

Instituto Tecnológico de Costa Rica

Jesús Andrés Cortés Álvarez

Carné: 2021579439

Lectura 1

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

- **Datos Estructurados:** Permiten tanto como el software, como personas accedan a estos de una forma eficaz y sencilla, mediante un formato estandarizado. Además son datos tabulares con filas y columnas, que definen a sus atributos. Es posible almacenar datos estructurados en bases de datos relacionales y administrarlos con el lenguaje de consulta estructurado como SQL. Ejemplos:
 - Bases de datos SQL
 - Directorios de productos
 - Control de inventario
- **Datos no estructurados:** Es información sin un modelo de datos establecido, mejor dicho, datos que no están ordenados por defecto. Es necesario diversas tecnologías o estrategias para realizar un análisis eficiente, dichos tipos de datos son utilizados en bases de datos NoSQL. Ejemplos:
 - Correos electrónicos
 - Imágenes
 - Archivos de texto
- **Datos semiestructurados:** Es posible mencionar que estos tipos de datos se encuentran en un punto intermedio al tipo datos ya mencionados, ya que estos carecen de un modelo de datos relacional o tabular específico. Como se menciona en la lectura con Amazon Redshift Spectrum, se admiten datos semiestructurados y amplía su almacén a su lago de datos, ya que esto le permite realizar análisis en datos de eventos de gran volumen. Ejemplos:
 - JSON
 - XML
 - Archivos de páginas web

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

- Proceso en el cual se analizan observaciones sobre valores que llega a tomar una variable en diferentes puntos del tiempo, según lo investigado los logs si se consideran datos de series de tiempo ya que estos disponen de acontecimientos de cambios de forma cronológica, como también actualizaciones, que han sucedido en el sistema, para de esta manera explicar su comportamiento.

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

- Según lo leído los Lake house permiten realizar consultas muy rápidas y profundas en almacenes de datos, warehouses y data lakes, que en otros modos no sería posible, además la posibilidad de almacenar datos en formato de archivos abiertos y esto permite mayor facilidad en la disponibilidad de datos para otras herramientas. Por otro lado, los Data warehouse poseen otra diferencia que es ejecutar análisis de forma rápida con volúmenes grandes de datos, además lo utilizan para realizar análisis fuera de línea. Y para finalizar el Data mart se utiliza en áreas específicas, además de que poseen una sencilla forma de ser diseñados, construidos y administrados.

¿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:** Las bases de datos orientadas a filas, almacenan las filas completas en un bloque físico, poseen un alto rendimiento en operaciones de lectura a través de índices de secundarios. Poseen una limitación en recursos disponibles para una sola máquina.
- **Column-oriented databases:** En este caso las bases de datos orientadas a columnas, organizan las columnas en su propio conjunto de bloques físicos, en lugar de empaquetar todas las filas, facilita mayor eficiencia en entrada y salida de consultas.
- La consulta mencionada en el enunciado, se verá beneficiada si se utiliza las bases de datos orientadas a columnas, ya que como se mencionó permite tener entrada y salida muy eficientes para consultas, ya que solo deben leer las columnas que se acceden mediante una consulta desde disco.

Referencias

Sitios web utilizados como referencia, para el desarrollo de la evaluación.

<https://aws.amazon.com/es/what-is/structured-data/>

<https://www.tibco.com/es/reference-center/what-is-time-series-analysis#:~:text=Datos%20de%20series%20temporales%3A%20Conjunto,mismo%20punto%20en%20el%20tiempo.>

<https://keepcoding.io/blog/que-son-logs-y-para-que-sirven/>