

Prueba Corta # 1

Fecha de entrega: 26/08/22 11:59 pm

Estudiante: David José Espinoza Soto - 2016012024

Explique en que consisten los siguientes conceptos

- Data Warehouse: Entiendo que es un tipo de base de datos optimizada para el almacenamiento y análisis de datos masivos.
- Data Lake: Es como un subtipo de data warehouse, ya que también almacena masivamente datos, solo que estos datos son raw data o datos crudos, y lo hace porque es más barato (no procesa los datos solo los manda a almacenar).
- Data Mart: Entiendo que es un tipo de data warehouse especializada en un tema o área de negocios dentro de una organización, y que su propocito es recolectar datos para la toma de deciciones.

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

Optimiza el rendimiento de las consultas analíticas, ya que reduce los requisitos generales de E/S de disco y reduce la cantidad de datos que se necesitan cargar desde el disco.

¿En que consiste streaming y batch processing?

Un streaming hace referencia a un flujo constante de datos

El procesamiento por lotes o batch processing tiene tres tipos de procesos: ETL, ELT y OLAP.

Extract Transform Load o ETL es el proceso de extraer mucha información de diferentes fuentes y cargarlos en el sistema para almacenarlos, como vienen de diferentes fuentes se modifican para limpiarlos, depurarlos y darles formato (basicamente los transforman).

Extract Load Transform o ELT, es una variacion del ETL ya que primero extrae los datos y los carga en el sistema, una vez en el sistema los datos se transforman.

Online Analytical Processing o OLAP permite la extracción de datos de esquemas multidimensionales (no sé que es un esquema multidimencional pero supongo que tiene que ver con la red o con un sistema especializado) para el desarrollo de informes y análisis.

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

Los datos estucturados siguen un esquema fijo o tabular, donde todos los registros tienen los mismos tipos de datos (serian los que meteriamos en una base de datos relacional). Los no estructurados no siguen ninguna estructura como por ejemplo texto, audio o video. Yo interpretaria pues que los datos semi estructurados tiene un esquema flexible, siguen un orden pero no tan estricto, como los documentos que se guardan en una base de datos no relacional.

Referencias bibliograficas

- colaboradores de Wikipedia. (2021, 12 octubre). Data mart. Wikipedia, la enciclopedia libre. Recuperado 26 de agosto de 2022, de 'https://es.wikipedia.org/wiki/Data_mart'
- Columnar storage - Amazon Redshift. (s. f.). Amazon Redshift Databas Developer Guide. Recuperado 26 de agosto de 2022, de 'https://docs.aws.amazon.com/redshift/latest/dg/c_columnar_storage_disk_mem_mgmt.html'
- Introducción a AWS Batch - AWS Batch. (s. f.). Amazon Redshift Databas Developer Guide. Recuperado 26 de agosto de 2022, de 'https://docs.aws.amazon.com/es-es/batch/latest/userguide/Batch_GetStarted.html'