Nurul Taufik Kurrahman

Task Part 1
Introduction to Data Warehouse

1. Sebutkan perbedaan antara data warehouse dan data lake!

   Data Warehouse:
   - Menyimpan data terstruktur dan terorganisir.
   - Cocok untuk analisis bisnis dan pelaporan.
   - Memerlukan transformasi data sebelum dimuat.
   - Biasanya kurang skalabel dan lebih mahal.

   Data Lake:
   - Menyimpan data dalam berbagai format, termasuk data mentah.
   - Cocok untuk analisis yang lebih luas, termasuk Big Data dan kecerdasan buatan.
   - Tidak memerlukan transformasi data sebelum penyimpanan.
   - Lebih fleksibel dan lebih ekonomis.

   Pilihan tergantung pada jenis data dan jenis analisis yang dibutuhkan. Data Warehouse lebih sesuai untuk data terstruktur dan analisis bisnis, sementara Data Lake lebih cocok untuk data beraneka ragam dan analisis yang lebih luas.

2. Apa yang membedakan teknologi database untuk datawarehouse (OLAP) dari teknologi database konvensional (OLTP)?

   Teknologi Database OLAP (Data Warehouse):
   - Tujuan Utama: Digunakan untuk analisis dan pelaporan data untuk mendukung pengambilan keputusan bisnis.
   - Desain Skema: Biasanya menggunakan skema bintang atau snowflake yang mendukung struktur data yang optimal untuk analisis.
   - Volume Data: Mengelola volume data yang besar dengan kueri kompleks dan agregasi data.
   - Kueri Complex: Mendukung kueri kompleks yang melibatkan agregasi, penggabungan data, dan pemrosesan berat.
   - Latensi Data: Toleran terhadap latensi data yang lebih tinggi karena data biasanya tidak selalu real-time.

   Teknologi Database Konvensional (OLTP):

   - Tujuan Utama: Digunakan untuk memproses transaksi bisnis sehari-hari, seperti penjualan atau pembaruan inventaris.
   - Desain Skema: Menggunakan skema relasional untuk memastikan integritas data dan efisiensi transaksi.
   - Volume Data: Mengelola volume data yang lebih rendah dibandingkan dengan OLAP.
   - Kueri Sederhana: Menangani kueri sederhana dengan tingkat latensi yang sangat rendah.
   - Latensi Data: Menyediakan data real-time untuk mendukung transaksi bisnis.

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

   - RDBMS (Relational Database Management System): RDBMS seperti Oracle, Microsoft SQL Server, dan Teradata sering digunakan sebagai basis data untuk Data Warehouse karena mendukung skema bintang atau snowflake untuk mengorganisir data dengan baik.

   - MOLAP (Multidimensional OLAP): Teknologi ini memungkinkan penyimpanan data dalam format multidimensional yang cocok untuk analisis cepat. Contohnya adalah Microsoft Analysis Services.

   - ETL (Extract, Transform, Load) Tools: Alat-alat seperti Informatica, Talend, dan Apache Nifi digunakan untuk mengekstrak, mengubah, dan memuat data dari berbagai sumber ke dalam Data Warehouse.

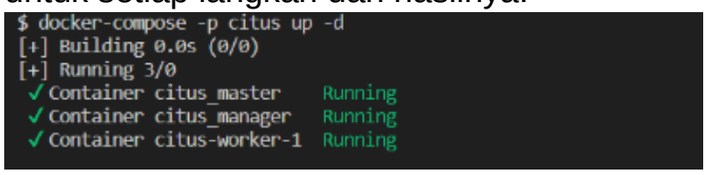
   - Columnar Databases: Database berbasis kolom seperti Amazon Redshift dan Google BigQuery sering digunakan untuk penyimpanan dan analisis data dalam format kolom yang efisien.

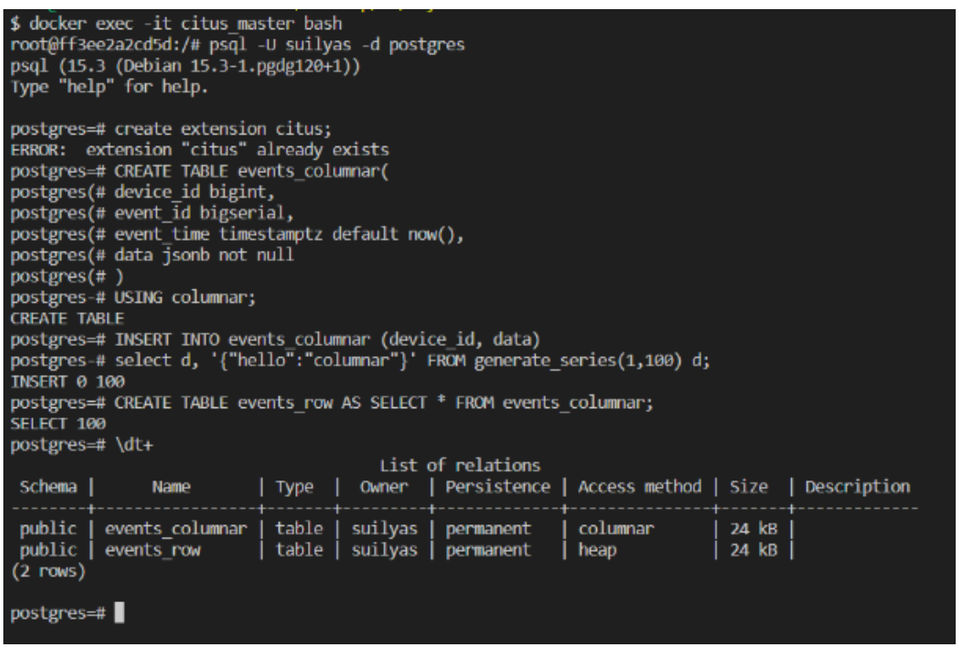
   - In-Memory Database: Database dalam memori seperti SAP HANA dan Exasol memungkinkan pemrosesan data yang sangat cepat dengan memanfaatkan RAM.

   - Big Data Technologies: Teknologi Big Data seperti Hadoop dan Spark sering digunakan untuk Data Warehouse skala besar yang memerlukan pemrosesan data distribusi dan analisis Big Data.

   - Data Warehousing Appliances: Solusi khusus seperti Netezza (IBM), Snowflake, dan Teradata menyediakan perangkat keras dan perangkat lunak yang dioptimalkan untuk Data Warehouse.

4. Tuliskan setiap perintah dari proses instalasi citus menggunakan docker compose sampai tabel terbentuk, berikan juga tangkapan layar untuk setiap langkah dan hasilnya!





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

   Access Method "Heap":

   - Penyimpanan Data: Data disimpan dalam format heap mirip dengan database PostgreSQL biasa.
   - Struktur Data: Tabel diatur dalam format baris (row-based), seperti pada database relasional biasa.
   - Performa Tulis: Cocok untuk operasi penulisan (INSERT, UPDATE) yang sering.
   - Kueri Selektif: Lebih baik untuk kueri yang memilih sedikit kolom dari banyak baris.
   - Ukuran Database: Biasanya menghasilkan ukuran database yang lebih besar karena data disimpan dalam format baris.

   Access Method "Columnar":

   - Penyimpanan Data: Data disimpan dalam format kolom, dengan kolom-kolom yang terpisah.
   - Struktur Data: Tabel diatur dalam format kolom (columnar), yang lebih efisien untuk kueri analitik.
   - Performa Kueri: Cocok untuk kueri yang memerlukan agregasi dan analisis data dengan selektivitas tinggi.
   - Kueri Selektif: Dapat memproses kueri yang memilih sebagian besar atau semua kolom dari sejumlah kecil baris.
   - Ukuran Database: Biasanya menghasilkan ukuran database yang lebih kecil karena data disimpan dalam format kolom.