

## BT lý thuyết b'3

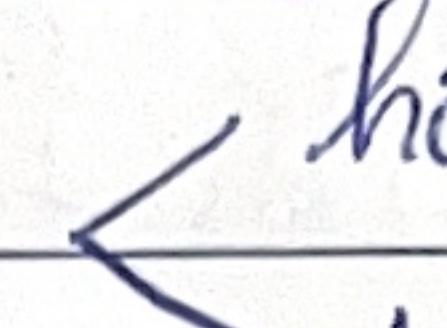
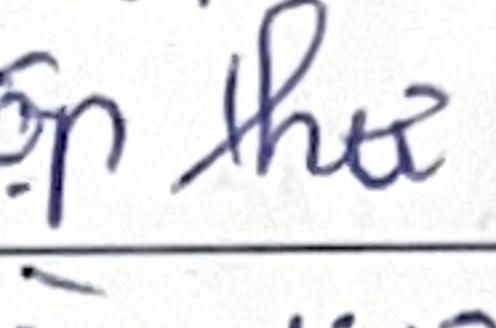
Nguyễn Thị Thùy Trang

12-DH-CNTT4

## ① Tên hệ quản trị CSDL

Apache Cassandra là 1 hệ quản trị CSDL NoSQL dạng Wide Column Store, dc thiết kế để xử lý dữ liệu lớn trong các hệ thống phân tán.

## ② Lls hình thành

Apache Cassandra dc pt ban đầu bởi Facebook năm 2008 nhằm phục vụ cho các tính năng như   tìm kiếm

đó, CSDL Cassandra dc công bố mã nguồn mở và chuyển sang qly bởi Apache Software Foundation. Qua n  
năm pt, ngày càng dc hoàn thiện và hiện đang dc  
sd rộng rãi (các hệ thống lớn, TG như Netflix ...)

trên

Instagram

Apple

Uber

## ③ Tác giả, tổ chức qly

• Tác giả ban đầu: Facebook

• Tổ chức qly htn: Apache Software Foundation

• Cassandra là dự án mã nguồn mở, dc cộng ~~đồng~~ toà<sup>đồng</sup>n  
củi đóng góp và pt.

## ④ Mô hình DL

Sử dụng mô hình lưu trữ dữ liệu dạng Wide Column

## keyspace

## Table

Partition Key : quyết định dữ liệu lưu trữ ở node nào

Clustering Key: dùng để sắp xếp dl () cùng 1 partition

Mô hình này giúp Cassandra tái sử dụng các truy vấn  
theo khóa và xử lý dữ liệu

## ⑤ Ngôn ngữ kinh doanh vs dl

Sđ cài (Cassandra Query language) để sao tac  
vs dl

Nd: SELECT \* FROM UserActivity WHERE user-id = 'u01';

Cần có hỗ trợ join phức tạp và các truy vấn phải dc  
hiệu kẽ phù hợp ngay từ đầu

⑥ Có chẽ phân tán dẽ

Apache Cassandra dc xây dựng theo ~~tree~~ Kiến trúc phân tán Peer - to - Peer, có node trung tâm

## Cách phân tán dl

- Cassandra sd Consistent Hashing để phân bô dl
- Mỗi node chịu trách nhiệm lưu trữ 1 p`c dữ liệu nhai định
- DL dc phân tán đều giữa các node

## Mô hình phân tán dữ liệu

- Khi thêm node ns, dl sẽ dc tự động cân - lai
- Ichong cần dùng f không khi mở rộng

## Mô phỏng dl lưu trữ dl phân tán

Giả sử hệ thống có 3 node

- Node A
- Node B
- Node C

## Cassandra sd Consistent Hashing để phân bô dl

Vd:

User_id	Hash Value	Node lưu trữ
U01	120	Node A
U02	350	Node B
U03	670	Node C

→ dữ liệu được phân tán đều, tránh quá tải 1 node

Ngoài ra, Cassandra hỗ trợ Replication (sao chép dữ liệu) nhằm đảm bảo:

- Tính sẵn sàng cao

- Hệ thống vẫn hoạt động khi 1 node gặp sự cố

### ⑦ Mô phỏng truy vấn phân tán

Khi người dùng gửi truy vấn, yêu cầu có thể được gửi đến bất kỳ node nào trong hệ thống. Node này sẽ đóng vai trò là Coordinator, xác định vị trí dữ liệu và gửi truy vấn đến các node chứa dữ liệu, sau đó tổng hợp kết quả và trả về.

I dùng

Vd truy vấn

SELECT \* FROM UserActivity WHERE user-id = 'u01';

Quá trình xử lý:

- Client gửi truy vấn đến bất kỳ node nào (Coordinator)

- Node Coordinator xác định partition key u01

- Gửi yêu cầu đến các node chứa dữ liệu (Node A, B)

- Thu nhận kết quả

- Trả kết quả về cho client

\* Client cần biết dữ liệu nằm đâu

Client



Coordinator (Node B)



Node A (chứa dữ liệu)



Coordinator



Client

- Tên CSDL : Online Learning DB

dùng để quản lý hoạt động học tập trực tuyến của sinh viên

- Các bảng mô (dl mô phỏng)

Bảng student

student_id	name	Class
SVO1	Nguyễn Văn A	CNTT2
SVO2	Trần Thị B	CNTT3
SVO3	Lê Văn C	CNTT4

Bảng CourseActivity

student_id	time	action	course
SVO1	08:00	login	CSDL
SVO2	08:10	view	CSDL
SVO3	09:00	login	Mạng HONG HA

- Mô phỏng lưu trữ phân tán

Giao thông có 3 node

Node

Dữ liệu lưu

Node A

SV01

Node B

SV02

Node C

SV03

→ Mỗi node lưu dl của 1 nhóm sv

- Mô phỏng truy vấn phân tán

yêu cầu: Lấy hoạt động của sinh viên SV01

1. I dùng gửi yêu cầu đến Node C Quy
2. Node C làm Coordinator
3. Node C xác định dl SV01 nằm ở node A trình
4. Node A gửi dl về Node C
5. Node C trả kp cho I dùng