

Prueba Corta # 2

1. Explique en qué consisten los siguientes conceptos:
 1. Data Warehouse: un repositorio central de información que viene de uno o más bancos de información.
 2. Data Lake: un repositorio central creado para guardar, procesar y asegurar grandes cantidades de información estructurada, semi-estructurada y no estructurada.
 3. Data Mart: una data warehouse dedicada a un área funcional o tema.
2. ¿De qué forma se benefician las aplicaciones del uso de Columnar Storage? Explique. R/ Las aplicaciones se benefician del uso de Columnar Storage porque, al guardarse gran cantidad de datos (ej. todas las mascotas disponibles para comprar en una tienda) en columnas, si se quisiera obtener un dato de todas estas mascotas (ej. la especie), es mucho más eficiente y rápido obtener solo este dato (o sea, solo la columna "especie"), comparado con buscar fila por fila solo para obtener uno de sus datos.
3. ¿En qué consiste streaming y batch processing? R/ Streaming consiste en un flujo constante de datos, a veces terabytes por hora, que se necesitan recolectar, guardar y procesar continuamente. Por otro lado, el batch processing, tal como el nombre lo dice, consiste en procesar grandes volúmenes de datos (batches de datos) recolectados a través de un tiempo.
4. ¿En qué consiste datos estructurados, semi estructurados y no estructurados?
 1. Datos estructurados: datos con un esquema bien definido (Ej. datos en bases de datos relacionales, datos en tablas, spreadsheets, etc.)
 2. Datos no estructurados: datos que no siguen ningún patrón (Ej. Documentos de texto, imágenes, videos, emails, etc.)
 3. Datos semi estructurados: datos que tienen un tipo de estructura, pero no siguen un esquema estricto, tiene características de datos estructurados y no estructurados (Ej. JSON, XML, etc.)