

Base de Datos II (IC4302) – Semestre 2, 2023

Lectura 1 – Data Warehousing on AWS

David Suárez Acosta – 2020038304

¿En qué consisten datos estructurados, semiestructurados y no estructurados? Comente ejemplos de estos tipos de datos.

- Los datos estructurados son datos que se rigen por cierta estructura establecida. Ejemplos de estos datos son los datos que se almacenan en tablas de filas o columnas.
- Los datos semiestructurados son datos que no siguen una estructura como la tabular pero utiliza etiquetas para separar los elementos y ordenar los datos. Un ejemplo de estos datos es el lenguaje XML.
- Los datos no estructurados son datos que no siguen ninguna forma de orden o modelo. Ejemplos de estos datos son los mensajes de texto o datos multimedia.

¿En qué consisten datos de series de tiempo? ¿Se consideran logs, datos de series de tiempo?

Los datos de series de tiempo (Streaming data) son volúmenes muy grandes de datos captados cada cierto tiempo (a veces terabytes por hora) los cuales necesitan ser recolectados, almacenados y procesados continuamente.

Los datos de series de tiempo si se pueden considerar logs, ya que el proposito de un log es ayudar a hacer troubleshoot de problemas, hacer revisiones y realizar analíticas. Los datos de serie de tiempo se consideran logs despues de ser procesados ya que cumplen con los propósitos del log ya mencionados.

¿Comente diferencias entre Lake house, Data warehouse y Data mart?

- Un Lake House es una combinación entre data warehouses y data lakes. Con un lake house se pueden consultar datos a traves de data warehouses, data lakes y bases de datos operacionales.
- Un Data Warehouse permite correr analíticas rápidas en volúmenes grandes de datos y permite encontrar patrones en los datos.
- Un Data Mart es una versión simple de data warehouses que se concentra en un área o tema específico.

¿En qué consiste Row-oriented Column-oriented databases? Suponiendo que existe una tabla en una base de datos relacional con 10 columnas cuyos nombres son column1, column2, ..., column10, ¿Una consulta como "SELECT column1, column2 FROM tabla" se vería mas beneficiada por Row-oriented o Column-oriented? Explique.

Row-oriented Databases son bases de datos que almacenan los datos en filas, las cuales se almacenan en un bloque físico. Column-oriented Databases son bases de datos que almacenan en columnas, donde cada columna se organiza en su propio set de bloques físicos en lugar de guardar toda la fila en un bloque.

Una consulta como un SELECT mencionado sería mas eficiente en un row-oriented database ya que en lugar de consultar todas las filas solamente para usar la información de las dos primeras columnas, el row-oriented

permite que solo se consulten las dos columnas que se van a usar y se ignoran las otras 8 columnas, haciendo que la consulta sea más rápida.