HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG
KHOA ĐÀO TẠO QUỐC TẾ VÀ SAU ĐẠI HỌC

BÀI TẬP MÔN AN TOÀN THÔNG TIN

ĐỀ TÀI: XÂY DỰNG HỆ THỐNG GIÁM SÁT LOG TẬP TRUNG THÔNG QUA BỘ CÔNG CỤ MÃ NGUỒN MỞ ELK

GVHD: PGS.TSKH. Hoàng Đăng Hải.
Lớp: M17CQCS01-B
Nhóm 2: **Nguyễn Hà Ly - Cao Quốc kiên**
Hoàng Minh Đức - Bùi TRần Tiến.

Hà Nội 06/2018

Mục lục

I. Đặt vấn đề 3

1. Mở đầu 3

2. Tổng quan về giám sát log tập trung 3

II. Phương pháp triển khai 6

1. Giới thiệu về ELK 6

2. Đặc điểm của ELK 7

3. Cấu trúc Logical 7

4. Nguyên lý hoạt động 9

III. Kết quả và nhận xét đánh giá 10

1. Yêu cầu chung 10

2. Các bước triển khai 10

3. Kết luận và đánh giá 14

IV. Tài liệu tham khảo 14

I. Đặt vấn đề

1. Mở đầu

Hệ thống giám sát an toàn trong các hệ thống mạng hiện nay đóng vai trò quan trọng, không thể thiếu trong hạ tầng công nghệ thông tin (CNTT) của các cơ quan, đơn vị, tổ chức. Hệ thống này cho phép thu thập, chuẩn hóa, lưu trữ và phân tích tương quan toàn bộ các sự kiện an toàn mạng được sinh ra trong hệ thống CNTT của tổ chức. Trong đó log là một thành phần cực kỳ hữu hiệu cho việc giám sát cũng như khắc phục các sự cố trong hệ thống mạng. Mỗi khi có sự cố xảy ra – và thực sự nó rất thường xuyên xảy ra, thì những thông tin này trở nên vô cùng quý giá trong việc khắc phục. Các thông tin trên thường được ghi lại dưới dạng văn bản, được gọi là cái file log. Hiện nay các hệ thống mạng thường dùng các giải pháp log nội bộ, bản thân một số máy chủ windows và linux đều có hệ thống ghi log của chính nó. Tuy nhiên các hệ thống ghi log này phải dùng dòng lệnh, điều này rất khó khăn trong việc sử dụng với người ít kinh nghiệm. Các tính năng cảnh báo qua chat, email,… phải tự tích hợp, viết nhiều lần. Cùng với việc chưa biến được các dữ liệu thô từ bản ghi log thành các thông tin dễ đọc, dễ hiểu. Chưa kể đến, với những hệ thống lớn, chạy nhiều ứng dụng, lượng truy cập cao thì công việc phân tích log thực sự là một điều vô cùng khó khăn.

Vậy nên các giải pháp log tập trung bắt đầu ra đời. Các server, bằng những cách khác nhau, sẽ đẩy các file log tại máy local tập trung về một máy log server. Các giải pháp log tập trung không chỉ giúp người quản trị có thể quản lý log của các máy client một cách dễ dàng hơn. Mà còn giúp người quản trị khai thác tối đa được lợi ích từ các file log.

Xuất phát từ thực tế trên, Nhóm thực hiện xin lựa chọn đề tài ***"Xây dựng hệ thống giám sát log tập trung thông qua bộ công cụ mã nguồn mở ELK"*** để trợ giúp người quản trị làm tốt hơn công việc giám sát hệ thống mạng.

2. Tổng quan về giám sát log tập trung

2. 1. Các thành phần cơ bản trong hệ thống mạng

Để một hệ thống mạng hoạt động tốt nó bao gồm rất nhiều thành phần, hoạt động trên các nền tảng và mỗi trường khác nhau:

- Các máy trạm
- Các máy chủ
- Các thiết bị hạ tầng mạng: Router, switch, Hub…

- Các thiết bị, hệ thống phát hiện và phòng chống xâm nhập: IDS/IPS, Snort, FireWall…

2. 2. Giám sát hệ thống mạng

Giám sát hệ thống mạng là việc sử dụng một hệ thống để liên tục theo dõi một mạng máy tính, xem xét tình trạng hoạt động của các thiết bị bên trong mạng máy tính đó, báo lại cho quản trị viên trường hợp mạng không hoạt động hoặc có các sự cố khác (tắc ngẽn, sập,...).

Một hệ thống giám sát gồm có nhiều thành phần: Máy trinh sát (Sensor), Máy thu thập (Collector), Cơ sở dữ liệu trung tâm và Công cụ phân tích (Analysis tool). Mỗi một thành phần bao gồm các chức năng riêng, cùng các phương pháp thu thập, phân tích và liệt kê nhằm đảm bảo đánh giá và phản hồi sự kiện xảy ra trong hệ thống mạng một cách nhanh chóng và chính xác nhất.

2. 3. Tổng quan về log

a) Khái niệm về log

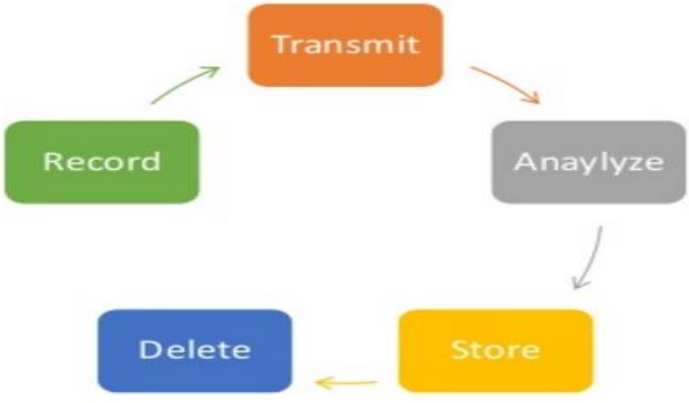
Log ghi lại liên tục các thông báo về hoạt động của cả hệ thống hoặc của các dịch vụ được triển khai trên hệ thống và file tương ứng. Log file thường là các file văn bản thông thường dưới dạng “clear text”, có thể dễ dàng đọc được bằng các trình soạn thảo văn bản (vi, vim, nano...) hoặc các trình xem văn bản thông thường (cat, tailf, head...) là có thể xem được file log.

Các file log có thể cung cấp các thông tin cần biết, để giải quyết các vấn đề với các ứng dụng, tiến trình được ghi vào log.

Tóm lại:

Log = Thời điểm + Dữ liệu.

Log ghi lại những hoạt động của hệ thống.



Hình 1.1 Vòng đời của chung của Log

Một vòng đời của Log bao gồm 5 bước chính được minh họa trong hình 1.2 cụ thể là:

- Đầu tiên log sẽ được ghi lại tại chính máy local sau đó nó sẽ được vận chuyển sang máy chủ quản lý log.
- Người quản trị mạng sẽ từ những bản ghi đó mà tiến hành phân tích, từ đó có thể giám sát được hoạt động của các máy client.
- Qua bước phân tích này mà người quản trị có thể phát hiện các hoạt động, hành vi xâm nhập không được phép.
- Sau khi phân tích, dữ liệu log sẽ được lưu trữ để sử dụng lại nếu cần.
- Bước cuối cùng là xóa, thường những tập tin log không cần thiết có thể được xóa bởi người quản trị nhằm giảm bớt lượng thông tin log không cần thiết.

b) Phân tích log

Phân tích các log hoặc các chuỗi thống kê là một nghệ thuật của việc trích dẫn đầy đủ ý nghĩa thông tin và đưa ra kết luận về một trạng thái an toàn từ các bản ghi thống kê những sự việc được sản sinh từ các thiết bị. Phân tích log không phải là 1 khoa học, nhưng ngày nay, việc tin tưởng vào kỹ năng phân tích độc lập và trực quan cũng như tính chất may mắn trong việc phân tích log chất lượng cũng là một khái niệm khoa học. Định nghĩa việc phân tích log có thể nghe rất khô khan, nhưng quan trọng là rút ra một "Kết luận có ý nghĩa". Nhìn một cách đơn giản vào các file log không phải là phân tích, bởi vì hiếm có những cái gì ngoài những sự nhàm chán và dường như chẳng liên quan gì đến nhau. Trong trường hợp một thiết bị 1 người sử dụng với rất ít các hoạt động, tất cả những bản ghi log mà chưa được nhìn trước là rất ít nghi ngờ, nhưng trong thực tế lại không dễ dàng như vậy.

c) Công dụng của log

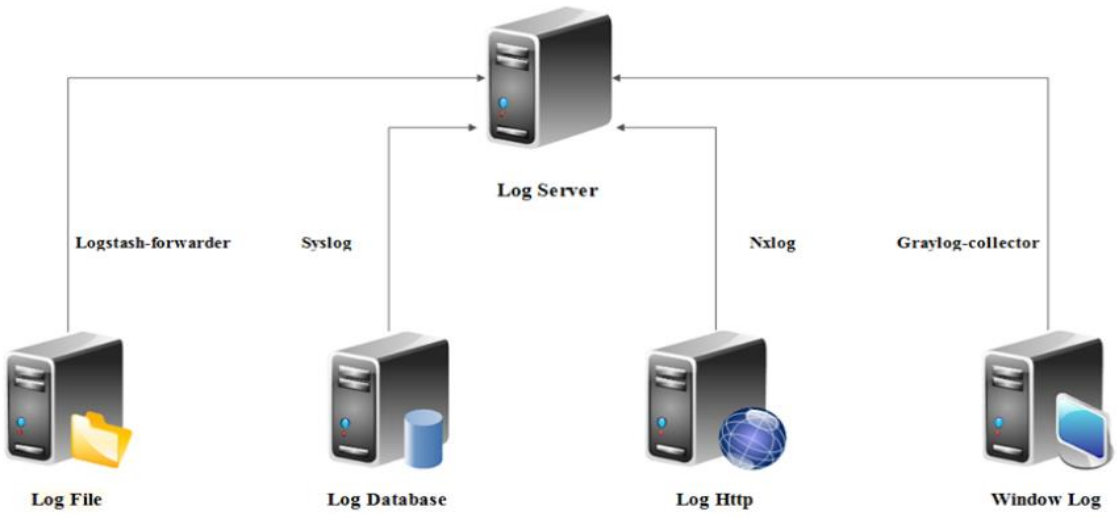
- Phân tích nguyên nhân khi có sự cố xảy ra.
- Giúp cho việc khắc phục sự cố nhanh hơn khi hệ thống gặp vấn đề.
- Giúp cho việc phát hiện, dự đoán một vấn đề có thể xảy ra đối với hệ thống.

2. 4. Giám sát log tập trung

Giám sát log tập trung là quá trình tập trung, thu thập, phân tích... các log cần thiết từ nhiều nguồn khác nhau về một nơi an toàn để thuận lợi cho việc phân tích, theo dõi hệ thống.

Lợi ích của giám sát log tập trung:

- Giúp quản trị viên có cái nhìn chi tiết về hệ thống -> có định hướng tốt hơn về hướng giải quyết.
- Mọi hoạt động của hệ thống được ghi lại và lưu trữ ở một nơi an toàn (log server) → đảm bảo tính toàn vẹn phục vụ cho quá trình phân tích điều tra các cuộc tấn công vào hệ thống.
- Log tập trung kết hợp với các ứng dụng thu thập và phân tích log khác nữa giúp cho việc phân tích log trở nên thuận lợi hơn → giảm thiểu nguồn nhân lực.



Hình 1.2. Mô hình Log tập trung

II. Phương pháp triển khai

1. Giới thiệu về ELK

ELK là viết tắt của tập hợp 3 phần mềm đi chung với nhau, phục vụ cho công việc thu thập log. Ba phần mềm này lần lượt là Elasticsearch, Logstash và Kibana.

Hiện nay phiên bản mới nhất của ELK là 6.2. Đây là bộ công cụ quản lý log 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 và triển khai trong hệ thống mạng của họ.

2. Đặc điểm của ELK

- Triển khai và cài đặt dễ dàng.
- ELK có thể nhận log từ rất nhiều thiết bị khác nhau: log của các server Linux, Windows, thiết bị mạng như router, switch, firewall, các thiết bị lưu trữ như CEPH,...
- Sử dụng công cụ chuyên dụng để tìm kiếm là Elasticsearch, giúp việc tìm kiếm các bản tin log được dễ dàng, nhanh chóng và chính xác.
- Phân tích được các số liệu từ các file log thành dạng số liệu, biểu đồ thống kê, tổng hợp lại trong các dashboard của Kibana.
- Cơ chế cảnh báo rất đa dạng, phong phú.
- Quản lý được hầu hết các bài toán được đưa ra trong giám sát hệ thống mạng.

3. Cấu trúc Logical

Hình 2.1. Cấu trúc Logic của ELK

ELK có 3 thành phần chính:

- 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.

a) Elasticsearch

Đầu tiên cần hiểu ElasticSearch là một công cụ tìm kiếm cấp doanh nghiệp (enterprise-level search engine). Mục tiêu của nó là tạo ra một công cụ, nền tảng hay kỹ thuật tìm kiếm và phân tích trong thời gian thực (ý nói ở đây là nhanh chóng và chính xác), cũng như cách để nó có thể áp dụng hay triển khai một cách dễ dàng vào nguồn dữ liệu (data sources) khác nhau.

Nguồn dữ liệu nói ở trên trên bao gồm các cơ sở dữ liệu nổi tiếng như MS SQL, PostgreSQL, MySQL, ...

Một số đặc điểm về ElasticSearch:

- Elasticsearch là một search engine.
- Elasticsearch được xây dựng để hoạt động như một server cloud theo cơ chế của RESTful.
- Kế thừa và phát triển từ Lucene Apache.
- Phát triển bằng ngôn ngữ Java.
- Là phần mềm open-source được phát hành theo giất phép của Apache License.
- Tương tự như Solr (Apache).
- Elasticsearch có thể tích hợp được với tất cả các ứng dụng sử dụng các loại ngôn ngữ: Java, JavaScript, Groovy, .NET, PHP, Perl, Python, Ruby

Cơ chế hoạt động của ElasticSearch:

- Sở dĩ Elasticsearch được gọi là "search & analyze in real time" là vì nó có khả năng trả về kết quả tìm kiếm một cách nhanh chóng và chính xác trong một nguồn dữ liệu lớn (big data source).
- Elasticsearch không chỉ tìm kiếm được các nguồn cơ sở dữ liệu nổi tiếng như MySQL, MS SQL, PostgreSQL, mà nó có thể là văn bản (text), pdf, doc, ...

Theo như cách thông thường tìm kiếm trong cơ sở dữ liệu database đều biết thì có hai cách là:

- Cách 1: Lật từng trang để tìm kiếm (No index).
- Cách 2: Lật tới phần mục lục để tìm kiếm.

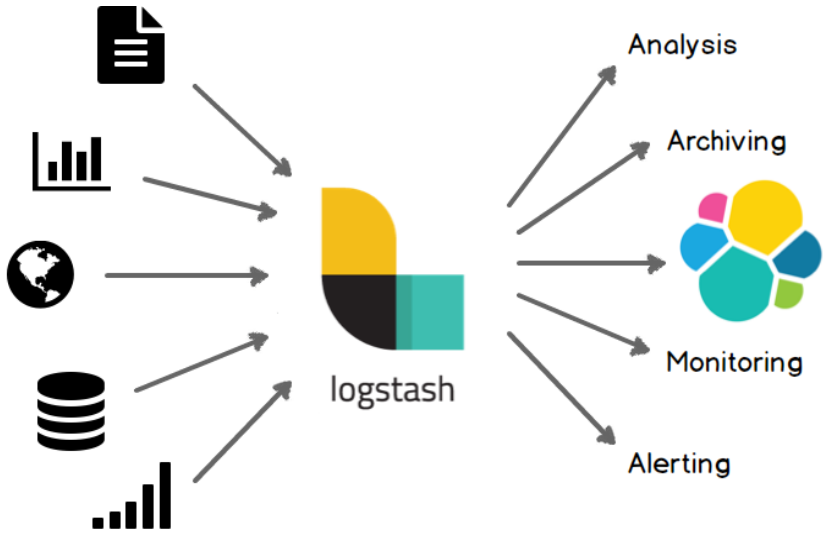
Về cơ bản thì ElasticSearch cũng áp dụng giải pháp giống Index. Tuy nhiên về mặt cơ chế xử lý và tìm kiếm thì có sự khác biệt, Index trong ElasticSearch được gọi là Inverted Index.

Inverted Index là kỹ thuật thay vì index theo từng đơn vị row (document) giống như mysql thì chúng ta sẽ biến thành Index theo đơn vị term. Cụ thể hơn, Inverted Index là một cấu trúc dữ liệu, nhằm mục đích map giữa term và các document chứa term đó.

b) Logstash

Logstash là một công cụ thu thập dữ liệu mã nguồn mở với khả năng pipelining thời gian thực. Logstash có thể tự động thu thập dữ liệu từ nhiều nguồn khác nhau và chuẩn hóa dữ liệu đó phụ thuộc vào đích đến của dữ liệu.

Ban đầu logstash chỉ đóng vai trò là một bộ thu thập log, nhưng khả năng của logstash hiện nay đã vượt qua cả vai trò đó. Bất kỳ một dạng sự kiện nào cũng đều có thể được logstash thu thập thông qua các plugins input và output, cùng với những code đã được đơn giản hóa giúp gia tăng khả năng nhập, xử lý và khai thác hiệu quả nhiều loại dữ liệu khác nhau.



Hình 2.2. Cấu trúc của Logstash

Logstash có một số lượng plugin đồ sộ (hơn 200) có thể đáp ứng bất kỳ dữ liệu nào được đưa đến đầu vào. Đơn giản nhất là log, metrics. Với web, logstash có thể biến các requests HTTP thành các sự kiện để phân tích. Hay có thể làm việc với NoSQL thông qua giao diện JDBC, cung cấp các cảm biến và IoT,…

c) Kibana

Kibana là một nền tảng phân tích và trực quan mã nguồn mở được thiết kế để làm việc với Elasticsearch. Ta sử dụng Kibana để tìm kiếm, xem và tương tác với dữ liệu được lưu trữ trong Elasticsearch. Từ đó dễ dàng thực hiện phân tích dữ liệu và trực quan hóa dữ liệu của mình thông qua biểu đồ, bảng.

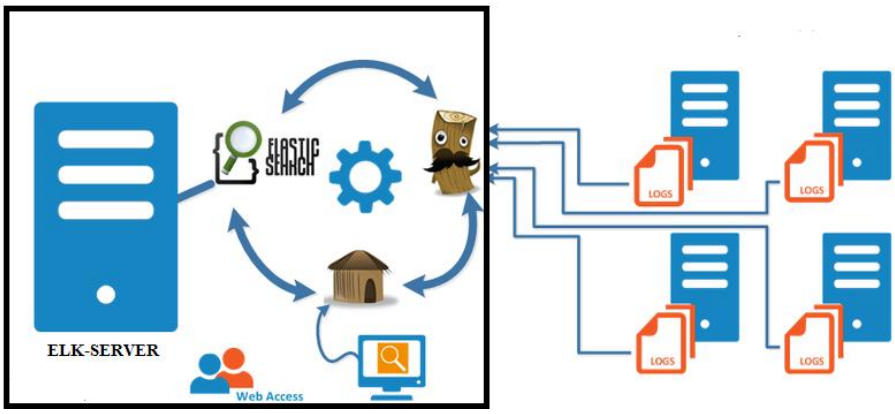
Kibana giúp nắm bắt nhanh chóng các dữ liệu có khối lượng lớn. Giao diện đơn giản, dựa vào trình duyệt cho phép nhanh chóng hiển thị các thay đổi khi truy vấn Elasticsearch trong thời gian thực.

4. Nguyên lý hoạt động

Hình 2.3. Nguyên lý hoạt động của ELK

1. 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)
2. *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*.
3. 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. Kibana hiển thị thông tin từ log cho người dùng.

III. Kết quả và nhận xét đánh giá



Hình 3.1. Mô hình triển khai ELK

1. Yêu cầu chung

Hệ thống bao gồm:

- 1 Server cài đặt ELK
- 1 Server cần thu thập log
- Cả 2 đều cài hệ điều hành CentOS7

Công cụ mô phỏng:

- Phần mềm Vmware

2. Các bước triển khai

Bước 1: Cài đặt hệ điều hành máy chủ CentOS7

Phần mềm ELK được cài đặt trên HĐH máy chủ Linux: CentOS-7 (64-bit). Các bước cài đặt HĐH CentOS-7 được triển khai bình thường.

Bước 2: Cấu hình các kết nối mạng (IP, Subnet, Gateway, DNS, Hostname)

- Cài đặt hostname

 # vi /etc/hostname

Đổi hostname thành bcvt.vn, cần *reboot* để server nhận hostname mới

- Cài đặt hosts

 # vi /etc/hosts

Thêm vào như sau

IP máy elk. bcvt.vn bcvt.vn

- Chỉnh sửa IP, Subnet, Gateway

 # vi /etc/sysconfig/network-scripts/ifccfg-ens33

 # systemctl restart network

Bước 3: Cập nhật hệ điều hành máy chủ, tắt SELinux, mở port firewall

- Cập nhật hệ điều hành máy chủ

 # yum update –y

- Tắt SELinux

 # vi /etc/sysconfig/selinux

Thay dòng "SELINUX=enforcing" thành "SELINUX=disabled"

- Firewall:

 # systemctl stop firewalld

 # systemctl disable firewalld

- Cài đặt Java hỗ trợ ELK

 # yum install java-1.8.0-openjdk.x86_64 -y

Bước 4: Cài đặt và cấu hình ELK

- Trước tiên các ta cần tải 3 file rpm này:

https://artifacts.elastic.co/downloads/elasticsearch/elasticsearch-6.2.2.rpm

https://artifacts.elastic.co/downloads/kibana/kibana-6.2.2-x86_64.rpm

https://artifacts.elastic.co/downloads/logstash/logstash-6.2.2.rpm

Sau đó sử dụng WinSCP, đưa 3 file vừa tải vào trong thư mục /opt

Rồi quay lại màn hình làm việc CentOS

- Cài đặt ELK

Ta vào trong thư mục /opt nơi vừa để 3 file ta tải và cài đặt ELK

 # cd /opt

 # yum localinstall elasticsearch-6.1.2.rpm kibana-6.1.2-x86_64.rpm logstash-6.1.2.rpm

Hệ điều hành sẽ tự động cài ELK.

- Cấu hình Elasticsearch

 # vi /etc/elasticsearch/elasticsearch.yml

Bỏ dấu # ở trước 2 dòng:

Network.host: localhost

http.port: 9200

 # systemctl start elasticsearch

 # systemctl enable elasticsearch

- Cấu hình Kibana

 # vi /etc/kibana/kibana.yml

Đổi thành: server.host: "0.0.0.0"

 # systemctl start kibana

 # systemctl enable kibana

Giờ ta có thể vào ELK thông qua giao diện Kibana qua: http://(IP máy):5601

Bước 5: Hướng dẫn đẩy syslog

- Cấu hình trên server ELK

```
# vim /etc/logstash/conf.d/linux.conf
input {
        tcp {
                port => 5000
                type => syslog
        }}
filter {}
output {
        elasticsearch {
                hosts => ["localhost:9200"]
        }}
```

 # systemctl restart logstash

- Cấu hình trên server cần lấy log

 # vi /etc/rsyslog.conf

Ta thêm: "." @@(ip server elk):5000 vào cuối file

 # setenforce 0

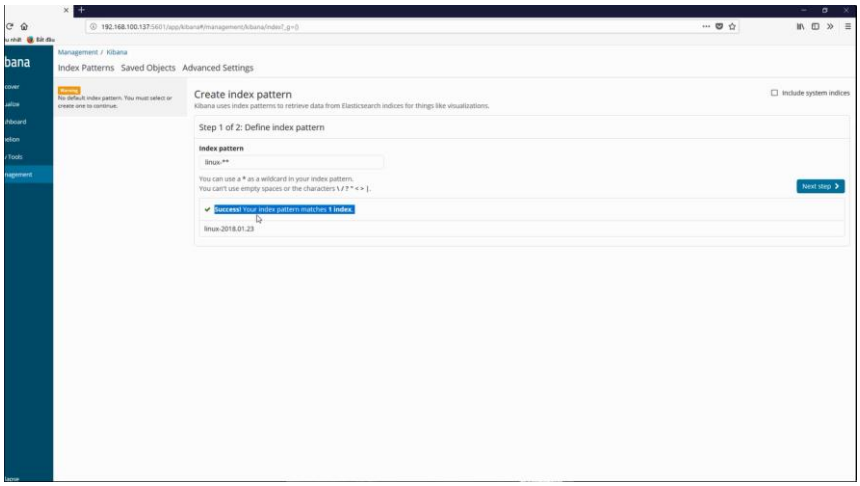
 # systemctl stop firewalld

 # systemctl disable firewalld

 # systemctl restart rsyslog

Vậy là ELK sẽ nhận được log của máy này

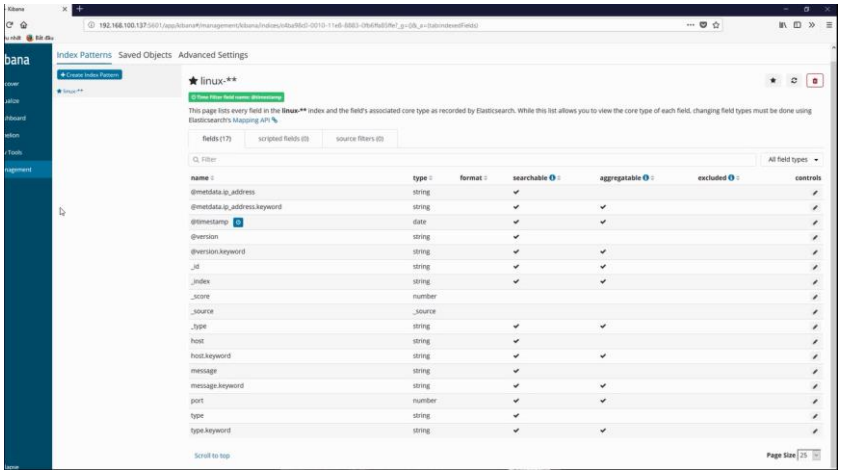
Trên Kibana ta cần nhập index (ở đây là tên file ta vừa tạo – linux)



Hình 3.2. Đẩy log ELK (1)

Chọn Time filter field name rồi ấn create index pattern

Đây là khi tạo thành công



Hình 3.3. Đẩy log ELK (2)

3. Kết luận và đánh giá

Đề tài đã nghiên cứu, làm chủ được công cụ ELK. Trình bày các lý thuyết về ELK, nguyên tắc hoạt động cũng như các mô hình triển khai, đồng thời xây dựng được mô hình mô phỏng khả năng ghi log cho hệ thống mạng của ELK. Từ kết quả của kịch bản thử nghiệm cho thấy hệ thống đáp ứng được yêu cầu đề ra của đề tài và đầy triển vọng ứng dụng để giải quyết các bài toán, có thể áp dụng cho nhiều các hệ thống mạng trong thực tế.

Trong phạm vi đề tài, báo cáo cơ bản đã đạt được các yêu cầu đặt ra. Tuy nhiên, các kết quả còn khá khiêm tốn. Trong thời gian tới, nếu có điều kiện em sẽ cố gắng phát triển thêm những nội dung sau:

- Áp dụng thu log được từ các thiết bị mạng như router, firewall, switch,…
- Xây dựng mô hình ứng dụng đầy đủ hơn và có thể phát triển cho doanh nghiệp.

IV. Tài liệu tham khảo

Tiếng Việt

[A] Hoàng Xuân Dậu (2007), Bài giảng an toàn bảo mật hệ thống thông tin, Học viện Công nghệ Bưu chính Viễn thông.

[B] http://tailieu.vn/tag/he-thong-giam-sat-mang.html

[C] https://www.wikipedia.org/

Tiếng Anh

[1] https://www.elastic.co/elk-stack/

[1.1] https://www.elastic.co/products