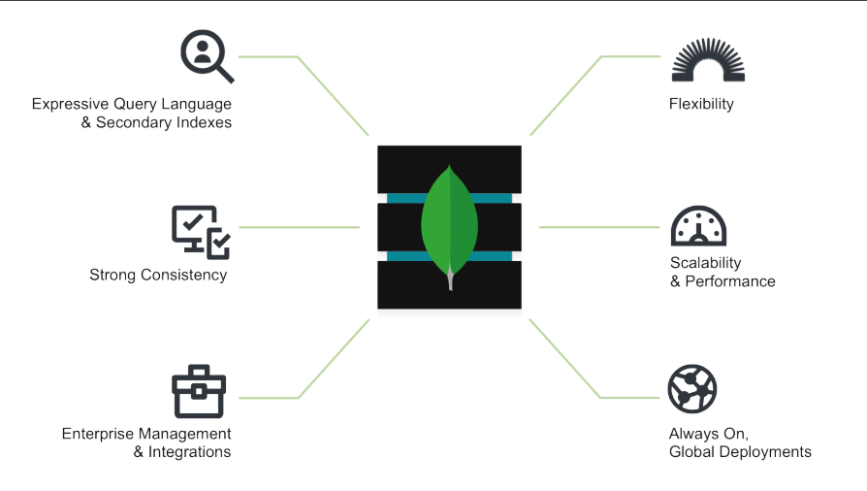
**Database Selection Report**

1. ***Chọn database cho đồ án:***

Sau khi tìm hiểu về cơ sở dữ liệu phù hợp cho đồ án này, chúng em quyết định chọn cơ sở dữ liệu MongoDB

1. ***Giới thiệu về MongoDB:*** 

* MongoDB là một hệ quản trị cơ sở dữ liệu mã nguồn mở.
* Kiến trúc: NEXUS ARCHITECTURE với triết lí tập trung kết hợp các ưu điểm của CSDL quan hệ và phi quan hệ



* Sự khác nhau giữa MongoDB với các hệ quản trị cơ sở dữ liệu quan hệ (RELATIONAL DATABASE MANAGEMENT SYSTEM – RDBMS) như MySQL, MS SQL… đó là lưu dữ liệu dạng NoSQL tức là không có sự ràng buộc giữa các bảng.
* MongoDb là document database – database hướng tài liệu. Nó lưu trữ dữ liệu trong các document dạng JSON với schema động rất linh hoạt. Nghĩa là bạn có thể lưu các bản ghi mà không cần lo lắng về cấu trúc dữ liệu như là số trường, kiểu của trường lưu trữ. Tài liệu MongoDB tương tự như các đối tượng JSON. So với RDBMS thì trong MongoDB **collection** ứng với **table**, còn **document** sẽ ứng với **row** , MongoDB sẽ dùng các document thay cho row trong RDBMS.



* Dữ liệu trong MongoDB lưu dưới dạng json. Do đó có thể hiểu document database ở đây nghĩa là dữ liệu lưu trữ kiểu text/json, chứ không lưu theo kiểu bảng, cột với kiểu dữ liệu byte, clob… như trong RDBMS.

1. ***Cách hoạt động:***

* Nguyên tắc hoạt động của MongoDB là dưới một tiến trình dịch vụ ngầm và mở một cổng (mặc định là cổng 27017), để có thể tiếp nhận các yêu cầu truy vấn, thao tác; sau đó tiến hành xử lý.
* Mỗi bản ghi của MongoDB (document) được gắn một trường có tên “\_id” nhằm xác định tính duy nhất của bản ghi. Có thể hiểu id này như tên gọi của một bản ghi và dùng phân biệt chúng với các bản ghi khác. Đồng thời, nó còn được sử dụng cho mục đích truy vấn hoặc tìm kiếm thông tin. Trường dữ liệu “\_id” được tự động đánh chỉ mục (index) để đảm bảo tốc độ truy vấn đạt hiệu suất tối ưu.
* Mỗi truy vấn dữ liệu đều được ghi đệm lên bộ nhớ RAM nên các truy vấn sau đó sẽ diễn ra nhanh hơn. Bởi nó không cần đọc dữ liệu từ ổ cứng.
* Khi thực hiện thêm, xóa hay sửa bản ghi thì MongoDB đều mất 60s để ghi các dữ liệu được thay đổi từ RAM xuống ổ cứng. Điều này nhằm mục đích đảm bảo hiệu suất mặc định của chương trình.

1. ***Các tính năng của MongoDB:***

* **Truy vấn ad hoc**

Đây được xem là tính năng tốt nhất của MongoDB. Nó có tác dụng giúp các trường truy vấn phạm vi, kết hợp cùng việc tìm kiếm biểu thức nhằm trả về kết quả tài liệu cụ thể với kích thước nhất định. Ngoài ra, các trường dùng trong MongoDB còn được dùng để lập chỉ mục chính, phụ.

* **Nhân rộng**

MongoDB cung cấp Replica Set cho phép nhân 2 hoặc nhiều bản sao của dữ liệu. Đồng thời, mỗi bản sao lại đóng vai trò chính và phụ.

- Khi nhân rộng, toàn bộ dữ liệu khi ghi và đọc được thực hiện trên bản sao chính.

- Bản sao thứ cấp sẽ dùng bản sao tích hợp để có thể duy trì các bản sao dữ liệu.

Trong trường hợp có bất kỳ bản sao chính nào bị thất bại thì Replica set sẽ chọn một bản sao thứ cấp để thay thế làm bản sao chính tiếp theo. Trong quá trình nhân rộng, Replica thứ cấp được tùy ý chọn các hoạt động nhưng dữ liệu cuối cùng vẫn phải tuân theo mặc định.

* **Cân bằng tải**

MongoDB sử dụng Sharding nhằm chia tỷ lệ theo chiều ngang và xác định dữ liệu phân phối trong collection. Điều này giúp người dùng có thể chọn một Shard key.

Nói tóm lại, MongoDB cân bằng tải bằng cách dựa vào các Shard key để chia dữ liệu thành các phạm vi và phân phối đồng đều. Chúng có thể chạy trên nhiều máy chủ khác nhau và thực hiện chức năng sao chép dữ liệu hay cân bằng tải nhằm giữ hệ thống hoạt động liên tục trong trường hợp phát sinh lỗi về phần cứng.

* **Lưu trữ tệp**

Khi tìm hiểu hệ cơ sở dữ liệu MongoDB thì bạn sẽ thấy, tính năng lưu trữ tệp được dùng như một hệ thống tệp (gọi là GridFS) đóng vai trò cân bằng tải, đồng thời, sao chép dữ liệu trên nhiều máy tính. Cụ thể, GridFS chia một tệp ra làm nhiều phần và lưu trữ thành các tài liệu riêng biệt. Sau đó, người dùng dễ dàng truy cập GridFS thông qua Mongofiles hay các plugin sử dụng cho Nginx và Lighttpd.

* **Tập hợp**

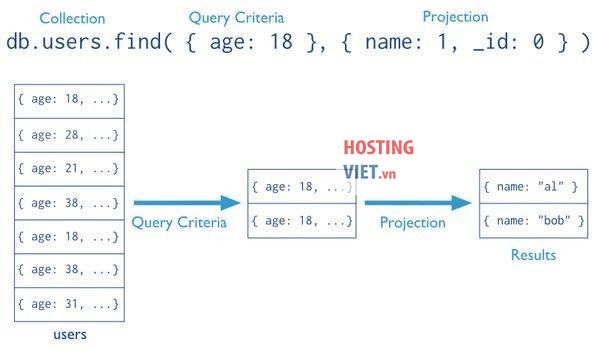
Tính năng này chính là chương trình mang đến ba giải pháp để thực hiện tập hợp gồm Aggregation Pipeline, Mapreduce và Single-purpose Aggregation. Trong đó, Aggregation Pipeline được đánh giá là có hiệu suất tốt nhất.

* **Giới hạn kích thước collection**

Các collection được MongoDB hỗ trợ thường có kích thước cố định. Vì thế, người ta gọi chúng là collection giới hạn. Với kích cỡ cố định, kết hợp cùng việc theo sau thứ tự chèn giúp tăng hiệu suất của các hoạt động liên quan đến dữ liệu. Và khi dữ liệu vượt giới hạn thì những tài liệu cũ hơn sẽ tự động bị xóa mà bạn không cần thực hiện thao tác thêm bất kỳ dòng lệnh nào.

* **Giao dịch**

Bắt đầu từ phiên bản 4.0 ra mắt vào tháng 6 năm 2018, MongoDB được bổ sung thêm tính năng hỗ trợ giao dịch ACID đa tài liệu.



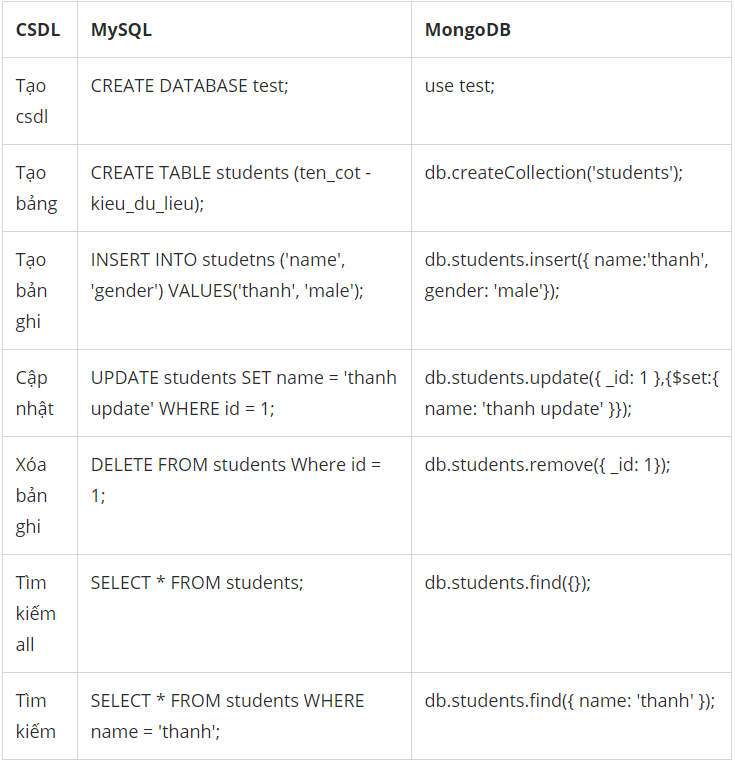
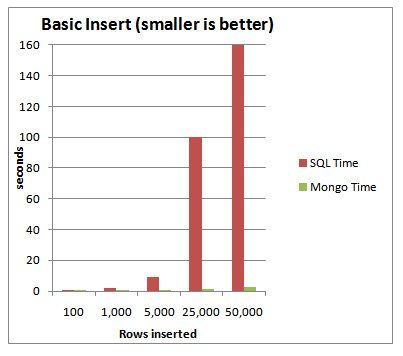


Figure 1 Một số câu lệnh cơ bản



1. ***Lý do chọn MongoDB:***

* Là một CSDL mã nguồn mở, đa nền tảng và miễn phí, có nền Javascript phù hợp với frontend(ReactJS) và backend(NodeJS) mà chúng em sử dụng.
* Có cộng đồng lớn sử dụng nên có thể dễ dàng tìm kiếm sự giúp đỡ nếu gặp khó khăn.
* Tốc độ truy vấn nhanh hơn nhiều so với hệ quản trị cơ sở dữ liệu quan hệ RDBMS do dữ liệu truy vấn được cached lên bộ nhớ RAM để lượt truy vấn sau diễn ra nhanh hơn mà không cần đọc từ ổ cứng.



* Hỗ trợ realtime tốt.
* Là một CSDL phi quan hệ(NoSQL) nên khác với RDBMS, dữ liệu trong đây không có sự ràng buộc và không có yêu cầu tuân theo khuôn khổ nhất định, điều này giúp tiết kiệm thời gian cho việc kiểm tra sự thỏa mãn về cấu trúc nếu muốn chèn, xóa, cập nhật hay thay đổi các dữ liệu trong bảng.
* Trường dữ liệu “\_id” luôn được tự động đánh index (chỉ mục) để tốc độ truy vấn thông tin đạt hiệu suất cao nhất.
* MongoDB dễ dàng mở rộng hệ thống bằng cách thêm node vào cluster – cụm các node chứa dữ liệu giao tiếp với nhau - hiệu năng cao, tính sẵn sàng cao.
* Tính linh hoạt lưu trữ dữ liệu theo các kích cỡ khác nhau, dữ liệu dưới dạng hướng tài liệu JSON nên bạn có thể chèn vào thoải mái bất cứ thông tin gì bạn muốn. Và mongoDB lưu kiểu dữ liệu là key:value giống JSON, ngoài ra, chúng em cũng sử dụng NodeJS cho đồ án của mình và nó cũng trả về kiểu JSON.
* So sánh giữa MySQL và NoSQL

|  |  |  |
| --- | --- | --- |
| **Tính năng** | **CSDL quan hệ** | **NoSQL** |
| **Hiệu suất** | Kém hơn  SQL  Quan hệ giữa các bảng | Cực tốt  Bỏ qua SQL  Bỏ qua các ràng buộc dữ liệu |
| **Khả năng mở rộng** | Hạn chế về lượng | Hỗ trợ một lượng rất lớn các node |
| **Hiệu suất đọc ghi** | Kém do thiết kế để đảm bảo sự ra vào/ra liên tục của dữ liệu | Tốt với mô hình xử lý lô và những tối ưu về đọc ghi dữ liệu |
| **Thay đổi số node trong hệ thống** | Phải tắt cả hệ thống. Việc thay đổi số node phức tạp | Không cần tắt cả hệ thống.  Việc thay đổi số node đơn giản, không ảnh hưởng đến hệ thống |
| **Phần cứng** | Đòi hỏi cao về phần cứng | Đòi hỏi thấp hơn về giá trị và tính đồng nhất của phần cứng |