**Part 1 - Introduction to Data Warehouse**

1. Sebutkan perbedaan antara data warehouse dan data lake

* Data warehouse adalah sistem penyimpanan yang digunakan untuk kueri dan analisis. Sedangkan data lake adalah repositori penyimpanan untuk semua jenis data.
* Menyimpan data terstruktur dengan skema tetap (schema-on-write). Data disusun dalam tabel-tabel dan relasi, mirip dengan database relasional. Menyimpan data dalam berbagai bentuk, termasuk terstruktur, semi-terstruktur, dan tidak terstruktur. Menggunakan skema saat membaca (schema-on-read).

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

|  | OLAP (Online Analytical Processing) | OLTP (Online Transaction Processing) |
| --- | --- | --- |
| **Tujuan** | Analisis data, pelaporan bisnis, pengambilan keputusan | Transaksi sehari-hari dan operasi bisnis |
| **Data** | Denormalisasi, data historis | Ternormalisasi, data real-time |
| **Volume Data** | Volume besar, data historis | Volume kecil per transaksi, jumlah transaksi tinggi |
| **Operasi Dominan** | Kueri baca kompleks, agregasi data | Transaksi baca/tulis sederhana |
| **Contoh Teknologi** | Amazon Redshift, Google BigQuery, Snowflake | MySQL, PostgreSQL, Oracle Database, SQL Server |

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

### Platform Data Warehouse

1. Amazon Redshift
2. Google BigQuery
3. Snowflake
4. Microsoft Azure Synapse Analytics
5. Teradata
6. Oracle Autonomous Data Warehouse
7. IBM Db2 Warehouse

* Alat ETL (Extract, Transform, Load)

1. Informatica PowerCenter
2. Talend
3. Apache NiFi
4. Microsoft SQL Server Integration Services (SSIS)
5. AWS Glue

* Alat BI (Business Intelligence)

1. Tableau
2. Power BI
3. QlikView
4. Looker

### Teknologi Pendukung Lainnya

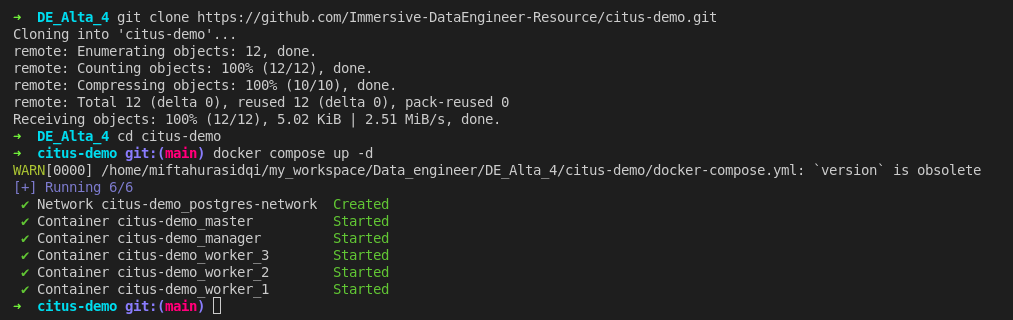
1. Apache Hadoop
2. Apache Spark
3. Presto

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

**langkah - langkah install citus**

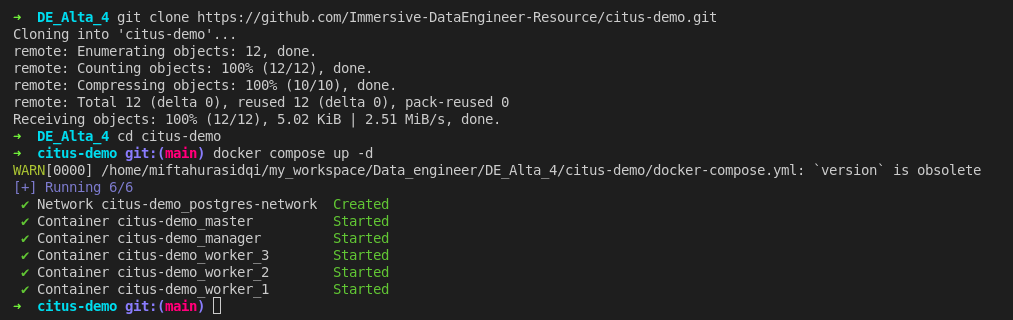
1. clone repo giithub

git clone https://github.com/Immersive-DataEngineer-Resource/citus-demo.git

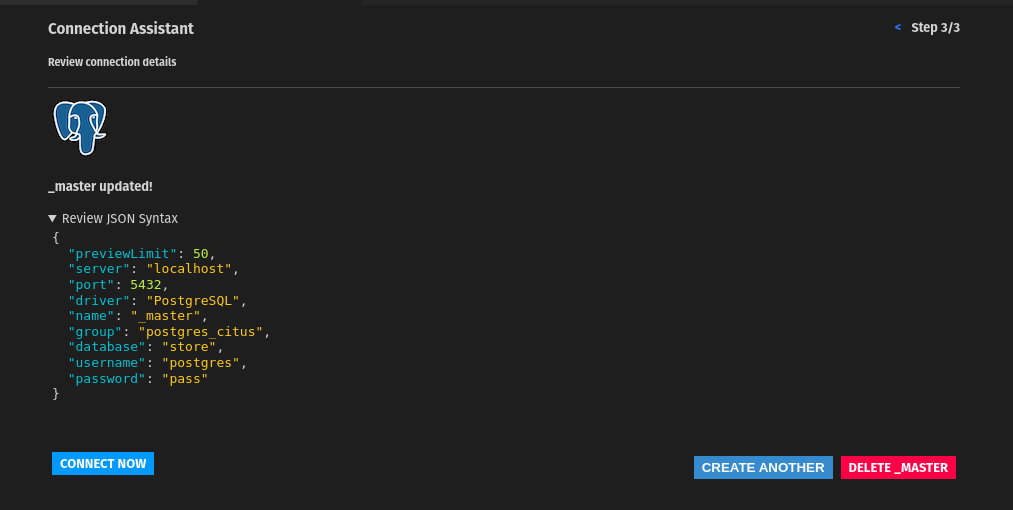


1. menjalankan docker compose

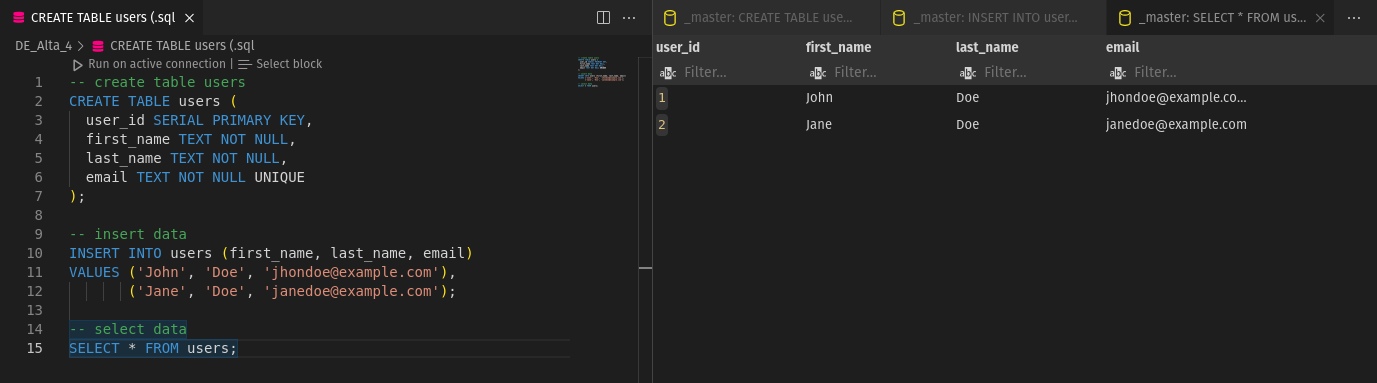
docker compose up -d



1. koneksi ke postgres



1. membuat dan menjalankan kueri



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

Heap Access Method

* Metode heap adalah cara tradisional penyimpanan data dalam PostgreSQL dan banyak sistem database relasional lainnya.
* Data disimpan dalam blok-blok pada disk, dengan setiap blok menyimpan beberapa baris dari tabel. Setiap baris dalam tabel disimpan secara utuh, dengan semua kolomnya disimpan bersama dalam satu blok.
* Kueri yang hanya membutuhkan beberapa kolom bisa kurang efisien, karena seluruh baris tetap harus dibaca dari disk.

Columnar Access Method

* Metode columnar, atau kolom, adalah cara penyimpanan data di mana setiap kolom dari tabel disimpan secara terpisah.
* Data disimpan dalam blok-blok pada disk, tetapi setiap blok hanya menyimpan data untuk satu kolom dari tabel, memungkinkan akses langsung ke kolom yang dibutuhkan.
* Kueri yang hanya membutuhkan beberapa kolom dapat dieksekusi lebih efisien, karena hanya kolom yang relevan yang perlu dibaca dari disk.