1. Perbedaan singkat antara data lake dan data warehouse:

- Data Lake:

Jenis Data: Semua jenis data (terstruktur, semi-terstruktur, tidak terstruktur).

Tujuan: Analisis data, machine learning, big data.

Fleksibilitas: Sangat fleksibel, tidak perlu ETL ketat.

Pengguna: Data scientist, data analyst, engineer.

Biaya: Relatif murah.

- Data Warehouse:

Jenis Data: Data terstruktur.

Tujuan: Laporan bisnis, analitik operasional.

Fleksibilitas: Kurang fleksibel, perlu ETL.

Pengguna: Business proffesionals

Biaya: Relatif mahal.

2. Perbedaan OLAP dan OLTP:

Tujuan:

OLAP: Analisis data dan pelaporan.

OLTP: Operasi transaksi harian.

Desain Data:

OLAP: Skema bintang atau salju, teroptimasi untuk query analitis.

OLTP: Skema relasional yang dinormalisasi, teroptimasi untuk transaksi cepat.

Volume dan Struktur Data:

OLAP: Menangani data historis besar, fokus pada agregasi.

OLTP: Menangani data terkini dan transaksi rutin.

Query dan Operasi:

OLAP: Query kompleks dan read-intensive.

OLTP: Query sederhana dan read-write intensive.

Indeks dan Penyimpanan:

OLAP: Indeks bitmap, penyimpanan kolom.

OLTP: Indeks B-tree, penyimpanan row-based.

Arsitektur:

OLAP: Multi-tier dengan ETL dan alat BI.

OLTP: Client-server atau peer-to-peer untuk transaksi operasional.

3. Teknologi yang biasanya dipakai untuk data warehouse

- AWS Redshift

- Google Big Query

- Clickhouse

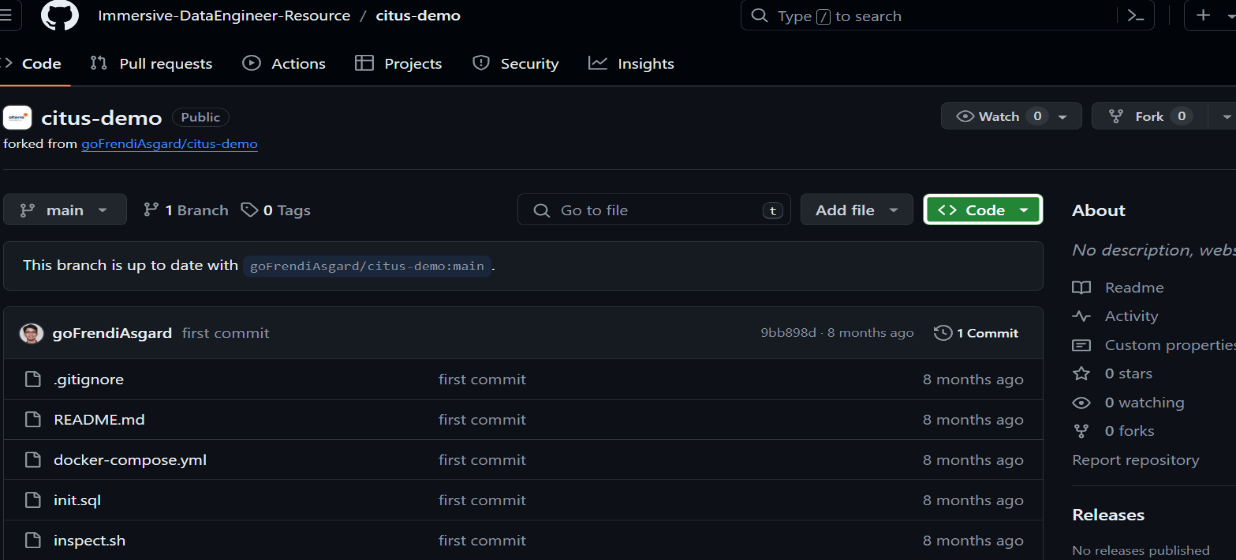
- Snowflake

- Databricks

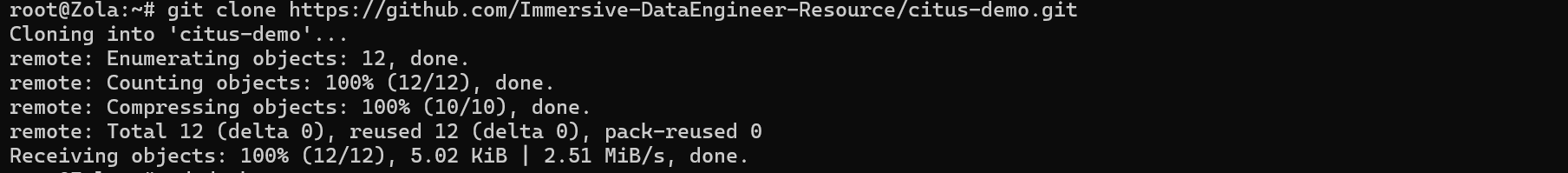
- Apache Dorris

- Postgre (with Citus extension)

4. Langkah-langkah instalasi citus:

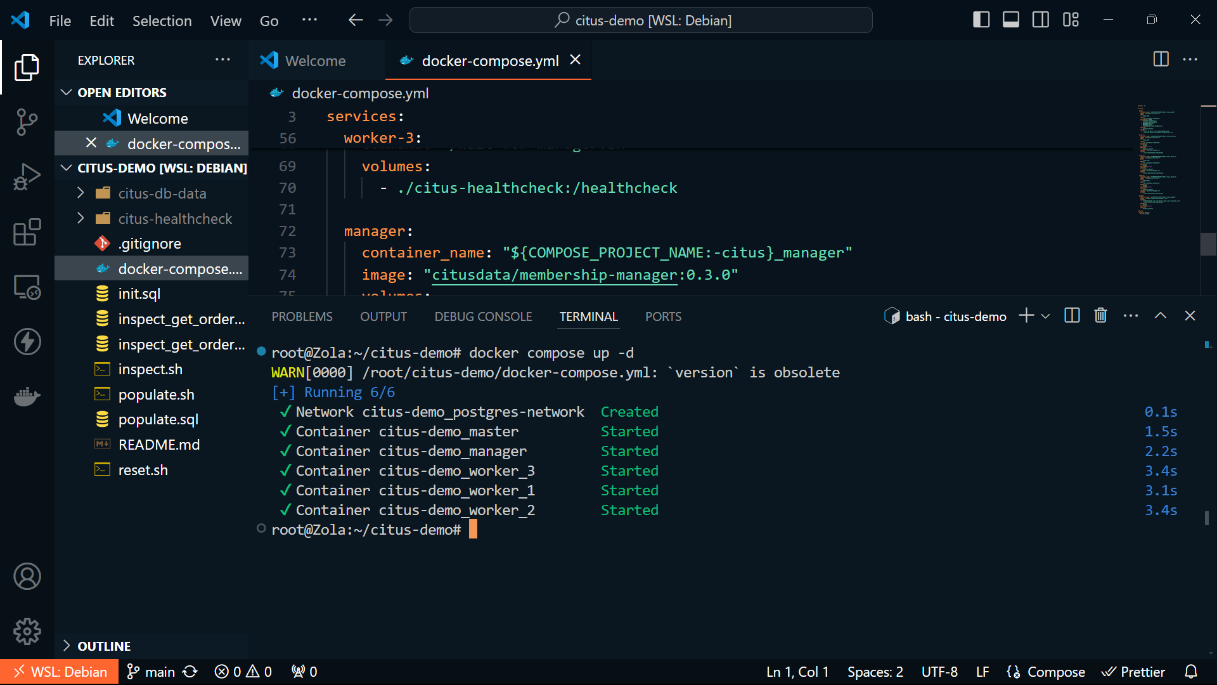
1). Buka repo github: [https://github.com/Immersive-DataEngineer Resource/citus-demo](https://github.com/Immersive-DataEngineer%20Resource/citus-demo)

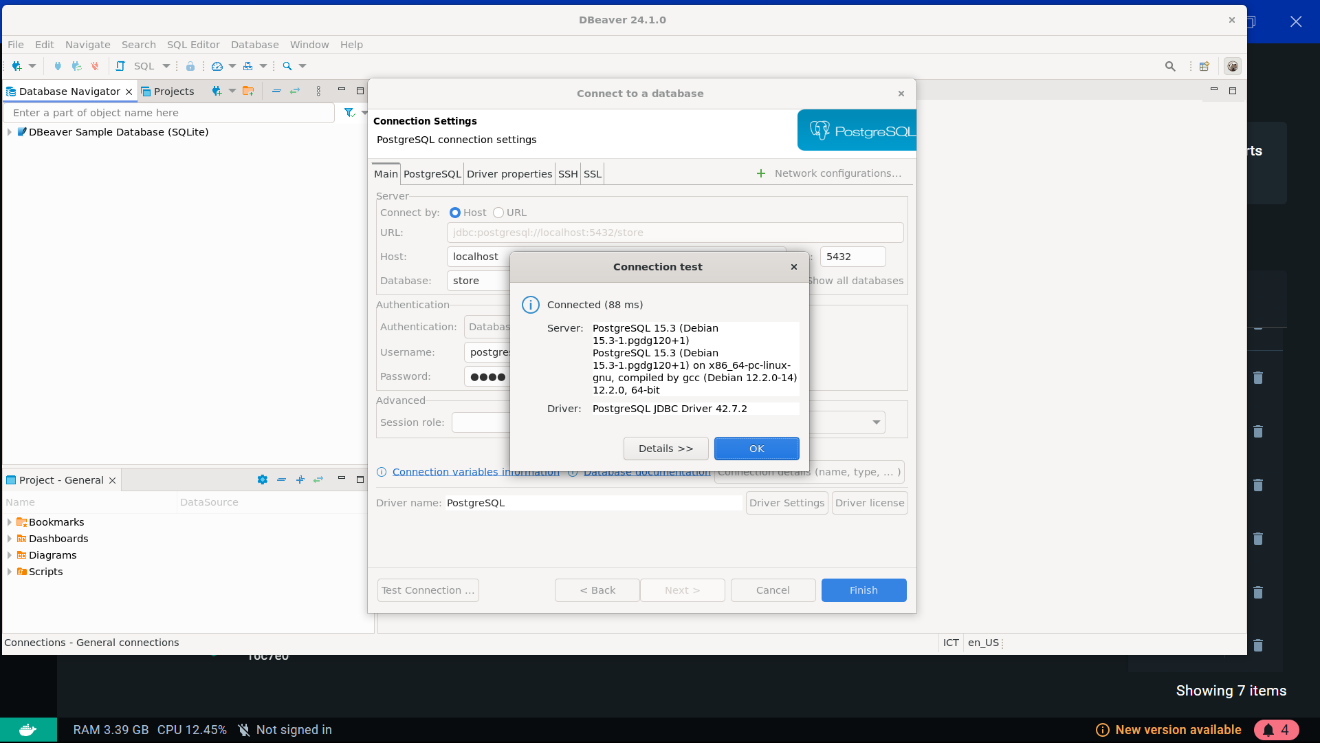
2). clone repository citusdata:

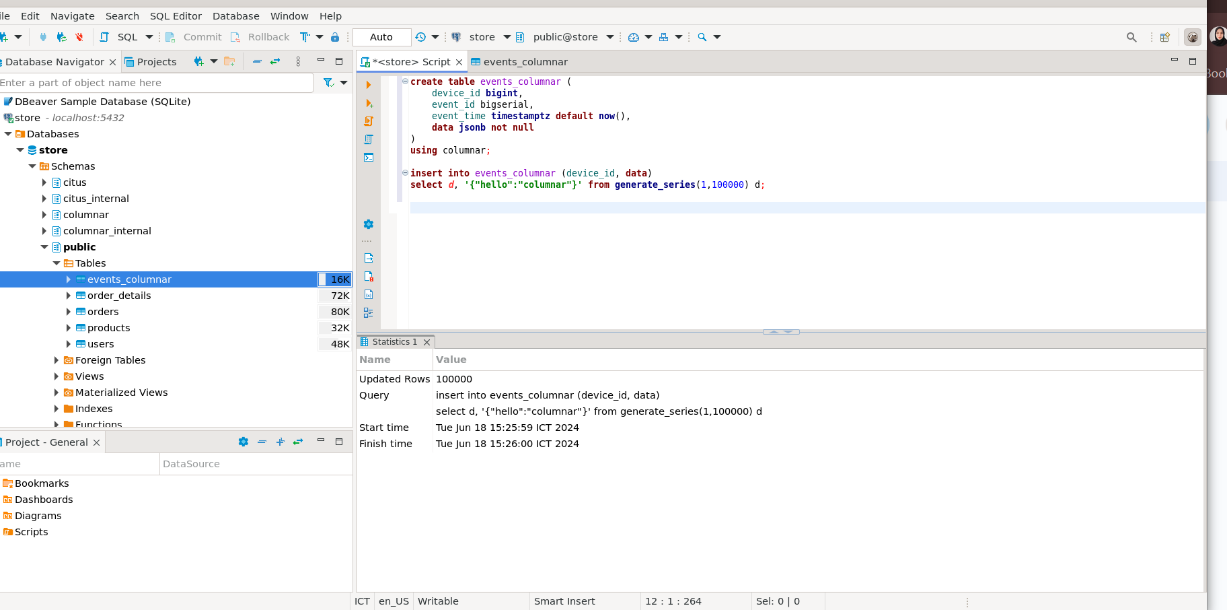


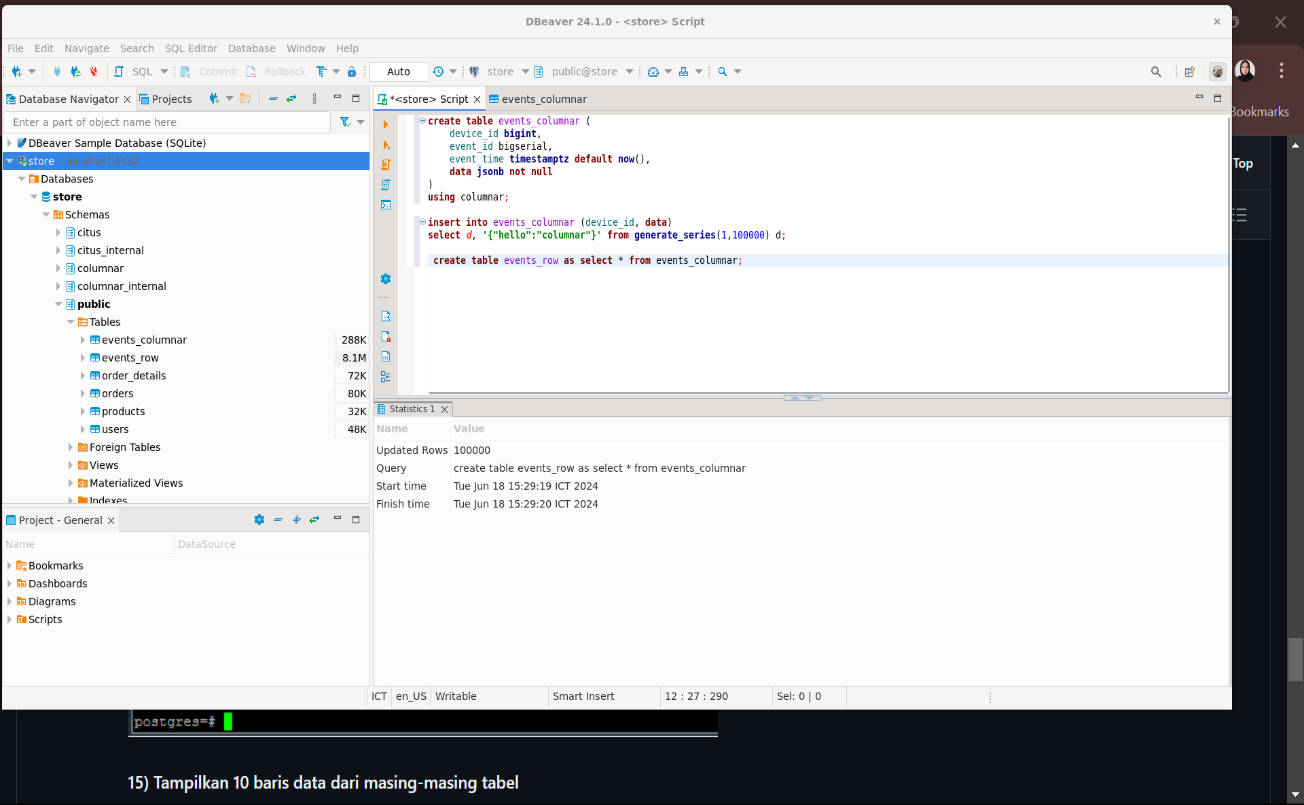
3). Jalankan docker compose:

buka file docker yang sudah diclone, kemudian buka terminal, ketik:

docker compose up -d

4). Connect dbeaver dengan database PostgreSQL yang ada di dalam container docker:

5). Buat tabel events\_columnar dengan menggunakan metode penyimpanan columnar

6). Buat tabel events\_row tanpa menggunakan metode penyimpanan columnar untuk perbandingan.

5. Perbedaan access method heap dan columnar:

1). Heap Storage (Row-Oriented)

- Data disimpan per baris.

- Keuntungan: Cepat untuk operasi yang mengakses seluruh atribut baris.

- Kekurangan: Lambat untuk query agregat dan analitik.

2). Columnar Storage (Column-Oriented)

- Data disimpan per kolom.

- Keuntungan: Cepat untuk query agregat dan analitik; kompresi data lebih baik.

- Kekurangan: Lambat untuk akses seluruh atribut baris; penulisan data lebih lambat.

Penggunaan:

- Heap Storage: Cocok untuk OLTP (transaksi online).

- Columnar Storage: Cocok untuk OLAP (analitik online).