

Erick Madrigal Zavala
2018146983

1) Conceptos:

a. Data Warehouse.

Un Data Warehouse es una base de datos que recopila información de muchas fuentes diferentes y la guarda en un lugar para que sea fácil de analizar. El Data Warehouse se enfoca en almacenar datos para fines analíticos

b. Data Lake

Un Data Lake es un depósito de datos que almacena información en su forma original, sin procesar ni estructurada, permitiendo una mayor flexibilidad que un Data Warehouse. Los datos pueden ser estructurados, semiestructurados o no estructurados y se almacenan en su forma original, lo que significa que no es necesario transformarlos antes de almacenarlos. El Data Lake es útil para empresas que necesitan almacenar grandes cantidades de datos y no están seguras de cómo los utilizarán en el futuro.

c. Data Mart

Un Data Mart es una base de datos enfocada en una función o departamento de una empresa. Contiene información específica, optimizada para las necesidades de los usuarios finales de ese departamento. Se construye a partir de una copia del Data Warehouse y proporciona acceso rápido y fácil a la información para la toma de decisiones informadas.

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

El almacenamiento columnar mejora la compresión ya que cada columna esta guardada su propio set de bloques, cada bloque físico contiene el mismo tipo de datos. Gracias a usar el mismo tipo de dato las bases de datos pueden usar algoritmos de compresión extremadamente eficientes. Gracias a esto se necesita mucho menos espacio para almacenar los datos y además se ve mejorada la velocidad de lectura y escritura

3) ¿En qué consiste streaming y batch processing?

a. Streaming: es un flujo constante de datos en tiempo real.

b. Batch processing: cuando los datos son extremadamente grandes para poder analizarlos lo que se hace es establecer “ventanas” ya sea de cantidad de datos o de tiempo en las cuales se toma la información y se analiza

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

a. Estructurados: son datos que se organizan y almacenan en una estructura predefinida y organizada. Esta estructura permite una fácil identificación de los datos y cómo se relacionan entre sí.

b. Semi estructurados: son datos que no se organizan en una estructura predefinida, pero aún tienen alguna forma de organización que permite su procesamiento.

c. No estructurados: son datos que no tienen ninguna estructura predefinida. Estos datos pueden ser de cualquier tipo, como texto, imágenes, videos y archivos de audio.