

**BỘ THÔNG TIN VÀ TRUYỀN THÔNG
HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN
THÔNG**



BÁO CÁO THỰC TẬP CƠ SỞ

TÌM HIỂU VỀ MONGODB

Giảng viên hướng dẫn : Kim Ngọc Bách

Họ và tên : Ngô Vũ Minh Quý

Mã sinh viên : B22DCVT427

Lớp : E22CQCN02-B

MỤC LỤC

I. Giới thiệu về MongoDB

II. Ứng dụng MongoDB trong Big Data

III. Cài đặt MongoDB

IV. Các thao tác cơ bản với MongoDB

V. Đánh giá và kết luận

I. Giới thiệu về MongoDB



MongoDB là một hệ quản trị cơ sở dữ liệu NoSQL mã nguồn mở, được phát triển bởi MongoDB Inc. Điểm nổi bật của MongoDB là cách lưu trữ dữ liệu dưới dạng tài liệu (document) theo định dạng BSON (Binary JSON), giúp linh hoạt hơn trong việc xử lý dữ liệu phi cấu trúc hoặc bán cấu trúc – vốn rất phổ biến trong các hệ thống Big Data.

MongoDB cho phép mở rộng theo chiều ngang thông qua kỹ thuật sharding, hỗ trợ tính sẵn sàng cao (high availability) thông qua replication và cung cấp khả năng truy vấn mạnh mẽ cùng với các công cụ như MongoDB Aggregation Framework – rất phù hợp để xử lý các tập dữ liệu lớn và phức tạp.

II. Ứng dụng MongoDB trong Big Data

+ Lý do lựa chọn MongoDB cho Big Data

Trong các hệ thống Big Data, dữ liệu thường có khối lượng rất lớn, thay đổi nhanh và không tuân theo một cấu trúc cố định. MongoDB đáp ứng tốt những yêu cầu này nhờ:

- **Tính mở rộng cao:** MongoDB có khả năng mở rộng ngang (horizontal scaling) bằng cách phân tán dữ liệu qua nhiều node.

- **Hiệu suất cao:** Với cơ chế lưu trữ dữ liệu dạng BSON và cơ sở dữ liệu tài liệu, MongoDB cho phép truy xuất dữ liệu nhanh chóng và linh hoạt.
- **Tính linh hoạt về mô hình dữ liệu:** Cho phép lưu trữ nhiều định dạng dữ liệu khác nhau, kể cả dữ liệu lồng nhau, rất phù hợp với dữ liệu phi cấu trúc như JSON, logs, hoặc dữ liệu từ các thiết bị IoT.
- **Hệ sinh thái phong phú:** MongoDB có thể tích hợp dễ dàng với các công cụ trong hệ sinh thái Big Data như Apache Spark, Hadoop, Kafka, v.v.

+ Ứng dụng thực tiễn

MongoDB được sử dụng rộng rãi trong nhiều ứng dụng Big Data như:

- **Phân tích log hệ thống:** Lưu trữ và xử lý log từ hàng ngàn máy chủ, cung cấp báo cáo theo thời gian thực.
- **Quản lý người dùng và hành vi khách hàng:** Ghi nhận lịch sử truy cập và hành vi người dùng từ website hoặc ứng dụng.
- **Hệ thống khuyến nghị (Recommendation Systems):** Lưu trữ dữ liệu sản phẩm, người dùng và tương tác để đưa ra gợi ý.
- **Dữ liệu IoT:** Thu thập và xử lý dữ liệu từ hàng triệu thiết bị cảm biến với tốc độ cao.
- **Thương mại điện tử và tài chính:** Phân tích dữ liệu giao dịch để phát hiện gian lận và tối ưu trải nghiệm khách hàng.



III. Cài đặt MongoDB

Việc cài đặt MongoDB có thể thực hiện theo nhiều cách: qua gói cài đặt chính thức, sử dụng Docker, hoặc dùng dịch vụ đám mây MongoDB Atlas. Ví dụ dưới đây hướng dẫn cài đặt MongoDB trên hệ điều hành Windows:

1. Tải MongoDB Community Edition từ trang chủ
<https://www.mongodb.com/try/download/community>
2. Cài đặt theo hướng dẫn và chọn tùy chọn "Complete".
3. Khởi động MongoDB bằng cách sử dụng **mongod** để chạy server và **mongo** để mở client.

IV. Các thao tác cơ bản với MongoDB

Trong một dự án Big Data, thao tác với MongoDB không chỉ dừng lại ở việc tạo và quản lý dữ liệu đơn giản mà còn bao gồm các kỹ thuật nâng cao phục vụ cho truy vấn hiệu quả, xử lý dữ liệu lớn và tích hợp với các hệ thống khác. Dưới đây là một số thao tác cơ bản và nâng cao với MongoDB:

+ Khởi tạo cơ sở dữ liệu và collection

MongoDB không yêu cầu tạo trước cơ sở dữ liệu hay collection — chúng được tạo tự động khi bạn chèn dữ liệu lần đầu. Tuy nhiên, bạn cũng có thể tạo chúng thủ công:

```
>_MONGOSH  
  
test> use BigDataProject // chuyển hoặc tạo database  
      db.createCollection("SensorData") // tạo collection
```

+ Thêm dữ liệu vào collection

```
>_MONGOSH  
  
< switched to db BigDataProject  
BigDataProject> // Thêm một tài liệu  
                db.SensorData.insertOne({  
                  device_id: "sensor_001",  
                  timestamp: ISODate("2025-05-14T08:00:00Z"),  
                  temperature: 28.4,  
                  humidity: 65,  
                  location: { lat: 10.762622, lng: 106.660172 }  
                })  
  
// Thêm nhiều tài liệu  
                db.SensorData.insertMany([  
                  {  
                    device_id: "sensor_002",  
                    timestamp: ISODate("2025-05-14T08:01:00Z"),  
                    temperature: 27.9,  
                    humidity: 66,  
                    location: { lat: 10.7627, lng: 106.6601 }  
                  },  
                  {  
                    device_id: "sensor_003",  
                    timestamp: ISODate("2025-05-14T08:02:00Z"),  
                    temperature: 29.0,  
                    humidity: 64,  
                    location: { lat: 10.7630, lng: 106.6605 }  
                  }  
                ])  
                |
```

+ Truy vấn

MongoDB hỗ trợ cú pháp truy vấn linh hoạt tương tự như JSON, bao gồm cả toán tử so sánh và logic.

```
}
BigDataProject> // Lấy tất cả các dữ liệu
db.SensorData.find()

// Truy vấn với điều kiện
db.SensorData.find({ temperature: { $gt: 28 } })

// Truy vấn nhiều điều kiện
db.SensorData.find({ temperature: { $gt: 27 }, humidity: { $lt: 66 } })

// Truy vấn với AND, OR
db.SensorData.find({
  $or: [
    { temperature: { $lt: 27 } },
    { humidity: { $gt: 70 } }
  ]
})

// Giới hạn và sắp xếp kết quả
db.SensorData.find().sort({ timestamp: -1 }).limit(5)
```

+ Cập nhật dữ liệu

Cập nhật thông tin tài liệu phù hợp với thay đổi dữ liệu theo thời gian thực:

```
BigDataProject> // Cập nhật một trường
db.SensorData.updateOne(
  { device_id: "sensor_001" },
  { $set: { temperature: 30.2 } }
)

// Cập nhật nhiều tài liệu
db.SensorData.updateMany(
  { temperature: { $gt: 30 } },
  { $set: { status: "warning" } }
)

// Tăng giá trị của trường
db.SensorData.updateOne(
  { device_id: "sensor_002" },
  { $inc: { temperature: 1 } }
)
```

+ Sử dụng Aggregation Framework để phân tích dữ liệu

Aggregation là một công cụ mạnh mẽ của MongoDB để xử lý dữ liệu Big

Data theo dạng pipeline (chuỗi bước).

```
BigDataProject> // Tính nhiệt độ trung bình theo từng thiết bị
db.SensorData.aggregate([
  { $group: { _id: "$device_id", avgTemp: { $avg: "$temperature" } } }
])

// Thống kê số lượng bản ghi theo từng ngày
db.SensorData.aggregate([
  {
    $project: {
      day: { $dateToString: { format: "%Y-%m-%d", date: "$timestamp" } },
      temperature: 1
    }
  },
  {
    $group: {
      _id: "$day",
      count: { $sum: 1 },
      avgTemp: { $avg: "$temperature" }
    }
  },
  { $sort: { _id: 1 } }
])
```

V.Đánh giá và kết luận

MongoDB là một giải pháp mạnh mẽ và linh hoạt để xử lý dữ liệu lớn trong các hệ thống Big Data hiện đại. Với khả năng mở rộng vượt trội, mô hình dữ liệu linh hoạt và hiệu năng xử lý tốt, MongoDB đang ngày càng được các doanh nghiệp tin dùng để xây dựng các hệ thống phân tích dữ liệu thời gian thực và có khả năng mở rộng cao. Đặc biệt, khi kết hợp với các công nghệ như Apache Spark hoặc Kafka, MongoDB giúp mở rộng đáng kể năng lực xử lý dữ liệu phức tạp.

Kết luận: MongoDB là sự lựa chọn phù hợp cho các hệ thống Big Data cần xử lý dữ liệu phi cấu trúc hoặc dữ liệu thời gian thực, và có thể tích hợp tốt trong kiến trúc dữ liệu hiện đại.

CÁC TÍNH NĂNG CHÍNH CỦA MONGODB

