## **Tìm hiểu và triển khai Elastic Search**

1. **Thành viên**

Nhóm 5 – XIZOT

|  |  |  |  |
| --- | --- | --- | --- |
| **MSSV** | **Họ & Tên** | **Nội dung công việc** | **Hoàn thành** |
| 20424027 | Phạm Thị Minh Hậu | - Elasticsearch là gì?  - Các thông tin chi tiết về Elasticsearch.  - Ưu và nhược điểm của Elasticsearch.  - Soạn báo cáo. | 33.3% |
| 20424030 | Đặng Trung Hiếu | - Cách cài đặt (trong slide) & demo.  - Cách hoạt động. | 33.3% |
| 20424058 | Nguyễn Văn Nhật | - Các khái niệm liên quan.  - Ưu và nhược điểm của Elasticsearch.  - Soạn slide. | 33.3% |

1. **Nội dung tìm hiểu**

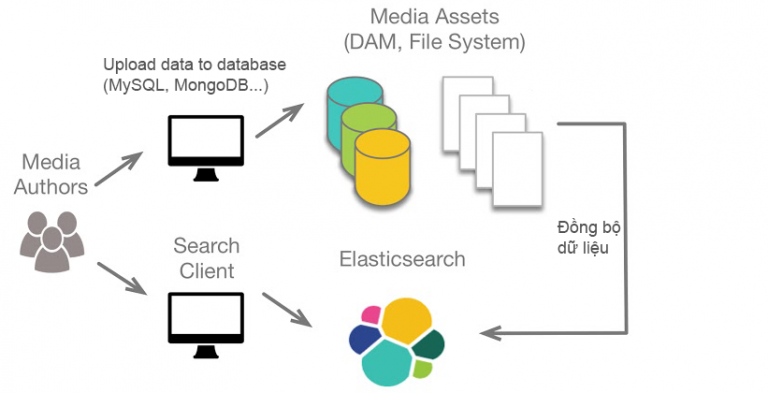
## **1/ Elasticsearch là gì?**

* Là một công cụ tìm kiếm dựa trên nền tảng Apache Lucene. Nó cung cấp một bộ máy tìm kiếm dạng phân tán, có đầy đủ công cụ với một giao diện web HTTP có hỗ trợ dữ liệu JSON

## **2/ Chi tiết về Elasticsearch (ES)**

* Là một search engine.
* Được kế thừa từ Lucene Apache.
* Thực chất hoạt động như 1 web server có khả năng tìm kiếm nhanh chóng (near realtime) thông qua giao thức RESTfull.
* Có khả năng phân tích và thống kê dữ liệu.
* Chạy trên server riêng và đồng thời giao tiếp thông qua RESTfull do vậy nó không phụ thuộc vào client viết bằng gì hay hệ thống hiện tại viết bằng gì, dễ dàng tích hợp vào hệ thống.
* Là hê thống phân tán và có khả năng mở rộng cao(horizontal scalability). Lắp thêm node cho nó là tự động mở rộng.
* Open source được phát triển bằng Java.

**3/ Cách hoạt động Elasticsearch (ES)**



Chạy dưới port: (local default 9200) dùng như 1 server riêng để phục vụ mục đích tìm kiếm và được dùng song song với 1 DB chính khác.

**4/ Ưu điểm và khuyết điểm của Elasticsearch (ES)**

*a) Ưu điểm.*

* Truy vấn search bằng ES sẽ cho ra các kết quả chính xác nhất vì vậy sẽ tăng performance.
* Khả năng mở rộng theo chiều ngang.
* Hỗ trợ Structured Query DSL (Domain-Specific Language ), cung cấp việc đặc tả những câu truy vấn phức tạp một cách cụ thể và rõ ràng bằng JSON.
* Có khả năng phân tích dữ liệu (Analysis data).

*b) Nhược điểm*

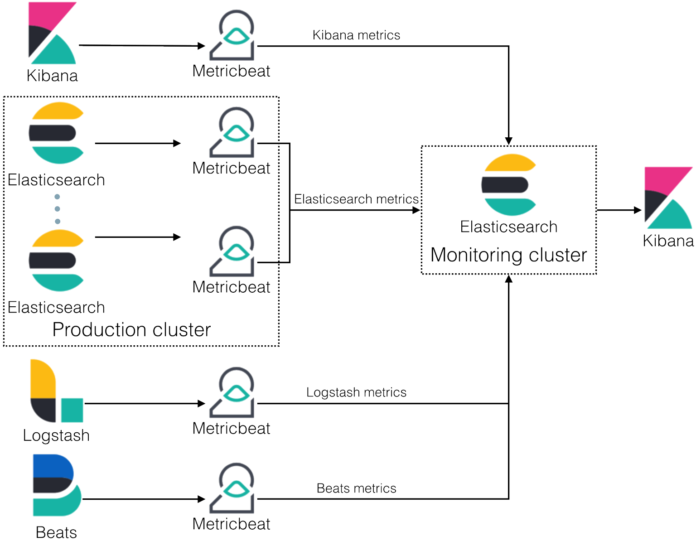
* Do là một search engine nên nó không có nhiều các thao tác truy vấn CRUD nên không đủ mạnh để có thể trở thành 1 db chính.
* Elasticsearch không có khái niệm Database transaction, nghĩa là nó sẽ không đảm bảo trọn vẹn dữ liệu khi Insert, Update, Delete .Khi thay đổi bản ghi nếu có lỗi thì làm cho mất mát dữ liệu.
* Không thích hợp với những hệ thống thường xuyên cập nhật dữ liệu. Tốn kém cho việc đánh index dữ liệu.

**5/ Các khái niệm cần biết về Elasticsearch (ES)**

*a) Document trong Elasticsearch là gì?*

* Document là một JSON object với một số dữ liệu. Đây là basic information unit trong ES. Hiểu 1 cách cơ bản thì đây là đơn vị nhỏ nhất để lưu trữ dữ liệu trong ES.

*b) Index là gì?*

* Nó được thiết kế để cho phép tìm kiếm full-text search.
* 

*c) Node là gì?*

* Là trung tâm hoạt động của Elasticsearch. Là nơi lưu trữ dữ liễu, tham gia thực hiện đánh index của cluster cũng như thực hiện thao tác tìm kiếm.
* Mỗi node được định danh bằng 1 **unique name**.

*d) Cluster là gì?*

* Tập hợp các nodes hoạt động cùng nhau, chia sẽ phụ thuộc cluster.name, Cluster sẽ được xác định bằng 1 ‘unique name’. Việc định danh cá cluster trùng tên sẽ gây nên lỗi cho các node vì vậy khi setup các bạn cần hết sức chú ý điểm này.
* một cluster sẽ gồm nhiều Nodes hoạt động trên các server khác nhau để đảm bảo nếu 1 server gặp sự cố thì server khác (node khác) có thể hoạt động đầy đủ chức năng so với khi có 2 servers. các Nodes có thể tìm thấy nhau để hoạt động trên cùng 1 cluster qua giao thức unicast.

**6/ Các tài liệu tham khảo**

* <https://www.elastic.co/guide/index.html>

- <https://topdev.vn/blog/elasticsearch-la-gi/>

* <http://www.elasticsearchtutorial.com/elasticsearch-in-5-minutes.html>