

Prueba corta #2

Max Richard Lee Chung - 2019185076

1. Explique en que consisten los siguientes conceptos:

a. Data Warehouse: Depósito central de información provenientes de una o más fuentes de datos, ya sean de sistemas de procesos u otras bases de datos. Las empresas lo utilizan para el propósito de reportes y análisis. Rápido análisis de información en cantidades masivas.

b. Data Lake: Conjunto de almacenamientos de datos con fácil acceso (escritura) y salida (consulta) de archivos abiertos.

c. Data Mart: Parte o área específica de un almacén de datos. Alivian el trabajo por medio de "sharding"

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

Las bases orientadas a columnas organiza cada columna en un conjunto de bloques, esto permite que sea mucho más eficiente en las entradas y salidas de información. La compresión de datos es más eficiente porque cada bloque almacena un único tipo de dato.

3. ¿En qué consiste streaming y batch processing?

Streaming proccessing consisten en la recolección, almacenamiento y procesamiento de grandes cantidades de datos de forma continua por medio de bloques o intervalos de tiempos predefinidos. Batch proccessing tiene tres clases diferentes, las cuales son "extract transform load", "extract load transform" y "online analytical proccessing". De forma secuencial, el primero consiste en la extracción continua de datos de múltiples fuentes de una datawarehouse, el segundo consiste en el proceso de cargar los datos almacenados a un sistema y por último, consiste en almacenar datos históricos agregados en esquemas multidimensionales.

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

Los datos estructurados, semi estructurados y no estructurados consisten en formas de almacenamiento, procesamiento y uso de los datos, ya que de esta forma, se define cuál tipo de bases de datos utilizar dependiendo de las características necesarias (tales como velocidad de extracción o carga de datos, procesamiento de datos continuos, entre otros). De esta forma se encuentran las estructuras relacionales (estructuras complejas) o no relacionales (estructuras no complejas) con el fin de crear una visualización adecuada para el caso, necesidad o problema en cuestión.

Referencias

- AWS (2023). Data Warehousing on AWS.