

Nama : Kharisma Novi Chandramukti

Kelas/Batch : Data Engineering 4

Part 1 – Introduction to Data Warehouse

1. Sebutkan perbedaan antara data warehouse dan data lake

JAWAB :

Perbedaan antara data warehouse dan data lake :

- Data warehouse :
 - Karakteristik : Metode akses heap adalah metode akses default pada Citus, yang menyimpan data dalam bentuk tabel yang tidak terstruktur
 - Keuntungan : Lebih cocok untuk operasi penulisan dan penghapusan data, serta lebih mudah untuk ekstensi
 - Kerugian : Lebih lambat untuk operasi baca karena data harus diindeks ulang setiap kali ada perubahan
 - Penggunaan : Cocok untuk aplikasi yang memiliki lebih banyak operasi penulisan dan penghapusan data
- Data lake :
 - Karakteristik : Metode akses heap adalah metode akses default pada Citus, yang menyimpan data dalam bentuk tabel yang tidak terstruktur
 - Keuntungan : Lebih cocok untuk operasi penulisan dan penghapusan data, serta lebih mudah untuk ekstensi
 - Kerugian : Lebih lambat untuk operasi baca karena data harus diindeks ulang setiap kali ada perubahan
 - Penggunaan : Cocok untuk aplikasi yang memiliki lebih banyak operasi penulisan dan penghapusan data

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

JAWAB :

1. Tujuan OLAP dan OLTP sebagai berikut :

- OLTP (Online Transaction Processing) : Dirancang untuk mendukung dan mengelola transaksi harian yang terjadi dalam bisnis, seperti operasi penjualan, pembelian, dan interaksi pelanggan lainnya. Fokus utamanya adalah pada ketersediaan data real-time, pengolahan transaksi dengan cepat, dan pengelolaan volume data transaksional yang tinggi
- OLAP (Online Analytical Processing) : Dirancang untuk analisis data dan pelaporan yang mendalam. Biasanya digunakan untuk menganalisis data historis dalam skala besar, melakukan operasi agregasi kompleks, dan memfasilitasi pemahaman pola data untuk keperluan bisnis strategis. OLAP tidak memerlukan respons real-time seperti yang dibutuhkan dalam OLTP

2. Berdasarkan struktur datanya :

- OLTP : Menangani transaksi individual dalam jumlah besar yang terjadi dalam waktu singkat. Basis data OLTP biasanya diatur dalam bentuk normalisasi agar transaksi dapat dijalankan dengan efisien dan konsisten
- OLAP : Terfokus pada analisis data dalam jumlah besar dan kompleksitas tinggi. Data dalam data warehouse (yang digunakan untuk OLAP) sering kali disimpan dalam bentuk denormalisasi atau semi-denormalisasi untuk memudahkan akses dan kinerja analisis

3. Akses Data:

- OLTP : Memungkinkan akses langsung ke data transaksional individual. Query dalam OLTP biasanya singkat dan sederhana, ditujukan untuk mengambil atau memperbarui data spesifik
- OLAP : Memungkinkan akses yang kompleks dan analisis yang mendalam terhadap data agregat dan historis. Query dalam OLAP sering kali rumit, melibatkan operasi agregasi seperti sum, average, atau pengelompokan data berdasarkan dimensi tertentu

4. Pemodelan Data :

- OLTP : Pemodelan data terfokus pada entitas bisnis seperti pelanggan, produk, dan pesanan, dengan penekanan pada konsistensi dan integritas data
- OLAP : Pemodelan data terfokus pada fakta (misalnya, penjualan bulanan) dan dimensi (seperti waktu, lokasi, produk), dengan penekanan pada analisis dan pemahaman pola data

5. Optimalisasi Kinerja :

- OLTP : Didesain untuk kinerja tinggi dalam transaksi data yang singkat dan langsung. Biasanya menggunakan indeks untuk mempercepat akses data
- OLAP : Didesain untuk kinerja tinggi dalam analisis data yang kompleks, sering kali melibatkan operasi agregasi dan pencarian data dalam skala besar. Pengoptimalan termasuk pengindeksan yang sesuai dan strategi penyimpanan yang mendukung akses data yang efisien

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

JAWAB :

- a. Platform Data Warehouse tradisional (Oracle Exadata, IBM Db2 Warehouse)
- b. Platform Data Warehouse berbasis Cloud (Amazon Redshift, Google BigQuery, Snowflake, Microsoft Azure Synapse Analytics)
- c. Platform Open Source (Apache Hive, Apache Impala)
- d. ETL (Extract, Transform, Load) Tools (Informatica PowerCenter, Talend, Apache Nifi)
- e. Data Integration and Processing Tools (Apache Spark, Apache Kafka)
- f. Business Intelligence (BI) Tools (Tableau, Power BI, Qlik Sense)
- g. Data Modeling Tools (Erwin Data Modeler, IBM InfoSphere Data Architect)
- h. Metadata Management (Apache Atlas, Informatica Metadata Manager)

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

JAWAB :

- Git clone repository dari immersive-data, bertujuan untuk mengkloning repository yang ada ke repository milik kita

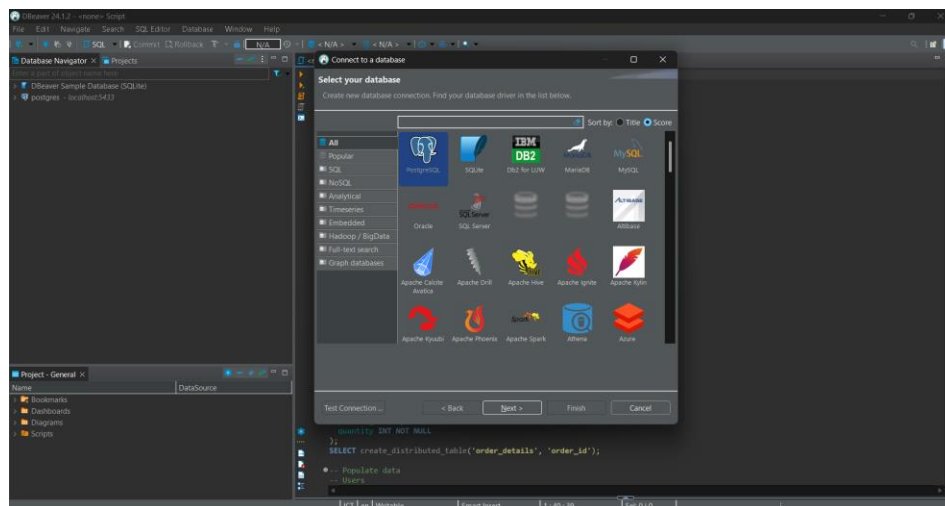
```
rismanovic@LAPTOP-U6LIFQG1:~/unit2$ git clone https://github.com/Immersive-DataEngineer-Resource/citus-demo.git
Cloning into 'citus-demo'...
remote: Enumerating objects: 12, done.
remote: Counting objects: 100% (12/12), done.
remote: Compressing objects: 100% (10/10), done.
remote: Total 12 (delta 0), reused 12 (delta 0), pack-reused 0
Receiving objects: 100% (12/12), 5.02 KiB | 2.51 MiB/s, done.
rismanovic@LAPTOP-U6LIFQG1:~/unit2$ code .
rismanovic@LAPTOP-U6LIFQG1:~/unit2$ |
```

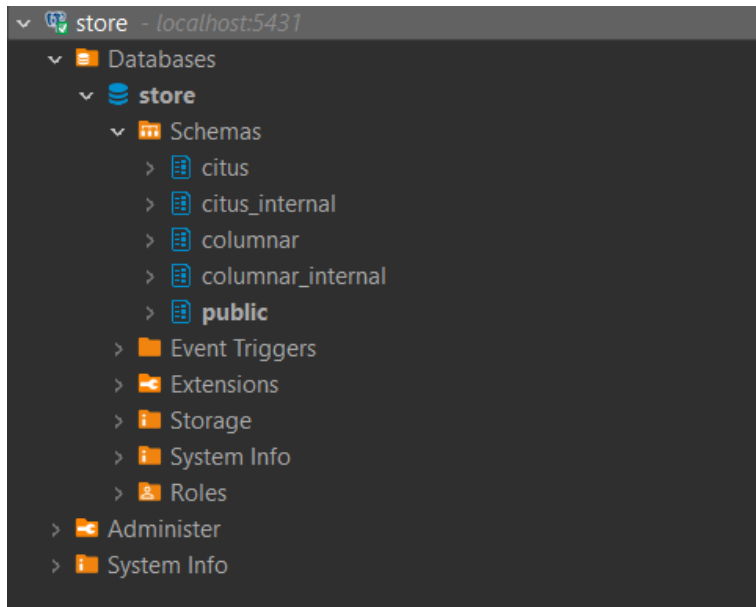
- Melakukan docker compose up -d untuk menjalankan container docker

```
rismanovic@LAPTOP-U6LIFQG1:~/unit2/citus-demo$ docker compose up -d
WARN[0000] /home/rismanovic/unit2/citus-demo/docker-compose.yml: 'version' is obsolete
[+] Running 31/5
  ✓ worker-2 Pulled                                231.1s
  ✓ manager Pulled                                146.6s
  ✓ worker-3 Pulled                                231.1s
  ✓ master Pulled                                 231.1s
  ✓ worker-1 Pulled                                231.1s
[+] Running 6/6
  ✓ Network citus-demo_postgres-network Created      0.1s
  ✓ Container citus-demo_master Started              1.1s
  ✓ Container citus-demo_manager Started             1.2s
  ✓ Container citus-demo_worker_3 Started            2.0s
  ✓ Container citus-demo_worker_2 Started            1.9s
  ✓ Container citus-demo_worker_1 Started            1.9s
rismanovic@LAPTOP-U6LIFQG1:~/unit2/citus-demo$
```

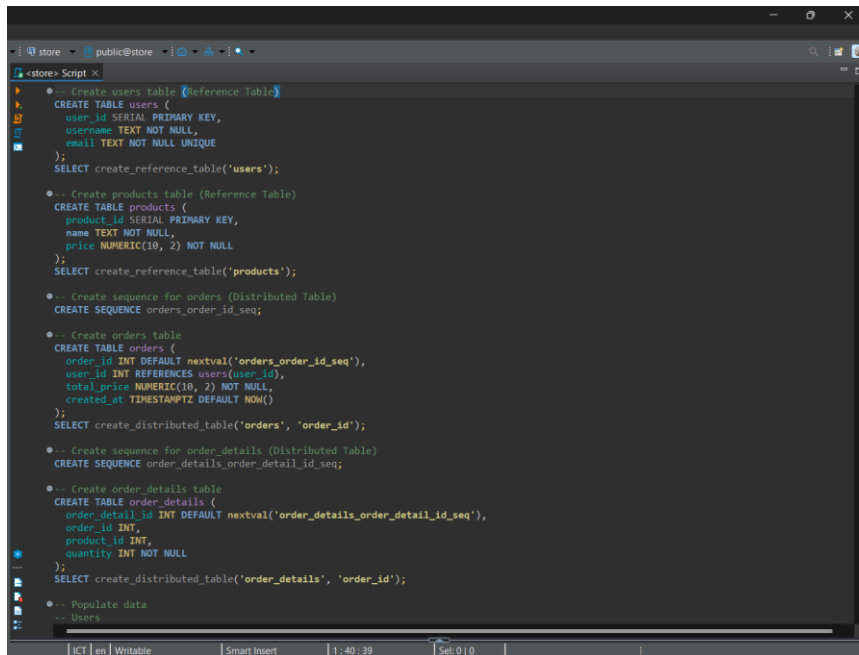
<input type="checkbox"/>		citus-demo	Running (5/5)	6.95%	2 minutes ago			
<input type="checkbox"/>		citus-demo_manager	citusdata/membership-manager:0.3.0 Running	0%	2 minutes ago			
<input type="checkbox"/>		citus-demo_worker_2	citusdata/citus:12.0.0 Running 5434:5432	0.09%	2 minutes ago			
<input type="checkbox"/>		citus-demo_worker_3	citusdata/citus:12.0.0 Running 5435:5432	0.11%	2 minutes ago			
<input type="checkbox"/>		citus-demo_worker_1	citusdata/citus:12.0.0 Running 5433:5432	0.12%	2 minutes ago			
<input type="checkbox"/>		citus-demo_master	citusdata/citus:12.0.0 Running 5431:5432	6.63%	2 minutes ago			

- Membuat connection baru dengan port 5431 dengan PostgreSQL dengan database name yaitu store

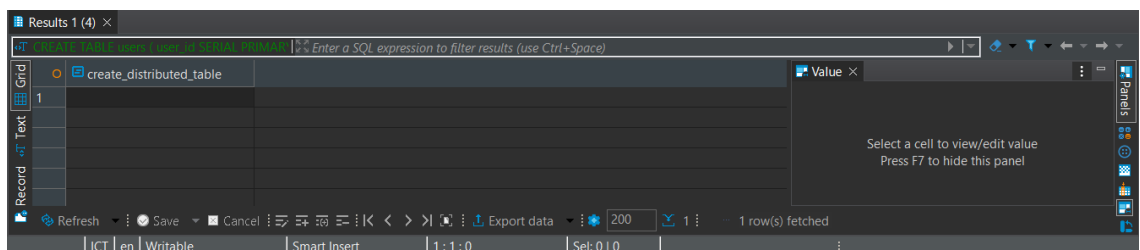




- Menyalin code dari repository citus-demo yaitu populate.sql ke dbeaver



- Membuat tabel



Results 1 x

Enter a SQL expression to filter results (use Ctrl+Space)

	order_id	shard_id	real_table_name	node_name
1	1	102,013	orders_102013	citrus-demo_worker_2
2	5	102,018	orders_102018	citrus-demo_worker_1
3	4	102,020	orders_102020	citrus-demo_worker_3
4	3	102,027	orders_102027	citrus-demo_worker_1
5	2	102,036	orders_102036	citrus-demo_worker_1

Value x

1

Refresh Save Cancel Export data 200 5 5 row(s) fetched - 0.009s, on 2024-07-19 at 20:35:44

ICT en Writable Smart Insert 1:1:0 Sel: 0 | 0

Results 1 x

Enter a SQL expression to filter results (use Ctrl+Space)

	real_table_name
1	orders_102013
2	orders_102036
3	orders_102027
4	orders_102020
5	orders_102018

Value x

orders_102013

Refresh Save Cancel Export data 200 5 5 row(s) fetched - 0.021s, on 2024-07-19 at 20:35:03

ICT en Writable Smart Insert 9:1:200 Sel: 0 | 0

- Hasil

DBViewer 24.12 - users

Database Navigator

store public@store

users

	user_id	username	email
1	1	JohnDoe	john.doe@example.com
2	2	JaneSmith	jane.smith@example.com

Value x

1

Refresh Save Cancel Export data 200 2 2 row(s) fetched - 0.016s, on 2024-07-20 at 20:14:38

ICT en

DBViewer 24.12 - products

Database Navigator

store public@store

products

	product_id	name	price
1	1	Laptop	1000
2	2	Phone	500
3	3	Headphones	200
4	4	Monitor	300

Value x

1

Refresh Save Cancel Export data 200 4 4 row(s) fetched - 0.020s, on 2024-07-20 at 20:15:01

ICT en

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

JAWAB :

a. Heap (Access Method Heap)

- a. Karakteristik : Metode akses heap adalah metode akses default pada Citus, yang menyimpan data dalam bentuk tabel yang tidak terstruktur
- b. Keuntungan : Lebih cocok untuk operasi penulisan dan penghapusan data, serta lebih mudah untuk ekstensi
- c. Kerugian : Lebih lambat untuk operasi baca karena data harus diindeks ulang setiap kali ada perubahan
- d. Penggunaan : Cocok untuk aplikasi yang memiliki lebih banyak operasi penulisan dan penghapusan data

b. Columnar (Access Method Columnar)

- e. Karakteristik : Metode akses heap adalah metode akses default pada Citus, yang menyimpan data dalam bentuk tabel yang tidak terstruktur.
- f. Keuntungan : Lebih cocok untuk operasi penulisan dan penghapusan data, serta lebih mudah untuk ekstensi.
- g. Kerugian : Lebih lambat untuk operasi baca karena data harus diindeks ulang setiap kali ada perubahan.
- h. Penggunaan : Cocok untuk aplikasi yang memiliki lebih banyak operasi penulisan dan penghapusan data