

1.

---

a.

Los data warehouse consisten en un almacen de datos en donde se pueden realizar operaciones de escritura y lectura de grandes volúmenes de datos. Para beneficiarse de este almacen, se haría uso de un data pipeline para extraer los datos de un sistema y guardarlos en el almacen.

b.

Los data lake contienen grandes cantidades de datos en su formato original que están guardados en repositorios para poder analizar los datos y así utilizar las tecnologías de Big Data.

c.

Los data mart son un data warehouse especializado en un área de funcionalidad específica, como en negocios y marketing.

2.

---

Las consultas de entrada y salida son más eficientes ya que la base de datos solo tiene que leer los datos de las columnas que se piden en la consulta. Esto nos garantiza que todos los datos de una columna que nos devuelve sean del mismo tipo, por lo que podemos realizar operaciones más eficientes para los datos de la columna solicitada.

3.

---

Streaming consiste en la recolección, almacenamiento y procesamiento de datos de las aplicaciones web, celulares, aplicaciones y servicios.

Batch processing se divide en dos procesos. Extract Transform Load es el proceso de extraer los datos de un sistema para procesarlos y guardarlos en un sistema de almacén de datos para ser analizados. Extract Load Transform guarda los datos extraídos en el sistema de almacén de datos y luego los procesa.

4.

---

Los datos estructurados son los datos que siguen un esquema fijo, por ejemplo una tabla de una base de datos relacional. Los datos semi estructurados son los datos que no siempre tienen los mismos atributos, por ejemplo los archivos xml y json. Los datos no estructurados son los datos que no siguen un patrón, por ejemplo los textos.