

Nama : Karno

Jurusan : Data Engineer 01

1. Perbedaan replication dan sharding :

- Replication adalah strategi dimana data disalin atau didistribusikan ke beberapa lokasi/node atau worker untuk tujuan backup data. Ini berguna apabila salah satu worker mati maka ada worker lain yang melakukan backup. Pada replication masing-masing node/worker memiliki data yang sama dengan node/worker lainnya.
- Sharding adalah strategi dimana data dibagi menjadi beberapa bagian dan dimasukkan kedalam node/worker yang berbeda. Ini bertujuan untuk mempercepat pembacaan data. Pada sharding data yang ada di node/worker yang satu berbeda dengan yang ada di node/worker lain.

2. Melakukan percobaan untuk membuat referensi table

```
EXPLORER
  CITUS-DEMO
    citus-db-data
    citus-healthcheck
    .gitignore
    docker-compose.yml
    init.sql
    inspect_get_order_shard_placement.sql
    inspect_get_order_shard_table_names.sql
    inspect.sh
    populate.sh
    populate.sql
    README.md
    reset.sh

$ inspect.sh

SQL="SELECT count(1) FROM orders"
echo "$SQL"
docker exec -it "$WORKER_CONTAINER" psql -U "$USER_NAME" -d "$DB_NAME" -c "$SQL";

for SHARD_TABLE_NAME in $SHARD_TABLE_NAMES
do
  SQL="select * from $SHARD_TABLE_NAME order by order_id limit 5"
  echo "$SQL"
  docker exec -it "$WORKER_CONTAINER" psql -U "$USER_NAME" -d "$DB_NAME" -c "$SQL"
done

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

$ ./inspect.sh

=====
MASTER
=====
Active worker count:
3

Get nodes:
nodeid | groupid | nodename | nodeport | noderack | hasmetadata | isactive | noderole | nodecluster | m
-----+-----+-----+-----+-----+-----+-----+-----+-----+---
1 | t | citus-demo_worker_3 | 5432 | default | t | t | primary | default | t
2 | t | citus-demo_worker_1 | 5432 | default | t | t | primary | default | t
3 | t | citus-demo_worker_2 | 5432 | default | t | t | primary | default | t
(3 rows)

select * from orders_102016 order by order_id limit 5
order_id | user_id | total_price | created_at
-----+-----+-----+-----
5 | 2 | 509.27 | 2023-11-04 12:39:42.770383+00
53 | 2 | 158.35 | 2023-11-04 12:39:42.770383+00
112 | 1 | 514.24 | 2023-11-04 12:39:42.770383+00
163 | 2 | 985.70 | 2023-11-04 12:39:42.770383+00
208 | 1 | 72.44 | 2023-11-04 12:39:42.770383+00
(5 rows)

=====
WORKER citus-demo_worker_2
=====
SELECT count(1) FROM orders
count
-----
2000
(1 row)

select * from orders_102011 order by order_id limit 5
order_id | user_id | total_price | created_at
-----+-----+-----+-----
1 | 2 | 43.53 | 2023-11-04 12:39:42.770383+00
15 | 1 | 134.41 | 2023-11-04 12:39:42.770383+00
```

- Di node/worker mana sajakah product headphone tersimpan, tunjukkan shard_id nya
Product headphone tersimpan di semua node dengan shard_id seperti digambar

orders 1 pg_dist_shard_placement(+) 2 X

WITH placement AS (SELECT shardid as Enter a SQL expression to filter results (use Ctrl+Space)

	node_name	shard_id	name	order_detail_id	order_id	product_id
1	citrus-demo_worker_1	102,010	Headphones	8	8	3
2	citrus-demo_worker_1	102,010	Headphones	686	686	3
3	citrus-demo_worker_1	102,010	Headphones	903	903	3
4	citrus-demo_worker_2	102,011	Headphones	25	25	3
5	citrus-demo_worker_2	102,011	Headphones	220	220	3
6	citrus-demo_worker_2	102,011	Headphones	390	390	3
7	citrus-demo_worker_2	102,011	Headphones	405	405	3
8	citrus-demo_worker_2	102,011	Headphones	640	640	3
9	citrus-demo_worker_2	102,011	Headphones	670	670	3
10	citrus-demo_worker_2	102,011	Headphones	675	675	3
11	citrus-demo_worker_2	102,011	Headphones	984	984	3
12	citrus-demo_worker_2	102,014	Headphones	223	223	3
13	citrus-demo_worker_2	102,014	Headphones	339	339	3
14	citrus-demo_worker_2	102,014	Headphones	520	520	3
15	citrus-demo_worker_2	102,014	Headphones	931	931	3
16	citrus-demo_worker_2	102,014	Headphones	987	987	3
17	citrus-demo_worker_1	102,013	Headphones	94	94	3
18	citrus-demo_worker_1	102,013	Headphones	258	258	3
19	citrus-demo_worker_1	102,013	Headphones	384	384	3
20	citrus-demo_worker_1	102,013	Headphones	444	444	3
21	citrus-demo_worker_1	102,013	Headphones	493	493	3
22	citrus-demo_worker_3	102,015	Headphones	109	109	3
23	citrus-demo_worker_3	102,015	Headphones	466	466	3

- Di node/worker mana saja order dengan id 13 tersimpan
Order dengan id 13 tersimpan di node “citrus-demo-worker_3” dengan shard_id 102,033 seperti pada gambar

WITH placement AS (SELECT shardid as Enter a SQL expression to filter results (use Ctrl+Space)

	order_id	shard_id	real_table_name	node_name
8	8	102,010	orders_102010	citrus-demo_worker_1
9	9	102,038	orders_102038	citrus-demo_worker_2
10	10	102,014	orders_102014	citrus-demo_worker_2
11	11	102,040	orders_102040	citrus-demo_worker_1
12	12	102,040	orders_102040	citrus-demo_worker_1
13	13	102,033	orders_102033	citrus-demo_worker_3
14	14	102,018	orders_102018	citrus-demo_worker_3
15	15	102,011	orders_102011	citrus-demo_worker_2
16	16	102,020	orders_102020	citrus-demo_worker_2
17	17	102,019	orders_102019	citrus-demo_worker_1

- Replication sebaiknya digunakan Ketika kita menjalankan suatu system yang harus terus menyediakan data tanpa henti. Sehingga Ketika salah satu server atau system mengalami error maka server yang lain siap digunakan tanpa harus mengganggu data utama. Replication bertindak dengan cara menyediakan data cadangan supaya siap digunakan Ketika dibutuhkan.
- Sharding sebaiknya digunakan Ketika kita memiliki dataset yang sangat besar atau Ketika computer memiliki beban kerja yang tinggi. Sharding bekerja dengan cara membagi data ke beberapa server (shard) sehingga masing-masing shard dapat mengatasi Sebagian masalahnya.