

Bài tập lý thuyết tuần 5

Người Thành Viên CNTT, 1252080231

- Tên hệ cơ sở: Apache Cassandra

Thuộc nhóm wide Column Store, được liệt kê là phân loại trên website NOSQL Databases - From Basic to Enterprise Use

- Lịch sử hình thành

+ Apache Cassandra được phát triển lần đầu vào năm 2008 bởi Facebook

+ Mục đích ban đầu là phục vụ hệ thống Inbox Search với yêu cầu dữ liệu lớn, phân tán và chịu lỗi cao.

+ Năm 2009, Cassandra được mở nguồn để và chuyển giao cho Apache Software Foundation quản lý.

+ Hiện nay, Cassandra là 1 trong những hệ quản trị NOSQL phát triển phổ biến trong doanh nghiệp

- Tác giả, tổ chức quản lý

+ Tác giả ban đầu: Facebook

+ Tổ chức quản lý hiện tại: Apache Software Foundation

+ Giấy phép: Apache License

- Mô hình hệ trữ

Loại hình NOSQL: wide column store

Đặc điểm

→ Dữ liệu được lưu trữ theo hàng nhưng không giống RDBMS
+ Với bảng gồm:

Partition key, Clustering key

+ không hỗ trợ join

+ Schema linh hoạt, phù hợp dữ liệu lớn.

→ Ngôn ngữ thao tác dữ liệu

Cassandra sử dụng CQL (Cassandra Query Language)

Có phép join giới hạn truy vấn thông

→ Cơ chế phân tán dữ liệu

Kiểm tra phân tán:

+ Cassandra sử dụng mô hình Peer-to-Peer

+ không có node Master

+ Mỗi node có vai trò ngang nhau

Phân mảnh dữ liệu

+ Dữ liệu được phân tán theo Partition key

+ Áp dụng thuật toán Consistent Hashing

Sao chép dữ liệu

Mọi bản ghi có thể tồn tại trên nhiều node

Grup hệ thống:

+ Chịu lỗi tốt

+ Không giới hạn chỉ node gặp lỗi cơ

- Mô phỏng lưu trữ & truy vấn phân tán

Bối cảnh mô phỏng:

Xét hệ thống quản lý danh ký học phần của sinh viên đại học.

Dữ liệu sinh viên được lưu trữ và phân tán theo nhóm điểm bài học này mô phỏng và tính toán tổng cao của hệ thống.

- Thiết kế bảng dữ liệu trong Cassandra

Create table dangkyhopphan (

choa text,

masinhvien text,

masinhvien text,

hocky text,

primary key (choa, masinhvien)

);

- Mô phỏng lưu trữ dữ liệu phân tán

Ở đây hệ thống Cassandra gồm 3 node:

Node1: Dữ liệu sinh viên khoa CNTT

Node2: Dữ liệu sinh viên khoa Sinh học

Node3: Dữ liệu sinh viên khoa Sư phạm

- ~~SQL~~
- ~~SQL~~ theo tài ghi chú bộ phận
insert into danhlyhocpho (chon, masinhien, masinhoc,
hocng)

values ('CNTT', 'SV301', 'K+201', '2024-1');

- 5. ~~Điểm~~

+ thời gian mở rộng ngày

+ chủ hợp với hệ thống dữ liệu là truy cập liên tục

+ điểm bất biến sẵn sàng cho mạng mở rộng, phân tán

- ~~Nhược điểm~~

+ không hỗ trợ phép nối join

+ không phù hợp với các giao dịch phức tạp phân tán (3PC)
quan hệ.

- Truy vấn phân tán không có Partition key

select * from danhlyhocpho

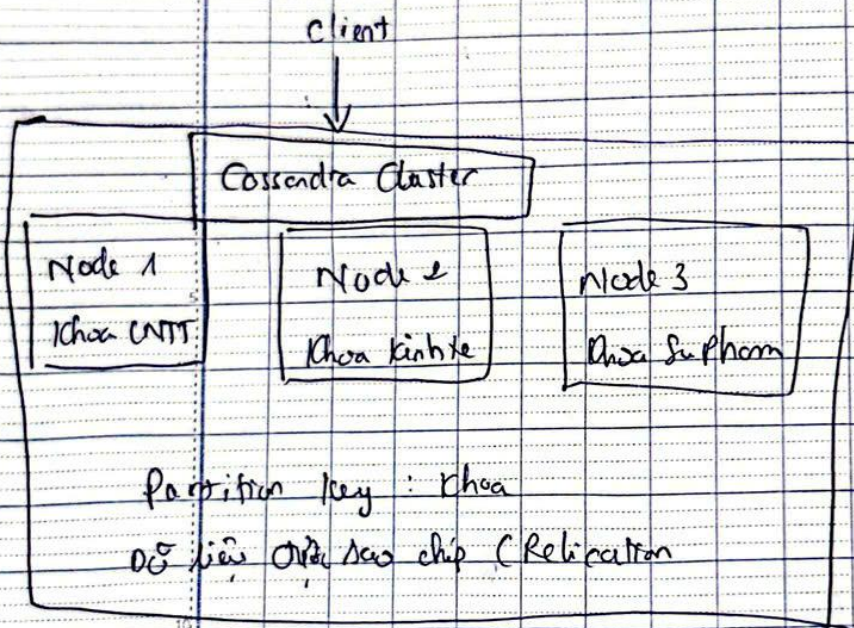
where masinhoc = 'CS101';

- Truy vấn phân tán có Partition key

select * from danhlyhocpho

where faculty = 'CNTT';

- mô hình dữ liệu phân tán Cassandra



- Dữ liệu được phân tán trên nhiều node trong Cassandra Cluster dựa trên Partition key (khoa)
- Dữ liệu được sao chép giữa các node nhằm đảm bảo tính sẵn sàng và khả năng chịu lỗi