**PART 3 - Replication + Sharding**

1. Jelaskan perbedaan antara replication dan sharding

Replication

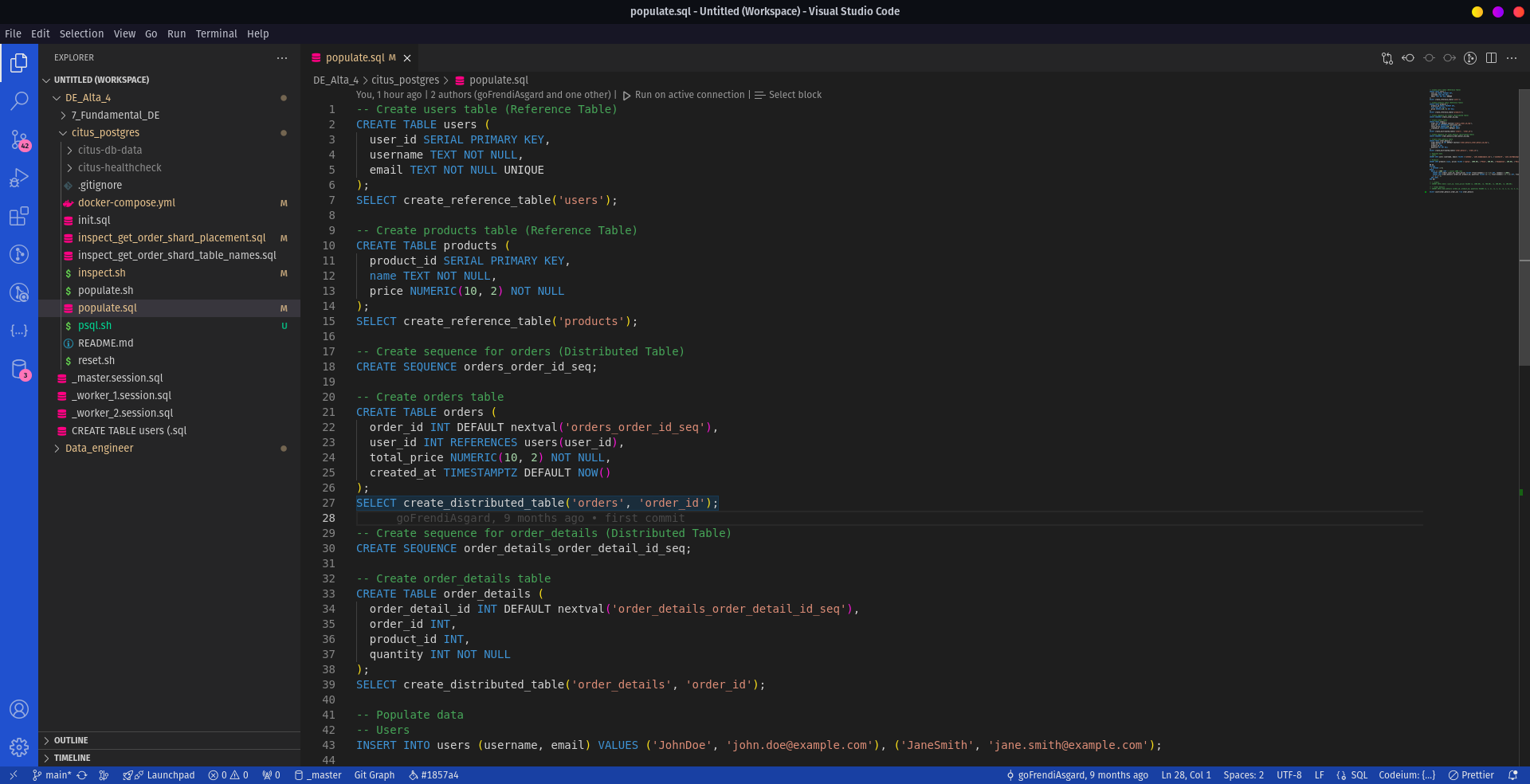
Replication dibuat untuk meningkatkan ketersediaan data dan ketahanan terhadap kegagalan node. Replikasi data yang sama disalin ke beberapa node. Setiap replika menyimpan salinan lengkap dari data. Replika dapat digunakan untuk backup data.

Sharding

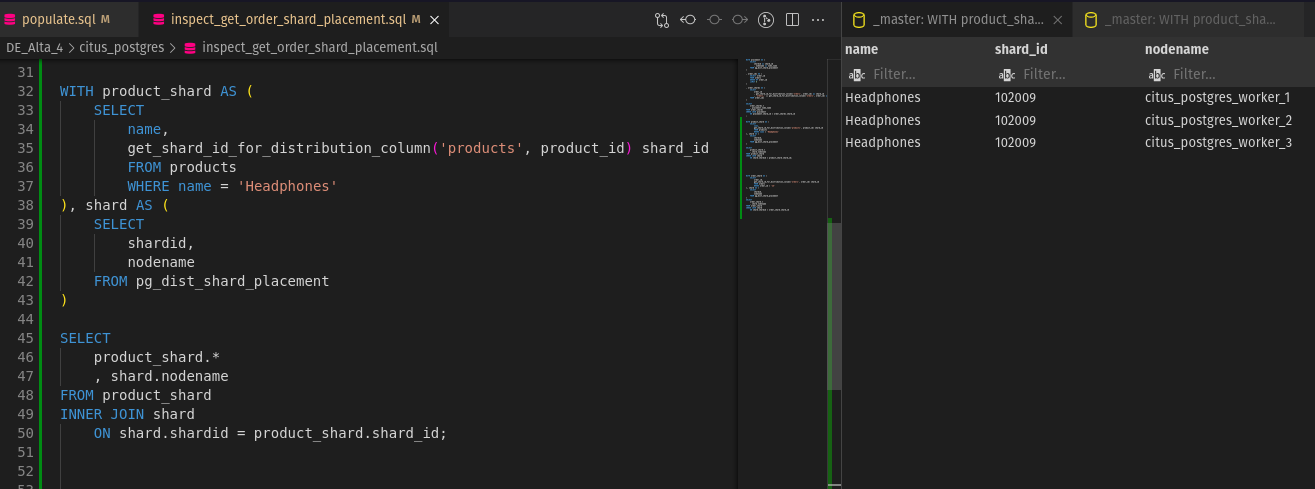
Membagi data besar ke beberapa node untuk meningkatkan kapasitas penyimpanan dan kinerja. Data dibagi menjadi beberapa subset yang disebut shard. Setiap shard disimpan di node yang berbeda. Data dipecah berdasarkan kriteria tertentu (misalnya, range, hash, list). Setiap shard hanya menyimpan sebagian data. Menangani data dalam jumlah besar dengan menambahkan lebih banyak node dan membagi beban penyimpanan dan pemrosesan.

2. Lakukan percobaan untuk membuat reference table + distributed table seperti pada

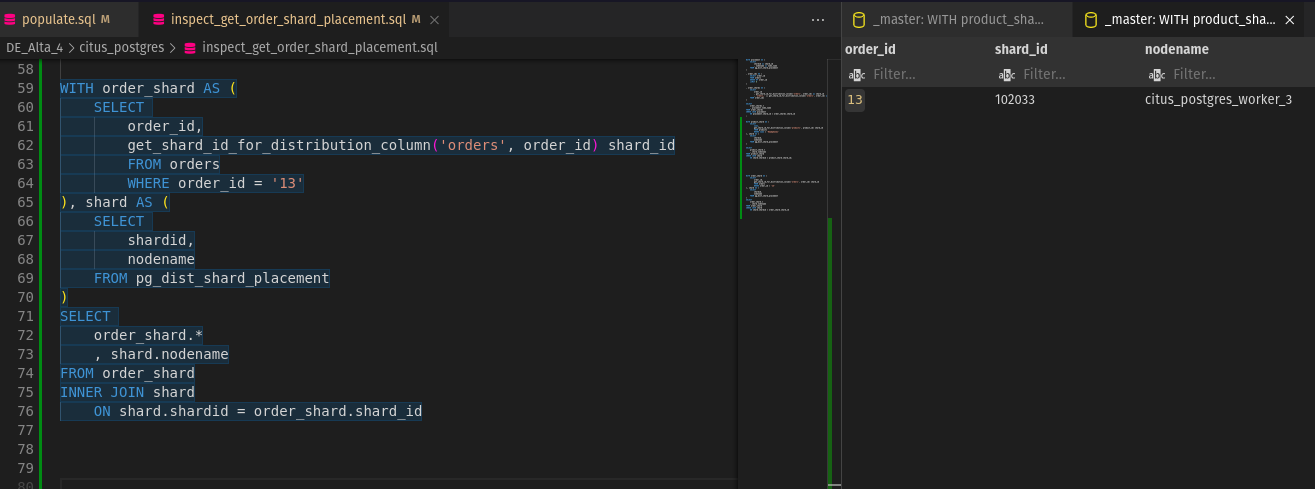
repo<https://github.com/Immersive-DataEngineer-Resource/citus-demo> .



3. Di node/worker mana saja product "Headphone" tersimpan? Tunjukkan shard id nya



4. Di node/worker mana saja order dengan id 13 tersimpan? Tunjukkan shard id nya



5. Kapan sebaiknya kita menggunakan replication? . Kapan sebaiknya kita menggunakan sharding?

1. Replikasi digunakan ketika membutuhkan salinan data tersedia di beberapa lokasi, dan performa pembacaan tinggi.
2. Sharding digunakan ketika membutuhkan skalabilitas horisontal, kinerja penulisan tinggi, dan pengelolaan dataset yang sangat besar.