**CÔNG TY CỔ PHẦN CHỨNG KHOÁN BOS**



# TÀI LIỆU TRIỂN KHAI HỆ THỐNG LƯU TRỮ NHẬT KÝ HỆ THỐNG ELK

# (Elasticsearch – Logstash – Kibana)

1. **Lịch sử sửa đổi tài liệu**

| Version | Ngày | Người viết | Ghi chú |
| --- | --- | --- | --- |
| 1.0 | 02/09/2019 | Bùi Công Hoàn | Bản khởi tạo |

1. **Mục lục:**

Tổng quan về hệ thống lưu trữ nhật ký tập chung ELK

Sơ đồ kết nối và các thiết bị cần chuẩn bị

Cài đặt và cấu hình hệ thống lưu trữ nhật ký tập chung ELK

1. Nội Dung Chính
2. Tổng quan về hệ thống lưu trữ nhật ký tập chung ELK

ELK Stack là tập hợp 3 phần mềm đi chung với nhau, phục vụ cho công việc logging. Ba phần mềm này lần lượt là:

* Elasticsearch: Cơ sở dữ liệu để lưu trữ, tìm kiếm và query log.
* Logstash: Tiếp nhận log từ nhiều nguồn, sau đó xử lý log và ghi dữ liệu vào Elasticsearch.
* Kibana: Giao diện để quản lý, thống kê log. Đọc thông tin từ Elasticsearch.

Điểm mạnh của ELK là khả năng thu thập, hiển thị, truy vấn theo thời gian thực. Có thể đáp ứng truy vấn một lượng dữ liệu cực lớn.

Hình dưới đây là cơ chế hoạt động của ELK stack :

A picture containing object

Description automatically generated

* Đầu tiên, log sẽ được đưa đến Logstash. (Thông qua nhiều con đường, ví dụ như server gửi UDP request chứa log tới URL của Logstash, hoặc Beat đọc file log và gửi lên Logstash).
* Logstash sẽ đọc những log này, thêm những thông tin như thời gian, IP, parse dữ liệu từ log (server nào, độ nghiêm trọng, nội dung log) ra, sau đó ghi xuống database là Elasticsearch.
* Khi muốn xem log, người dùng vào URL của Kibana. Kibana sẽ đọc thông tin log trong Elasticsearch, hiển thị lên giao diện cho người dùng query và xử lý.

Các thành phần trong ELK Stack.

* 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).
  + 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.
  + Việc tìm kiếm trong một lượng ít dữ liệu rất dễ dàng, nhưng nếu có 1 tỷ dòng dữ liệu thì thế nào? Elasticsearch cho phép bạn có cái nhìn để khai thác khuynh hướng và các mẫu trong dữ liệu.
  + 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:
    - Khả năng co giãn, Độ sẵn sàng cao.
    - Dự đoán trước, tin cậy.
    - Đơn giản, trong suốt

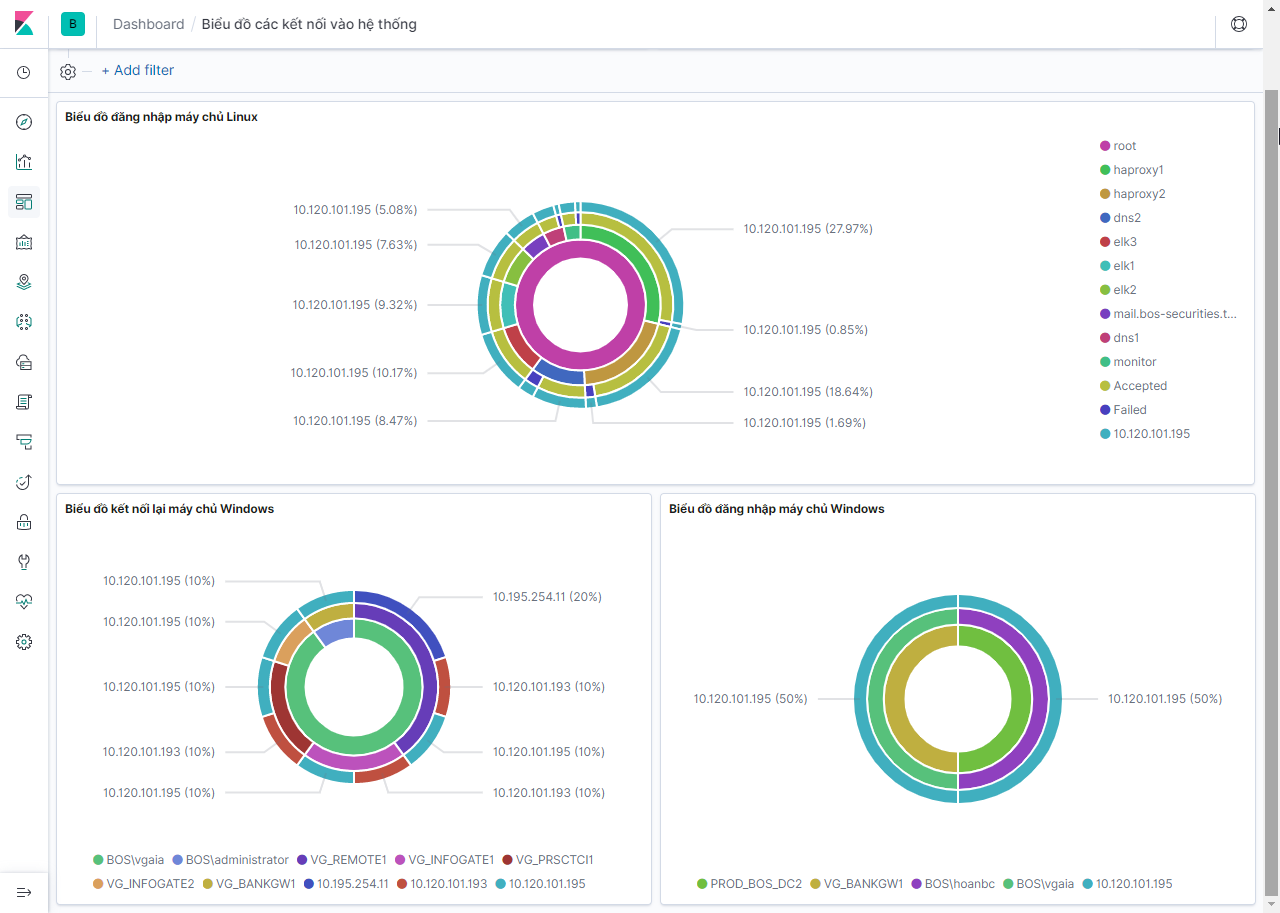
Các vai trò trên Elasticsearch

M: Master Node chịu trách nghiệm cho các hoạt động quản lý tạo, xóa Index, sắp xếp share, thêm hoặc xóa một Node Elasticsearch khỏi Cluster, tại một thời điểm chỉ có một Node trong cụm có quyền Master

D: Data Node chịu trách nghiệm cho việc lưu trữ dữ liệu ở các Shared và thực hiện các hoạt động liên quan đến dữ liệu. Tại một thời điểm có thể có nhiều Data Node trong một cụm

I: Ingest Node sẽ hỗ trợ bạn thực hiện hoạt động xử lý các documents trước khi quá trình index bắt đầu

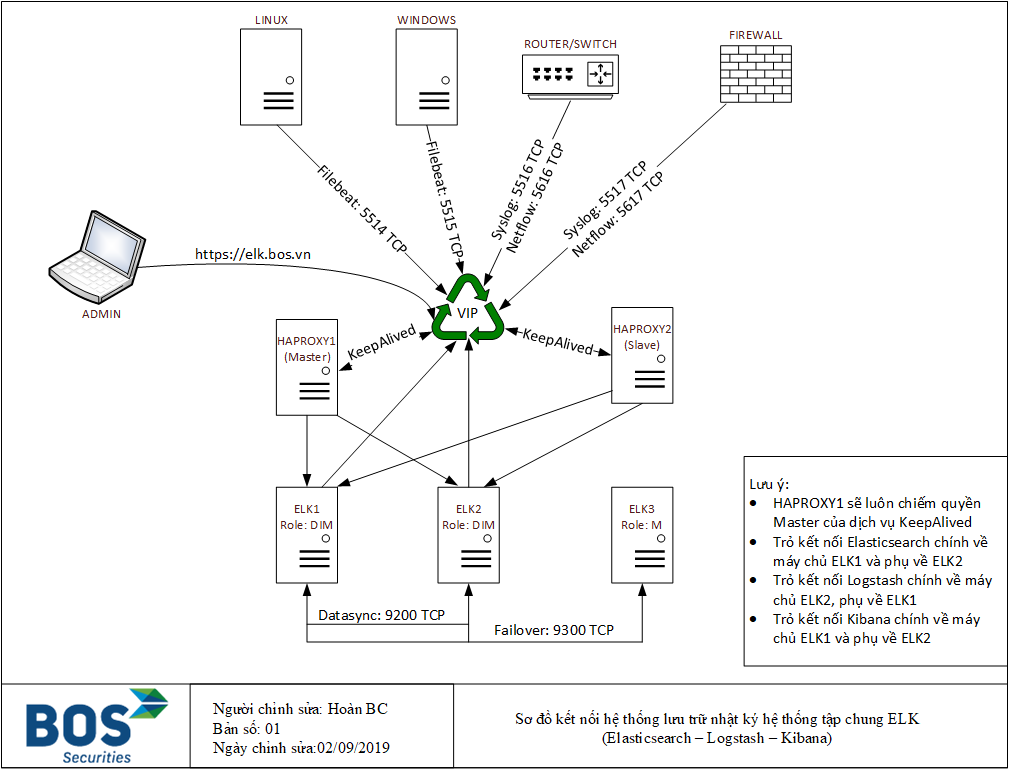
* Logstash.
  + Logstash có chức năng phân tích cú pháp của các dòng dữ liệu. Việc phân tích làm cho dữ liệu đầu vào ở một dạng khó đọc, chưa có nhãn thành một dạng dữ liệu có cấu trúc, được gán nhãn.
  + Khi cấu hình Logstash luôn có 3 phần: Input, Filter, Output.
  + Bình thường khi làm việc với Logstash, sẽ phải làm việc với Filter nhiều nhất. Filter hiện tại sử dụng Grok để phân tích dữ liệu.
* Kibana.
  + 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: theo các dạng biểu đồ.



Các trường hợp sử dụng ELK stack

* Với các hệ thống hoặc ứng dụng nhỏ, ta không cần sử dụng ELK stack làm gì, cứ dùng thư viện ghi log đi kèm với ngôn ngữ, sau đó ghi log ra file rồi đọc bình thường.
* Tuy nhiên, với những hệ thống lớn nhiều người dùng, có nhiều service phân tán (microservice), có nhiều server chạy cùng lúc… thì việc ghi log xuống file không còn hiệu quả nữa. Lúc này số lượng máy chủ trên hệ thống là lớn và nhiều do đó chúng ta không thể dùng cách thủ công là remote vào từng máy rồi đọc log của từng server được, lúc này ELK stack sẽ giải quyết vấn đề đó. ELK stack sẽ ghi log tập chung vào một chỗ khiến chúng ta có thể dễ dàng quản lý log trên toàn hệ thống.

1. Sơ đồ kết nối và các thiết bị cần chuẩn bị
   1. Sơ đồ kết nối



* 1. Các thiết bị cần chuẩn bị

Danh sách máy chủ

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| STT | Địa chỉ máy chủ | Tên máy chủ | Hệ điều hành | CPU/RAM | DISK | Ghi chú |
| 1 | 10.200.9.55 | Haproxy1 | CentOS 7 Minimal | 2 Core / 4GB | 40GB | VIP: 10.200.9.51 |
| 2 | 10.200.9.56 | Haproxy2 | CentOS 7 Minimal | 2 Core / 4GB | 40GB |
| 3 | 10.200.9.57 | Elk1 | CentOS 7 Minimal | 4 Core / 8GB | 300GB | Role: DIM |
| 4 | 10.200.9.58 | Elk2 | CentOS 7 Minimal | 4 Core / 8GB | 300GB | Role: DIM |
| 5 | 10.200.9.59 | Elk3 | CentOS 7 Minimal | 2 Core /1.5GB | 40GB | Role: M |

Cổng kết nối

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Địa chỉ nguồn | Địa chỉ đích | Cổng |
| 1 | Any | 10.200.9.51 | 5514, 5515, 5516, 5517 |
| 2 | 10.200.9.57  10.200.9.58 | 9200 |
| 3 | 10.200.9.57  10.200.9.58  10.200.9.59 | 10.200.9.57  10.200.9.58  10.200.9.59 | 9200,9300 |
| 4 | 10.200.9.55  10.200.9.56 | 10.200.9.57  10.200.9.58 | 5514, 5515, 5516, 5517, 5601, 9200 |

Chỉnh sửa file /etc/hosts trên các máy chủ ELK

10.200.9.57 elk1

10.200.9.58 elk2

10.200.9.59 elk3

Cài đặt và cấu hình hệ thống lưu trữ nhật ký tập chung ELK

1. Cài đặt hệ thống
   * 1. Cài đặt KeepAlived trên máy chủ HAPROXY1, HAPROXY2

yum install -y gcc kernel-headers kernel-devel keepalived

Cấu hình KeepAlived trên Máy chủ HAPROXY1

* + 1. Cài đặt và cấu hình dịch vụ Haproxy trên máy chủ HAPROXY1, HAPROXY2
    2. Cài đặt JAVA trên máy chủ ELK1, ELK2, ELK3
* Tải Java SE Development Kit 11với định dạng nén GZ về máy chủ

Giải nén thư mục cài đặt JAVA

tar xzf jdk-11.0.4\_linux-x64\_bin.tar.gz

mv jdk-11.0.4 /opt/

Chọn JAVA mặc định cho hệ thống

cd /opt/jdk-11.0.4/

alternatives --install /usr/bin/java java /opt/jdk-11.0.4/bin/java 2

alternatives --config java

There are 1 programs which provide 'java'.

Selection Command

-----------------------------------------------

\* 1 /opt/jdk-11.0.4/bin/java

Enter to keep the current selection[+], or type selection number: 1

Thay thế đường dẫn cho Java

alternatives --install /usr/bin/jar jar /opt/jdk-11.0.4/bin/jar 2

alternatives --install /usr/bin/javac javac /opt/jdk-11.0.4/bin/javac 2

alternatives --set jar /opt/jdk-11.0.4/bin/jar

alternatives --set javac /opt/jdk-11.0.4/bin/javac

Kiểm tra phiên bản Java đã cài

[root@elk]# java -version

java version "11.0.4" 2019-07-16 LTS

Java(TM) SE Runtime Environment 18.9 (build 11.0.4+10-LTS)

Java HotSpot(TM) 64-Bit Server VM 18.9 (build 11.0.4+10-LTS, mixed mode)

Thiết lập các biến môi trường Java

export JAVA\_HOME=/opt/jdk-11.0.4

export JRE\_HOME=/opt/jdk-11.0.4/jre

export PATH=$PATH:/opt/jdk-11.0.4/bin:/opt/jdk-11.0.4/jre/bin

* + 1. Cài đặt Elasticsearch Cluster trên máy chủ ELK1, ELK2 và ELK3
* Tải signing key:

rpm --import <https://artifacts.elastic.co/GPG-KEY-elasticsearch>

* Thêm thư viện Repo

vi /etc/yum.repos.d/elasticsearch.repo

[elasticsearch-7.x]

name=Elasticsearch repository for 7.x packages

baseurl=https://artifacts.elastic.co/packages/7.x/yum

gpgcheck=1

gpgkey=https://artifacts.elastic.co/GPG-KEY-elasticsearch

enabled=1

autorefresh=1

type=rpm-md

* Cài đặt

sudo yum install -y elasticsearch

* Cấu hình

Trên máy chủ ELK1 chỉnh sửa file elasticsearch.yml tại /etc/elasticsearch với nội dung như sau:

# ---------------------------------- Cluster -----------------------------------

cluster.name: BOS-ELK

node.master: true

node.data: true

node.ingest: true

# ------------------------------------ Node ------------------------------------

node.name: node-1

# ----------------------------------- Paths ------------------------------------

path.data: /var/lib/elasticsearch

path.logs: /var/log/elasticsearch

# ----------------------------------- Memory -----------------------------------

# ---------------------------------- Network -----------------------------------

network.host: elk1

http.port: 9200

transport.host: elk1

transport.tcp.port: 9300

# --------------------------------- Discovery ----------------------------------

discovery.zen.ping.unicast.hosts: ["elk1", "elk2", "elk3"]

discovery.zen.minimum\_master\_nodes: 2

discovery.zen.ping\_timeout: 30s

Trên máy chủ ELK2 chỉnh sửa file elasticsearch.yml tại /etc/elasticsearch với nội dung như sau:

# ---------------------------------- Cluster -----------------------------------

cluster.name: BOS-ELK

node.master: true

node.data: true

node.ingest: true

# ------------------------------------ Node ------------------------------------

node.name: node-2

# ----------------------------------- Paths ------------------------------------

path.data: /var/lib/elasticsearch

path.logs: /var/log/elasticsearch

# ----------------------------------- Memory -----------------------------------

# ---------------------------------- Network -----------------------------------

network.host: elk2

http.port: 9200

transport.host: elk2

transport.tcp.port: 9300

# --------------------------------- Discovery ----------------------------------

discovery.zen.ping.unicast.hosts: ["elk1", "elk2", "elk3"]

discovery.zen.minimum\_master\_nodes: 2

discovery.zen.ping\_timeout: 30s

Trên máy chủ ELK3 chỉnh sửa file elasticsearch.yml tại /etc/elasticsearch với nội dung như sau:

# ---------------------------------- Cluster -----------------------------------

cluster.name: BOS-ELK

node.master: true

node.data: false

node.ingest: false

# ------------------------------------ Node ------------------------------------

node.name: node-3

# ----------------------------------- Paths ------------------------------------

path.data: /var/lib/elasticsearch

path.logs: /var/log/elasticsearch

# ----------------------------------- Memory -----------------------------------

# ---------------------------------- Network -----------------------------------

network.host: elk3

http.port: 9200

transport.host: elk3

transport.tcp.port: 9300

# --------------------------------- Discovery ----------------------------------

discovery.zen.ping.unicast.hosts: ["elk1", "elk2", "elk3"]

discovery.zen.minimum\_master\_nodes: 2

discovery.zen.ping\_timeout: 30s

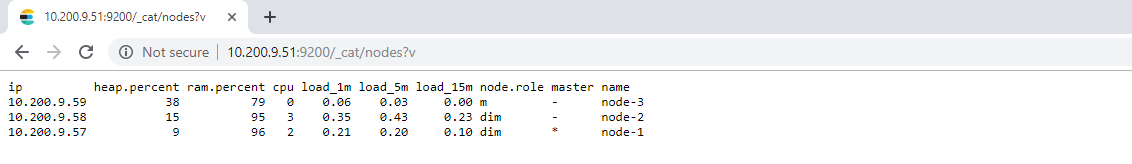
Tiền hành start dịch vụ Elasticsearch theo thứ tự từ ELK1 tới ELK2 và ELK3 bằng lệnh

Systemctl start elasticsearch

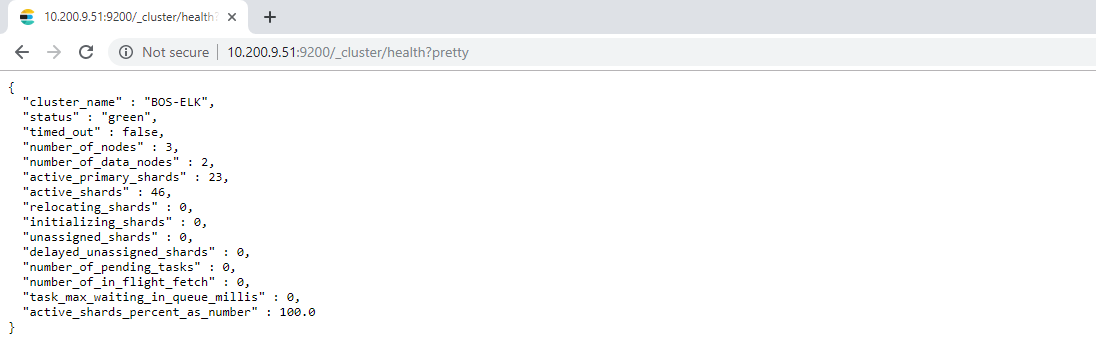
Khởi động dịch vụ Elasticsearch cùng với hệ điều hành

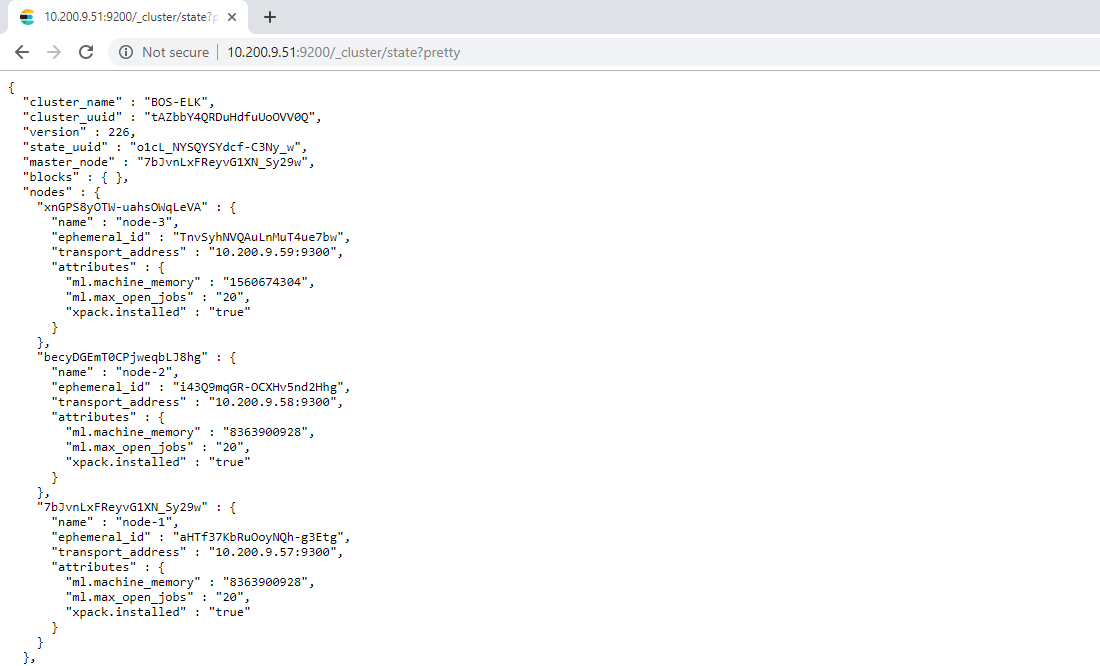
Systemctl enable elasticsearch

Tiến hành kiểm tra trạng thái của Elasticsearch Cluster bằng các lệnh sau:

<http://10.200.9.51:9200/_cat/nodes>

<http://10.200.9.51:9200/_cluster/health?pretty>



<http://10.200.9.51:9200/_cluster/state?pretty>

* + 1. Cài đặt Kibana trên máy chủ ELK1, ELK2
* Tải signing key

rpm --import <https://artifacts.elastic.co/GPG-KEY-elasticsearch>

* Thêm thư viện Repo

Vi /etc/yum.repos.d/kibana.repo

[kibana-7.x]

name=Kibana repository for 7.x packages

baseurl=https://artifacts.elastic.co/packages/7.x/yum

gpgcheck=1

gpgkey=https://artifacts.elastic.co/GPG-KEY-elasticsearch

enabled=1

autorefresh=1

type=rpm-md

* Cài đặt Kibana

sudo yum install -y kibana

* Cấu hình Kibana

Trên máy chủ ELK1 chỉnh sửa file kibana.yml tại /etc/kibana với nội dung như sau:

server.port: 5601

server.host: "10.200.9.57"

server.name: "elk1"

elasticsearch.hosts: ["http://10.200.9.51:9200"]

Trên máy chủ ELK2 chỉnh sửa file kibana.yml tại /etc/kibana với nội dung như sau:

server.port: 5601

server.host: "10.200.9.58"

server.name: "elk2"

elasticsearch.hosts: ["http://10.200.9.51:9200"]

Tiền hành start dịch vụ Kibana theo thứ tự từ ELK1 tới ELK2 bằng lệnh

Systemctl start kibana

Khởi động dịch vụ Elasticsearch cùng với hệ điều hành

Systemctl enable kibana

* + 1. Cài đặt Logstash trên máy chủ ELK1, ELK2
* Tải signing key

rpm --import <https://artifacts.elastic.co/GPG-KEY-elasticsearch>

* Thêm thư viện Repo

vi /etc/yum.repos.d/ logstash.repo

[logstash-7.x]

name=Elastic repository for 7.x packages

baseurl=https://artifacts.elastic.co/packages/7.x/yum

gpgcheck=1

gpgkey=https://artifacts.elastic.co/GPG-KEY-elasticsearch

enabled=1

autorefresh=1

type=rpm-md

Cài đặt Logstash

sudo yum install -y logstash

Tiền hành start dịch vụ Logstash theo thứ tự từ ELK1 tới ELK2 bằng lệnh

Systemctl start logstash

Khởi động dịch vụ Logstash cùng với hệ điều hành

Systemctl enable logstash

* + 1. Kiểm tra tính sẵn sàng cao của hệ thống
  1. Cấu hình cơ bản cho hệ thống
     1. Tạo Index Pattents
     2. Tạo Visualizations
     3. Tạo Dashboard