

Hoàng Công Toàn - 10CN TT4 - 1250080204

1. Tên HQT CSDL

- MongoDB

2. Lịch sử hình thành

- Khởi nguồn: được phát triển vào năm 2007 bởi công ty 10gen (sau này đổi tên thành MongoDB Inc.) có trụ sở tại New York.
- Mục đích ban đầu: Ban đầu, nó do thiết kế như 1 phần của giới pháp nền tảng đám mây (Platform as a Service), nhưng sau đó nhóm phát triển này nhận thấy tiềm năng của nó nên đã tách ra thành 1 sản phẩm độc lập.
- Mã nguồn mở: Năm 2009, MongoDB chính thức được công bố là dự án mã nguồn mở (open source) cho phép cộng đồng sử dụng và đóng góp miễn phí.

3. Tác giả và tổ chức quản lý

- Tác giả chính: DKight Merriman, Eliot Horowitz và Kevin Ryan (những người sáng lập 10gen / MongoDB Inc.).
- Tổ chức quản lý hiện tại: MongoDB Inc. (công ty niêm yết trên sàn NASDAQ)



4. Mô hình lưu trữ (~~Storage~~ Storage Model)

- MongoDB sử dụng mô hình Document - Oriented (hướng tài liệu), khác hoàn toàn với mô hình bảng (table) của SQL truyền thống:
 - + BSON (Binary JSON): chữ viết để lưu trữ dưới dạng các tài liệu (Document) có cấu trúc tương tự JSON nhưng ở dạng nhị phân để tối ưu tốc độ và không gian.
 - + Schema-less (không định nghĩa trước cột hay kiểu dữ liệu): Không cần định nghĩa trước cột hay kiểu dữ liệu. Mỗi tài liệu trong cùng 1 bộ sưu tập (collection) có thể có các trường (field) khác nhau.
 - + Phân cấp: Dữ liệu có thể lồng nhau (Nested Document).

5. Ngôn ngữ thao tác dữ liệu

- Sử dụng MQL (MongoDB Query Language).
Một ngôn ngữ trung gian dựa trên cấu trúc JSON / Java Script, rất trực quan cho lập trình viên.

6. Cơ chế phân tán

- Đây là phần quan trọng nhất của NoSQL để xử lý Big Data. Mongo DB sử dụng cơ chế Sharding (phân mảnh) để phân tán dữ liệu.

+ Khái niệm Sharding: là quá trình chia nhỏ một tập dữ liệu lớn (Big Data) ra nhiều phần nhỏ (gọi là Shard) và lưu trữ trên nhiều máy chủ vật lý phân tán.

+ Mô hình kiến trúc phân tán (Cluster):

- 1.) Shard (máy chủ chứa dữ liệu)
- 2.) Config Server (máy chủ cấu hình)
- 3.) Mongos (Router)

+ Cách thức hoạt động:

• Hệ thống sẽ chọn 1 máy chủ lưu làm shard key

• Dựa vào giá trị của shard key, hệ thống sẽ tự động điều chuyển dữ liệu sang máy chủ tương ứng.



Mô phỏng: hệ thống giao dịch ngân hàng

[người dùng]

↓ gửi yêu cầu tìm kiếm

[Router Manager]

Kịch bản 1: Truy vấn nhanh

Yêu cầu: "Tìm giao dịch ở Hà Nội"

↓ Router biết địa chỉ
→ đi tìm chủ 1 máy

[Máy chủ 1 (Hà Nội)]

Kịch bản 2: Truy vấn chậm

Yêu cầu: "Tìm giao dịch 5-triệu"

↓ Router không biết địa chỉ
→ tìm All máy

[Máy chủ 1 (HN)]

[Máy chủ 2 (TP.HCM)]

↓ Tổng hợp kết quả từ 2 máy

Trả kết quả cho người dùng