

Họ và Tên: Nguyễn Phương Thiên Vũ

MSSV: 1250080237

Lớp: 12_ĐH_CNTT5

BÀI TẬP LÝ THUYẾT TUẦN 3

Tìm hiểu, trình bày về một trong các hệ quản trị cơ sở dữ liệu NoSQL ScyllaDB

1. Tên hệ quản trị CSDL

ScyllaDB

2. Lịch sử hình thành

ScyllaDB được thành lập vào năm 2015 bởi công ty ScyllaDB, do Avi Kivity và Dor Laor sáng lập.

Mục tiêu của nhóm phát triển là xây dựng một hệ quản trị cơ sở dữ liệu:

- Tương thích với Apache Cassandra
- Nhưng được viết hoàn toàn bằng C++ thay vì Java

Việc sử dụng C++ giúp ScyllaDB:

- Khai thác tối đa CPU đa nhân
- Giảm độ trễ
- Tránh các vấn đề về Garbage Collection của Java
- Đạt hiệu năng cao hơn Cassandra trong môi trường hệ thống phân tán

ScyllaDB vì vậy được xem là một giải pháp NoSQL hiện đại, hiệu năng cao, thay thế Cassandra trong các hệ thống xử lý dữ liệu lớn.

3. Tác giả, Tổ chức quản lý

ScyllaDB được phát triển bởi công ty **ScyllaDB Inc.**

ScyllaDB là phần mềm mã nguồn mở

4. Mô hình lưu trữ dữ liệu

ScyllaDB sử dụng mô hình: Wide Column Store (CSDL cột rộng) / Column Families

Dữ liệu được lưu theo: Key → Column Family → Columns

Trong đó:

Key (Partition Key) là:

Mã dùng để xác định dữ liệu nằm ở node nào trong hệ phân tán

Column Family = Table trong SQL

Columns là các thuộc tính của dữ liệu

5. Ngôn ngữ thao tác dữ liệu

ScyllaDB dùng: CQL – Cassandra Query Language

Cú pháp giống SQL ở các lệnh cơ bản như Create table, Insert, select

```
CREATE TABLE users (
```

```
id INT PRIMARY KEY,
```

```
name TEXT,
```

```
email TEXT
```

```
);
```

```
INSERT INTO users (id, name, email)
```

```
VALUES (1, 'A', 'a@gmail.com');
```

```
SELECT * FROM users WHERE id = 1;
```

6. Cơ chế phân tán dữ liệu của ScyllaDB

ScyllaDB là CSDL phân tán ngang (Peer-to-peer)

Không có server trung tâm, mỗi máy trong hệ thống gọi là Node.

Sử dụng kiến trúc shard-per-core:

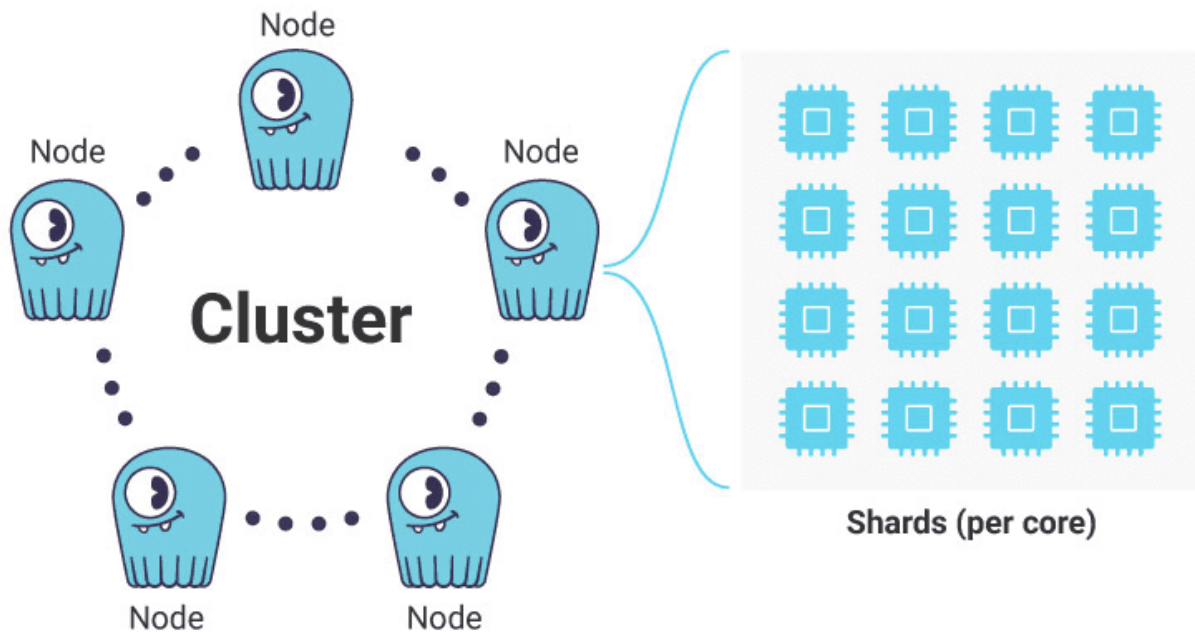
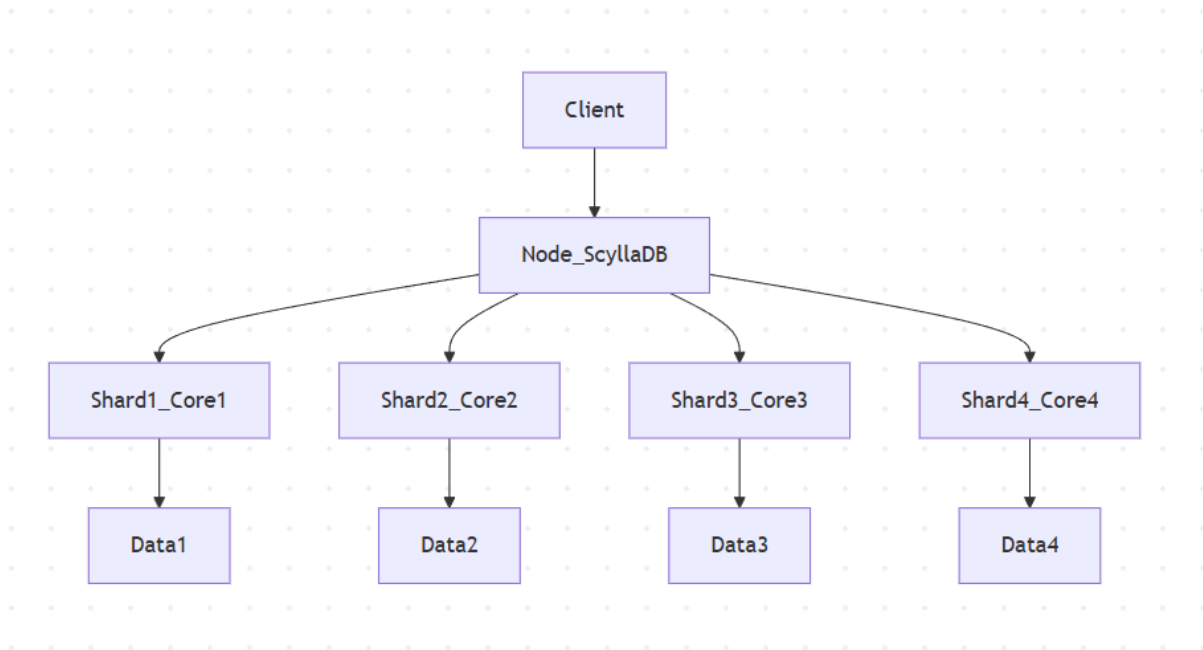
Mỗi lõi CPU (core) = một shard (một đơn vị xử lý dữ liệu độc lập)

Giả sử máy chủ có: 8 CPU cores

ScyllaDB sẽ tạo: 8 shards

Mỗi shard:

- Chạy trên 1 core riêng
- Có bộ nhớ riêng
- Có dữ liệu riêng
- Không chia sẻ tài nguyên với shard khác



7. Mô phỏng truy vấn phân tán

Ví dụ có 3 node:

Node A	Node B	Node C
--------	--------	--------

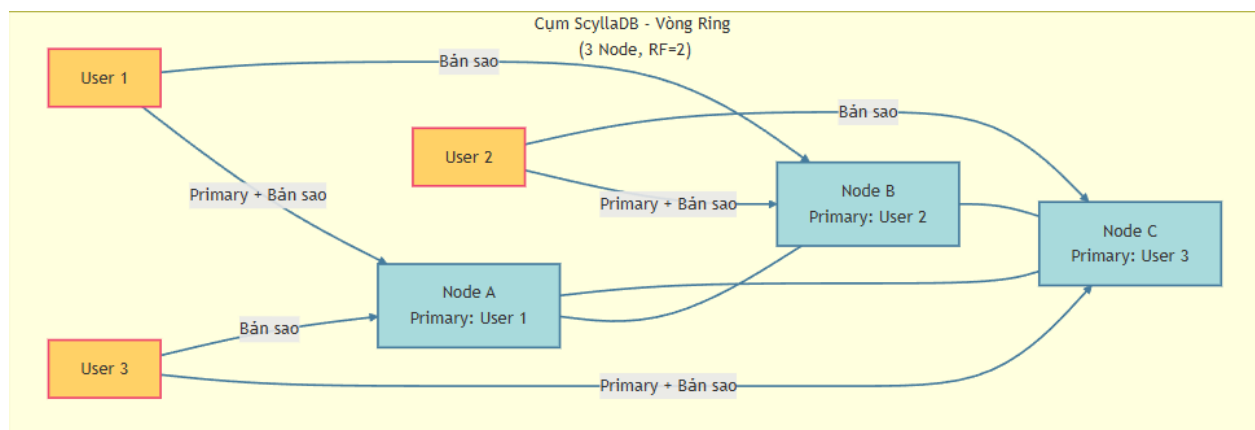
Dữ liệu:

User 1	Node A
User 2	Node B
User 3	Node C

Thì ScyllaDB sao chép dữ liệu:

User 1	Node A + Node B
User 2	Node B + Node C
User 3	Node C + Node A

Một node hỏng thì node khác vẫn có dữ liệu



Khi người dùng truy vấn dữ liệu:

SELECT * FROM users WHERE id = 2;

1. Biết rằng User 2 nằm ở Node B và Node C
2. Gửi truy vấn đến Node B (hoặc C)
3. Trả kết quả về cho người dùng

8. Vì sao ScyllaDB là NoSQL?

Vì ScyllaDB không tuân theo mô hình cơ sở dữ liệu quan hệ (Relational Database) như Oracle, SQL DataBase.

Trong ScyllaDB:

- Không có FOREIGN KEY

- Không có JOIN
- Mỗi bảng hoạt động độc lập