

Prueba Corta #2 Bases de datos II

Isaac Araya Solano 2018151703

1. Explique en qué consisten los siguientes conceptos:

a. Data Warehouse

Es un sistema de almacenamiento de datos que se usa para el manejo de grandes cantidades de datos empresariales. Permite un fácil acceso a información analítica y Business Intelligence. Los data warehouses almacenan los datos optimizados para consultas y análisis y pueden ser usados para análisis predictivos, minería de datos y visualización de tendencias, entre otras cosas.

b. Data Lake

Es un sistema de almacenamiento de datos que permite almacenar cantidades grandes de datos sin necesidad de predefinir un modelo de datos. Los datos pueden ser procesados posteriormente dependiendo de la necesidad de los análisis y los usuarios finales. Los datos que se ingresen pueden ser estructurados o no estructurados.

c. Data Mart

Es una versión simplificada y específica de un Data Warehouse. Se enfoca en un área en particular y es usado por usuarios que tengan necesidades analíticas similares.

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

Explique.

El columnar storage beneficia a las aplicaciones ya que, al almacenar los datos por columnas en lugar de por filas permite que las consultas sean más rápidas y eficientes ya que hace que no sea necesario acceder a toda la información de una fila sino al dato que se necesita en específico, haciendo que la eficiencia en general de la aplicación aumente. Además, permite que los datos

sean comprimidos de una mejor manera ya que los datos dentro de cada columna tienen datos similares y suelen ser repetitivos. Por otro lado, facilita el procesamiento en paralelo aumentando la escalabilidad y permite un análisis mas eficiente de los datos.

3. ¿En qué consiste streaming y batch processing?

El streaming consiste en el traspaso de información o un flujo de información constante a través del tiempo. El streaming permite recibir información en pequeños fragmentos y permite acceder al contenido de forma inmediata sin necesidad de esperar a que se descargue la totalidad de los datos.

El batch processing es un método de procesamiento de datos en el que se procesan grandes cantidades de datos de forma automatizada en lotes, en lugar de procesarlos de forma individual. En este método se agrupan datos similares para procesarlos juntos y se procesan todos los lotes de datos de una sola vez, sin interacción humana.

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

Datos estructurados: Son datos que son almacenados siguiendo un esquema y/o formato. Tienen reglas para la estructura de los datos y sus relaciones. Suelen ser altamente organizados y fáciles de buscar.

Datos semi estructurados: Son datos que tienen cierta estructura, pero no necesariamente siguen un formato estricto. Ejemplos de formatos semiestructurados son XML, JSON O JTML.

Datos no estructurados: Son datos que no tienen estructuras definidas y se presentan en formatos de texto, imagen, videos, grabaciones de audios, etc.