de datos distribuida tiene la comitiva de dispersar la información a través de múltiples PCs o sistemas informáticos, interconectadas mediante una red. Cada ubicación almacena una parte de la base de datos o una copia completa de la misma, lo que ofrece una mejora en la gestión de fallos ya que, si un nodo falla, los demás pueden seguir trabajando.

La escalabilidad es otra ventaja clave, ya que se pueden añadir más nodos al sistema para aumentar la capacidad de almacenamiento y procesamiento.

Las bases de datos en la nube son una extensión de las bases de datos distribuidas. Siendo que ofrecen escalabilidad elástica y reducción de costos, ya que los recursos se ajustan constantemente según la demanda. sin dejar de lado que la seguridad y la privacidad de los datos son puntos críticos en este punto.

Garcia-Molina, Ullman y Widom en *Database Systems: The Complete Book* (2da edición) coinciden en que las bases de datos en la nube son una evolución natural de las bases de datos distribuidas, pero añaden que "La dependencia de un proveedor externo puede ser un riesgo, especialmente en términos de disponibilidad y control sobre los datos" (p. 780). [7]

- ***Arquitecturas de Bases de Datos Distribuidas.***

- 1. Arquitectura Cliente-Servidor:**

Consiste en un servidor central que gestiona la base de datos y múltiples clientes que solicitan datos. El servidor procesa las consultas y devuelve resultados. El servidor es el único punto de acceso y puede volverse un cuello de botella.

- 2. Arquitectura Peer-to-Peer (P2P):**

Todos los nodos tienen igual responsabilidad (sin servidor central). Los datos se distribuyen y replican entre pares. Gracias a la descentralización, tiene mayor tolerancia a fallos. Se ocupa un consenso complejo y requiere protocolos como Paxos o Raft para coherencia.

- 3. Arquitectura Sharding (Fragmentación):**

Divide la base de datos en fragmentos (shards) distribuidos en diferentes nodos. Cada shard almacena un subconjunto de datos. Permite manejar grandes volúmenes de datos gracias a su escalabilidad horizontal

- 4. Arquitectura fragmentada y replicada:**

Los datos pueden dividirse (fragmentación horizontal o vertical) y/o replicarse entre nodos para mejorar el rendimiento y la disponibilidad. "Las decisiones de fragmentación y replicación son esenciales para optimizar el rendimiento en sistemas distribuidos" (Navathe & Elmasri, 2016). [3]

- 5. Arquitectura Federada:**

Cada nodo mantiene su autonomía administrativa y lógica, pero colabora en consultas globales. Es común cuando se integran bases de datos heterogéneas. "Las bases federadas son útiles para integración de datos, pero sacrifican rendimiento por flexibilidad" (Elmasri & Navathe, 2016, p. 890). [3] Las consultas globales son ineficientes por lo que se genera el problema Overhead

- ***Estrategias para distribuir Bases de Datos.***

Al diseñar una base de datos distribuida, es esencial considerar cómo se distribuirán y replicarán los datos. Las principales estrategias incluyen:

- 1. Fragmentación***

Consiste en dividir una base de datos en fragmentos que se almacenan en diferentes nodos y nos permite mejorar el rendimiento y la localización de los datos, pero requiere mecanismos para recomponer los datos cuando sea necesario. Existen varios tipos:

- **Horizontal:** Cada fragmento contiene un subconjunto de filas de una tabla.
- **Vertical:** Cada fragmento contiene un subconjunto de columnas de una tabla.
- **Mixta:** Combina fragmentación horizontal y vertical. [9]

2. Replicación

Implica mantener copias de los datos en múltiples nodos para mejorar la disponibilidad y la tolerancia a fallos, esta puede ser sincrónica o asincrónica. Esta se puede observar de diferentes tipos:

- **Completa:** Cada nodo tiene una copia completa de la base de datos.
- **Parcial:** Solo se replican partes específicas de la base de datos.
- **Selectiva:** Se replican datos según criterios específicos. [8]

3. Híbrida

Combina fragmentación y replicación, permitiendo una distribución más flexible y adaptada a las necesidades específicas del sistema. [9]

• Gestores y funciones para Bases de Datos Distribuidas Relacionales.

1. PostgreSQL:

Ofrece replicación en caliente y lógica, permitiendo la replicación de datos en tiempo real y la replicación selectiva de tablas o esquemas específicos. Además, existen herramientas como BDR (Bi-Directional Replication) para replicación bidireccional. [10]

2. MySQL:

Soporta replicación maestro-esclavo y maestro-maestro. La replicación puede ser configurada para ser sincrónica o asincrónica, y permite la replicación de bases de datos completas o tablas específicas. [11]

3. Oracle Database:

Ofrece soluciones avanzadas de replicación, incluyendo Oracle GoldenGate para replicación en tiempo real y Oracle Data Guard para alta disponibilidad y recuperación ante desastres. [12]

4. IBM Db2:

Proporciona capacidades de replicación y soporte para bases de datos relacionales distribuidas, permitiendo que los gestores de bases de datos se comuniquen y cooperen entre sí para acceder a datos distribuidos en diferentes sistemas. [13]

Referencias Bibliográficas

- [1] Db2 11.1. (2022, noviembre 28). Ibm.com. <https://www.ibm.com/docs/es/db2/11.1?topic=managers-designing-distributed-databases>
- [2] Ozsu, M. T., & Valduriez, P. (2011). *Principles of Distributed Database Systems* (3rd ed.). Springer.
- [3] Elmasri, R., & Navathe, S. B. (2016). *Fundamentals of Database Systems* (7th ed.). Pearson.
- [4] Rahimi, S. K., & Haug, F. S. (2010). *Distributed database management systems: A practical approach*. Wiley-IEEE Computer Society Pr.
- [5] Ray, C. (2009). *Distributed database systems*. Pearson Education India.
- [6] Kroenke, D. M., & Auer, D. J. (2016). *Database Processing: Fundamentals, Design, and Implementation* (14th ed.). Pearson.
- [7] Garcia-Molina, H., Ullman, J. D., & Widom, J. (2008). *Database Systems: The Complete Book* (2nd ed.). Pearson.
- [8] Haider, K. (2020, febrero 13). ¿Qué es la replicación de datos? Beneficios, tipos, ejemplos y casos de uso. Astera. <https://www.astera.com/es/type/blog/data-replication/>
- [9] Wikipedia contributors. (s/f). Base de datos distribuida. Wikipedia, The Free Encyclopedia. https://es.wikipedia.org/w/index.php?title=Base_de_datos_distribuida&oldid=166281669
- [10] PostgreSQL. (2025, mayo 3). PostgreSQL. <https://www.postgresql.org/>
- [11] (S/f). Mysql.com. Recuperado el 3 de mayo de 2025, de <https://www.mysql.com/>
- [12] Database. (s/f). Oracle.com. Recuperado el 3 de mayo de 2025, de <https://www.oracle.com/database/>
- [13] IBM Db2. (2024, junio 6). Ibm.com. <https://www.ibm.com/db2>