

### 1. Explique en que consisten los siguientes conceptos:

- a. **Data Warehouse:** Datos procesados y estructurados con una razón, estos realizan operaciones de escritura por lotes y lectura de grandes volúmenes de datos. Emplean un esquema desnormalizado por el requisito de alto rendimiento de datos.
- b. **Data Lake:** Es un patrón arquitectónico que combina los mejores elementos de los *data warehouses* y los *data lake*. Asimismo, es un lugar donde los datos no han sido procesados, aún sin propósito específico
- c. **Data Mart:** Es una forma simple de un *data warehouse* centrado en un área funcional o de tema específico. Se pueden construir *data marts* de grandes *data warehouse*, así como almacenes operativos, o un híbrido de los dos. Estos son fáciles de diseñar, construir y administrar.

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

Les permite ser más eficiente en los elementos de entrada/salida (I/O) para consultas de solo lectura. Además, que cuando se utiliza la orientación a columnas los datos se empaquetan en un propio conjuntos de bloques y estos solo contienen el mismo tipo de dato, por lo tanto, la base de datos podría realizar algoritmos de comprensión muy eficientes por los tipo de datos.

### 3. ¿En que consiste streaming y batch processing?

**Streaming:** Es el procesamiento de datos con un flujo continuo de datos a medida que se generan. Se pueden procesar de manera secuencial e incrementarse registro por registro.

**Batch processing:** Se utilizan para manejar grandes cantidades de datos o cuando las fuentes de datos son heredados que no pueden entregar datos en flujos. El procesamiento de datos puede ser de hasta minutos, horas o días.

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

**Estructurados:** Información que ha sido formateada y transformada en un modelo de datos bien definido.

**Semi-estructurados:** Son un tipo de dato que tienen algunas características consistentes y definidas. Estos no se limitan a una estructura rígida como lo que en las bases de datos relacionales.

**No-estructurados:** Los datos se representan de manera absoluta sin procesar, estos datos son difíciles de procesar por su compleja organización y formato. Están presentes muchos campos que conocemos por ejemplo, redes sociales, chats, imágenes satelitales, entre otros.