

Nama: Qorina M. H. Mumtaza

Kelas: DE-1

1. Jelaskan perbedaan antara replication dan sharding!

Jawab:

Replication dan sharding adalah dua Teknik yang digunakan dalam manajemen database untuk mengalamatkan skalabilitas, ketersediaan, dan distribusi data dengan karakteristik berbeda sebagai berikut:

- Replication berfokus pada peningkatan ketersediaan data dan kinerja baca, sementara sharding diarahkan untuk meningkatkan skalabilitas tulis dan mengelola kumpulan data besar.
- Replication menyimpan salinan data yang sama di beberapa server, artinya semua server memiliki data yang sama. Sementara sharding mendistribusikan subset data yang berbeda ke server yang berbeda, yang memungkinkan setiap server berisi subset data berbeda berdasarkan kunci shard.
- Replication lebih cocok untuk scenario di mana konsistensi data sangat penting (misalnya dalam konfigurasi master-slave), sementara sharding mungkin memerlukan pertimbangan tambahan untuk menjaga konsistensi data di seluruh shard.

2. Lakukan percobaan untuk membuat reference table + distributed table seperti pada repo <https://github.com/Immersive-DataEngineer-Resource/citus-demo>

```
nodeid | groupid | nodename | nodeport | noderack | hasmetadata | isactive | noderole | nodecluster | metadatasynced | shouldhaveshards
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----
1 | 0 | localhost | 5432 | default | t | t | primary | default | t | t
(1 row)

Order shard placement:
order_id | shard_id | real_table_name | node_name
-----+-----+-----+-----
1 | 102012 | orders_102012 | localhost
5 | 102017 | orders_102017 | localhost
4 | 102019 | orders_102019 | localhost
3 | 102026 | orders_102026 | localhost
2 | 102035 | orders_102035 | localhost
(5 rows)

Order shard table names:
orders_102012
orders_102035
orders_102026
orders_102019
orders_102017

SELECT count(1) FROM orders
count
-----
1000
(1 row)
```

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

select * from products where name = 'Headphones';

Results 1 X

select * from f | Enter a SQL expression to filter results (use Ctrl+Space)

	123 proc	ABC nam	123 price	Value X
1	3	Headphone	200	3

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

select * from orders where order_id = '13';

Results 1 X

select * from orders where order_id = '1': Enter a SQL expression to filter results (use Ctrl+Space)

Grid	123 order_id	123 user_id	123 total_price	created_at	Value X
1	13	1	467.16	2023-11-07 17:57:57.548 +0700	13

5. Kapan sebaiknya kita menggunakan replication?

Secara keseluruhan, replication berfungsi dalam situasi diperlukannya ketersediaan data, meningkatkan kinerja baca, dan meningkatkan toleransi kesalahan sistem. Namun perlu diingat bahwa replication bukanlah solusi yang universal dan harus dikonfigurasi serta dikelola dengan cermat berdasarkan kasus penggunaan yang spesifik. Pilihan jenis replication (misalnya master-slave, master-master) dan jumlah replication bergantung pada requirements aplikasi dan keseimbangan antara konsistensi data dan performa.

6. Kapan sebaiknya kita menggunakan sharding?

Apabila data yang diolah memiliki ukuran besar atau data sejenis transaksi, maka lebih baik memakai metode ini. Hal ini dikarenakan sharding melakukan pengelompokan pemecahannya berdasarkan metode yg sudah dilakukan (hashing).