

Explique:

Data Warehouse: Son repositorios centrales de datos de diferentes fuentes. Es usado para almacenamiento, análisis y reportes. Son una parte fundamental de la inteligencia de negocios.

Data Lake: Un patrón arquitectural que combina los mejores elementos de data warehouse y data mart. Permiten generar consultas a través de los diferentes tipos de almacenamiento de datos (Warehouse, Mart, base de datos operacional). Se pueden guardar datos en formato abierto.

Data Mart: Una forma simple de warehouse concentrada en un área funcional o tema específico. Estos son sencillos de diseñar, crear y administrar.

¿De que forma se benefician las aplicaciones del uso de Columnar Storage?

Esté es más eficiente en entradas y salidas para consultas de lectura. Esto debido a que tienen que leer únicamente las columnas relevantes para la consulta específica, en vez de buscar en la fila de forma secuencial por la info. Esto permite mayor velocidad de procesamiento.

¿En que consiste streaming y batch processing?

Estos son dos tipos de flujos de trabajo para el procesamiento de datos. El streaming consiste del procesamiento de datos de forma secuencial e incremental a tiempo real. Esto permite visibilidad y manejo de información actual para aspectos de negocio y actividad de usuario. El batch processing consiste del procesamiento de muchos datos agrupados en lotes de forma periódica y automatizada. Estos toman más tiempo para procesar los archivos y son menos flexibles en visibilidad y manejo.

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

Estos describen la organización y formato de los datos. Los datos estructurados son aquellos que se almacenan en una base de datos relacional, y se componen de tablas con filas y columnas bien definidas. Los datos semi-estructurados son aquellos que no se ajustan a una estructura fija como la de los datos estructurados, pero aún tienen cierta organización y pueden ser consultados. Estos datos a menudo se almacenan en formatos como XML, JSON o YAML. Los datos no estructurados son aquellos que no tienen una estructura predefinida, lo que los hace más difíciles de almacenar, procesar y analizar. No hay organización específica para los datos.