

1. Explique en qué consisten los siguientes conceptos:**a. Data Warehouse**

Podemos referirnos a él como un almacén de datos o repositorio. Este es una base de datos que esta diseñada para almacenar y analizar grandes cantidades de datos, o mejor dicho se encuentran optimizados para realizar operaciones de escritura por lotes y lectura de grandes volúmenes de datos. Los Data Warehouse por lo general emplean esquemas desnormalizados como el esquema de estrella y el esquema de copo de nieve, suelen estar almacenados en la nube y aprovechan herramientas de BI.

b. Data Lake

Al igual que Data Warehouse se emplea para el almacenamiento de grandes cantidades de datos, sin embargo posee la diferencia con respecto a Data Warehouse que no requiere una estructura predefinida. Permite almacenar cualquier tipo de datos como datos estructurados, semiestructurados y no estructurados. Por su estructura permite que los datos tengan mayor facilidad para ser consultados.

c. Data Mart

Esta es una manera más sencilla para almacenar datos enfocados en una área específica. Se pueden construir a partir de un gran almacén, y estos son fáciles de diseñar, construir y administrar. Sin embargo se debe contemplar que puede llegar complicar las consultas si se realizan entre distintas áreas específicas.

2. ¿De qué forma se benefician las aplicaciones del uso de Columnar Storage? Explique.

Por su arquitectura de almacenar los datos en este caso en columnas, nos brinda mayor eficiencia en cuanto a las consultas porque se accede solo a las columnas correspondientes evitando recorrer todos los datos como en el caso en una base de datos que almacena la información en filas, lo que implica en que el tiempo de respuesta es menor. Otro aspecto en que se ve favorecido es en cuanto a la escalabilidad, que no dificultaría crecer en un momento que sea necesario

También la comprensión de datos es más efectiva y nos reduce el espacio que es necesario así como mejor rendimiento en la lectura de los datos

3. ¿En que consiste streaming y batch processing?

Streaming lo podemos considerar como una técnica de procesamiento de datos que procesa pequeñas cantidades de datos a medida que se van recibiendo, gracias a ello permite tomar

decisiones de manera rápida es decir una acción inmediata. Es de gran utilidad implementarlo en aplicaciones de análisis de datos en tiempo real.

Por su parte batch processing procesa los datos en grandes cantidades reuniéndolos en lotes. Por ejemplo se establece que el lote al alcanzar un tamaño de 5 GB se procesarán los datos, sin importar el tiempo transcurrido hasta el momento de alcanzar ese tamaño se ejecutará el procesamiento. Por esa razón se implementa batch processing cuando no es requerido que los datos reciban una respuesta en tiempo real.

4. ¿ En que consiste datos estructurados, semi estructurados y no estructurados?

Datos estructurados: Este tipo de datos cuenta con una estructura definida y están organizados ya sea en un esquema o tabla. Este tipo de datos se encuentran bien organizados y permite ser analizados fácilmente.

Datos semi estructurados: Este tipo de dato posee una estructura definida pero parcialmente lo que permite algún dato de distinta estructura. El ejemplo por excelencia para este tipo son los archivos JSON.

Datos no estructurados: Por ultimo para este tipo no cuenta con ninguna estructura definida lo que provoca que su análisis sea más complejo como si lo es un dato estructurado. Este tipo es muy común y se pueden encontrar en fuentes como por ejemplo redes sociales, documentos etc.