

Instituto Tecnológico de Costa Rica

Ingeniería en Computación.

Bases de Datos II.

1er semestre

Prueba Corta #2

ESTUDIANTE:

Edgar André Araya Vargas

2020142856

PROFESOR:

Gerardo Nereo Campos Araya

Grupo: 1

Fecha de entrega:

Domingo 19 de Marzo del 2023.

1.Explique en que consisten los siguientes conceptos:

a. Data Warehouse

Un Data Warehouse busca proporcionar una vista consolidada, consistente y coherente de los datos empresariales que se pueden usar para análisis y generación de informes. Los datos almacenados en un almacén de datos se estructuran en dimensiones y métricas que permiten a los usuarios analizarlos y obtener información comercial valiosa. En pocas palabras, es una base de datos diseñada para almacenar y gestionar grandes cantidades de datos que se utilizan para la toma de decisiones.

b. Data Lake

A diferencia de los Data Warehouses, que almacenan datos predefinidos y estructurados para análisis específicos, los lagos de datos permiten que las organizaciones almacenen, procesen y analicen aún más los datos de una variedad de formatos y fuentes. Esto significa que los datos pueden usarse para diferentes propósitos y analizarse de diferentes maneras. Un Data Lake puede integrar varias tecnologías como Hadoop y Spark para almacenar y procesar grandes cantidades de datos.

c. Data Mart

Los Data Marts están diseñados para satisfacer las necesidades de información específicas de un departamento o grupo de usuarios en particular, esto mediante una estructura de almacenamiento hecha para proporcionar un acceso rápido y específico a la información relacionada con un área específica de la organización.

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

Las bases de datos en columnas son particularmente eficientes en el manejo de grandes cantidades de datos, lo que las convierte en una herramienta valiosa para aplicaciones que requieren el análisis de grandes cantidades de información. La columnarización de

datos permite una mayor compresión y una menor lectura de datos del disco, lo que da como resultado una velocidad de consulta significativamente mejorada y una mayor eficiencia de procesamiento, lo que genera beneficios para las empresas y organizaciones que utilizan estos sistemas, especialmente en términos de ahorro de tiempo y recursos y la capacidad de administrar de manera efectiva grandes cantidades de datos.

3. ¿En que consiste streaming y batch processing?

El streaming implica el procesamiento continuo de datos en tiempo real a medida que se reciben, lo que puede producir resultados más rápidos y precisos. El procesamiento por lotes o batching se refiere al procesamiento de una gran cantidad de datos en conjunto, que luego se procesa en lotes posteriores.

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

Una tabla de base de datos o un archivo CSV son ejemplos de datos estructurados, que tienen un formato bien definido y organizado en filas y columnas. Los datos semi estructurados no tienen una estructura definida, pero tienen cierta organización en forma de etiquetas o metadatos. Los documentos HTML, XML y JSON son ejemplos de datos semi estructurados. Finalmente, los datos no estructurados son completamente desorganizados y no tienen ninguna estructura definida, por esto, los mas dificiles de procesar.