



VIETTEL AI RACE	TD194
GIỚI THIỆU VỀ VỊ TRÍ MONGODB	Lần ban hành: 1

1. MONGODB là gì?

MongoDB là một phần mềm mã nguồn mở dùng để quản trị cơ sở dữ liệu NoSQL.

NoSQL (Not only SQL) được sử dụng thay thế cho cơ sở dữ liệu quan hệ (Relational Database – RDB) truyền thống. Cơ sở dữ liệu NoSQL khá hữu ích trong khi làm việc với các tập dữ liệu phân tán lớn. MongoDB là một công cụ có thể quản lý thông tin hướng document cũng như lưu trữ hoặc truy xuất thông tin.

Trong khi đó, ngôn ngữ truy vấn có cấu trúc (SQL) là ngôn ngữ lập trình được tiêu chuẩn hóa, dùng để quản lý cơ sở dữ liệu quan hệ. Dữ liệu được chuẩn hóa SQL dưới dạng schema và table và mọi table đều có cấu trúc cố định.

Hiện nay, có nhiều công ty toàn cầu sử dụng MongoDB để lưu trữ lượng dữ liệu “khổng lồ” của họ như Facebook, Nokia, eBay, Adobe, Google,...

2. CÔNG DỤNG CỦA MONGODB

MongoDB giúp các tổ chức lưu trữ lượng lớn dữ liệu trong khi vẫn hoạt động nhanh chóng. Ngoài lưu trữ dữ liệu, MongoDB còn được sử dụng trong các trường hợp sau:

- Tích hợp một lượng lớn dữ liệu đa dạng
- Mô tả các cấu trúc dữ liệu phức tạp, biến hoá
- Cung cấp dữ liệu cho các ứng dụng hiệu suất cao
- Hỗ trợ các ứng dụng đám mây lai và đa đám mây
- Hỗ trợ phương pháp phát triển Agile

Thay vì sử dụng các table và row như trong cơ sở dữ liệu quan hệ, vì là cơ sở dữ liệu NoSQL, MongoDB được tạo thành từ collection và document.

Document được tạo thành từ các cặp khóa-giá trị (là đơn vị dữ liệu cơ bản của MongoDB). Còn collection, tương đương với table trong SQL, là nơi chứa các bộ document.

3. CÁC THUẬT NGỮ MONGODB THƯỜNG DÙNG

3.1 _ID

_id là một trường bắt buộc trong mọi document của MongoDB. _id được sử dụng để đại diện cho tính duy nhất của một document trong một collection.

Trường _id hoạt động giống như khóa chính (primary key) của document. _id là một số thập lục phân 12 byte đảm bảo tính duy nhất của mọi document. Bạn có thể cung cấp _id trong khi chèn document. Trong 12 byte này:

	VIETTEL AI RACE	TD194
	GIỚI THIỆU VỀ VỊ TRÍ MONGODB	Lần ban hành: 1

- 4 byte đầu tiên đại diện cho thời điểm hiện tại (dựa trên hệ giây của Unix Epoch);
- 3 byte tiếp theo cho id máy;
- 2 byte tiếp theo cho process id của máy chủ MongoDB;
- 3 byte cuối cùng là giá trị tăng đơn giản.

Nếu bạn không cung cấp được số id thì MongoDB sẽ tự động cung cấp một id duy nhất cho document của bạn.

3.2 Document

Document là đơn vị lưu trữ dữ liệu cơ bản trong cơ sở dữ liệu MongoDB. Document mang vai trò tương tự như row trong các hệ thống cơ sở dữ liệu quan hệ truyền thống.

Document là một cách để sắp xếp và lưu trữ dữ liệu dưới dạng một tập hợp các cặp field-value. Document trong MongoDB không cần phải có cùng một bộ field hoặc cấu trúc với các document khác trong cùng một collection.

Đồng thời, các field chung trong document của một collection có thể chứa các loại dữ liệu khác nhau.

3.3 Collection

Collection là một tập hợp các document MongoDB. Collection tương tự như table trong hệ thống cơ sở dữ liệu quan hệ. Các collection có tính chất schema less, do đó các document trong cùng một collection có thể có các trường khác nhau.

Thông thường, một collection chứa các document có mục đích tương tự hoặc liên quan với nhau.

3.4 Database

Trong MongoDB, database là một container vật lý chứa tập hợp các collection. Một database có thể chứa 0 collection hoặc nhiều collection.

Một phiên bản máy chủ MongoDB có thể lưu trữ nhiều database và không có giới hạn về số lượng database có thể được lưu trữ trên một phiên bản, nhưng giới hạn ở không gian bộ nhớ ảo có thể được phân bổ bởi hệ điều hành.

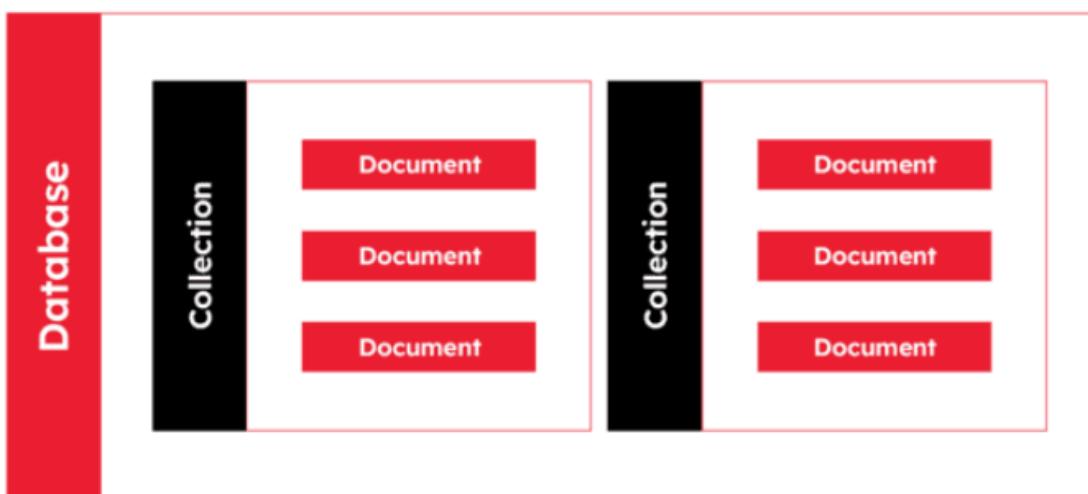
4. MONGODB hoạt động như thế nào?

4.1 MongoDB lưu trữ dữ liệu như thế nào?

Như chúng ta biết rằng MongoDB là một máy chủ cơ sở dữ liệu và dữ liệu được lưu trữ trong các cơ sở dữ liệu này. Hay nói cách khác, môi trường

MongoDB cung cấp cho bạn một máy chủ mà bạn có thể khởi động và sau đó tạo nhiều cơ sở dữ liệu trên đó bằng MongoDB.

Nhờ vào cơ sở dữ liệu NoSQL, dữ liệu được lưu trữ dưới dạng collection và document. Do đó, cơ sở dữ liệu, collection và document có mối liên hệ với nhau như hình dưới đây:



Trong máy chủ MongoDB, bạn có thể tạo nhiều cơ sở dữ liệu và nhiều collection.

Cách cơ sở dữ liệu MongoDB chứa các collection cũng giống như cách cơ sở dữ liệu MySQL chứa các table.

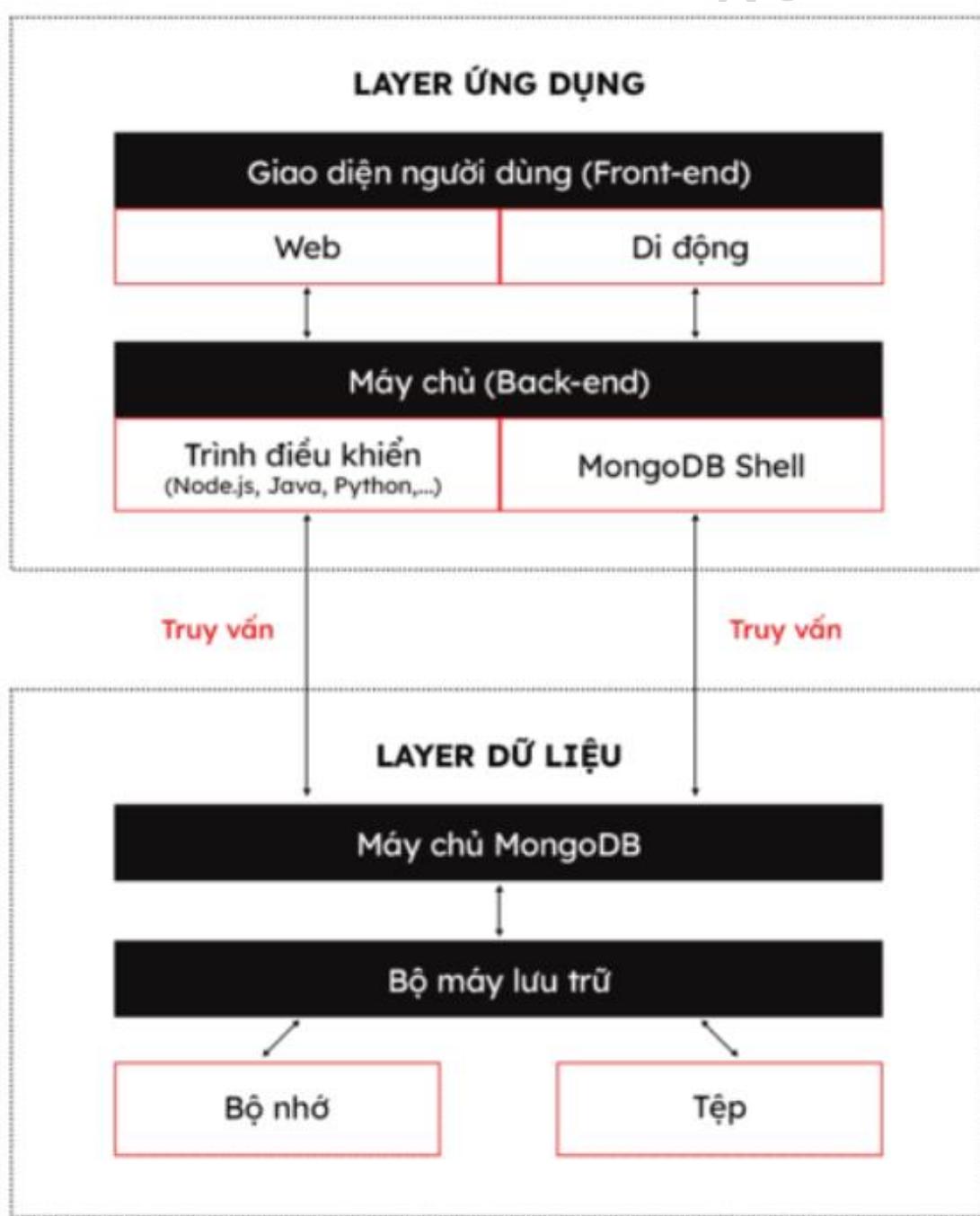
Bên trong collection, chúng ta có document. Các document này chứa dữ liệu mà bạn muốn lưu trữ trong cơ sở dữ liệu MongoDB và một collection có thể chứa nhiều document. Đồng thời, với tính chất schema-less (không cần một cấu trúc lưu trữ dữ liệu), document này không nhất thiết phải giống với document khác.

Các document được tạo bằng cách sử dụng các field (trường). Các field là các cặp khóa-giá trị trong document, giống như các column trong cơ sở dữ liệu quan hệ. Giá trị của các field có thể là bất kỳ loại dữ liệu BSON nào như double, string, boolean, v.v.

MongoDB lưu trữ dữ liệu ở định dạng BSON document. Ở đây, BSON là đại diện cho định dạng mã hoá nhị phân của các tài liệu JSON (chữ B trong BSON là viết tắt của Binary). Hay nói cách khác, trong phần backend, máy chủ MongoDB chuyển đổi dữ liệu JSON thành dạng nhị phân, được gọi là BSON, và BSON này có thể được lưu trữ và truy vấn hiệu quả hơn. Kích thước tối đa của BSON document là 16 MB.

Trong MongoDB document, bạn được phép lưu trữ dữ liệu lồng nhau. Việc lồng dữ liệu này cho phép bạn tạo các mối quan hệ phức tạp giữa dữ liệu và lưu trữ chúng trong cùng một document, giúp cho quá trình làm việc và tìm nạp dữ liệu hiệu quả hơn so với SQL.

4.2 MongoDB hoạt động như thế nào?



MongoDB hoạt động với hai layer:

- Layer Ứng dụng
- Layer Dữ liệu

Layer Ứng dụng còn được gọi là Layer Trùu tượng Cuối cùng (Final Abstraction Layer), gồm hai phần, đầu tiên là Front-end (Giao diện người dùng) và thứ hai là Back-end (máy chủ):

- Giao diện người dùng là nơi người dùng sử dụng MongoDB với sự trợ giúp của Web hoặc Di động. Web và thiết bị di động này bao gồm các trang web, ứng dụng di động, ứng dụng mặc định của Android, ứng dụng iOS, v.v.



VIETTEL AI RACE	TD194
GIỚI THIỆU VỀ VỊ TRÍ MONGODB	Lần ban hành: 1

- Phần back-end chứa một máy chủ được sử dụng để thực hiện logic phía máy chủ và cũng chứa trình điều khiển hoặc MongoDB shell để tương tác với máy chủ MongoDB với sự trợ giúp của truy vấn.

MongoDB shell là một bảng điều khiển JavaScript tương tác với MongoDB mà bạn có thể sử dụng để kết nối với máy chủ cơ sở dữ liệu và thực thi các lệnh trên đó, cho phép bạn thực hiện các tác vụ quản trị và đọc, viết hoặc thao tác dữ liệu trực tiếp.

Các truy vấn này được gửi đến máy chủ MongoDB thuộc Layer Dữ liệu. Nay, máy chủ MongoDB nhận các truy vấn và chuyển các truy vấn đã nhận tới công cụ lưu trữ vì bản thân máy chủ MongoDB không trực tiếp đọc hoặc ghi dữ liệu vào tệp hoặc đĩa hoặc bộ nhớ.

Sau khi chuyển các truy vấn nhận được tới bộ máy lưu trữ, bộ máy lưu trữ chịu trách nhiệm đọc hoặc ghi dữ liệu trong tệp hoặc bộ nhớ.