

Prueba Corta 1

Gabriela Gutiérrez Valverde - 2019024089

Explique en que consisten los siguientes conceptos:

1. **Data Warehouse:** sirve para hacer análisis (insights) sobre un gran volumen de datos y descubrir los patrones que existan.
2. **Data Lake:** sistema que permite almacenar, proteger y modificar grandes cantidades de datos (Simple Storage Service), incluyendo los datos crudos.
3. **Data Mart:** es un sistema simple, parecido a un data warehouse pero se encuentra centrado en un área o tema específico. La consulta puede volverse compleja debido a esta distribución de los datos.

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

Un sistema orientado en columnas es más eficiente en algunos casos ya que genera una respuesta más rápida al realizar lecturas. Esto porque se llama solamente la columna deseada, en lugar de tener que llamar todas las filas. Además se puede tener una mejor comprensión, ya que cada bloque físico contiene el mismo tipo de datos, esto permite utilizar algoritmos más eficientes.

¿En que consiste streaming y batch processing?

Ambos se tratan de procesamiento de datos para poder obtener y analizar aquellos datos que tengan una importancia para el negocio. Streaming significa que hay un flujo constante de datos (real time), registro por registro o en ventanas de tiempo móviles. Batch significa que los datos se encuentran en disco y esta se divide en ETL (Extract, Transform, Load), ELT (Extract, Load, Transform) y el OLAP(Online Analytical Processing).

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

- **Datos estructurados:** siguen un esquema definido fijo o tabular. Los datos se encuentran estandarizados.
- **Datos no estructurados:** No siguen patrones, carecen de estructura o arquitectura identificable. NLP (Natural Language Processor).
- **Datos semi estructurados:** No necesariamente deben tener los mismos atributos, por ejemplo, un archivo Json.