

Jose Julian Gutierrez Badilla, carnet: 2021072199

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

Estos terminos se refieren a la forma en la cual los datos son organizados. Por un lado los datos estructurados son aquellos que siguen un formato y que ademas son faciles de almacenar, analizar y consultar, por ejemplo: un registro de los productos de una tienda, el cual se almacenan en tablas con campos como nombre, codigo, precio, etc.

Por otro lado los datos semiestructurados no siguen un esquema fijo como los datos estructurados, en este caso, los datos pueden ser almacenados en documentos tipo XML, JSON, etc. Por ejemplo: un archivo XML de los usuarios pertenecientes a la comunidad de un videojuego.

Finalmente, los datos no estructurados no poseen una organizacion ni esquema especificos, y son bastante mas complicados de analizar y almacenar que los datos estructurados y semiestructurados. Por ejemplo: videos e imagenes.

En que consisten datos de series de tiempo? Se consideran logs, datos de series de tiempo?

Los datos de series de tiempo consisten en informacion capturada y ordenada en funcion del tiempo, lo que representa mediciones o valores utilizados para estudiar tendencias, comportamientos, entre otras, en determinado rango de tiempo. Por ende, los logs si se podrian considerar datos de series de tiempo, ya que es informacion que el propio sistema genera para poder solucionar problemas, hacer analisis, etc.

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

Por un lado, un Lake house combina los elementos del data warehouse y data lake, obteniendo flexibilidad y capacidad para almacenar datos en su forma original, ya sean datos estructurados, semiestructurados o no estructurados, sin la necesidad de un esquema, y a la vez, puede proporcionar un esquema organizado y estructurado para el almacenamiento de los datos, lo que permite que las consultas sean mas rapidas y eficientes.

En que 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 veria mas beneficiada por Row-oriented o Column-oriented? Explique

Por un lado, las Row-oriented databases son orientadas a filas. Los datos se almacenan en filas, es decir, un registro es representado por una fila completa, con todos sus atributos. Por otro lado, las Column-oriented databases son orientadas a columnas, lo que quiere decir que un registro es representado por una columna completa. Por ende, la consulta especificada se veria mas beneficiada por un column-oriented database, ya que solo es necesario consultar la informacion de dichas columnas, no los atributos de sus filas.