

**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG KHOA CÔNG NGHỆ THÔNG TIN**

Học phần: **An toàn hệ điều hành**

Bài báo cáo:

**Chủ đề số 7**

**Tìm hiểu về công cụ xử lý log ELK (Elastic Search, Logtash và Kibana)**

**Giảng viên hướng dẫn**: Hoàng Xuân Dậu

**Sinh viên thực hiện** : Nguyễn Đức Nhật (B19DCA131)

Nguyễn Minh Nhật (B19DCAT132)

Nguyễn Thị Cẩm Nhung (B19DCAT133)

Ngô Khắc Phương (B19DCAT140)

Nguyễn Minh Phương (B19DCAT141)

Nguyễn Công Phượng (B19DCAT143)

Nguyễn Văn Quốc (B19DCAT147)

Hà Nội 2022

Mục lục

[**A. Giới thiệu 3**](#_Toc97928615)

[**I, Giới thiệu về công cụ xử lý ELK (Elastic Search, Logtash và Kibana). 3**](#_Toc97928616)

[**1, Tổng quan về ELK 3**](#_Toc97928617)

[**2, Ưu điểm phần mềm ELK stack: 3**](#_Toc97928618)

[**B. Nội Dung 3**](#_Toc97928619)

[**I, Kiến trúc của công cụ xử lý log ELK (Elastic Search, Logtash và Kibana). 4**](#_Toc97928620)

[**II, Các thành phần trong ELK stack 4**](#_Toc97928621)

[**1. Elasticsearch. 4**](#_Toc97928622)

[**2. Logstash. 5**](#_Toc97928623)

[**3. Kibana. 7**](#_Toc97928624)

[**4. Beats 8**](#_Toc97928625)

[**4.1 Tìm kiếm và phân tích 8**](#_Toc97928626)

[**4.2 Lưu trữ dữ liệu 9**](#_Toc97928627)

[**4.3 Thu thập và cấu hình log 9**](#_Toc97928628)

[**4.4 Trực quan hoá dữ liệu 9**](#_Toc97928629)

[**4.5 Bảo mật và giám sát 10**](#_Toc97928630)

[**4.6 Quản lý 10**](#_Toc97928631)

[**IV, Cài đặt và thử nghiệm phân tích một số file log 10**](#_Toc97928632)

[**1. Chuẩn bị: 10**](#_Toc97928633)

[**2. Thực hiện : 10**](#_Toc97928634)

[**C, Kết Luận: 14**](#_Toc97928635)

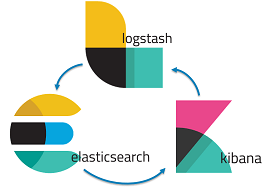
[**Tài liệu tham khảo: 14**](#_Toc97928636)

# Giới thiệu

## **I, Giới thiệu về công cụ xử lý ELK (Elastic Search, Logtash và Kibana).**

### **1, Tổng quan về ELK**

ELK là viết tắt của tập hợp 3 phần mềm cốt lõi đi kèm với nhau, phục vụ cho công việc giám sát hệ thống. Ba phần mềm này lần lượt là Elasticsearch, Logstash và Kibana. ELK được phát triển từ đầu những năm 2000 và cho đến nay đã hơn 250 triệu lượt tải xuống và sử dụng. Hiện tại phiên bản mới nhất của ELK là 6.4. Đây là bộ công cụ giám sát tập trung mã nguồn mở rất mạnh, có thể xử lý rất nhiều bài toán quản lý hệ thống mạng nên rất được các công ty, tổ chức tin dùng.



### **2, Ưu điểm phần mềm ELK stack:**

- ELK stack là phần mềm mã nguồn mở, không tốn nhiều chi phí khi triển khai.

- ELK mới được phát triển mạnh gần đây, nên cộng đồng hỗ trợ rất đông đúc và mạnh mẽ.

- Thu thập được log từ rất nhiều nguồn khác nhau: log hệ thống, log ứng dụng, log thiết bị mạng, log snmp, log từ các hệ thống API (Application Programming Interface)…

- Khả năng tương thích cao, dễ dàng tích hợp với các hệ thống khác mà không gặp khó khăn.

- Giao diện Kibana trực quan, sinh động. - Phương tiện cảnh báo đa dạng, tích hợp với email, sms, slack, các ứng dụng OTP (One Time Password) như Telegram, Whatsapp,...

- Hiệu năng xử lý cao, có thể chịu tải được một lượng lớn dữ liệu (log) đẩy về mà không phải quá phụ thuộc vào khả năng xử lý của phần cứng.

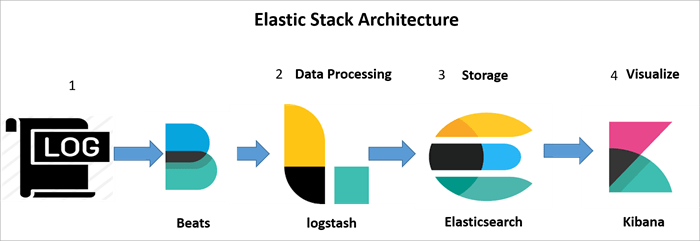
- Giải quyết được hầu hết các bài toán trong giám sát hệ thống mạng: giám sát hạ tầng, giám sát dịch vụ, giám sát an ninh, giám sát người dùng,... Đây là đặc điểm chính giúp cho ELK trong tương lai sẽ được các công ty và tổ chức sử dụng để triển khai hệ thống giám sát tập trung bên trong hệ thống mạng của họ.

# B. Nội Dung

## **I, Kiến trúc của công cụ xử lý log ELK (Elastic Search, Logtash và Kibana).**



* **Một kiến ​​trúc ngăn xếp ELK đơn giản bao gồm:**
* Logs (nhật ký): Từ máy chủ, nhật ký phải được xác định và phân tích.
* Logstash: Dự án này thu thập nhật ký và dữ liệu cho các sự kiện. Nó cũng biến đổi dữ liệu.
* Elasticsearch: Dữ liệu được Logstash chuyển đổi được lưu trữ, tìm kiếm và lập chỉ mục.
* Kibana: Cơ sở dữ liệu của Elasticsearch được sử dụng để khám phá, trực quan hóa và chia sẻ nhật ký.
* Tuy nhiên, cần thêm một thành phần nữa hay còn gọi là Thu thập dữ liệu được gọi là Beats. Điều này khiến Elastic đổi tên ELK thành Elastic Stack.



* Đây là một kiến ​​trúc đơn giản của Elastic Stack được các tổ chức nhỏ hơn sử dụng. Trong trường hợp có một lượng lớn dữ liệu, cấu trúc có thể liên quan đến nhiều bản ghi máy chủ ứng dụng và Logstash.

**II, Các thành phần trong ELK stack**

1. **Elasticsearch.**

* Elasticsearch là một RESTful distributed search engine. Hiểu nôm na là nó cung cấp khả năng tìm kiếm phân tán qua API. Lưu trữ dữ liệu theo dạng NoSQL database (cơ sở dữ liệu không có cấu trúc).
* Đầu tiên cần hiểu Elasticsearch là một công cụ tìm kiếm nguồn mở phân tán dựa trên Apache Lucene và được phát hành theo giấy phép Apache 2.0
* Elasticsearch cho phép bạn thực thi và kết hợp rất nhiều loại tìm kiếm: có cấu trúc, không cấu trúc, geo, metric theo cách bạn muốn.
* Elasticsearch rất nhanh, thực sự rất nhanh. Bạn có câu trả lời ngay tức thì với các dữ liệu thay đổi.
* Bạn có thể chạy nó trên hàng trăm server với hàng petabyte dữ liệu.
* Vận hành dễ dàng
* Elasticsearch sử dụng chuẩn RESTful APIs và JSON.
* Các khái niệm cơ bản cảu Elasticsearch:

+Document

+Index

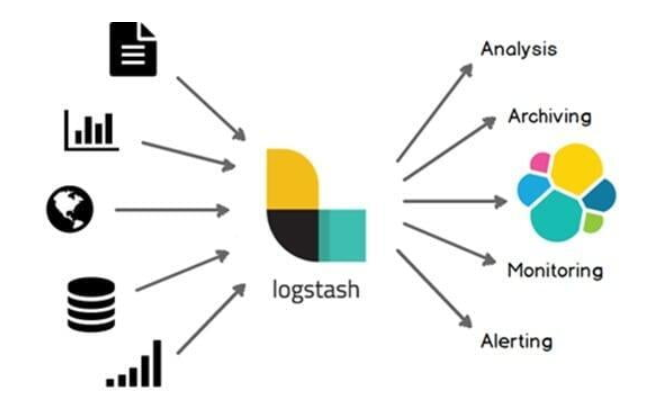
+Shard

+Node

+Cluster

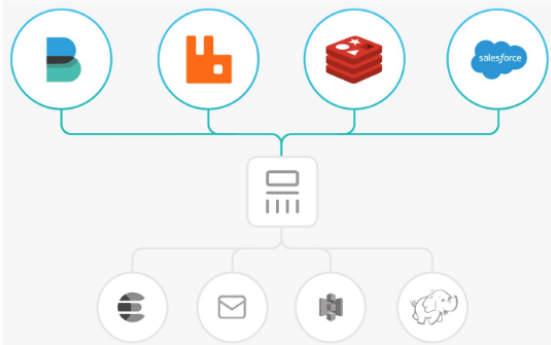


1. **Logstash.**

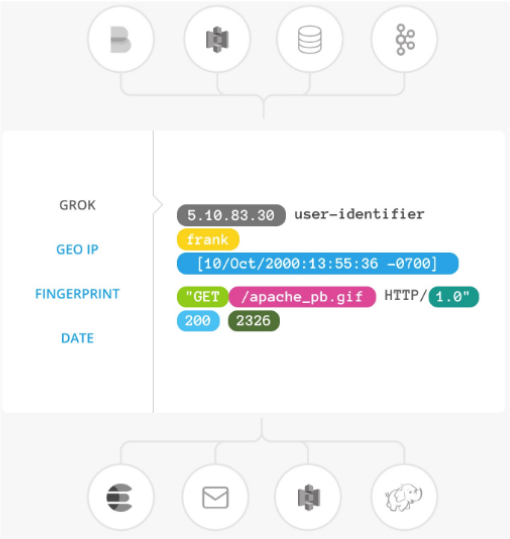


* Logstash là một đường ống dữ liệu giúp thu nhập, phân tích, xử lý log với một lượng lớn dữ liệu và sự kiện có cấu trúc được tạo ra trên các hệ thống khác nhau.
* Logstash có các chức năng như:
* Xử lí dữ liệu tập trung: Logstash giúp xây dựng một đường ống dữ liệu có thể tập trung xử lý dữ liệu. Với việc sử dụng nhiều plugin cho đầu vào và đầu ra, nó có thể chuyển đổi rất nhiều nguồn đầu vào khác nhau thành một định dạng chung.
* Hỗ trợ các định dạng log tùy chỉnh: Log được viết bởi các ứng dụng khác nhau thường có các định dạng cụ thể dành riêng cho ứng dụng. Logstash giúp phân tích cú pháp và xử lý các định dạng tùy chỉnh trên quy mô lớn. Nó cung cấp hỗ trợ để viết các bộ lọc của riêng bạn cho mã thông báo và cũng cung cấp các bộ lọc sẵn sàng sử dụng.
* Phát triển plugin: Các plugin tùy chỉnh có thể được phát triển và xuất bản, và có rất nhiều plugin được phát triển tùy chỉnh đã có sẵn.
* Mỗi dòng log của logtash được lưu dưới dạng json.
* Logstash giữ nhiệm vụ rất quan trọng bao gồm 3 giai đoạn trong chuỗi xử lý ELK pipeline, tương ứng với 3 module: Input, Filter, Output.

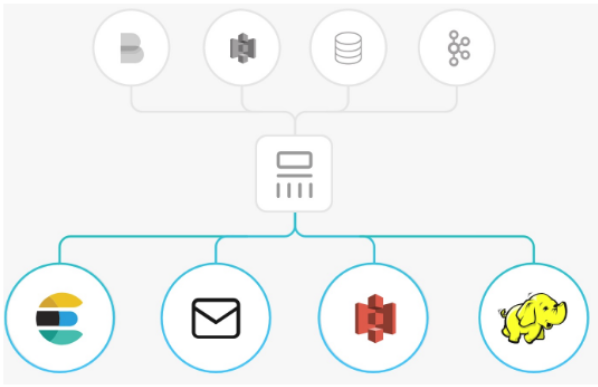
+ **INPUT**: tiếp nhận/thu thập dữ liệu logs event ở dạng thô từ các nguồn khác nhau



+Filter: sau khi tiếp nhận dữ liệu sẽ tiến hành thao tác dữ liệu logs event (như thêm, xóa,... nội dung log) theo cấu hình của quản trị viên để xây dựng lại cấu trúc dữ liệu log event theo ý của mình



**+Output**: sau cùng sẽ thực hiện chuyển tiếp dữ liệu logs event về các dịch vụ khác như Elasticsearch tiếp nhận và lưu trữ log.



* Bình thường khi làm việc với Logstash, sẽ phải làm việc với Filter nhiều nhất. Filter hiện tại sử dụng Grok để phân tích dữ liệu.

1. **Kibana.**

* Kibana là một nền tảng trực quan hóa dữ liệu được cấp phép Apache 2.0, giúp trực quan hóa mọi loại dữ liệu có cấu trúc và không cấu trúc được lưu trữ trong các chỉ mục của Elaticsearch.
* Kibana hoàn toàn được viết bằng HTML và JavaScript. Nó sử dụng các khả năng tìm kiếm và lập chỉ mục mạnh mẽ của Elaticsearch được thể hiện thông qua API RESTful của nó để hiển thị đồ họa mạnh mẽ cho người dùng cuối.
* Kibana được phát triển riêng cho ứng dụng ELK, thực hiển chuyển đổi các truy vấn của người dùng thành câu truy vấn mà Elasticsearch có thể thực hiện được. Kết quả hiển thị bằng nhiều cách, ví dụ theo các dạng biểu đồ.
* Kibana giúp dễ dàng hiểu được khối lượng dữ liệu lớn. Giao diện dựa trên trình duyệt đơn giản của nó cho phép bạn nhanh chóng tạo và chia sẻ bảng điều khiển động hiển thị các thay đổi đối với các truy vấn Elaticsearch trong thời gian thực.
* Một số chức năng chính của Kibana:
* Nó cung cấp nền tảng trực quan hóa cho doanh nghiệp
* Nó cung cấp phân tích thời gian thực, tóm tắt, biểu đồ và gỡ lỗi khả năng.
* Nó cung cấp một giao diện trực quan và thân thiện với người dùng, có thể tùy chỉnh thông qua một số tính năng kéo thả.
* Nó cho phép lưu bảng điều khiển và quản lý nhiều hơn một bảng điều khiển. Bảng điều khiển có thể dễ dàng chia sẻ và nhúng trong các hệ thống khác nhau.
* Nó cho phép chia sẻ ảnh chụp nhanh của log.

1. **Beats**

Beats là một tập hợp các công cụ thu thập thông tin chuyên dụng, được biết đến như là các Shipper (người vận chuyển) giúp thu thập và gửi dữ liệu từ Client tới máy chủ ELK. Ngoài ra, các beat này có thể được gửi thẳng trực tiếp lên Elasticsearch do bản thân các công cụ đã được chuẩn hóa sẵn, việc kết nối các beat đến logstash thường mang ý nghĩa bảo mật đối với các hệ thống tầm lớn khi họ muốn bảo vệ dịch vụ elasticsearch.

Trong Beats có rất nhiều các beats con như:

* FileBeat, Metricbeat, Paketbeat, Winlogbeat, Audibeat, Heartbeat



**III, Các tính năng của ELK**

* 1. **Tìm kiếm và phân tích**

*Tìm kiếm*

* Elasticsearch được biết đến với khả năng tìm kiếm mạnh mẽ trong nguồn dữ liệu lớn
* Sử dụng cấu trúc được gọi là chỉ mục đảo ngược, được thiết kế để cho phép tìm kiếm toàn văn bản một cách nhanh chóng và chính xác
* Elasticsearch không chỉ tìm kiếm được các nguồn cơ sở dữ liệu nổi tiếng như MySQL, MS SQL, mà nó có thể là văn bản (text), pdf, doc
* Sức mạnh của nó đến từ việc chấm điểm mức độ liên quan, mô hình xếp hạng
* Cung cấp DSL truy vấn nâng cao và một loạt các tính năng nâng cao tìm kiếm.

*Phân tích*

* Các tính năng phân tích mạnh mẽ của ELK cho phép bạn lấy dữ liệu bạn đã tìm kiếm và tìm ra ý nghĩa sâu sắc hơn
* Khung tổng hợp giúp cung cấp dữ liệu tổng hợp dựa trên truy vấn tìm kiếm, để xây dựng các bản tóm tắt phức tạp của dữ liệu.
* API khám phá đồ thị cho phép bạn trích xuất và tóm tắt thông tin về các tài liệu. Cách tốt nhất để hiểu hoạt động của API này là sử dụng Graph trong Kibana để khám phá các kết nối.
* Tạo cảnh báo ngưỡng để kiểm tra định kỳ khi dữ liệu trong các chỉ số Elasticsearch của bạn vượt lên trên hoặc dưới ngưỡng nhất định trong một khoảng thời gian nhất định
  1. **Lưu trữ dữ liệu**
* Elasticsearch hỗ trợ lưu trữ một số kiểu dữ liệu khác nhau cho các trường trong tài liệu và mỗi kiểu dữ liệu đó cung cấp nhiều kiểu con của riêng nó
* Elasticsearch hỗ trợ nhiều cách để đảm bảo dữ liệu không đánh cắp và rơi vào tay tin tặc.
  1. **Thu thập và cấu hình log**
* Beats là lựa chọn tuyệt vời để thu thập dữ liệu. Bất kể bạn có loại dữ liệu nào
* **Logstash**: Tiếp nhận/thu thập log ở dạng thô từ nhiều nguồn
* Thao tác dữ liệu sự kiện log (như thêm, xoá, thay thế,.. nội dung log) theo cấu hình của quản trị viên để xây dựng lại cấu trúc dữ liệu log event theo mong muốn. Biến dữ liệu thô thành thông tin có giá trị.
* Logstash có thể tự động thống nhất dữ liệu từ các nguồn khác nhau và chuẩn hóa dữ liệu thành các điểm đến mà bạn lựa chọn
  1. **Trực quan hoá dữ liệu**
* Tạo trực quan hóa dữ liệu trong các chỉ số Elasticsearch. Hình ảnh hóa Kibana dựa trên các truy vấn Elasticsearch.
* Trang tổng quan Kibana hiển thị một bộ sưu tập các hình ảnh và tìm kiếm. Bạn có thể sắp xếp, thay đổi kích thước và chỉnh sửa nội dung trang tổng quan
* Kibana Lens là một giao diện người dùng trực quan, dễ sử dụng, giúp đơn giản hóa quá trình hiển thị dữ liệu thông qua trải nghiệm kéo và thả
* Ứng dụng Bản đồ cho phép bạn phân tích dữ liệu địa lý của mình trên quy mô lớn, với tốc độ và thời gian thực. Tìm kiếm toàn cầu trên nhiều lớp, bạn có thể hiểu và theo dõi dữ liệu của mình một cách dễ dàng.
* Dễ dàng chia sẻ, cộng tác với hình ảnh trên Kibana
  1. **Bảo mật và giám sát**
* Trang bị cho các nhóm hoạt động bảo mật, tìm các mối đe dọa để ngăn chặn, phát hiện và ứng phó với các mối đe dọa. Nó ngăn chặn phần mềm độc hại tại máy chủ, tự động phát hiện các mối đe dọa và sự bất thường cũng như hợp lý hóa phản ứng với quy trình làm việc trực quan
* Các tính năng giám sát của ELK cung cấp cho bạn khả năng hiển thị về cách ELK của bạn đang chạy. Theo dõi hoạt động của nó để đảm bảo bạn đang tận dụng tối đa nó.
  1. **Quản lý**
* Elastic Stack đi kèm với nhiều công cụ quản lý, giao diện người dùng và API để cho phép toàn quyền kiểm soát dữ liệu, người dùng.

## **IV, Cài đặt và thử nghiệm phân tích một số file log**

### **1. Chuẩn bị:**

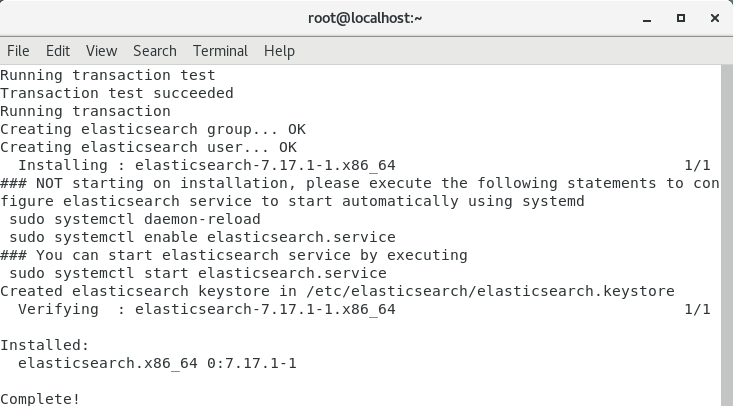
* Máy chủ Linux CentOS làm ELK Server.
* Máy client Ubuntu.

### **2. Thực hiện :**

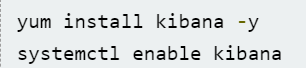
* Cài đặt trên máy chủ ELK :
* Cài đặt Elasticsearch :
* Thực hiện lệnh :



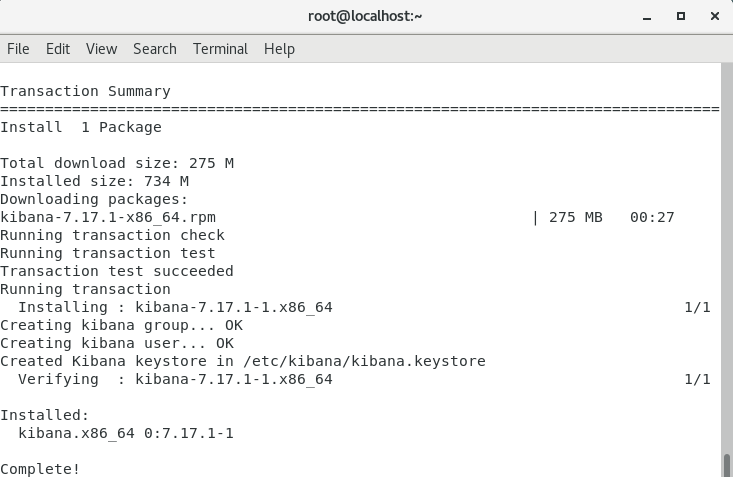
* Kết quả :



* Cài đặt Kibana :
* Thực hiện lệnh :



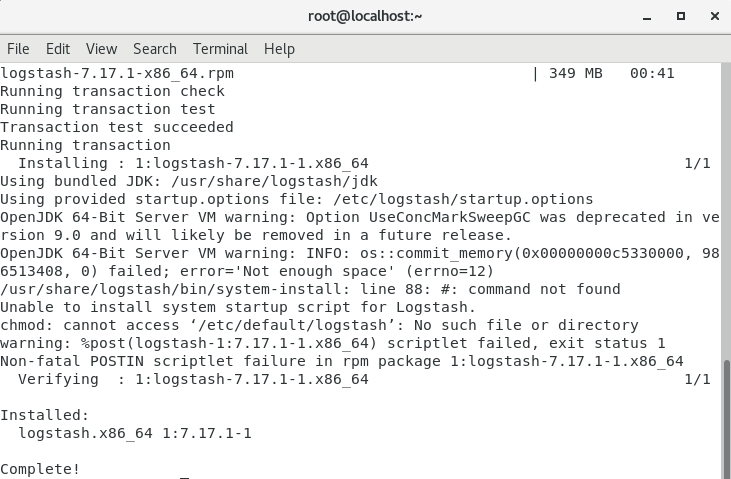
* Kết quả :



* Cài đặt Logstash CentOS :
* Thực hiện lệnh :



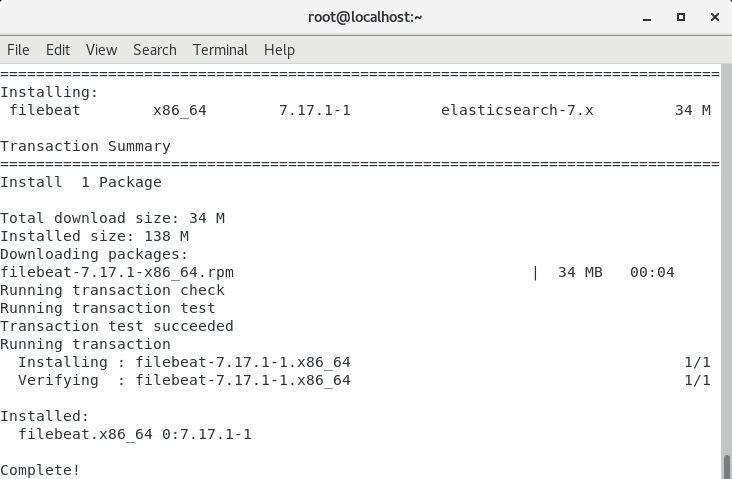
* Kết quả :



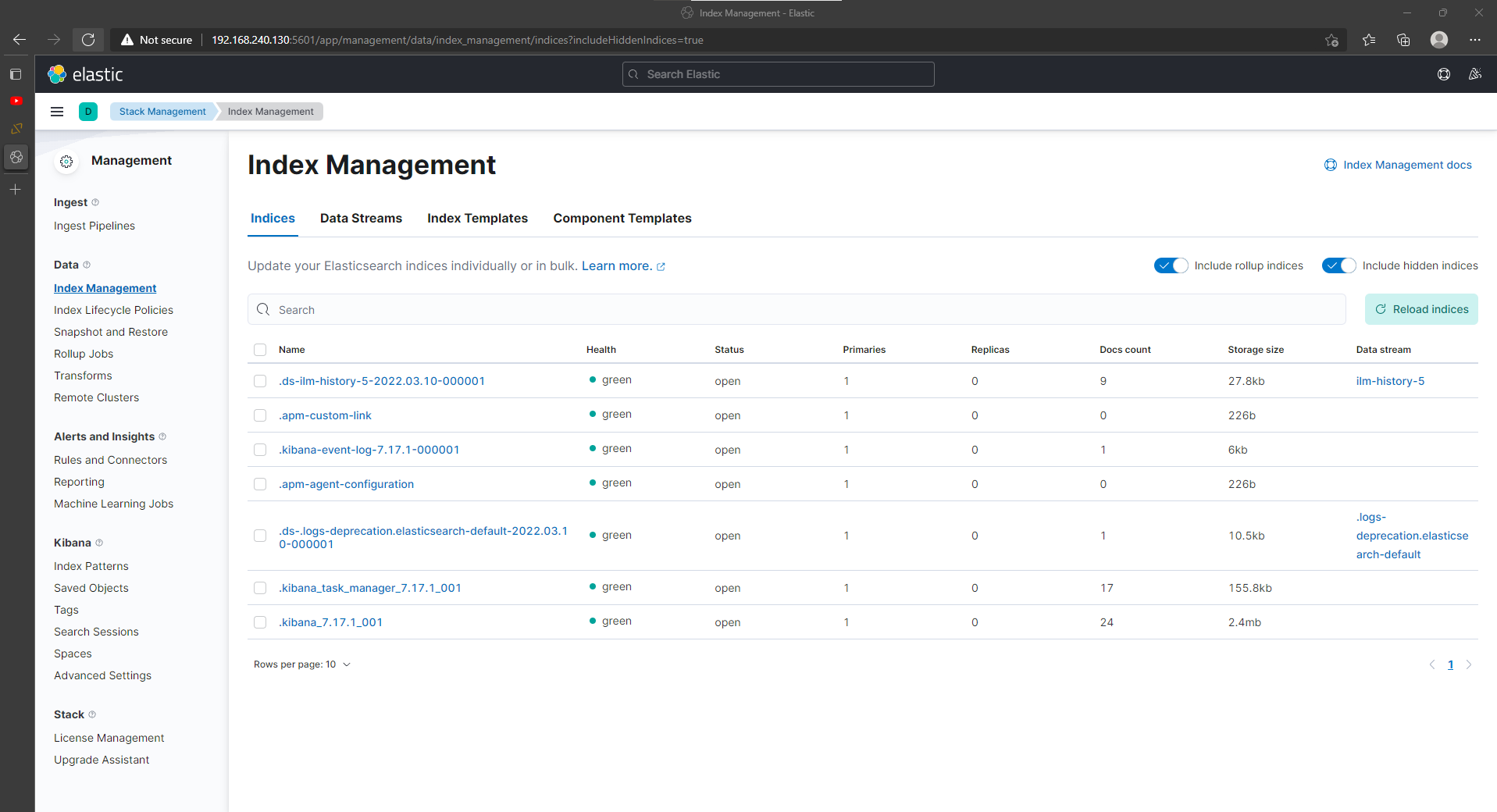
* Cài đặt Beats/Filebeat CentOS :
* Thực hiện lệnh:



* Kết quả:



* Xem log trong Kibana:

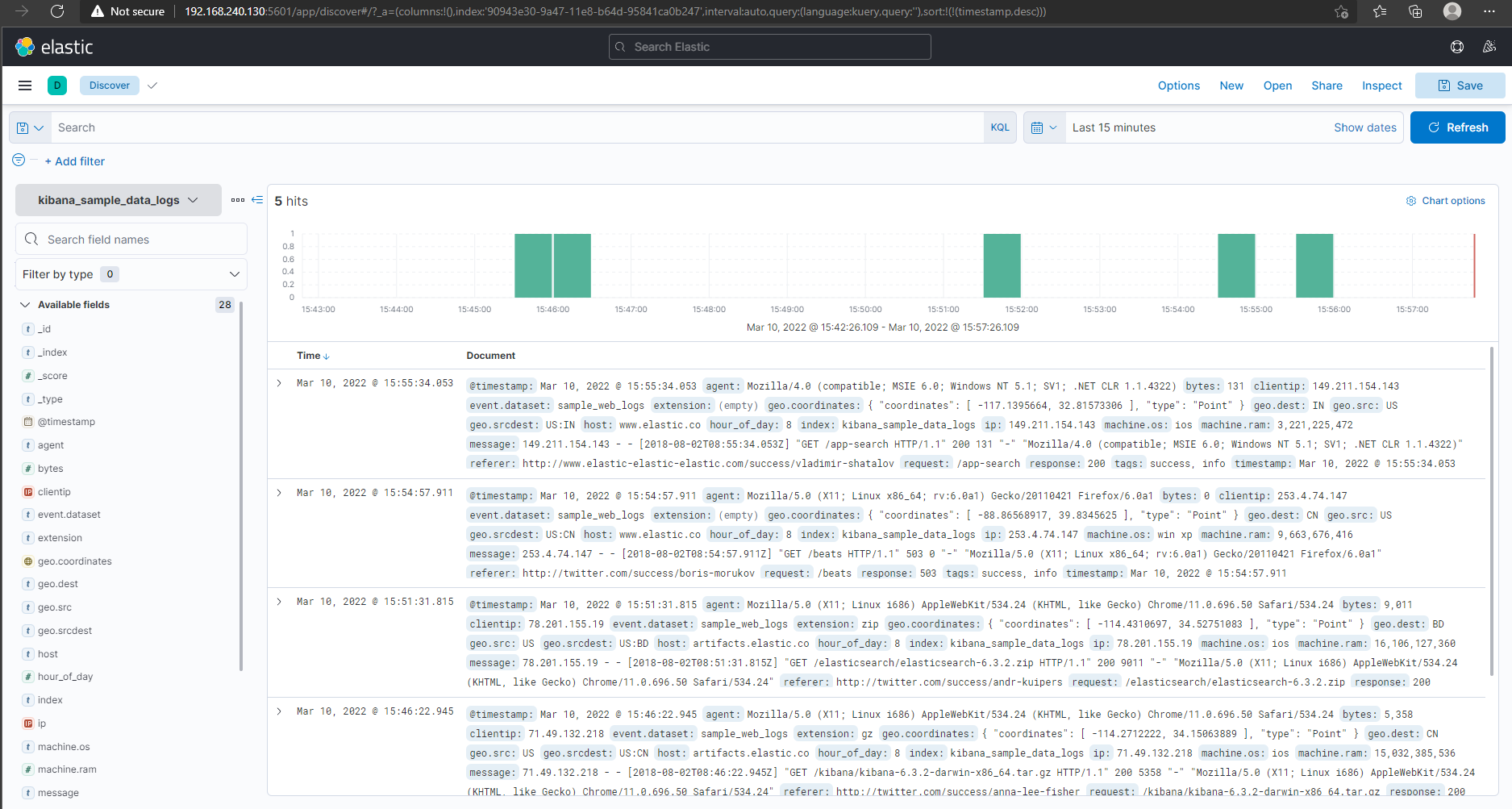


Để truy vấn bằng Kibana ta sẽ tạo các **Index patterns**, đó là truy vấn thông tin các index có tiền tố là **filebeat-**, nhấn vào **Index patterns** của Kibana, bấm vào **Create index pattern**

Điền **filebeat-\*** vào **index pattern**, rồi nhấn **Next Step**

Chọn **@timestamp** ở mục **Time Filter field name,** rồi nhấn **Create Index Pattern**

Sau đó chọn **Discover** để xem thông tin về các logs.

****

# C, Kết Luận:

Báo cáo đã thực hiện được những công việc sau:

Giới thiệu khái quát về công cụ xử lý log ELK (Elastic Search, Logtash và Kibana):

* + 1. Kiến trúc cơ bản của ELK
    2. Các thành phần của ELK thì gồm Elastic Search, Logtash, Kibana và có thêm cả Beat
    3. Các tính năng của ELK thì phổ biến là tìm kiếm và phân tích, lưu trữ dữ liệu, thu thập và cấu hình log, trực quan hóa dữ liệu, bảo mật giám sát và quản lý.
    4. Cài đặt và thử nghiệm phân tích được 1 số file log cơ bản.

Trên đây là báo cáo của nhóm G7, nội dung nhóm em hoàn thành còn nhiều thiếu xót. Chúng em mong nhận được những góp ý chỉnh sửa từ thầy để hoàn thiện nội dung đã tìm hiểu và học hỏi thêm.

Nhóm G7 trân thành cảm ơn thầy!

### **Tài liệu tham khảo:**

<http://infosecptit.com/at-hdh/INT1484-Bai%20giang%20An%20toan%20he%20dieu%20hanh.pdf>

<https://www.daihockhonggiay.com/blogs/post/gioi-thieu-ve-elk-stack>