Prueba Corta #1

Explique en que consisten los siguientes conceptos?

- Data Warehouse Es una base de datos corporativa en la que se integra informacion depurada de las multiples funtes que existen en la organizacion este mantiene una gran cantidad de informacion, estos deben ser almacenados de forma segura, fiable, facil de recuperar y facil de administrar. La informacion debe ser homogenea y fiable, este puede ser fisico o logico y da una ayuda en la captura de datos de diversas fuentes
- Data Lake Repositorio centralizado diseñado para almacenar, procesar y proteger grandes cantidades de datos estructurados, semiestructurados o sin estructurar. Estos proporcionan una plataforma escalable y segura que permite realizar las siguientes tareas a las empresas:
    - Almacenar cualquier tipo o volumen de datos con fidelidad absoluta
    - Procesar datos en tiempo real o en modo por lotes
    - Analizar datos mediante SQL, Python, R o cualquier otro lenguaje, datos de terceros o aplicaciones de estadísticas.
- Data Mart Esta es una forma sencilla de un almacenamiento de datos que se centra en un tema en particular o linea de negocio, como:
    - Ventas
    - Finanzas
    - Marketing

      Las fuentes de los data mart pueden incluir sistemas operativos internos, un almacen de datos central y datos externos

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

El proposito de una columnar es escribir y leer los datos de forma eficiente hacia y desde un disco duro para acelerar el tiempo que toma procesar la respuesta de un query. Estas bases de datos almacenan la data de una forma que optimiza el rendimiento de los inputs y outputs de un disco. El principal beneficio es que este tiene un rendimiento mas rapido en comparacion al row-oriented.

En que consiste streamig y batch processing?

- Batch Processing El procesamiento por Batch se refiere al procesamiento de un gran volumen de datos por Batch dentro de un período de tiempo específico. Procesa un gran volumen de datos a la vez, se usa cuando el tamaño de los datos es conocido y finito, se tarda un poco más de tiempo en procesar los datos. El procesador por Batch procesa los datos en varias pasadas. Cuando los datos se recopilan a lo largo del tiempo y los datos similares se agrupan/agrupan en Batch, en ese caso se utiliza el procesamiento por Batch.
- Stream Processing El procesamiento de Stream se refiere al procesamiento de un flujo continuo de datos inmediatamente después de su producción. Analiza los datos de transmisión en tiempo real, utiliza cuando el tamaño de los datos es desconocido, infinito y continuo. Se tarda unos segundos o milisegundos en

procesar los datos. En este la tasa de salida de datos es tan rápida como la tasa de entrada de datos, procesa los datos en pocas pasadas. Cuando el Stream de datos es continuo y requiere una respuesta inmediata, en ese caso se utiliza el procesamiento de Stream.

En que consiste dato estructurados y no estructurados?

- Dato Estructurado: Los datos estructurados son informacion que ha sido formateada y transformada en un modelo de datos bien definidos. Los datos sin procesar se pueden leer apartir de SQL, inclusive, las estructuras y funcionamiento de las bases relacionales SQL son el ejemploi mas acertado para explicar que es un dato estructurado.
- No Estructurado: Los datos no estructurados son aquellos que se pueden mostrar como datos presentes y sin procesar, es complejo debido a su compleja organizacion y formato. La gestión de datos no estructurados puede tomar datos de muchas formas, incluidas publicaciones en redes sociales, chats, imágenes satelitales, datos de sensores de IoT, correos electrónicos y presentaciones, para organizarlos de manera lógica y predefinida en un almacenamiento de datos.