

Prueba corta 2

Dillan Almendares Barrantes 2020033336

1. Explique en que consisten los siguientes conceptos:

a. Data Warehouse: Grandes bases de datos orientadas al análisis de los datos estructurados, no reciben escrituras constantes y se suelen hacer lecturas muy grandes, son principalmente usados para business intelligence.

b. Data Lake: Repositorio donde se almacena raw data de cualquier tamaño, permitiendo guardar los datos que no son compatibles con el data warehouse.

c. Data Mart: Es un tipo de data warehouse para un área funcional en específico para que sea más sencillo de crear y administrar.

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

Explique.

Aumenta el rendimiento en consultas de lectura porque no tiene que leer columnas que no fueron solicitadas, ya que cada columna está en un bloque de datos distinto y como cada bloque de datos le pertenece a una columna con un solo tipo de datos, se pueden comprimir de mejor manera.

3. ¿En que consiste streaming y batch processing?

El streaming se trata de un flujo constante de datos que debe ser procesada y el batch processing en procesar tareas en lotes cuando hay suficientes recursos disponibles.

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

- Los datos no estructurados carecen de un esquema bien definido, por ejemplo, los videos o imágenes.
- Los datos estructurados tienen nombre y tipo como las tablas sql.
- Los semi estructurados tienen un formato definido pero el tipo tiene cierta libertad como los JSON, donde se representan los datos como strings.