

Họ và tên: Nguyễn Trọng Thuận

Lớp: 12_DH_CNTT4

MSSV: 1250080193

Bài Tập Buổi 3

Tìm hiểu, trình bày về một trong các hệ quản trị cơ sở dữ liệu NoSQL tùy chọn tại website NoSQL Database-From Basics to Enterprise Use /nômô phỏng việc lưu trữ, truy vấn phân tán của chúng.

Bài làm

MongoDB là một hệ quản trị cơ sở dữ liệu NoSQL phổ biến, thuộc nhóm Document Databases, được thiết kế để lưu trữ và xử lý dữ liệu linh hoạt trong các hệ thống phân tán và quy mô lớn.

Theo trang website nosql-databases.org, MongoDB là đại diện tiêu biểu cho các hệ NoSQL hiện tại, phục vụ các ứng dụng web, microservices, big data và hệ thống toàn cầu.

- Lịch sử hình thành.

- + MongoDB được phát triển lần đầu vào năm 2009
- + Ban đầu do công ty 10gen phát triển, sau này đổi tên thành MongoDB Inc.
- + MongoDB ra đời nhằm khắc phục các hạn chế của hệ CSDL quan hệ truyền thống trong môi trường dữ liệu lớn và phân tán

- Tác giả / tổ chức quản lý
- + Tổ chức quản lý và phát triển: MongoDB Inc.
- + MongoDB cung cấp:
 - Community Edition (mã nguồn mở).
 - Enterprise Edition (thương mại, hỗ trợ doanh nghiệp).

- Mô hình lưu trữ dữ liệu

+ Đặc điểm chính:

- Dữ liệu được lưu dưới dạng document (JSON/BSON)
- Không yêu cầu schema cố định
- Mỗi document có thể có cấu trúc khác nhau.
- Phù hợp với dữ liệu bán cấu trúc và phi cấu trúc

- Ví dụ

Ví dụ document:

```
{
  "_id": 1,
  "ma_sv": "SV001",
  "mon_hoc": "CSDL Nang Cao",
  "diem": 8.5,
  "co_so": "HCM"
}
```

- Ngôn ngữ thao tác với dữ liệu
- + MongoDB sử dụng MongoDB Query Language (MQL) thay vì SQL truyền thống.

Ví dụ thao tác dữ liệu:

- Thêm dữ liệu:

```
db.ketqua.insertOne({
  ma_sv: "SV002",
  mon_hoc: "NoSQL",
  diem: 9,
  co_so: "HCM"
})
```

- Truy vấn dữ liệu:

```
db.ketqua.find({co_so: "HCM"})
```

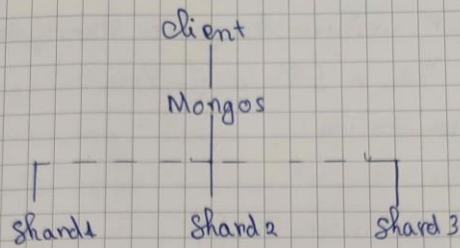
- Cơ chế phân tán trong MongoDB

Theo nosql-databases.org, MongoDB hỗ trợ horizontal scaling thông qua cơ chế sharding.

+ Các thành phần chính:

- Shard: lưu trữ dữ liệu thực tế.
- Config server: lưu thông tin phân mảnh.
- Mongos: định tuyến truy vấn từ client.

Sơ đồ kiến trúc phân tán :



- Mô phỏng lưu trữ dữ liệu phân tán

Bài toán mô phỏng :

Hệ thống quản lý kết quả học tập sinh viên theo

cơ sở đào tạo :

Chọn shard key :

shard key = co-so

Phân bố dữ liệu :

• shard 1: co-so = "HCM"

• shard 2: co-so = "HN"

• shard 3: co-so = "DN"

- Mô phỏng truy vấn phân tán :

+ Truy vấn có shard key :

db.ketqua.find({co-so: "HCM"}).

+ Phân tích :

- Mongodbi gửi gửi truy vấn đến shard HCM.
- Không gửi đến các shard khác.
- Giảm chi phí truyền dữ liệu.
- Thể hiện tới mức hoá toàn cục.

+ Truy vấn không có shard key.
db.ketqua.find({diem: {gte: 8}}).

+ Phân tích:

- Mongo gửi truy vấn đến tất cả các shard.
- Mỗi shard xử lý cục bộ.
- Kết quả được gom lại.
- Gây tốn tài nguyên mạng.

- Nhận xét và kết luận

+ MongoDB phù hợp cho hệ thống phân tán và dữ liệu lớn.

+ Hiệu quả truy vấn phụ thuộc vào thiết kế shard key.

+ Truy vấn không có shard key làm giảm hiệu năng.

+ Phù hợp với kiến trúc microservices và ứng dụng web hiện đại.

MongoDB là một hệ quản trị CSDL NoSQL. Tiêu biểu, hỗ trợ lưu trữ linh hoạt và truy vấn dữ liệu môi trường phân tán.