PART1

1. A. Data warehouse bertujuan untuk merancang analisis bisnis dan pelaporan, mengumpulkan dan menyimpan data dari berbagai sumber untuk analisis. Struktur data warehouse menggunakan skema yang telah ditentukan sebelumnya (predefined schema), seperti star schema atau snowflake schema, Data diubah (transformed) dan diproses sebelum dimasukkan ke dalam data warehouse (ETL - Extract, Transform, Load). Jenis datanya Biasanya menyimpan data terstruktur seperti tabel relasional dan data yang dioptimalkan untuk query SQL.

Data harus memenuhi standar kualitas dan konsistensi tinggi sebelum dimasukkan. Data warehouse lebih mahal untuk skala besar karena membutuhkan perangkat keras dan perangkat lunak khusus, Skalabilitas terbatas. Kecepatan pada data warehouse Dioptimalkan untuk query cepat dan analisis ad-hoc dengan kinerja tinggi, Indeks dan denormalisasi digunakan untuk mempercepat query. Biasanya memiliki kontrol keamanan yang ketat dan tata kelola data yang baik serta data yang dimasukkan telah melalui proses pembersihan dan validasi.

B. Data lake bertujuan menyimpan data dalam bentuk mentah atau semi-terstruktur.

Dirancang untuk menyimpan volume data yang sangat besar dan mendukung berbagai jenis analisis data, termasuk analitik besar (big data), machine learning, dan data mining.

Digunakan oleh data scientist dan pengembang yang membutuhkan akses ke data mentah untuk analisis mendalam. Struktur data lake menyimpan data dalam bentuk asli atau mentah tanpa skema yang telah ditentukan sebelumnya (schema-on-read). Jenis data lake menyimpan berbagai jenis data, termasuk terstruktur, semi-terstruktur (misalnya, JSON, XML), dan tidak terstruktur (misalnya, log file, media, teks).

Tidak ada persyaratan ketat untuk kualitas dan konsistensi data saat disimpan. Kecepatannya tidak dioptimalkan untuk query cepat karena data disimpan dalam bentuk mentah.

Analisis memerlukan pengolahan tambahan, yang bisa memakan waktu lebih lama. Keamanannya memerlukan pengaturan keamanan dan tata kelola yang baik karena menyimpan data mentah dari berbagai sumber.

Lebih rentan terhadap masalah kualitas data karena data disimpan dalam bentuk asli.

1. Teknologi database yang digunakan untuk data warehouse (OLAP) dan database konvensional (OLTP) memiliki perbedaan yang signifikan dalam hal desain, tujuan, dan cara pengoperasiannya seperti :
2. OLAP (Online Analytical Processing) :

-Dirancang untuk analisis data dan pelaporan.

-Mendukung query kompleks dan analisis historis untuk pengambilan keputusan bisnis.

-Contoh penggunaan: analisis tren penjualan, laporan keuangan, analisis pasar.

Struktur datanya Menggunakan skema data multidimensi seperti star schema atau snowflake schema. Data biasanya di-denormalisasi untuk mempercepat query analitik, dirancang untuk query baca (read-heavy).

Jenis querynya mendukung query yang kompleks dan sering kali membutuhkan pemrosesan agregat, pengelompokan, dan perhitungan. Query biasanya memerlukan waktu lebih lama untuk dieksekusi karena kompleksitasnya. Kinerjanya dioptimalkan untuk kinerja query baca yang cepat dan analisis data yang mendalam, menggunakan teknik seperti indexing, partitioning, dan materialized views untuk mempercepat query. OLAP menyimpan volume data yang besar, sering kali mencakup data historis dalam jangka waktu yang panjang. dirancang untuk analisis tren jangka panjang dan pelaporan historis. Keamanannya biasanya memiliki kontrol keamanan yang ketat untuk melindungi data analitik yang sensitif.

Konsistensi data sangat penting, tetapi tidak seketat pada sistem OLTP.

1. OLTP (Online Transaction Processing):

Bertujuan dirancang untuk menangani transaksi sehari-hari, mendukung operasi transaksi yang cepat dan efisien, seperti insert, update, dan delete.

Contoh penggunaan: sistem perbankan, aplikasi e-commerce, sistem manajemen pelanggan. Strukturnya menggunakan skema relasional yang terstruktur dan distandarisasi, data biasanya dinormalisasi untuk menghindari redundansi dan memastikan integritas data, dirancang untuk operasi baca-tulis (read-write) yang cepat dan efisien. Jenis querynya mendukung query sederhana yang cepat dieksekusi.

Fokus pada efisiensi dan kecepatan dalam operasi transaksi sehari-hari. Kinerjanya dioptimalkan untuk kecepatan dan efisiensi transaksi, menggunakan indeks dan normalisasi untuk memastikan kinerja optimal dalam operasi transaksi. OLTP menyimpan data transaksi saat ini dan biasanya hanya mencakup data dalam jangka waktu pendek.

Fokus pada penyimpanan data transaksi terbaru dan menghindari penyimpanan data historis yang besar. Keamanannya memiliki mekanisme keamanan dan integritas data yang kuat untuk memastikan setiap transaksi diproses dengan benar.

Konsistensi dan integritas data adalah prioritas utama.

1. Teknologi yang digunakan dalam datawarehouse seperti:

-Platform Data Warehouse Tradisional (Oracle Exadata, IBM Db2 Warehouse)

-Platform Data Warehouse Berbasis Cloud (Amazon Redshift, Google BigQuery, Snowflake, Microsoft Azure Synapse Analytics)

-Platform Open Source (Apache Hive, Apache Impala)

-ETL (Extract, Transform, Load) Tools (Informatica PowerCente, Talend, Apache Nifi)

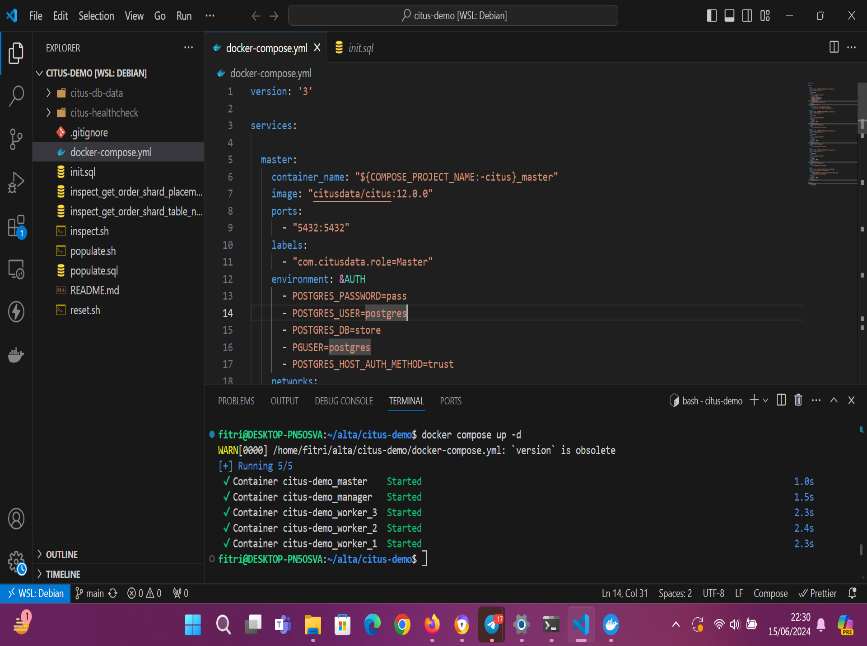
-Data Integration and Processing Tools (Apache Spark, Apache Kafka)

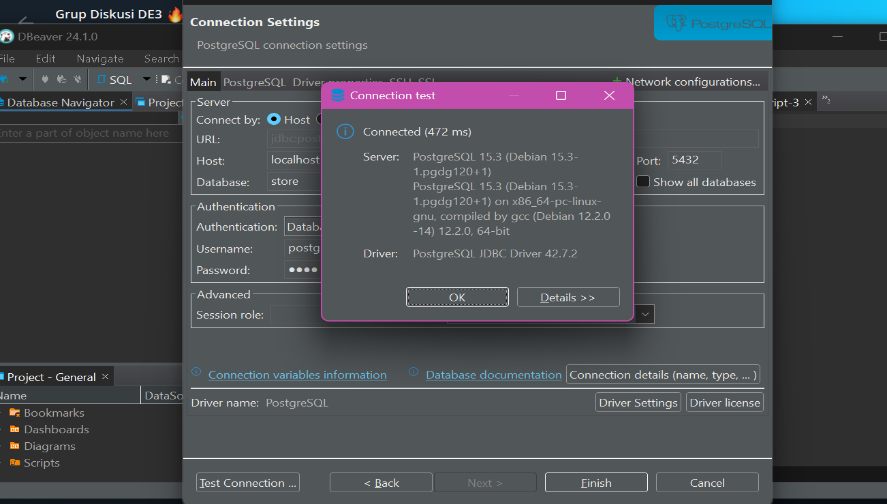
-Business Intelligence (BI) Tools (Tableau, Power BI, Qlik Sense)

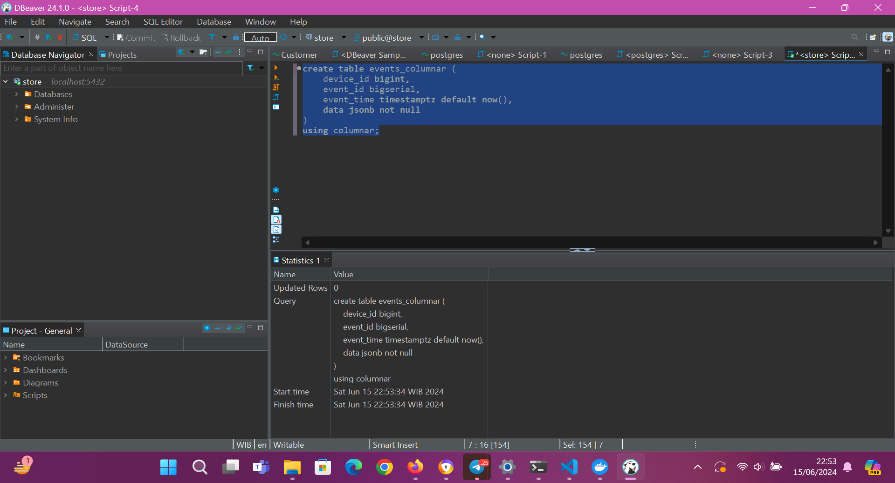
-Data Modeling Tools (Erwin Data Modeler, IBM InfoSphere Data Architect)

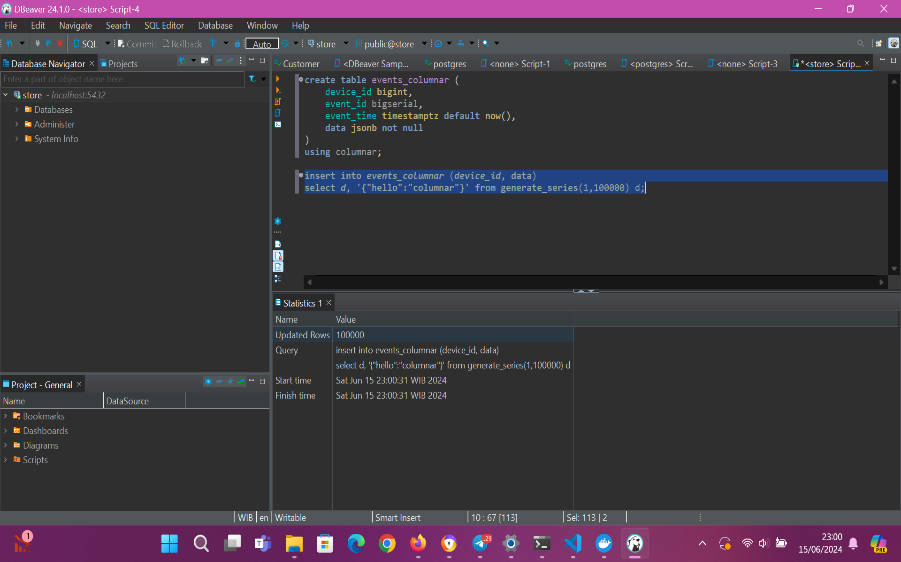
-Metadata Management (Apache Atlas, Informatica Metadata Manager).

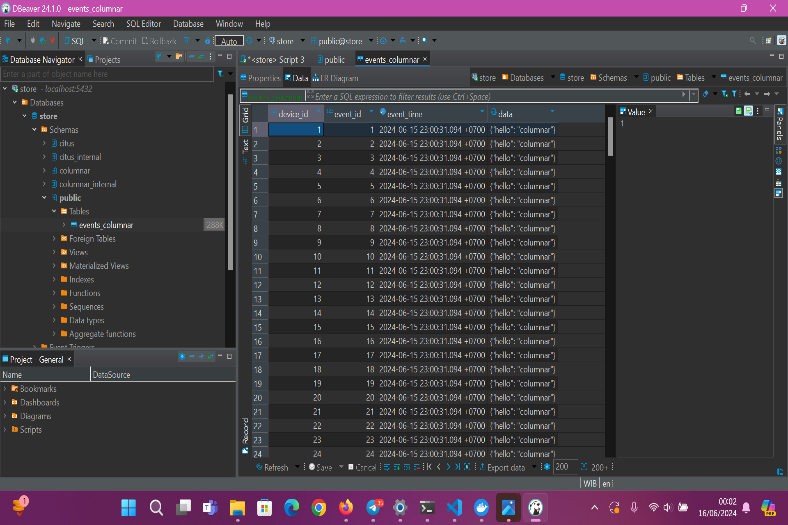
4. Berikut langkah proses instalasi citus menggunakan docker compose

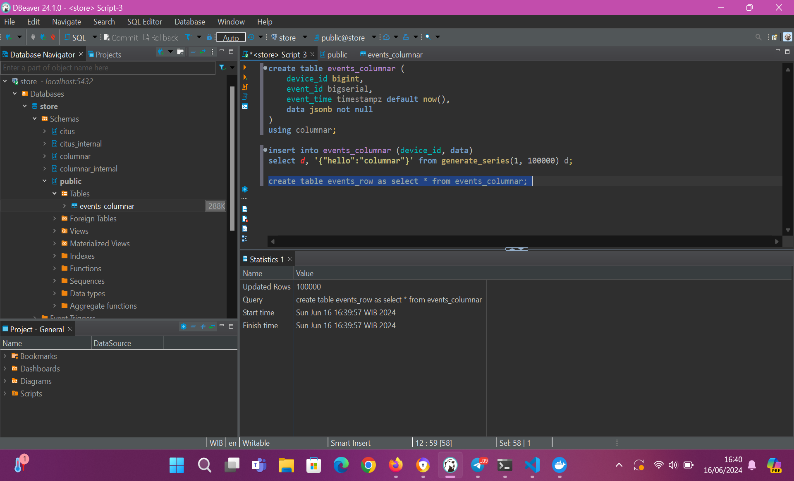


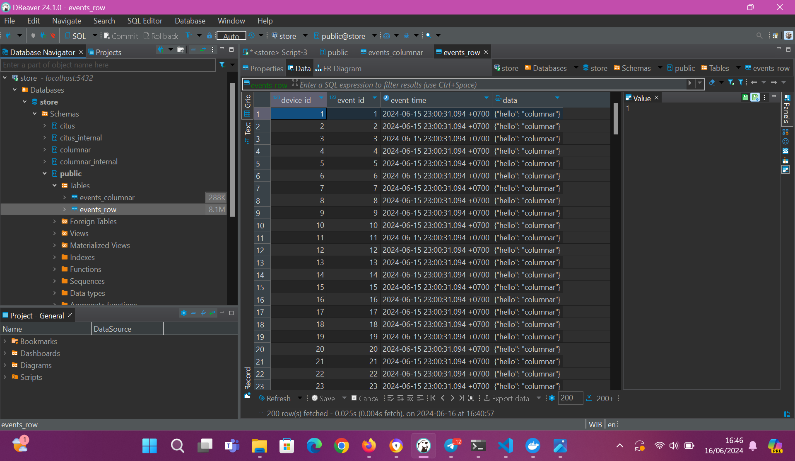


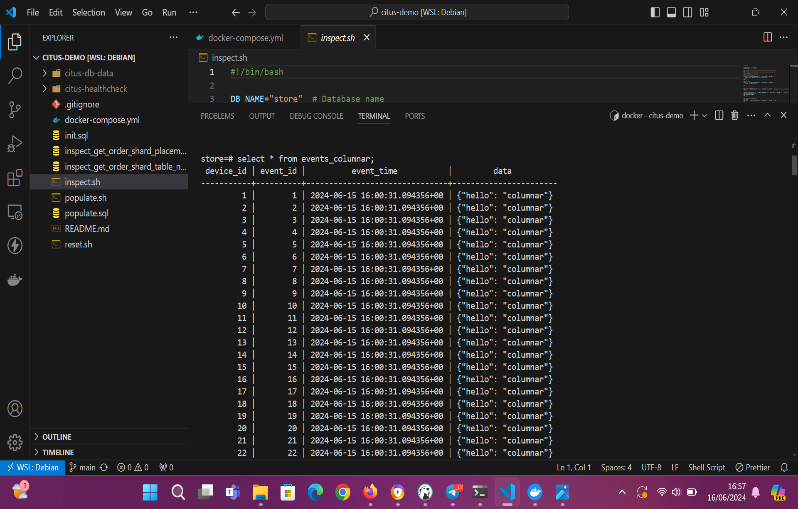


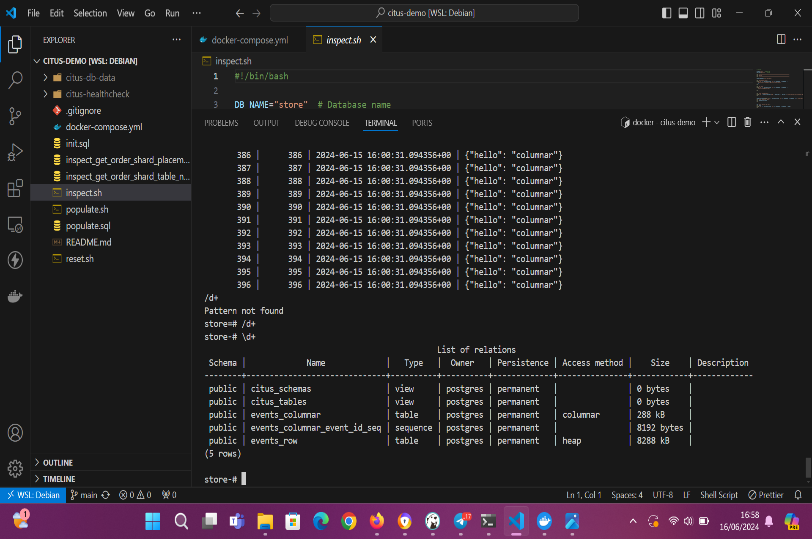












1. -Access method Heap ialah data disimpan dalam format baris, yang berarti semua kolom untuk setiap baris disimpan secara berurutan. Keuntungannya yaitu Cepat untuk operasi yang melibatkan banyak kolom dari beberapa baris, seperti SELECT \* atau INSERT sedangkan kelemahannya yaitu kurang efisien untuk query analitik yang hanya membutuhkan beberapa kolom dari banyak baris, karena semua data baris harus dibaca meskipun tidak semua kolom dibutuhkan.

Contoh heap SELECT \* FROM sales WHERE order\_id = 12345; Heap digunakan dalam tabel yang sering diakses untuk operasi penulisan dan query transaksi yang memerlukan akses cepat ke banyak kolom dalam satu baris.

-Columnar pada citus data disimpan dalam format kolom, yang berarti semua nilai dari satu kolom disimpan secara berurutan, keuntungannya sangat efisien untuk query analitik yang memerlukan agregasi atau pencarian hanya pada beberapa kolom dari banyak baris, karena hanya kolom yang dibutuhkan yang dibaca, sedangkan kelemahannya yaitu operasi penulisan (insert, update, delete) bisa lebih lambat karena data perlu disimpan dalam format kolom, dan penyisipan baris baru mungkin melibatkan beberapa operasi disk.

Contoh query SELECT SUM(revenue) FROM sales WHERE region = 'North America';

Penggunaan ruang biasanya lebih rendah karena kompresi yang lebih baik dan hanya kolom yang diperlukan yang dibaca. Dan columnar citus dsigunakan dalam tabel yang sering diakses untuk query analitik yang hanya memerlukan beberapa kolom dari banyak baris, mengoptimalkan query yang memerlukan agregasi atau pencarian kolom tertentu.