

PART 1 – Introduction Data Warehouse

1. Sebutkan perbedaan data warehouse dan data lake !

Jawaban :

- Data Warehouse Memiliki struktur yang terdefinisi dengan baik, biasanya dalam bentuk model data yang diatur sebelumnya seperti star schema atau snowflake schema. Data warehouse biasanya digunakan untuk Analisa bisnis, pelaporan, dan pemodelan prediktif.
- Data Lake tidak memiliki struktur schema. Data disimpan dalam format asli tanpa pemodelan dan pengolahan awal yang signifikan. Data lake digunakan untuk analisis yang lebih eksploratif dan fleksibel.

2. Apa yang membedakan teknologi database untuk data warehouse (OLAP) dengan teknologi database konvensional (OLTP)

Jawaban :

- Database OLTP digunakan untuk mendukung transaksi sehari-hari dan operasi bisnis yang cepat dan efisien. Tujuannya adalah untuk menyimpan, memperbarui, dan mengambil data transaksi secara efisien dalam waktu nyata. Database OLTP cenderung memiliki struktur data yang sederhana dan normalisasi, dengan fokus pada efisiensi transaksi dan konsistensi data.
- Database OLAP digunakan untuk analisis dan pelaporan data yang kompleks. Tujuannya adalah untuk menyediakan akses cepat dan efisien ke data historis, memungkinkan pengguna untuk menganalisis tren, melakukan pemodelan prediktif, dan membuat laporan bisnis. Database OLAP cenderung menggunakan skema data yang lebih denormalisasi untuk memfasilitasi analisis yang cepat dan kompleks.

3. Teknologi apa saja yang biasa dipakai untuk data warehouse ?

Jawab :

- Teknologi data warehouse terdapat beberapa yang biasa digunakan diantaranya, AWS Redshift, Google Big Query, Apache Cassandra, Apache HBase, MongoDB, Snowflake, Microsoft Azure Synapse Analytics. Apache Hive, Apache Druid, ClickHouse. Postgre, dll.

4. Tulis perintah dari proses instalasi citrus menggunakan docker compose !

Jawab :

- Maaf mas, kata mas bilal untuk nomor 4 di skip karena sama seperti part 2.

5. Jelaskan perbedaan antara access method heap dan columnar pada citrus !

Jawab :

- Heap Access Method : Metode akses ke data yang mirip dengan struktur "tumpukan" di mana data disimpan dalam bentuk baris-baris yang saling terkait. Setiap baris data disimpan secara berurutan dalam blok penyimpanan yang berdekatan di dalam tabel.
- Columnar Access Method : Metode akses ke data di mana data disimpan dalam bentuk kolom-kolom yang saling terkait, bukan baris-baris. Setiap kolom data disimpan secara berurutan dalam blok penyimpanan yang berdekatan, yang memungkinkan akses yang efisien ke kolom tertentu tanpa perlu membaca seluruh baris data.