

## 1. Explique en qué consisten los siguientes conceptos:

**a. Data Warehouse:** Un Data Warehouse es un repositorio central o almacén de información que proviene de diversas fuentes de datos.

**b. Data Lake:** Es un repositorio que almacena grandes cantidades de datos de distintos tipos en “crudo” que pueden ser estructurados, semiestructurados o no estructurados. De esta forma se pueden analizar y procesar después según se necesite.

**c. Data mart:** Es una que fragmento de un Data Warehouse que contiene datos, los cuales están enfocados en áreas específicas, por ejemplo, ventas. Esto permite un análisis mas detallado y profundo de dicha área.

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

- El Columnar Storage consiste en ordenar los datos en columnas y no en filas. En este tipo de almacenamiento aumenta la eficiencia para las operaciones de lectura, ya que en lugar de leer toda la fila como en Row-Oriented Databases solo lee las columnas que importan o que se necesitan. También ocupa menos espacio. En general, el Columnar Storage es una mejor opción para almacenar datos en Datawarehouses.

## 3. ¿En que consiste streaming y batch processing?

**a. Streaming:** Es una técnica para procesar datos que consiste en procesar los datos a medida van llegando de manera continua y en tiempo real, por lo que los resultados son obtenidos de forma casi inmediata.

**b. Batch Processing:** Es una técnica para procesar datos que consiste en procesar grandes cantidades de datos en una sola operación. Los datos se recolectan después de cierta ventana de tiempo o memoria definida y luego se procesan por grupos de

datos. El procesamiento no es en tiempo real y por lo tanto los resultados no se obtienen rápidamente.

#### **4. ¿En qué consiste datos estructurados, semi estructurados y no estructurados?**

**a. Estructurados:** Son datos que tienen un esquema bien definido. Por ejemplo, la información que se almacena en las bases de datos relacionales tiene esquema con datos como el id, nombre, fecha, etc.

**b. Semiestructurados:** Son datos que tiene cierta estructura o esquema definido, pero no son necesariamente estructurados y tampoco no estructurados. Ejemplo: JSON.

**c No estructurados:** Son datos que no tienen un esquema o patrón definido. Ejemplo: Imágenes o audios.