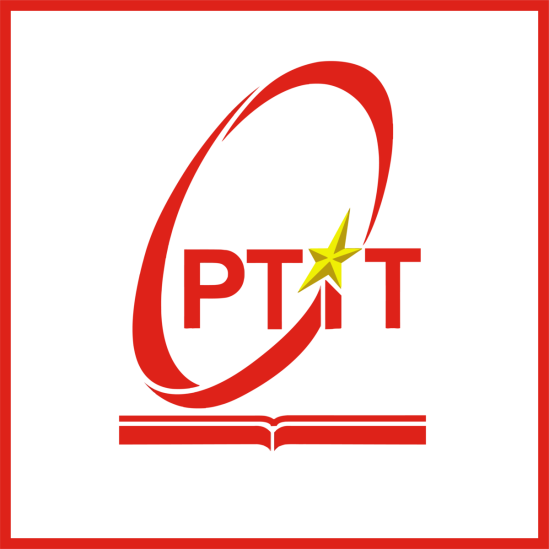
**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

**KHOA AN TOÀN THÔNG TIN**



Môn: An Toàn Hệ Điều Hành

**BÁO CÁO BÀI TẬP LỚN:**

**NHÓM 9:**

TÌM HIỂU VỀ SPLUNK VỀ THU THẬP VÀ PHÂN TÍCH DỮ LIỆU

Họ và tên sinh viên trong nhóm:

Đỗ Tiến Sĩ (B20DCAT153) Nhóm trưởng

Nguyễn Thị Quyên (B20DCAT152)

Đỗ Ngọc Sơn (B20DCAT154)

Trần Lê Cao Sơn (B20DCAT156)

Nguyễn Viết Nam (B20DCAT128)

Nguyễn Hà Thanh (B20DCAT173)

Họ và tên giảng viên:

Hoàng Xuân Dậu

Mục Lục

[I. Tổng quan về Splunk 3](#_Toc128861773)

[1. Khái niệm 3](#_Toc128861774)

[2. Tính năng 3](#_Toc128861775)

[II.Kiến trúc Splunk 5](#_Toc128861776)

[Kiến trúc của Splunk có 3 thành phần chính: 5](#_Toc128861777)

[**1.** **Splunk Forwarder.** 5](#_Toc128861778)

[**2.** **Splunk Search Head.** 6](#_Toc128861779)

[**3.** **Đánh giá** 9](#_Toc128861781)

[III. Cài Đặt 10](#_Toc128861782)

[**1.** **Cài đặt trên Window Server 2019** 10](#_Toc128861783)

[**2.** **Cài đặt trên Ubuntu 22.04(Linux)** 13](#_Toc128861784)

[IV. Cấu hình 15](#_Toc128861785)

[1. **Cấu hình input log từ syslog server vào splunk server (trên Linux)** 15](#_Toc128861786)

[V. Tổng kết 19](#_Toc128861787)

# I. Tổng quan về Splunk

## 1. Khái niệm

Splunk là hệ thống có thể captures, trích ra các dữ liệu thời gian thực có liên quan tới nhau từ đó có thể tạo ra các đồ thị, các báo cáo, các cảnh báo và các biểu đồ.

Mục đích của Splunk là giúp cho việc xác định mô hình dữ liệu và thu thập dữ liệu máy trên toàn hệ thống dễ dàng hơn. Nó cung cấp số liệu, chuẩn đoán các vấn đề xảy ra, phục vụ tốt cho hoạt động kinh doanh.

Splunk có thể tìm kiếm các sự kiện đã và đang xảy ra, đồng thời cũng có thể báo cáo và phân tích thống kê kết quả tìm được. Nó có thể nhập các dữ liệu của máy dưới dạng có cấu trúc hoặc không cấu trúc. Hoạt động tìm kiếm và phân tích sử dụng SPL (Search Processing Language), được tạo để quản lý Big Data. Do được phát triển từ Unix Piping và SQL nên Splunk có khả năng tìm kiếm dữ liệu, lọc, sửa đổi, chèn và xóa dữ liệu.

## 2. Tính năng

* **Định dạng Log:** Hỗ trợ hầu như tất cả các loại log của hệ thống, thiết bị hạ tầng mạng, phần mềm, Firewall, IDS/IPS, Log Event, Register của các máy trạm ….
* **Các hình thức thu thập dữ liệu:** Splunk có thể thực hiện việc thu thập log từ rất nhiều nguồn khác nhau. Từ một file hoặc thư mục (kể cả file nén) trên server, qua các kết nối UDP, TCP từ các Splunk Server khác trong mô hình Splunk phân tán, từ các Event Logs, Registry của Windows …Splunk kết hợp rất tốt với các công cụ thu thập log khác.
* **Cập nhật dữ liệu:** Splunk cập nhật dữ liệu liên tục khi có thay đổi trong thời gian thực. Giúp cho việc phát hiện và cảnh báo trong thời gian thực.
* **Đánh chỉ mục dữ liệu:** Splunk có thể đánh chỉ mục dữ liệu với một khối lượng dữ liệu rất lớn trong một khoảng thời gian ngắn. Giúp việc tìm kiếm diễn ra nhanh chóng và thuận tiện.

Diagram

Description automatically generated

Các loại data, log mà Splunk index được

* **Tìm kiếm thông tin:** Splunk làm việc rất tốt với dữ liệu lớn và cập nhật liên tục. Nó cung cấp cơ chế tìm kiếm với một “Splunk Language” cực kỳ thông minh bao gồm các từ khóa, các hàm và cấu trúc tìm kiếm giúp người sử dụng có thể truy xuất mọi thứ, theo rất nhiều tiêu chí từ tập dữ liệu rất lớn. Những nhà quản trị mạng cao cấp và chuyên nghiệp thường gọi Splunk với cái tên “Splunk toàn năng” hay “Splunk as Google for Log files” để nói lên sức mạnh của Splunk.
* **Giám sát và cảnh báo:** Splunk cung cấp cho người dùng một cơ chế cảnh báo dựa trên việc tìm kiếm các thông tin do chính người sử dụng đặt ra. Khi có vấn đề liên quan tới hệ thống phù hợp với các tiêu chí mà người dùng đã đặt ra thì hệ thống sẽ cảnh báo ngay tới người dùng (cảnh bảo trực tiếp qua giao diện, giử Email).
* **Khắc phục sự cố:** Splunk còn cung cấp một cơ chế tự động khắc phục với các vấn đề xảy ra bằng việc tự động chạy các file Script mà người dùng tự tạo (Ví dụ như: Chặn IP, dòng Port …) khi có các cảnh báo xảy ra.
* **Hiển thị thông tin:** Splunk cung cấp một cơ chế hiển thị rất trực quan giúp người sử dụng có thể dễ dàng hình dung về tình trạng của hệ thống, đưa ra các đánh giá về hệ thống. Splunk còn từ động kết xuất ra các báo cáo với nhiều loại định dạng một cách rất chuyên nghiệp.
* **Phát triển:** Cung cấp các API hỗ trợ việc tạo các ứng dụng trên Splunk của người dùng. Một số bộ API điển hình như Splunk SDK (cung cấp các SDK trên nền tảng Python, Java, JS, PHP), Shep (Splunk Hadoop Intergration - đây là sự kết hợp giữa Splunk và Hadoop), Shuttl (là một sản phẩm hỗ trợ việc sao lưu dữ liệu trong Splunk), Splunkgit (Giúp bạn hình dung dữ liệu tốt hơn), Splunk powershell resource Kit (Bộ công cụ hỗ trợ việc mở rộng và quản lý hệ thống).

# II.Kiến trúc Splunk

## Kiến trúc của Splunk có 3 thành phần chính:

* 1. **Splunk Forwarder.**

Nó liên quan đến việc chuyển tiếp dữ liệu. Thành phần này được sử dụng để thu thập tất cả dữ liệu của log được cài đặt trong dãy máy và dữ liệu log sẽ được chuyển tiếp tới bộ chỉ mục Splunk để lưu trữ và xử lý. Có hai loại Forwarder khác nhau:

* Universal Forwarder: không được xử lý mà chỉ xử lý dữ liệu thô được chuyển tiếp như hiện tại.
* Heavy Forwarder: Phân tích cú pháp và lập chỉ mục trên máy chủ tại nguồn.

Bộ lập chỉ mục lưu trữ và lập chỉ mục dữ liệu đến từ Forwarder. Phiên bản Splunk biến đổi dữ liệu đến thành các sự kiện và lưu trữ nó trong các chỉ mục để thực hiện các hoạt động tìm kiếm một cách hiệu quả. Nếu bạn đang nhận dữ liệu từ Universal Forwarder, thì trước tiên, trình lập chỉ mục sẽ phân tích cú pháp dữ liệu rồi lập chỉ mục cho nó. Phân tích cú pháp dữ liệu được thực hiện để loại bỏ dữ liệu không mong muốn. Tuy nhiên, nếu bạn đang nhận dữ liệu từ Heavy Forwarder, thì Indexer chỉ lọc chỉ mục cho dữ liệu. Indexer tạo các tệp ở nhiều định dạng và lưu trữ chúng trong các nhóm, chẳng hạn như:

* Nén dữ liệu thô
* Con trỏ chỉ mục và phía dữ liệu thô (. TSIDX files).
* Tệp siêu dữ liệu.

Splunk xử lý dữ liệu cho phép tìm kiếm và phân tích nhanh. Nó tăng cường dữ liệu theo nhiều cách khác nhau như:

* Tách luồng dữ liệu dữ liệu thành các sự kiện riêng lẻ, có thể tìm kiếm.
* Tạo hoặc xác định dấu thời gian.
* Trích xuất cá trường hợp máy chủ, nguồn và loại nguồn.
* Thực hiện các hành động người dùng xác định đối với dữ liệu đến, chẳng hạn như xác định các trường hợp tùy chỉnh, che dữ liệu nhạy cảm, viết khóa mới hoặc khóa đã sửa đổi, áp dụng quy tắc vi phạm cho các sự kiện nhiều dòng, lọ cá sự kiện không mong muốn và định tuyến các sự kiện đến các chỉ mục hoặc máy chủ được chỉ định. Quá trình lập chỉ mục này còn được gọi là sử lý sự kiện.

Một lợi ích khác của Splunk Indexer là sao chép dữ iệu. Bạn không cần lo lắng về việc mất dữ liệu vì Splunk giữ nhiều bản sao của dữ liệu được lập chủ mục. Quá trình này được gọi là sao chép chỉ mục hoặc phân cụm bộ chỉ mục.

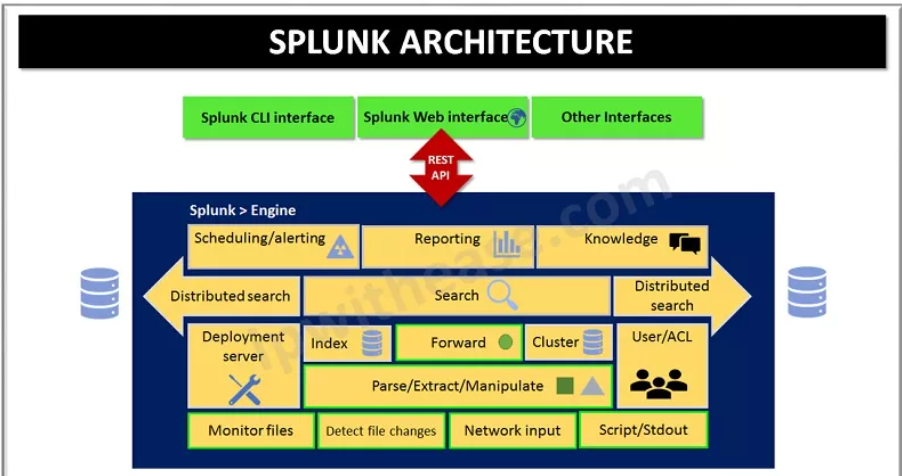
* 1. **Splunk Search Head.**

Đây là một GUI được sử dụng để tìm kiếm, phân tích và báo cáo các mục đích. Người dùng có thể thực hiện tìm kiếm và truy vấn dữ liệu Splunk và giao diện với chỉ mục để có quyền truy cập cho yêu cầu dữ liệu cụ thể.

Splunk Search Head có thể được cài đặt trên các máy chủ khác nhau và chỉ cần đảm bảo rằng dịch vụ Web Splunk đã bật trên máy chủ Splunk để các tương tác không bị dừng.

Có hai dạng Search Head:

* Search Head: Chính xác là giao diện người dùng nơi chỉ có thể truy xuất dữ liệu dựa trên các từ khóa và không có việc lập chỉ mục xảy ra với nó.
* Search Peer: là thứ mà nó có thể chứa cả kết quả tìm kiếm cũng như phục vụ cho việc lập chỉ mục.

****

* Mức thấp nhất của kiến trúc Splunk mô tả các phương thức nhập liệu khác nhau được hỗ trợ bởi Splunk. Những phương thức nhập này có thể được cấu hình để gửi dữ liệu trên các bộ phân loại Splunk.
* Trước khi dữ liệu đến được các bộ phân loại Splunk, nó có thể được phân tích cú pháp hoặc thao tác, có nghĩa là làm sạch dữ liệu có thể được thực hiện nếu cần.
* Một khi dữ liệu được lập chỉ mục trên Splunk, nó sẽ tiến hành đi vào cụ thể để phân tích dữ liệu.
* Splunk hỗ trợ hai loại triển khai: triển khai độc lập và triển khai phân tán. Tùy thuộc vào loại triển khai, tìm kiếm tương ứng được thực hiện. Công cụ Splunk có các thành phần bổ sung khác của quản lý dữ liệu, báo cáo và lên kế hoạch, và cảnh báo. Toàn bộ công cụ Splunk được tiếp xúc với người dùng thông qua Splunk CLI, Splunk Web Interface, và Splunk SDK, được hỗ trợ bởi hầu hết các ngôn ngữ.
* Splunk cài đặt một quy trình máy chủ phân tán trên máy chủ được gọi là splunkd. Quá trình này có trách nhiệm lập chỉ mục và xử lý một số lượng lớn dữ liệu thông qua các nguồn khác nhau. Splunk có khả năng xử lý số lượng lớn dữ liệu phát trực tuyến và lập chỉ mục cho phân tích thời gian thực trên một hoặc nhiều đường ống.
* Mỗi đường ống đơn bao gồm một loạt các bộ vi xử lý, dẫn đến xử lý dữ liệu nhanh hơn và hiệu quả hơn. Danh sách dưới đây là các khối kiến trúc splunk:
  + **Pipeline:** Đây là một quá trình cấu hình đơn luồng duy nhất nằm trong splunk.
  + **Bộ vi xử lý:** Chúng là những hàm số có thể tái sử dụng cá nhân hoạt động trên dữ liệu đi qua một đường ống. Đường ống trao đổi dữ liệu giữa họ thông qua một hàng đợi.

Diagram

Description automatically generated

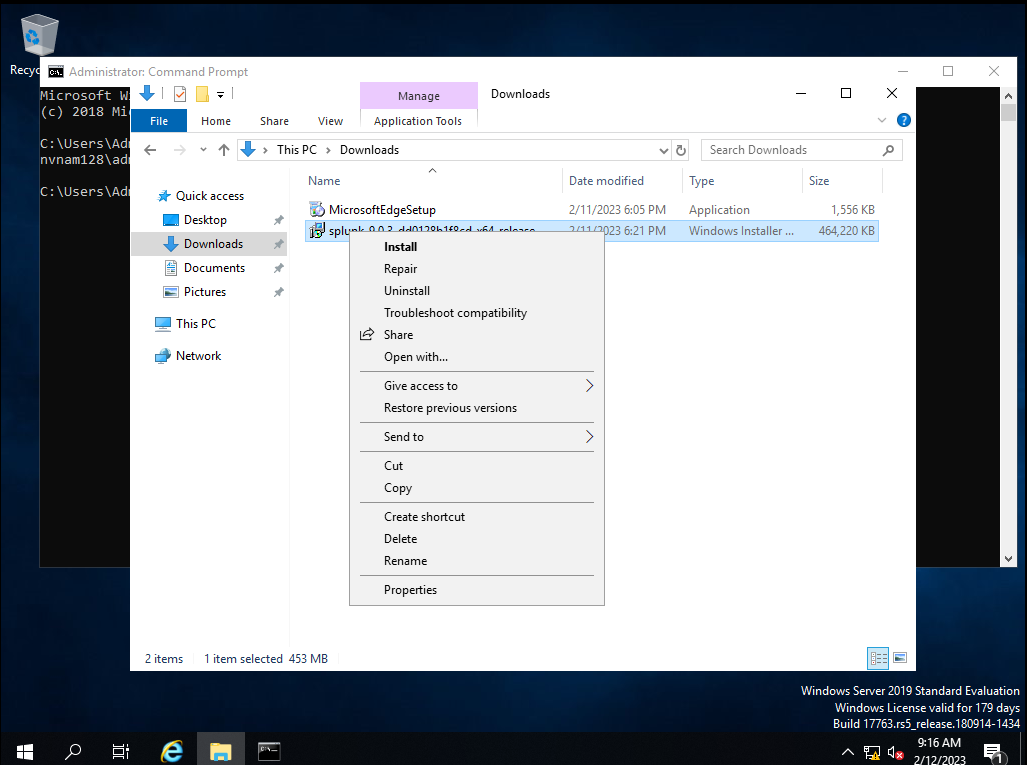
* Splunk cho phép người dùng tìm kiếm, điều hướng và quản lý dữ liệu trên Splunk Enterprise thông qua giao diện web được gọi là Splunk Web. Nó là một máy chủ ứng dụng web dựa trên Python cung cấp một giao diện web để sử dụng Splunk. Trong phiên bản trước của Splunk: splunkd và SplunkTeb là hai quy trình riêng biệt, nhưng từ Splunk 6, cả hai quy trình đã được tích hợp là 1. Nó cho phép người dùng tìm kiếm, phân tích và hình dung dữ liệu bằng cách sử dụng giao diện web. Giao diện Splunk Web có thể được truy cập bằng cách sử dụng cổng Web Splunk, và Splunk cũng cho thấy REST API để truyền thông thông qua cổng quản lý chia sẻ.
* Một trong những thành phần quan trọng của kiến ​​trúc của Splunk là kho dữ liệu. Nó có trách nhiệm nén và lưu trữ dữ liệu ban đầu (nguyên vẹn). Dữ liệu được lưu trữ trong các tệp Time Series Index (T SIDX). Một kho dữ liệu cũng bao gồm lưu trữ và lưu trữ dựa trên chính sách lưu giữ có thể cấu hình.
* Các triển khai của Splunk Enterprise có thể bao gồm từ việc triển khai các máy chủ đơn (có chỉ số vài gigabyte dữ liệu mỗi ngày và được truy cập bởi một vài người dùng đang tìm kiếm, phân tích và hình dung dữ liệu) tới các triển khai lớn của doanh nghiệp ở nhiều trung tâm dữ liệu, lập chỉ mục hàng trăm terabytes dữ liệu và tìm kiếm được thực hiện bởi hàng trăm người dùng. Splunk hỗ trợ truyền thông với một cá thể khác của một máy chủ Splunk thông qua TCP để chuyển tiếp dữ liệu từ một máy chủ Splunk sang một máy khác để lưu trữ dữ liệu và các yêu cầu phân phối và phân phối dữ liệu khác thông qua giao tiếp TCP Splunk-to-Splunk.
* Bundles là các thành phần của kiến trúc Splunk lưu trữ cấu hình dữ liệu đầu vào, tài khoản người dùng, ứng dụng Splunk, tiện ích và môi trường khác.
* Các mô-đun là những thành phần của kiến trúc Splunk được sử dụng để thêm các tính năng mới bằng cách sửa đổi hoặc tạo bộ xử lý và đường ống. Các mô-đun chỉ là các kịch bản tùy chỉnh và các phương pháp nhập dữ liệu hoặc phần mở rộng có thể thêm một tính năng mới hoặc sửa đổi các tính năng hiện có của Splunk.
  1. **Đánh giá**
* Splunk mạnh về khả năng phân tích và cảnh báo tuy nhiên nó lại không mạnh và không đảm bảo về việc thu thập và truyền tải log. Cụ thể là nó chưa có cơ chế bảo mật trên đường truyền, không phù hợp với những hệ thống đòi hỏi bảo mật cao.
* Chưa có cơ chế giúp tự động phát hiện ra các tấn công hay các vấn đề từ bên ngoài. Những điều này phụ thuộc vào kinh nghiệm sử dụng và vốn hiểu biết của người quản trị.
* Để triển khai được một hệ thống sử dụng Splunk hiệu quả chúng ta cũng cần có một hệ thống riêng, đây cũng là một trở ngại không nhỏ với các hệ thống có quy mô trung bình và nhỏ.

# III. Cài Đặt

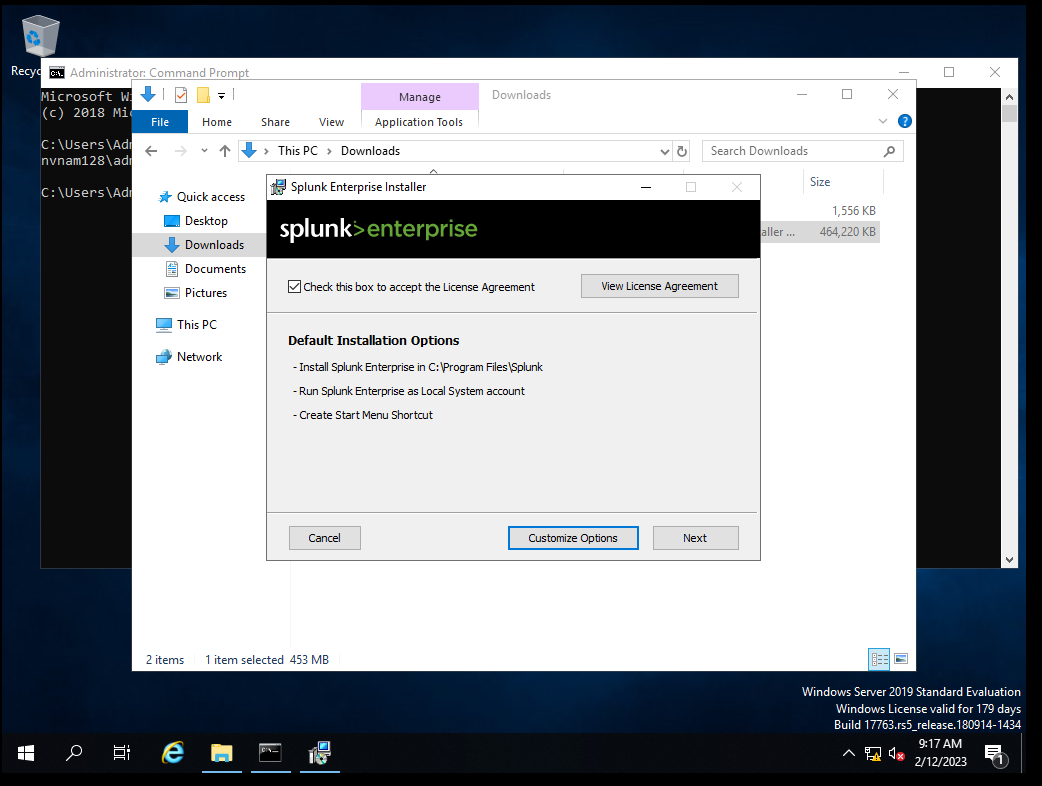
1. **Cài đặt trên Window Server 2019**

* Bước 1: tải phần mềm Splunk từ trang chủ của hãng
* Bước 2: Cài đặt:

Chọn Install

 Clicker vào Check this box to accept the license agreement.

Rồi Customize options.

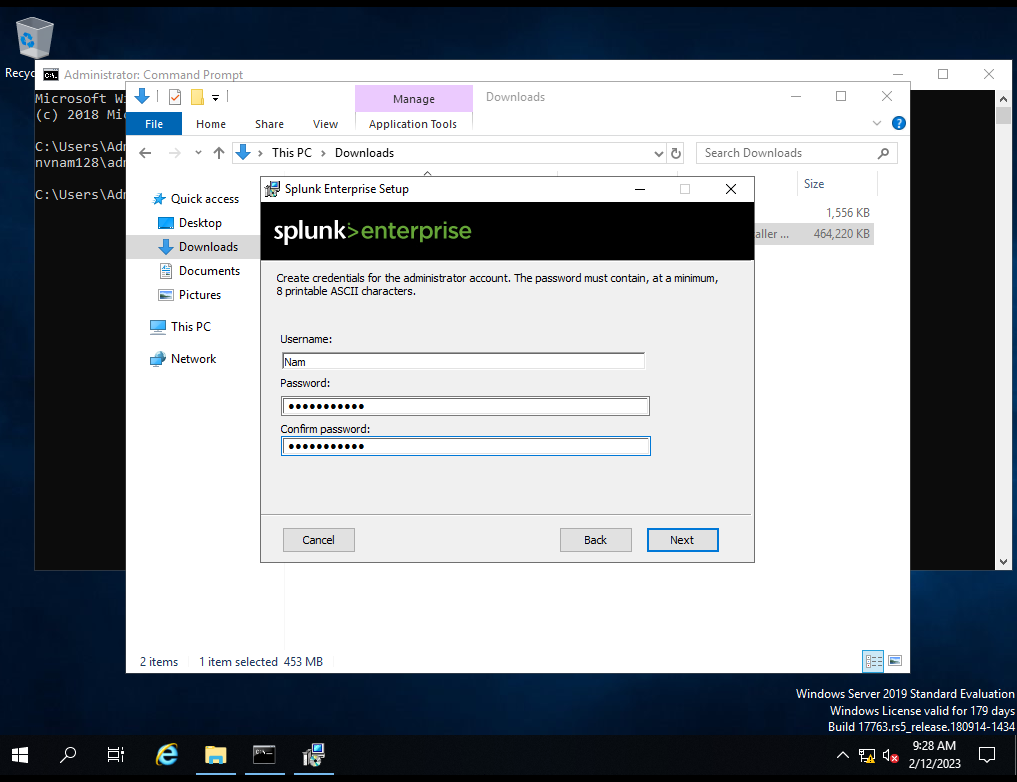


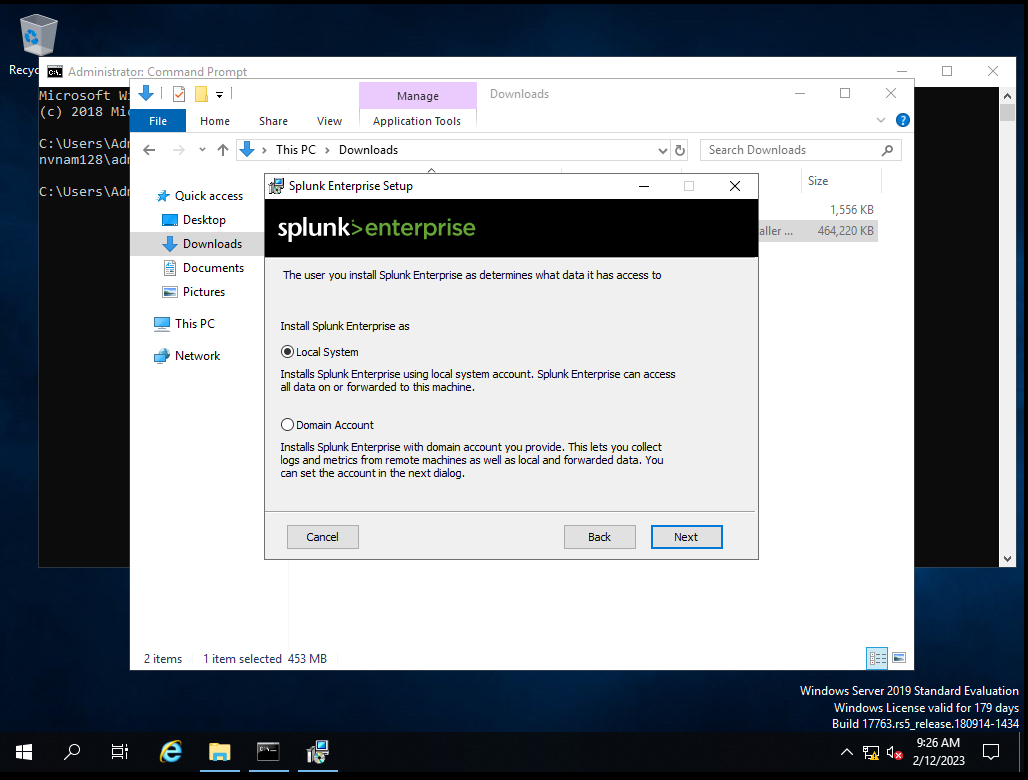
Thay đổi vị trí lưu Splunk (nếu cần) sau đó chọn Next.

Chọn local dùng cho mình

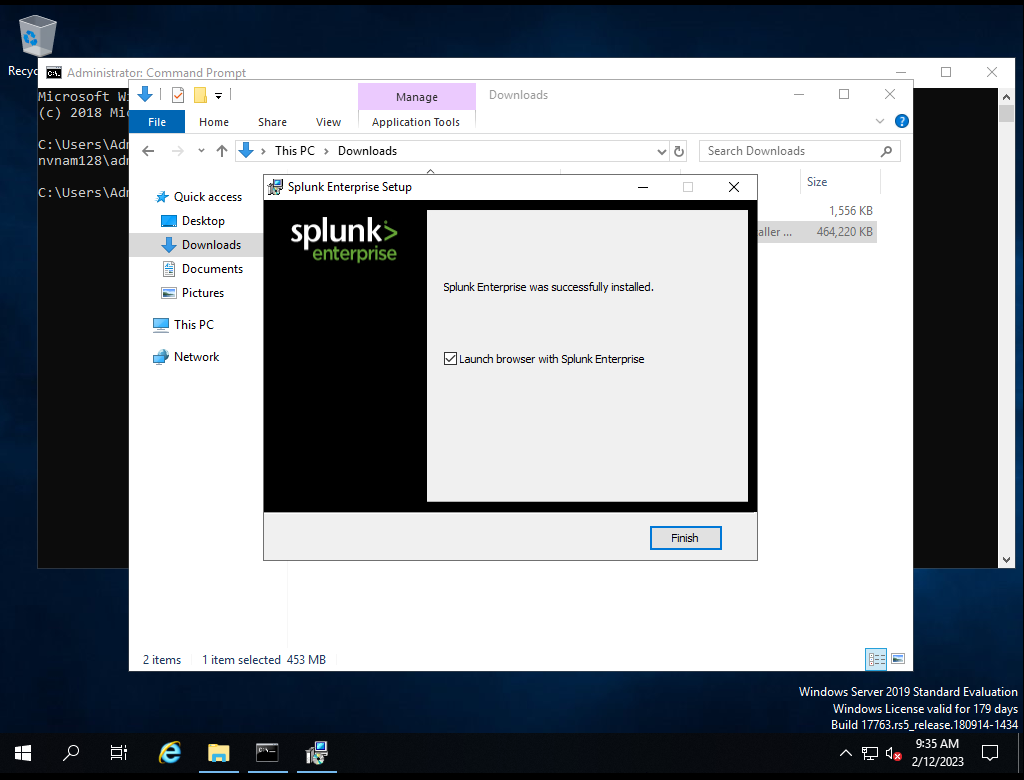
Còn dùng cho công ty hoặc doanh nghiệp chọn domain.

Ở đây mình chọn local -> ấn Next



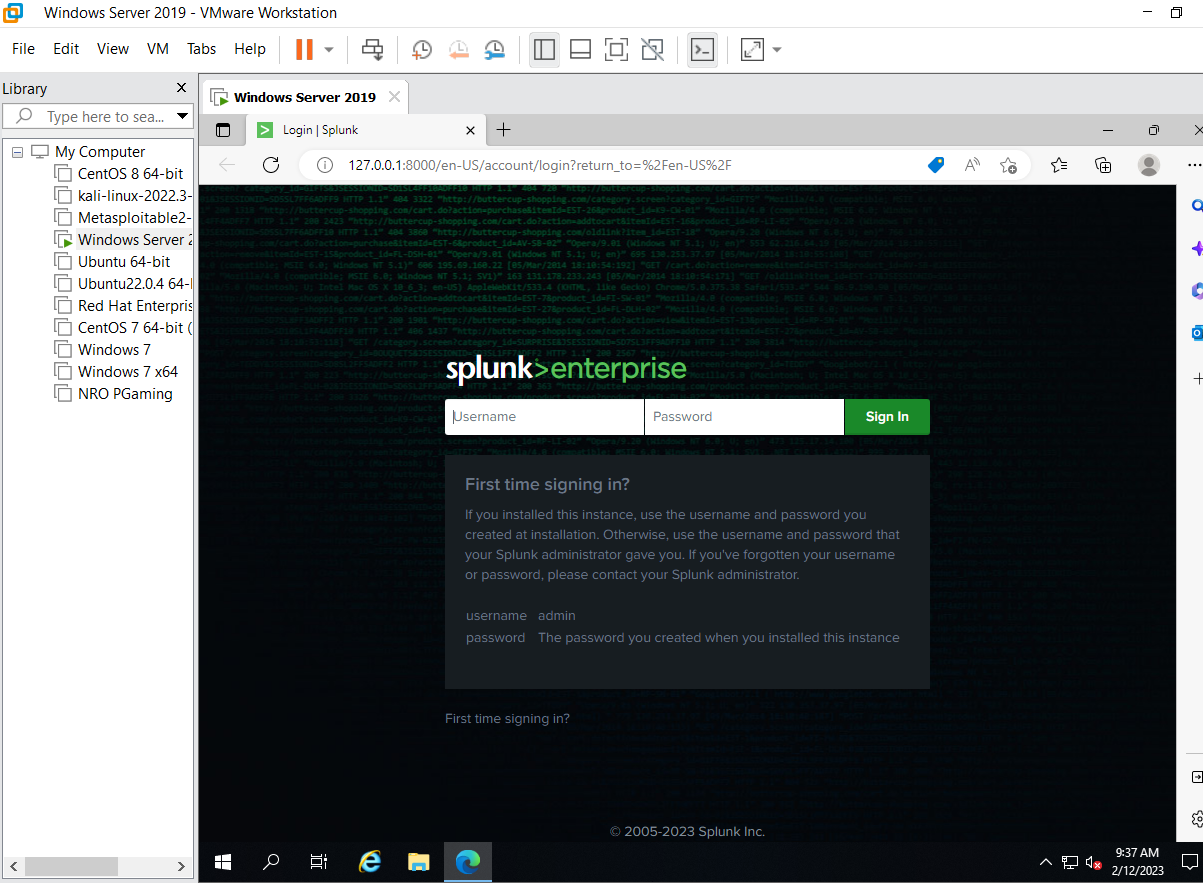


Cài đặt thành công -> Finish



Bây giờ chúng ta vô bằng ip với port 8000:

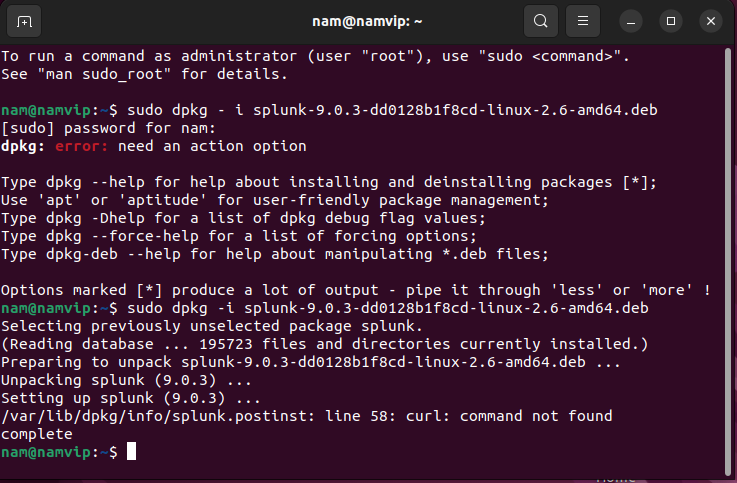
<http://localhost:8000>



1. **Cài đặt trên Ubuntu 22.04(Linux)**

* Bước 1: tài phiên bản thích hợp cho Ubuntu
* Bước 2: Cài đặt phiên bản deb

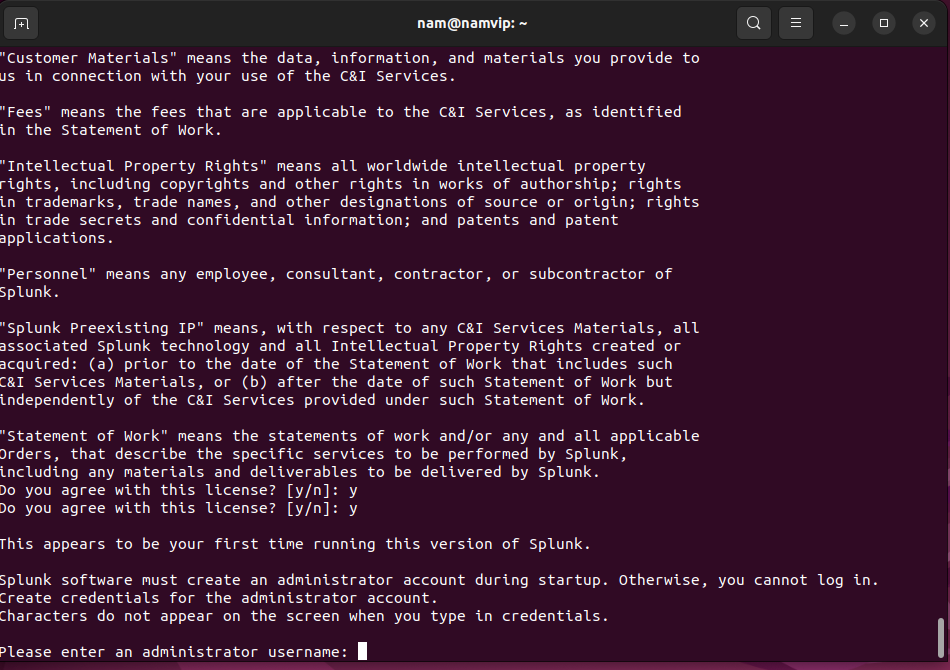
Trong CLI, nhập **dpkg -i *splunk\_package\_name*.deb(chạy bằng lệnh sudo): Sudo dpkg -i splunk-9.0.3-dd0128b1f8cd-linux-2.6-amd64.deb**

****

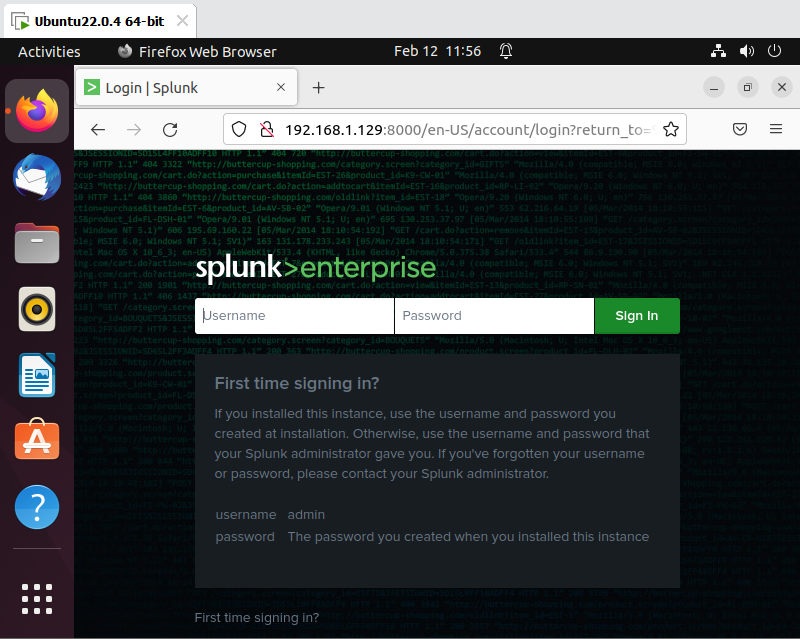
kích hoạt và khởi động lại splunk:

**sudo /opt/splunk/bin/splunk enable boot-start**

di chuyển xuống dưới và nhận y(2 lần)



Đăng nhập vào Splunk: Khi Splunk đã khởi động, Bạn có thể truy cập vào trang quản trị của nó từ trình duyệt bằng các nhập địa chỉ IP của máy chủ và cổng 8000. Ở đây IP máy chủ là 192.168.1.129 -> [**http://192.168.1.129:8000**](http://192.168.1.129:8000)



Cài đặt thành công.

# IV. Cấu hình

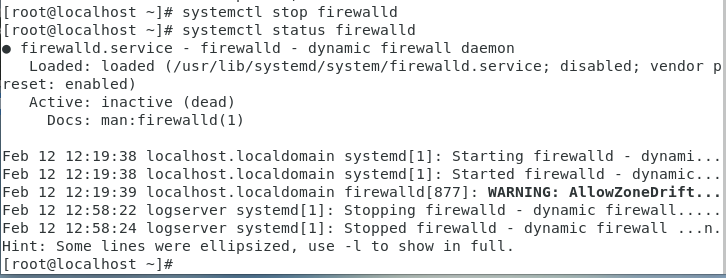
* 1. **Cấu hình input log từ syslog server vào splunk server (trên Linux)**

Chuẩn bị :

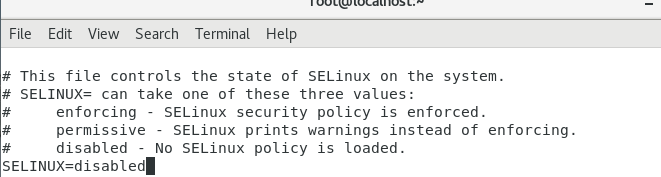
+1 máy centos hostname splunk.local , ip address 192.168.0.114 đóng vai trò là splunk server

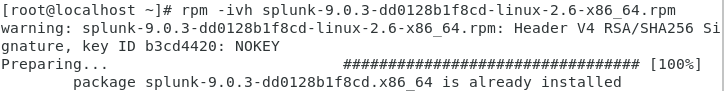
  
+1 máy centos client hostname là logserver, ip address 192.168.0.115 đóng vai trò là syslog server gửi log về cho splunk server.

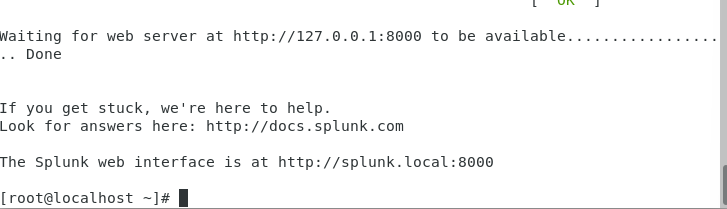
  
Tắt firewall, selinux trên cả 2 máy:  
+Tắt firewall : # systemctl stop firewalld

  
+Tắt Selinux: # vi /etc/sysconfig/selinux

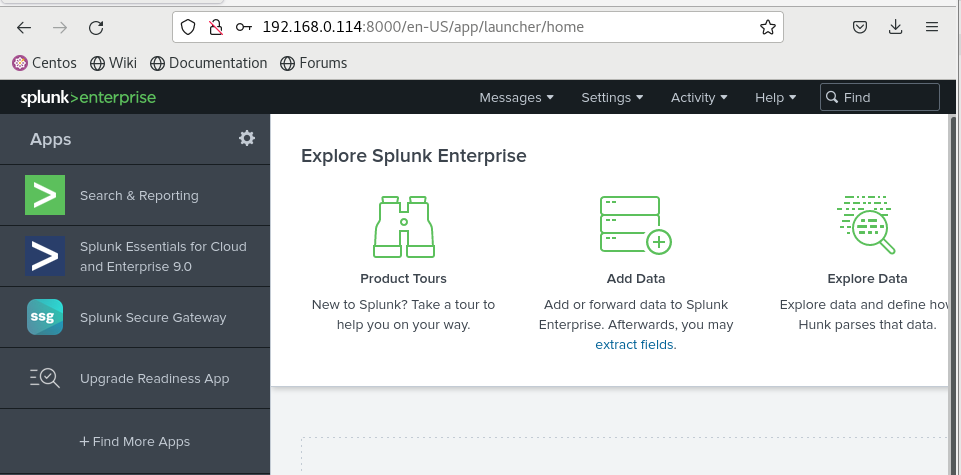
Sửa “SELINUX=enforcing” thành”SELINUX=disabled”

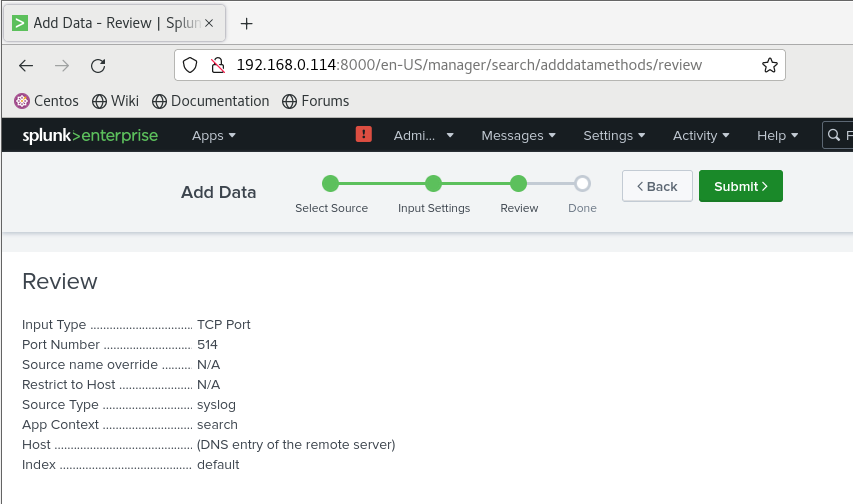


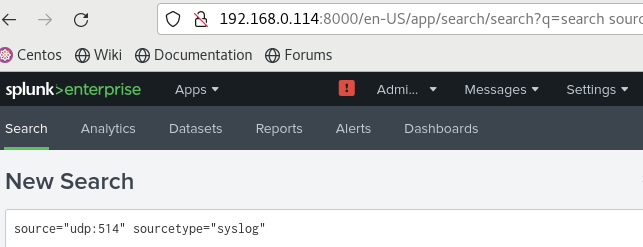
Trên máy Splunk server:  
Tải gói splunk rpm 64 bit từ trang web www.splunk.com và cài đặt dưới quyền admin  
  
Gõ đường dẫn để chạy splunk:  
# /opt/splunk/bin/splunk start  
Bảng license agreement hiện ra, chọn y để khởi động.

  
Splunk đã được cài đặt thành công  
Để splunk khởi động mỗi khi restart máy gõ lệnh:  
# /opt/splunk/bin/splunk enable boot-start

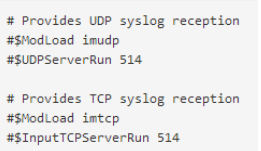
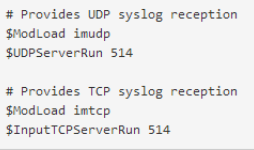
Vào giao diện web của splunk, chọn mở trình duyệt(firefox) rồi gõ đường dẫn 192.168.0.114:8000 để vào giao diện splunk web.

  
Sau khi truy cập vào giao diện web splunk, để nhận log từ syslog server , ta vào input data, chọn tcp, add  
port listen là 514 source type là syslog và có thể tùy chọn lắng nghe từ nhiều nguồn hoặc từ nguồn chỉ định(192.168.0.115).Sau đó save.

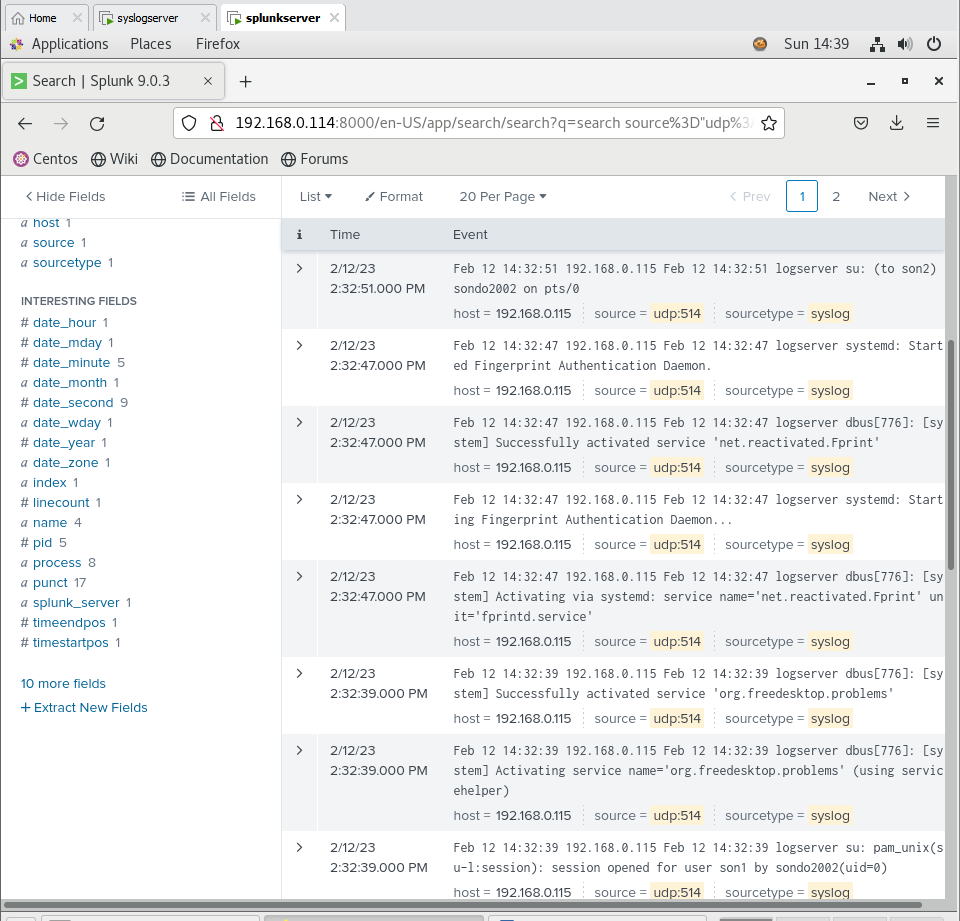
  
Vào input data nhấp chọn udp và làm tương tự như ở trên tcp

  
Hoàn tất quá trình cài đặt và cấu hình splunk.

Trên máy SyslogServer :  
Cài đặt syslog server  
# Yum install rsyslog  
Sau khi cài đặt xong, vào /etc/rsyslog.conf chỉnh sửa nội dung file rsyslog.conf để mở port 514 cho syslog(bỏ 2 dấu #)

Đồng thời thêm vào 2 dòng dưới section #####RULES#####  
\*.\* @192.168.0.114  
Mail.\* @192.168.0.114  
Sau đó thoát ra save file lại và restart rsyslog bằng lệnh :  
# service rsyslog restart  
Sau đó thử switch user trên syslog server để sản sinh log.Sau đó qua kiểm tra bên splunk server xem log đó đã được send qua hay chưa



# V. Tổng kết

- Tóm lại, qua quá trình tìm hiểu Splunk, chúng tôi đã hiểu được về nền tảng phân tích dữ liệu mạnh mẽ này và cách sử dụng nó để giúp giải quyết các vấn đề trong hệ thống. Splunk cho phép thu thập và phân tích dữ liệu từ nhiều nguồn khác nhau, giúp người dùng tìm kiếm và giám sát các sự cố trong hệ thống một cách dễ dàng. Ngoài ra, Splunk còn cung cấp khả năng phân tích và trực quan hóa dữ liệu để đưa ra các quyết định thông minh cho doanh nghiệp.

- Splunk là một công cụ quan trọng và được sử dụng rộng rãi trong các lĩnh vực như an ninh mạng, IoT và đám mây. Chúng ta có thể sử dụng Splunk để giám sát, quản lý và tối ưu hóa hệ thống của mình, giúp nâng cao hiệu suất và giảm thiểu thời gian gián đoạn.

- Tuy nhiên, việc phát triển Splunk cũng có thể gặp phải một số định thức, đặc biệt là trong quá trình xử lý các lượng dữ liệu lớn. Chúng ta cần có kiến thức chuyên sâu và kinh nghiệm thực tế để có thể tận dụng tối đa các tính năng của Splunk.

- Nhìn chung, Splunk là một công cụ phân tích dữ liệu mạnh mẽ và hữu ích cho các doanh nghiệp và các chuyên gia phân tích dữ liệu. Việc ôm giữ và sử dụng hiệu quả Splunk sẽ giúp chúng ta tối ưu hóa hệ thống hóa và đưa ra các quyết định kinh doanh thông minh.

Phụ lục

Tài liệu tham khảo:

Trang gốc về splunk: <https://www.splunk.com/>

https://en.wikipedia.org/wiki/Splunk

<https://viblo.asia/p/phat-hien-xam-nhap-voi-splunk-OeVKB8MdlkW#_iii-phat-hien-xam-nhap-voi-splunk-6>