Tarea 1 Lunes 10 de marzo de 2025

1. **Conceptos de Bases de datos distribuidas (guardar las referencias)**

Las bases de datos distribuidas son un sistema de bases de datos en el que los datos se distribuyen entre múltiples ubicaciones físicas, que pueden estar en diferentes computadoras o servidores, pero se presentan al usuario como una sola base de datos lógica. A continuación, te detallo algunos conceptos clave relacionados con las bases de datos distribuidas:

**1. Fragmentación (Sharding)**

La fragmentación es el proceso de dividir una base de datos en fragmentos más pequeños, que se distribuyen entre varios nodos de la red. Los fragmentos pueden ser:

* **Fragmentación horizontal:** Se distribuyen filas de una tabla entre varios nodos.
* **Fragmentación vertical:** Se distribuyen las columnas de una tabla entre varios nodos.
* **Fragmentación mixta:** Se combina la fragmentación horizontal y vertical.

**2. Replicación**

La replicación consiste en mantener copias exactas de los datos en diferentes nodos. Los datos replicados ayudan a mejorar la disponibilidad, la tolerancia a fallos y el rendimiento de las consultas. Los tipos de replicación incluyen:

* **Replicación maestro-esclavo:** Un nodo maestro controla las actualizaciones, y los esclavos solo leen datos.
* **Replicación multi-maestro:** Varios nodos pueden ser responsables de escribir y actualizar datos, con mecanismos para resolver conflictos.

**3. Transacciones Distribuidas**

En un sistema de bases de datos distribuidas, las transacciones se pueden realizar sobre diferentes nodos. Asegurar la coherencia y la integridad de las transacciones es un desafío, y se requiere un sistema para coordinar las transacciones distribuidas, como el **protocolo de consenso** y el **protocolo de dos fases (2PC)**.

**4. Consistencia y CAP Theorem**

El **teorema CAP** establece que, en una base de datos distribuida, es imposible garantizar al mismo tiempo:

* **Consistencia:** Todos los nodos tienen los mismos datos en todo momento.
* **Disponibilidad:** El sistema responde a todas las solicitudes, incluso si algunos nodos están caídos.
* **Tolerancia a particiones:** El sistema sigue funcionando incluso si hay fallos en la red que separan a los nodos. Dependiendo del caso de uso, un sistema de base de datos distribuida priorizará una o más de estas propiedades.

**5. Coherencia Eventual**

Este concepto hace referencia a la idea de que, aunque los datos pueden no estar actualizados de inmediato en todos los nodos, con el tiempo todos los nodos eventualmente tendrán la misma versión de los datos. Esto es común en sistemas que priorizan la disponibilidad y la tolerancia a particiones (como bases de datos NoSQL).

**6. Modelo de Datos Distribuidos**

Los modelos de datos pueden ser **relacionales** o **NoSQL**, dependiendo de la estructura de los datos y el tipo de consultas. Las bases de datos distribuidas pueden ser:

* **SQL distribuido:** Utilizan el modelo relacional y están basadas en SQL (por ejemplo, PostgreSQL en un entorno distribuido).
* **NoSQL distribuido:** Utilizan otros modelos como clave-valor, documento, columna, o grafos, y están diseñadas para escalabilidad horizontal (por ejemplo, MongoDB, Cassandra).

**7. Nodos y Redes de Comunicación**

Un nodo es una entidad en la red que almacena datos. Los nodos pueden ser máquinas físicas o virtuales interconectadas, y la comunicación entre ellos es crucial para la sincronización de datos, replicación y consultas distribuidas.

**8. Control de Concurrencia Distribuido**

Cuando múltiples usuarios acceden y modifican los datos simultáneamente, se deben aplicar mecanismos de control de concurrencia para evitar inconsistencias y garantizar que las transacciones sean correctas. Esto puede implicar el uso de bloqueos, versiones de datos y otros mecanismos.

**9. Escalabilidad**

Las bases de datos distribuidas están diseñadas para **escalar horizontalmente**, es decir, añadir más nodos para manejar mayores volúmenes de datos y más tráfico. Esta característica es esencial para aplicaciones que crecen rápidamente.

**10. Tolerancia a Fallos**

En un sistema distribuido, algunos nodos pueden fallar, pero el sistema debe seguir funcionando. La tolerancia a fallos se logra a través de técnicas como la **replicación de datos** y el **mantenimiento de copias de seguridad** en nodos diferentes.

**11. Sincronización de Datos**

La sincronización de datos entre los diferentes nodos puede ser compleja, especialmente cuando se actualizan los datos en múltiples nodos al mismo tiempo. Los sistemas de bases de datos distribuidas usan algoritmos de consenso para asegurarse de que los nodos se mantengan consistentes entre sí.

**12. Bases de Datos Distribuidas Multi-Modelo**

Algunas bases de datos distribuidas permiten utilizar más de un modelo de datos, como relacional y NoSQL, de manera combinada. Estas bases de datos son útiles para aplicaciones complejas que requieren diferentes tipos de almacenamiento y consultas.

[Datos sacados del ChatGPT horrendo despues buscare la información de libros pertinentes]

1. **Diferencia entre Base de Datos Centralizada y una Base de datos distribuida (bases de datos en la nube)**
2. **Arquitectura de base de datos distribuidas**
3. **Estrategias para distribuir una base de datos**
4. **Gestores y funciones para bases de datos distribuidas (relacionales) (NoSQL) (Replicacion)**
5. **Referencias**

Hay que repartirse la investigacion despues e ir repondiendo