

Mục Lục

Chú thích hình ảnh

Hình 1 : Sơ đồ triển khai Splunk phổ biến trong doanh nghiệp.

Hình 2 : Các loại data , log mà Splunk index được.

Hình 3 : Hệ thống index và tìm kiếm được phân phối giữa nhiều core Splunk.

Hình 4 : Sơ đồ hoạt động của Mapreduce.

Hình 5 : Ví dụ minh họa cách mà Mapreduce hoạt động.

Hình 6 : Phân nhỏ dữ liệu đầu vào.

Hình 7 : Sao chép chương trình.

Hình 8 : Thực hiện hàm Map cho ra kết quả <key,value>.

Hình 9 : Thực hiện hàm Reduce và thông báo kết quả cho Master.

Hình 10 : Thông báo chương trình mapreduce hoàn tất và kết quả được lưu trữ trên R tập tin.

Hình 11 : Ví dụ tìm các sự kiện xảy ra trong 60 phút trước.

Hình 12 : Lưu và chia sẻ kết quả tìm.

Hình 13 : Kết quả có thể được chia sẻ dưới dạng link.

Hình 14 : Lưu kết quả tìm kiếm.

Hình 15 : Những kết quả phải thỏa những điều kiện được thiết lập mới được lưu.

Hình 16 : Kết quả tìm kiếm sẽ xuất hiện trong menu Search & Report.

Hình 17 : Tạo một alert.

Hình 18 : Đặt tên alert và điều kiện để kích hoạt alert.

Hình 19 : Chạy kết quả tìm kiếm event mỗi giờ, khởi động alert khi kết quả tìm kiếm lớn hơn 0.

Hình 20 : Nếu số lượng event tìm được trong 5 phút bé hơn 5 thì kích hoạt alert.

Hình 21 : Các option trong Alert.

Hình 22 : Một table dạng số.

Hình 23 : Một table dạng chart.

Hình 24 : Biểu đồ chart dữ liệu nhận được trong một khoảng thời gian.

Hình 24 : Các tùy chọn formatting của chart.

Hình 25 : Ví dụ về một dashboard cơ bản.

Hình 26 : Tắt selinux.

Hình 27 : Cấu hình mặc định trong file rsyslog.conf.

Hình 28 : Cấu hình để mở port 514 cho syslog.

Hình 29 : Giao diện web của Splunk.

Hình 30 : Giao diện splunk đã có thêm add-on Windows.

Hình 31 : Cấu hình Forwarding and Receiving.

Hình 32 : Tùy chọn các loại log mà universalforwarder sẽ gửi.

Hình 33 : Splunk đã nhận được log của Windows.

Hình 34 : Menu chính của Splunk.

Hình 35 : Tạo một Dashboard mới.

Hình 36 : Tùy chỉnh kiểu Dashboard sẽ xuất ra.

Hình 37 : Biểu đồ biểu diễn log hệ thống Window dạng pie.

Hình 37 : Biểu đồ biểu diễn log hệ thống Window dạng cột.

Hình 38 : Thêm ghép nhiều biểu đồ sẽ trở thành một dashboard.

Chú thích thuật ngữ

Big Data: Là tập hợp các dữ liệu lớn từ nhiều nguồn như hệ thống máy tính, mysql, các ứng dụng.v.v.v

Map Reduce : là một thuật toán giúp các ứng dụng xử lý nhanh một lượng dữ liệu lớn.

UniversalForwarder : là một phiên bản của splunk nhưng chỉ có tính năng thu thập và gửi dữ liệu.

Light Forwarder : là một phiên bản của Splunk , không có tính năng phân tích mà chỉ forward dữ liệu. Ít được sử dụng ở các phiên bản splunk 6.0 .

Heavy Forwarder : Là một phiên bản của Splunk, có thể phân tích và gửi nhưng không có khả năng tìm kiếm phân phối dữ liệu.

Dashboard : Là một bảng bao gồm nhiều biểu đồ với nhiều kiểu định dạng khác nhau.

Pfsense : Phần mềm firewall mã nguồn mở.

1 Đặt Vấn Đề

Trong mọi doanh nghiệp, hệ thống công nghệ thông tin là hệ thống vô cùng quan trọng. Ngày nay với mức độ phát triển công nghệ nhanh chóng, thì ngoài việc đảm bảo khả năng vận hành, hoạt động liên tục và chính xác thì việc đảm bảo an ninh thông tin là một thách thức lớn.

- Nguy cơ bên ngoài: Tin tặc bên ngoài lợi dụng lỗ hổng hệ thống để đột nhập
- Nguy cơ bên trong: do hành vi người dùng, ý thức về mức độ an toàn dữ liệu còn chưa được cao.
- Tính thống nhất trong quản trị : Khi hệ thống càng lớn thì mức độ phức tạp trong quản lý cũng sẽ tăng cao.

SIEM là một giải pháp hoàn chỉnh, đầy đủ cho phép các tổ chức thực hiện việc giám sát các sự kiện cho một hệ thống. Các thành phần chính của SIEM bao gồm: thành phần thu thập nhật ký, thành phần phân tích, thành phần lưu trữ, thành phần quản trị tập trung. Ngoài ra còn có các thành phần khác như: thành phần giám sát Network ở mức lớp 7 trong mô hình OSI, các module tạo báo cáo (Compliance Report, Dashboard)

Giải pháp SIEM có những ưu điểm sau:

- Hỗ trợ thu thập, phân tích các sự kiện theo thời gian thực được thu thập từ các hệ thống gửi về, được kết hợp cùng với các thông tin liên quan đến người dùng, các thành phần trong hệ thống và dữ liệu.
- Cung cấp khả năng lưu trữ log dài, toàn diện (log management) và khả năng phân tích theo ngữ cảnh (Correlation).
- Cung cấp các chức năng được xây dựng sẵn và cho phép thay đổi (Customized) theo các yêu cầu của các tổ chức.
- Dễ dàng triển khai và sử dụng.

Splunk là một cầu nối giữa việc quản lý log một cách đơn giản và bảo mật thông tin, thu thập sự kiện. Cái mà phân biệt ở Splunk so với các server Syslog hay các công cụ SIEM khác là Splunk Apps. Một thư viện quản lý hơn 200 add-on khác nhau. Chính vì điều đó đã làm cho Splunk trở nên khác biệt, tăng khả năng thu thập thông tin các loại log khác nhau, có giao diện gần gũi và thân thiện, cung cấp những tính năng tìm kiếm và phân tích dữ liệu thu được.

2 Tổng quan về Splunk

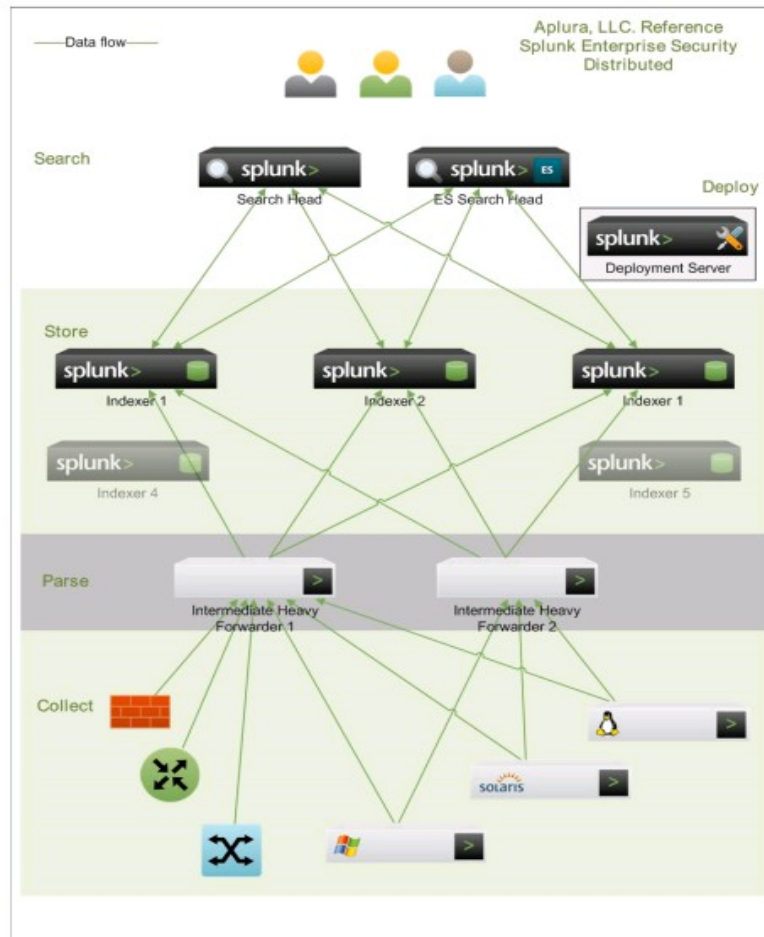
2.1 Splunk là gì?

Splunk là hệ thống có thể captures, trích ra các dữ liệu thời gian thực có liên quan tới nhau từ đó nó 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, chẩn 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 các 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.

2.2 Sơ đồ Splunk phổ biến



Hình 1 : Sơ đồ triển khai Splunk phổ biến trong doanh nghiệp

Mô hình trên bao gồm các thành phần như:

+Nhiều thiết bị Forwarders trung gian phục vụ cho quá trình load, tính sẵn sàng cao, và cải thiện tốc độ xử lý các event sắp tới.

+Một Indexer liên kết với nhiều hệ thống. Với nhiều search-peer(indexer) cải thiện hiệu năng của quá trình nhập dữ liệu và tìm kiếm. Nó giúp giảm thời gian tìm kiếm và cung cấp tính dự phòng cao.

+Có nhiều đầu tìm kiếm. Những hệ thống riêng biệt này sẽ phân phối bất kỳ yêu cầu tìm kiếm trên tất cả các search-peer đã cấu hình trước đó để cải thiện hiệu năng tìm kiếm.

+Đầu tìm kiếm riêng biệt được thể hiện ở đây để hỗ trợ ứng dụng Splunk's Enterprise Security(ES).

+Server triển khai. Hệ thống nay có thể được tích hợp với các dịch vụ Splunk khác, hoặc triển khai độc lập. Nếu muốn triển khai hệ thống lớn, một hệ thống độc lập là rất quan trọng.

2.3 Splunk thu thập những gì?

*Splunk thu thập dữ liệu hệ thống do máy móc tạo ra

Dữ liệu hệ thống bao gồm nhiều hạng mục record của tất cả các hoạt động và hành vi- hành vi của khách hàng, giao dịch của user, hành vi của hệ thống.

2.4 Splunk có thể làm gì?

-Server Metrics	-Vulnerability Data
-Custom Applications	-Physical Security
-Windows registries	-Scripts
-Card key	-Patch Mgmt
-Server Logs	-Host Config
-DNS Logs	-Virtual Logs
-Host ID	-Database Logs
-Router	-Email Logs
-RAS VPN	-Application Logs

2.5 Splunk cung cấp cho chúng ta những gì?

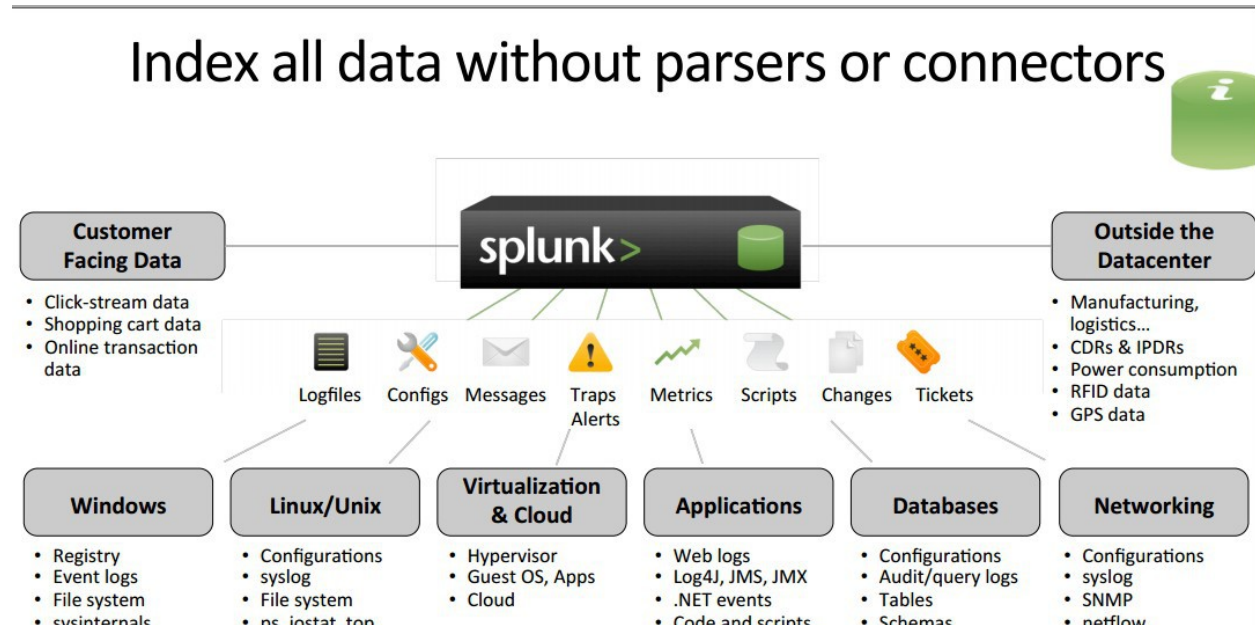
Splunk cung cấp 1 giao diện chung cho tất cả dữ liệu IT như tìm kiếm dữ liệu, những cảnh báo, những báo cáo(report), hay chúng ta có thể chia sẻ dữ liệu đó cho một ai đó. Splunk cung cấp giải pháp tìm kiếm tối ưu.

2.6 Splunk, Giải pháp tối ưu cho Big Data?

-Splunk tìm kiếm những dữ liệu có liên quan với nhau, giúp thu hẹp phạm vi tìm kiếm , tiết kiệm thời gian, và làm cho công tác quản trị mạng tốt hơn.

2.7 Tại sao chọn Splunk?

Splunk còn được gọi là Google của log, có công cụ search mạng mẽ nó chấp nhận dữ liệu ở bất kỳ định dạng nào.



Hình 2 : Các loại data , log mà Splunk index được.

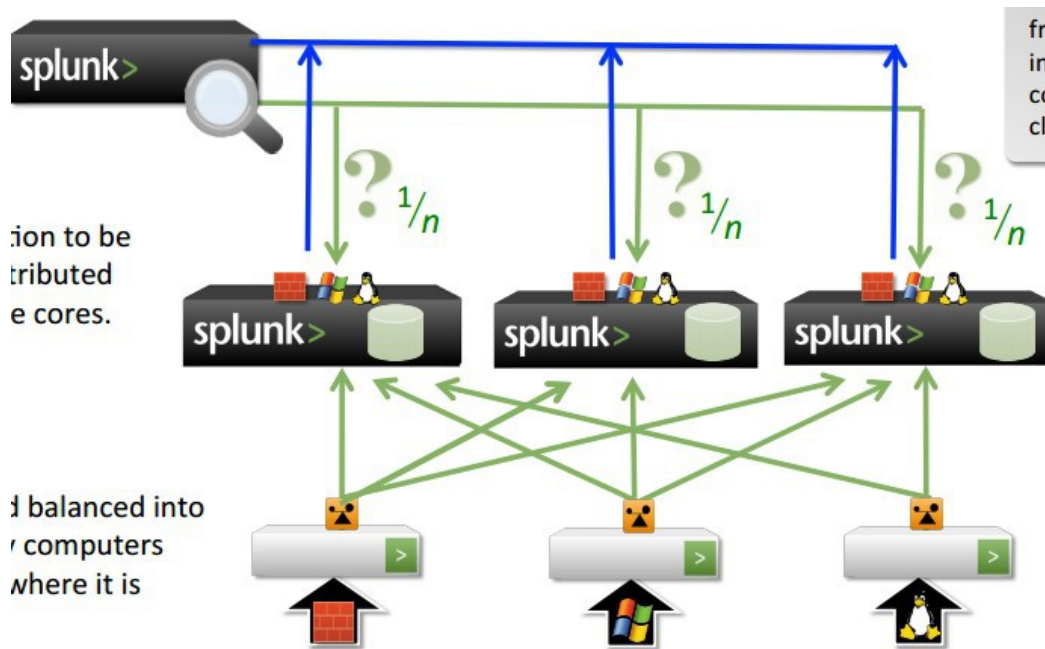
Splunk tự động list ra thời gian cụ thể của từng sự kiện xảy trong hệ thống mà nó đang giám sát.

Cảnh báo trong thời gian thực. Ta có thể chỉnh tùy chọn, định nghĩa các loại cảnh báo và có thể chỉ định ai nhận được cảnh báo đó.

Splunk cung cấp thông tin tìm kiếm thông minh: Kết quả tìm kiếm được sắp xếp hợp lý, có liên quan với nhau, khả năng hiển thị thời gian thực, phân tích lịch sử các sự kiện đã xảy ra.

Splunk có thể lưu trữ khối lượng dữ liệu lớn của hệ thống IT và dữ liệu này có thể có cấu trúc bất kỳ, song tốc độ truy vấn dữ liệu nhanh.

Tìm kiếm phân tán sử dụng Map Reduce(1 phần mềm của Google, phục vụ cho việc tính toán phân tán các tập dữ liệu lớn trên các cụm máy tính)



Hình 3 : Hệ thống index và tìm kiếm được phân phối giữa nhiều core Splunk.

Dữ liệu cần tìm kiếm được phân phối giữa nhiều cores

Mỗi indexer xử lý tập hợp con của toàn bộ dữ liệu và tạo ra một phần của kết quả tìm kiếm tổng thể rồi đưa nó vào vào đầu của quá trình tìm kiếm để giảm tải.

Tham khảo:

<http://docs.splunk.com/Documentation/Splunk/6.0.2/installation/RunSplunkasadifferentnon-rootuser>

3 Giải pháp với Splunk

3.1 Quản lý các ứng dụng:

3.1.1 Giải quyết vấn đề nhanh hơn, giảm thời gian bị downtime:

-Troubleshoot vấn đề 1 cách nhanh chóng, giảm chi phí và giảm thời gian để điều tra và khắc phục sự cố tới 70%.

-Giảm sự phức tạp bằng cách cung cấp cho các nhà phát triển được truy cập vào log của ứng dụng thông qua 1 vị trí trung tâm mà không cần quyền truy cập vào hệ thống đó.

-Giám sát toàn bộ môi trường ứng dụng của chúng ta trong thời gian thực để ngăn chặn các vấn đề ảnh hưởng tới người dung, giữ lại log từ các sự kiện định kỳ để ngăn ngừa mất mát.

-Nhắm được hoạt động của toàn bộ ứng dụng:

-Truy vết và giám sát các giao dịch của ứng dụng thông qua các tầng của kiến trúc phân tán và từ nhiều nguồn dữ liệu.

-Phát hiện các bất thường hoặc các vấn đề trong hoạt động, thời gian đáp ứng và chủ động giải quyết chúng trước khi nó ảnh hưởng tới người dung ứng dụng.

-Theo dõi số liệu hoạt động quan trọng như thời gian đáp ứng end-to-end, độ dài thông điệp hàng đợi và đếm số lần giao dịch thất bại để đảm bảo ứng dụng đáp ứng được nhu cầu cần thiết.

-Nhận được toàn bộ hoạt động của ứng dụng trong thời gian thực trên toàn bộ cơ sở hạ tầng ứng dụng của chúng ta.

-Đạt được cái nhìn toàn diện về cách mà người dùng sử dụng dịch vụ của chúng ta, từ đó có thể cung cấp dịch vụ tốt hơn.

-Làm phong phú hệ thống của chúng ta bằng cách thêm các nguồn phi CNTT như giá cả cơ sở dữ liệu, thông tin khách hàng và thông tin vị trí.

3.1.2 Tại sao Splunk là giải pháp tốt cho việc quản lý ứng dụng

Không giống các công cụ quản lý truyền thống, splunk có thể index, phân tích, khai thác dữ liệu từ bất kỳ tầng ứng dụng nào. Nó cung cấp 1 góc nhìn trung tâm về toàn bộ hệ thống cơ sở hạ tầng của chúng ta.

Ngôn ngữ tìm kiếm trong splunk giúp người sử dụng so sánh các sự kiện, các giao dịch và chỉ số hoạt động quan trọng khác.

Quyền điều khiển được trao cho nhiều nhóm trong một tổ chức. Những hiểu biết về dữ liệu ứng dụng có thể kết hợp với thông tin có cấu trúc như thông tin user hoặc giá cả thông tin để doanh nghiệp quyết định tốt hơn.

Nhà sản xuất quản lý hoạt động ứng dụng AppDynamics và Extrahop đã phát triển ứng dụng Splunk để giúp khách hàng quản lý tốt hơn các dữ liệu ứng dụng như log, các sự kiện, hoạt động của cơ sở hạ tầng và nhiều hơn thế nữa.

3.2 Quản lý hoạt động IT

Trung tâm IT dữ liệu trên toàn thế giới đang trở nên cực kỳ phức tạp, với hàng trăm công nghệ khác nhau và thiết bị ở nhiều layer. Ảo hóa và điện toán đám mây cũng đang trở nên phức tạp, đặc biệt là các vấn đề liên quan đến hiệu suất hoạt động. Đội ngũ quản trị và quản lý CNTT lãng phí nhiều thời gian trong việc di chuyển từ một giao diện điều khiển tới giao diện điều khiển khác, cố gắng theo dõi các dữ liệu cần thiết để đảm bảo hiệu suất và tính sẵn sàng cao.

Splunk cung cấp 1 cách tiếp cận tốt hơn mà không cần phải phân tích cú pháp hay tùy chỉnh nó. Splunk thu thập và lập indexes chứa tất cả dữ liệu được tạo ra bởi hệ thống IT của chúng ta (hệ thống mạng, server, OS, ảo hóa, v.v.) . Nó hoạt động với bất kỳ dữ liệu mà máy tạo ra, bao gồm log, file cấu hình, số liệu hiệu suất, SNMP trap và các ứng dụng log tùy chỉnh.

+ Giải quyết vấn đề nhanh hơn, giảm thời gian Downtime:

Giúp nắm bắt được hoạt động ảo hóa, hệ thống cloud private và public từ 1 giao diện trung tâm.

Giúp tìm được nguồn gốc của vấn đề nhanh hơn 70% mà không cần phải tìm kiếm trong hệ thống, server hay máy ảo.

Quản lý hệ thống của chúng ta trong thời gian thực, ngăn ngừa vấn đề xảy ra trước khi nó ảnh hưởng tới người dùng và có thêm kinh nghiệm xử lý các sự kiện xảy ra định kỳ để tránh mất mát.

Chỉ cần 1 người quản lý có quyền truy cập trực tiếp, đảm bảo an toàn cho dữ liệu, giúp tránh leo thang đặc quyền.

+Tương quan các sự kiện ở tất cả các tầng layer của hệ thống:

Tìm các liên kết giữa người sử dụng, hiệu suất các sự kiện liên quan tới cơ sở hạ tầng được cung cấp bởi splunk

Kết hợp phân tích dữ liệu thời gian thực tương quan, so sánh với hàng triệu terabytes dữ liệu lịch sử. Phân tích phát hiện thành phần khả nghi có thể giúp dự đoán và ngăn ngừa mất mát hoặc vấn đề về hiệu năng.

Tồn tại dữ liệu từ khắp nơi trên mỗi tầng của trung tâm dữ liệu. Quản lý môi trường của chúng ta để nhận biết được sự thay đổi, so sánh ngay lập tức để biết độ thiếu hụt hiệu năng của hệ thống, những vấn đề có sẵn hoặc vấn đề bảo mật, an ninh.

+ Giảm chi phí cung cấp dịch vụ CNTT:

Sử dụng sức mạnh và khả năng mở rộng của splunk không chỉ cho hoạt động quản lý CNTT mà còn dùng để hỗ trợ kiểm toán, an ninh.

Giảm số lượng các công cụ và kỹ năng cần thiết để duy trì quản lý cơ sở hạ tầng phức tạp của chúng ta.

3.2.1 Phân tích hoạt động IT:

Splunk dùng trong hoạt động phân tích IT cung cấp những hiểu biết toàn diện theo nhiều tầng giúp cho định hướng của doanh nghiệp tốt hơn tùy theo từng trường hợp cụ thể.

Chủ động trong việc nhận diện và khắc phục lỗi dịch vụ để đảm bảo sự hài lòng của khách hàng và giúp tăng số lượng khách hàng sử dụng.

Đạt hiệu quả trong quá trình hoạt động do nắm bắt được những nguy hiểm tiềm tàng trong quá trình hoạt động kinh doanh.

Giúp đạt được các mục tiêu kinh doanh bằng cách cung cấp tầm nhìn toàn diện trên toàn hệ thống công nghệ không đồng nhất, các dịch vụ, cách quản lý, lên kế hoạch về dung lượng, phân tích mức sử dụng của người dùng và nhiều hơn nữa.

3.2.2 Giám sát cơ sở hạ tầng:

+Máy chủ: Với Splunk, chúng ta có thể

Chủ động giám sát các máy chủ và hiểu biết sâu hơn về hiệu suất, cấu hình, truy cập và các lỗi phát sinh.

Tương quan hiệu suất máy chủ, các lỗi và dữ liệu sự kiện với người dùng, ảo hóa và ứng dụng thành phần để ngăn ngừa và khắc phục lỗi.

Phân tích và tối ưu hóa chi phí cho việc theo dõi dung lượng máy chủ, báo cáo an ninh trong thời gian thực.

+Hệ thống lưu trữ: Với Splunk, chúng ta có thể

Tương quan log, số liệu hiệu suất và các sự kiện từ hệ thống lưu trữ của chúng ta với máy chủ, mạng và dữ liệu từ các ứng dụng để giải quyết các vấn đề và làm tăng sự hài lòng của khách hàng.

Sử dụng công cụ phân tích mạnh mẽ để khắc phục sự cố trong thời gian thực và phân tích hiệu suất hệ thống lưu trữ của chúng ta.

Giảm thời gian phát triển và cắt giảm chi phí bằng việc dễ dàng tích hợp với các nhà cung cấp dịch vụ lưu trữ, như NetApp và EMC.

+Hệ thống mạng: Với Splunk, chúng ta có thể:

Giám sát và theo dõi dữ liệu mạng từ các thiết bị không dây, switch, router, firewall và trên những thiết bị khác bằng cách sử dụng SNMP, Netflow, syslog, PCAP, v.v.

Chủ động nhận diện các vấn đề an ninh mạng và thực hiện phân tích vấn đề. Tương quan dữ liệu mạng với các ứng dụng, hệ thống lưu trữ và phân tích máy chủ để giữ cho mạng của chúng ta an toàn và hoạt động mọi lúc.

Đạt được chỉ số ROI tối đa bằng cách tối ưu hóa dung lượng mạng lưới của chúng ta, xác định độ trễ, quản lý băng thông, xác định top 10 tài nguyên mạng thường được sử dụng và mô hình sử dụng.

3.2.3 Splunk cho hệ điều hành

Splunk và ứng dụng của splunk có thể giúp chúng ta:

Tương quan số liệu hệ thống và dữ liệu sự kiện với cả dữ liệu ở các tầng công nghệ khác một cách dễ dàng.

Tìm liên kết giữa vấn đề hiệu suất ứng dụng và hệ điều hành, ảo hóa, hệ thống lưu trữ, mạng, và cơ sở hạ tầng máy chủ.

Nắm được toàn bộ hoạt động hệ thống bằng cách cung cấp bảng điều khiển trung tâm sức khỏe hệ thống xuyên suốt môi trường không đồng bộ.

Nắm được năng lực hạn chế của hệ thống hoặc tình trạng nhàn rỗi.

Theo dõi những thay đổi và đảm bảo an ninh cho môi trường của chúng ta bằng cách giám sát môi trường để phát hiện những hoạt động bất ngờ, thay đổi vai trò của người sử dụng, truy cập trái phép, v.v..

3.2.4 Quản lý ảo hóa

Cơ sở hạ tầng ảo hóa tạo ra môi trường năng động, nơi mà tài nguyên máy tính như máy chủ, storage, phần cứng mạng được ảo hóa từ các ứng dụng, hệ điều hành và người sử dụng. Môi trường ảo phức tạp đòi hỏi cách tiếp cận mới với các dịch vụ IT truyền thống như xử lý sự cố hiệu suất, quản lý và phân tích rủi ro.

Ứng dụng ảo hóa của Splunk kết hợp sức mạnh và tính năng của Splunk Enterprise được thiết kế dành riêng cho công nghệ ảo hóa. Nó giúp tăng tốc dữ liệu thu thập được cơ sở hạ tầng ảo. Kết hợp dữ liệu hạ tầng ảo hóa với dữ liệu tầng công nghệ khác sẽ cho 1 góc nhìn bao quát hơn về hệ thống trung tâm dữ liệu.

Splunk App cho ảo hóa có thể tương thích và thu thập dữ liệu ảo hóa từ các công nghệ ảo hóa như VMware vSphere, Citrix XenServer và Microsoft Hyper-V, và công nghệ ảo hóa máy tính bàn như Citrix XenApp và Citrix XenDesktop.

Nó tạo các báo cáo đa dạng, đồng nhất về các công nghệ ảo hóa từ tất cả các lớp ứng dụng và cơ sở hạ tầng của chúng ta.

Giúp chủ động ngăn chặn, quản lý vấn đề hiệu suất, tắc nghẽn cổ chai, những sự kiện bất ngờ, những thay đổi và lỗi an ninh bảo mật nguy hiểm. Nó phân tích và báo cáo chính xác giúp cho người dùng có trải nghiệm tối ưu.

Tương quan dữ liệu ảo hóa, giúp việc tìm ra các sự kiện có liên quan một cách dễ dàng hơn, tương quan các vấn đề về hiệu năng, mạng và kiến trúc hệ thống máy chủ.

Giữ lại số liệu về hiệu suất hoạt động của máy để theo dõi và phân tích. Thu thập dữ liệu có chiều sâu từ máy chủ, máy ảo, hệ thống máy tính. Cung cấp khả năng hiển thị hoạt động và phân tích hoàn chỉnh bằng cách xác định khả năng của máy chủ, các máy ảo nhàn rỗi, các máy chủ sử dụng đúng mức, sức chứa dữ liệu, theo dõi thống kê hiệu suất để tìm mô hình sử dụng và tránh khả năng tắt nghẽn có thể.

Theo dõi những thay đổi và báo cáo về tài sản. theo dõi chi tiết sự thay đổi mà người dùng thực hiện, tự động hóa các tác vụ của vSphere cũng như báo cáo tình trạng các thành phần ảo.

Cải thiện an ninh bằng cách giám sát môi trường để tìm các hoạt động đáng ngờ, vai trò của người sử dụng bị thay đổi, truy cập trái phép và nhiều hơn nữa.

Với VMware vSphere

-Splunk App cho VMware cung cấp khả năng hiển thị các hoạt động 1 cách chi tiết, hiệu suất, log, các tác vụ, sự kiện và lưu đồ từ máy chủ, các máy ảo và các trung tâm ảo hóa. Cung cấp hình ảnh bao quát và chính xác về tình trạng sức khỏe của môi trường ảo hóa, chủ động xác định các vấn đề về hiệu suất, bảo mật, khả năng hoạt động và những thay đổi của máy ảo.

- Nhận được thông tin sức khỏe máy ảo trong thời gian thực. Có thể xác định lập tức khu vực máy ảo, máy chủ có vấn đề. Phân tích dữ liệu theo thời gian để xác định xem nó có ảnh hưởng đến cấu hình tài nguyên. Nhận báo cáo chi tiết dựa trên mỗi 20s. Khám phá lỗi và các trường hợp ngoại lệ bằng việc chỉ ra các sự kiện có liên quan tới nhau bằng dữ liệu log VC và ESXi trong một giao diện điều khiển duy nhất.

-Có thể biết được tình trạng sức khỏe của từng máy ảo. Tăng tốc độ troubleshoot nhờ vào việc so sánh giữa các máy ảo với nhau.

-Chủ động trong việc quản lý hành vi mờ ám của user, các cuộc tấn công tiềm năng bằng những báo cáo an ninh

-Nhận bắt được thông tin CPU, bộ nhớ, ổ đĩa và dung lượng disk sử dụng trong thời gian thực. Chủ động cảnh báo khi thiếu hụt dung lượng xảy ra. Lấy lại không gian lưu trữ không sử dụng để cho người dùng có trải nghiệm tối ưu. Sử dụng xu hướng theo thời gian và tối ưu hóa dựa trên tiêu thụ. Dự báo thông tin CPU, bộ nhớ, nhu cầu ổ cứng cần thiết và hiệu năng của từng máy chủ, máy ảo VMs thông qua lịch sử sử dụng tài nguyên của máy ảo.

Với Citrix XenServer và Microsoft Hyper-V:

-Cung cấp góc nhìn theo thời gian thực về các yếu tố như hiệu năng, chỉ số tiêu thụ tài nguyên, cấu trúc liên kết trên nền tảng máy chủ ảo hóa bằng cách sử dụng một khuôn khổ báo cáo chung. Nó bao gồm một chuỗi các biểu đồ liên quan đến hoạt động IT, giám sát hoạt động, lên kế hoạch khả năng chịu tải, và thay đổi theo dõi.

-Dashboard Out-of-the-box cho 1 cái nhìn cụ thể trong thời gian thực về tình trạng sức khỏe của máy ảo và máy chủ.

-Đào sâu vào lưu đồ để cho cái nhìn chuyên sâu về hiệu năng, log, thay đổi về cấu hình, các cảnh báo và hơn nữa.

-Truy cập vào lịch sử dữ liệu cho việc phân tích và xử lý sự cố.

-Cấu hình cảnh báo dựa trên kịch bản có sẵn cho các vấn đề thường gặp như bộ nhớ, CPU, dung lượng ổ đĩa thấp.

-Giám sát và theo dõi tài nguyên mà máy ảo tiêu thụ để hỗ trợ cho việc lên kế hoạch về khả năng hoạt động của máy ảo.

-Cho góc nhìn 360 độ về khả năng hiển thị máy ảo với dữ liệu từ tầng công nghệ khác giúp giải quyết và xử lý sự cố nhanh hơn.

3.3 An ninh trong lĩnh vực IT

3.3.1 Mỗi đe dọa an ninh ngày một tăng:

Hiện tại các phần mềm malware đã trở nên “tàng hình”, và thường trông giống như một dịch vụ hay 1 ứng dụng bình thường nào đó. Nó được xây dựng để lây lan trên toàn bộ hệ thống. Kẻ tấn công có thể tùy ý nghiên cứu chỉnh sửa hệ thống của chúng ta, nếu bị phát hiện, kẻ tấn công có thể kích hoạt malware khác để tiếp tục thu thập dữ liệu. Splunk có thể thu thập và index bất kỳ dữ liệu nào mà không quan tâm đến định dạng hoặc kích cỡ và thực hiện tìm kiếm tự động trên hàng petabyte dữ liệu. Splunk có một ngôn ngữ lệnh phân tích mạnh mẽ, thông minh, giúp các nhà phân tích đặt ra những câu hỏi về bảo mật dựa trên dữ liệu của chúng ta. Cách tiếp cận đặc biệt này giúp chúng ta chủ động trong việc tìm kiếm mối đe dọa bằng cách kiểm tra hoạt động của dữ liệu trong môi trường hoạt động bình thường.

3.3.2 Quản lý log:

Phần mềm Splunk giúp khách hàng cải thiện vấn đề phân tích dữ liệu log để quản lý việc kinh doanh của họ tốt hơn. Splunk tự động index dữ liệu, bất kể có cấu trúc hay không cấu trúc. , cho phép chúng ta nhanh chóng tìm kiếm, báo cáo, và chẩn đoán các hoạt động và các vấn đề an ninh một cách ít tốn kém hơn. Với Splunk-việc quản lý log của chúng ta sẽ dễ hơn bao giờ hết.

3.3.3 Ứng dụng Splunk dành cho an ninh:

Với ứng dụng an ninh của Splunk chúng ta có thể sử dụng số liệu thống kê trên bất kỳ dữ liệu nào để tìm kiếm các mối đe dọa tiềm ẩn, trong khi vẫn có thể giám sát liên tục các mối đe dọa đã bị phát hiện bởi những sản phẩm an ninh truyền thống.

Ứng dụng an ninh Splunk chạy ở phía trên Splunk Enterprise và cung cấp công cụ để giám sát, cảnh báo và phân tích cần thiết để xác định và giải quyết các mối đe dọa đã biết và chưa biết. Nó phù hợp với đội ngũ an ninh nhỏ hoặc một trung tâm hoạt động bảo mật.

Bảng điều khiển an ninh cung cấp một cách xem hoàn toàn tùy biến với các từ khóa bảo mật quan trọng trong lĩnh vực an ninh domain. Ứng dụng an ninh Splunk chứa 1 thư viện dựng sẵn các số liệu an ninh để hỗ trợ người dùng nhận diện được các tình huống và giám sát liên tục các nguy cơ bảo mật trên domain. Và tất cả thông tin đó đều được thể hiện rõ trên bảng điều khiển Dash board.

Tính năng xem xét lại các sự kiện đã xảy ra: Cung cấp chi tiết quy trình công việc phân tích cần thiết để các ưu tiên của vụ việc, bối cảnh của sự cố, loại của nó và các máy chủ có liên quan. Chỉ một click chuột và chúng ta có thể thấy được các dữ liệu thô mà ứng dụng an ninh splunk lưu trữ.

Tính năng bảo vệ tài sản và điều tra nhận dạng mối nguy hiểm cung cấp cho nhà phân tích an ninh khả năng xem xét các mối đe dọa dựa trên một loạt các sự kiện an ninh. Đơn giản chỉ cần chọn một khung thời gian sự kiện hoặc nhiều sự kiện đại diện cho những hoạt động đáng ngờ và Splunk sẽ tự động hiển thị một bản tóm tắt mô hình an ninh. Với 1 cú click chuột, chúng ta có thể xem tất cả các dữ liệu thô được đặt ra theo thứ tự thời gian, đưa ra 1 cái nhìn trực tiếp cho đồng nghiệp hoặc tạo ra một tìm kiếm mới để xem các sự kiện đã xuất hiện này có tiếp tục xuất hiện hay không.

Phân tích và dự đoán: Bảng điều khiển phân tích cung cấp một điểm. Nhấp vào điểm đó sẽ hiện các giải pháp để biết được hướng đi tương lai của điểm đó và dự báo giá trị dựa trên mô hình dữ liệu. Chỉ cần

chọn kiểu dữ liệu, bất kỳ đối tượng chứa kiểu dữ liệu đó, kiểu hàm trình diễn, thuộc tính và chu kỳ phân tích mà chúng ta muốn tạo.

Danh sách các mối đe dọa: Splunk cung cấp dịch vụ out-of-the-box hỗ trợ cho 18 mã nguồn mở đe dọa tới dữ liệu nhằm tăng thêm tính bảo mật cho hệ thống của chúng ta. Splunk cho phép chúng ta thêm mã nguồn mở của riêng chúng ta và nguồn cung cấp dữ liệu thanh toán chỉ với vài click chuột mà không cần một cam kết dịch vụ. Splunk còn công tác với trung tâm bảo mật Norse Security, 1 trung tâm bảo mật uy tín toàn cầu. Splunk còn cho khách hàng cảm giác trải nghiệm dịch vụ an ninh Splunk cho hệ thống doanh nghiệp trong vòng 30 ngày.

4 Các tính năng chính trong hoạt động Giám sát mạng của Splunk

4.1 Map Reduce

4.1.1 Map reduce là gì?

Mapreduce là 1 phương thức thực thi để giúp các ứng dụng có thể xử lý nhanh 1 lượng dữ liệu lớn (big data). Các dữ liệu này được đặt tại các máy tính phân tán. Các máy tính này sẽ hoạt động song song độc lập với nhau. Điều này làm rút ngắn thời gian xử lý toàn bộ dữ liệu. Dữ liệu đầu vào có thể là dữ liệu có cấu trúc (dữ liệu lưu trữ dạng bảng quan hệ 2 chiều) hoặc dữ liệu không cấu trúc (dữ liệu dạng tập tin hệ thống)

4.1.2 Ưu điểm của mapreduce

Xử lý tốt bài toán về lượng dữ liệu lớn có các tác vụ phân tích và tính toán phức tạp không lường trước được

Có thể tiến hành chạy song song trên các máy phân tán 1 cách chính xác và hiệu quả. Dữ liệu hoạt động một cách độc lập, không cần phải theo dõi xử lý các tác vụ, xử lý lỗi.

Có thể thực hiện mô hình Mapreduce trên nhiều ngôn ngữ (Java,C++,Python,Perl,Ruby,C) với các thư viện tương ứng.

4.1.3 Nguyên tắc hoạt động của Mapreduce

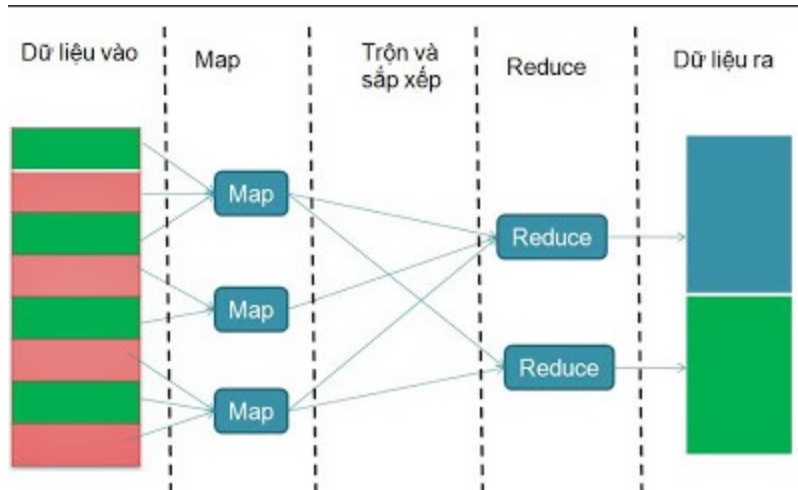
Mapreduce hoạt động gồm 2 quá trình thực hiện 2 hàm "Map" và "Reduce"

Ý tưởng chính của Mapreduce chính là thực hiện việc "Chia để trị"

- Chia vấn đề cần xử lý (dữ liệu) thành các phần nhỏ để xử lý
- Xử lý các vấn đề nhỏ đó 1 cách song song trên các máy tính phân tán hoạt động độc lập
- Tổng hợp các kết quả thu được để đưa ra kết quả cuối cùng

Như vậy toàn bộ quá trình mapreduce có thể hiểu như sau

- Đọc dữ liệu đầu vào
- Thực hiện xử lý các phần dữ liệu vào (xử lý từng phần một) (Thực hiện hàm Map)
- Trộn và sắp xếp các kết quả thu được từ các máy tính làm sao để được kết quả tiện lợi nhất so với mục đích của quá trình
- Tổng hợp các kết quả trung gian thu được từ các máy tính phân tán (Thực hiện hàm reduce)
- Đưa ra kết quả cuối cùng



Hình 4 : Sơ đồ hoạt động của Mapreduce

4.1.4 4.Chi tiết 2 hàm Map và Reduce

Thay vì định nghĩa dữ liệu dưới dạng bảng giá trị có quan hệ, Mapreduce thực hiện định nghĩa dữ liệu dưới dạng các cặp gồm $\langle \text{key}, \text{value} \rangle$

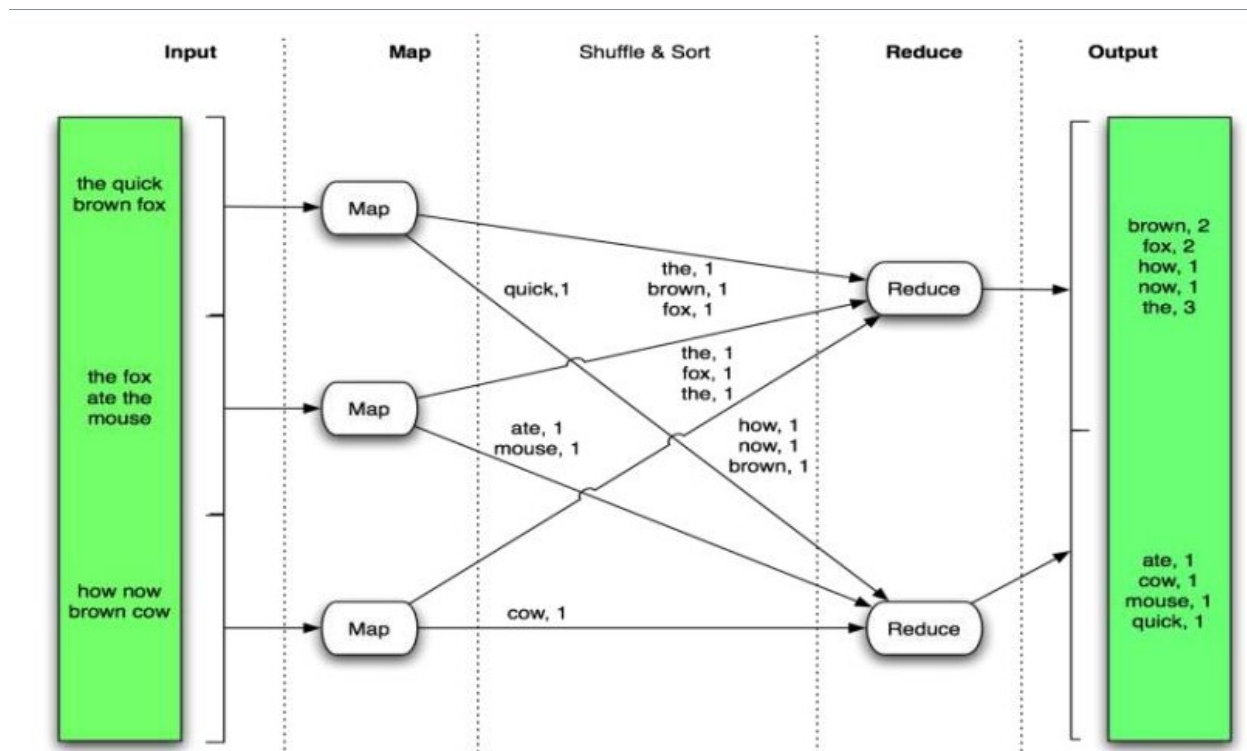
Đối với 1 tệp tin "key" có thể là tên của tệp tin đó còn "value" có thể là nội dung của tệp. Một ví dụ khác "key" là địa chỉ 1 trang web còn value là số lần người dùng truy cập trang web đó. Hai hàm Map và Reduce tập trung xử lý dữ liệu dưới dạng các cặp $\langle \text{key}, \text{value} \rangle$ như trên

Hàm Map: Dữ liệu được đưa vào hàm map là các dữ liệu đã được phân nhỏ thành các phần. Đầu vào của hàm Map là các cặp $\langle k_1, v_1 \rangle$. Sau khi xử lý toàn bộ dữ liệu đầu vào (gồm nhiều phần sau khi được phân nhỏ) kết quả thu được là tập hợp gồm các cặp $\langle k_2, v_2 \rangle$. Các dữ liệu này được gọi là các dữ liệu trung gian

Các dữ liệu trung gian này có thể được ghép lại với nhau theo danh sách các khóa để thuận tiện cho quá trình reduce sau này

Hàm Reduce: Từ dữ liệu đầu ra của hàm map (gồm danh sách các cặp $\langle k_2, v_2 \rangle$) của các máy tính phân tán, hàm reduce thực hiện việc tổng hợp các giá trị này lại. Kết quả đầu ra là các cặp $\langle k_3, v_3 \rangle$ đã được xử lý

Quá trình thực hiện mapreduce với bài toán "WordCount"



Hình 5 : Ví dụ minh họa cách Mapreduce hoạt động.

Hàm Map:

Input: 1 dòng văn bản

Output: Danh sách các cặp <key,value> ứng với từng chữ trong dòng văn bản đó. Trong đó "key" là chữ, value=1.

Hàm Reduce:

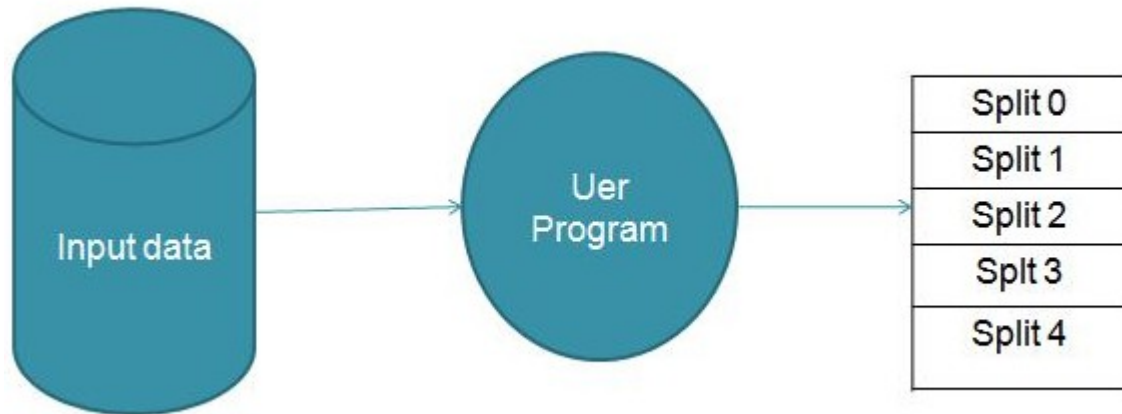
Input : danh sách các cặp key, giá trị đếm được của mỗi từ.

Output: key=từ trong cả đoạn, value=số lượng từ tương ứng trong đoạn.

4.1.5 Thực thi Mapreduce trong hệ thống

-Phân nhỏ dữ liệu đầu vào

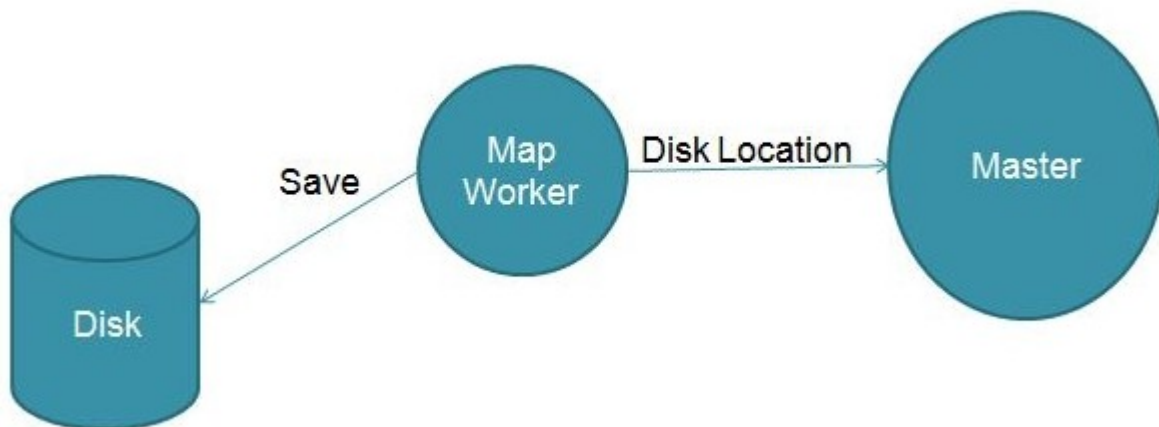
Thông qua thư viện Mapreduce ứng với từng ngôn ngữ , chương trình có nhiệm vụ phân mảnh tệp dữ liệu đầu vào. Dữ liệu vào được chia thành các phần nhỏ.



Hình 6 : Phân nhỏ dữ liệu đầu vào

-Sao chép chương trình

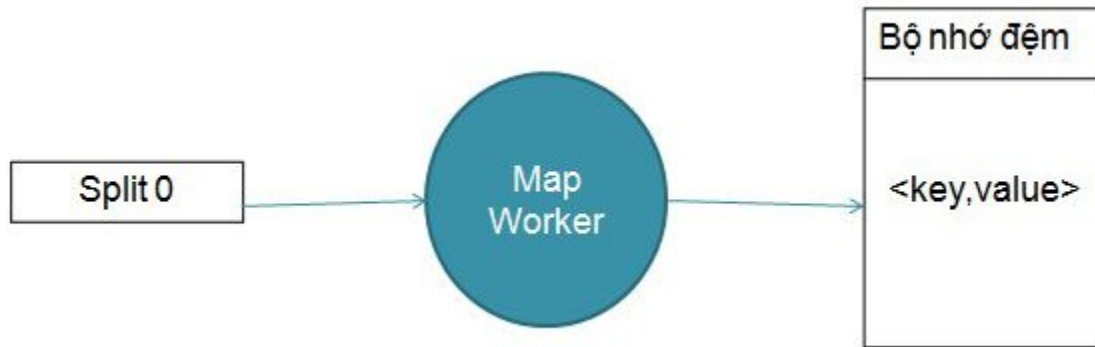
Chương trình mapreduce làm nhiệm vụ sao chép chương trình chạy thành các tiến trình song song lên các máy tính phân tán. Các máy gồm có Master và Worker. Trong đó máy Master làm nhiệm vụ điều phối sự hoạt động của quá trình thực hiện Mapreduce trên các máy Worker. Các máy Worker làm nhiệm vụ thực hiện quá trình Map và Reduce với dữ liệu mà nó nhận được



Hình 7 : Sao chép chương trình

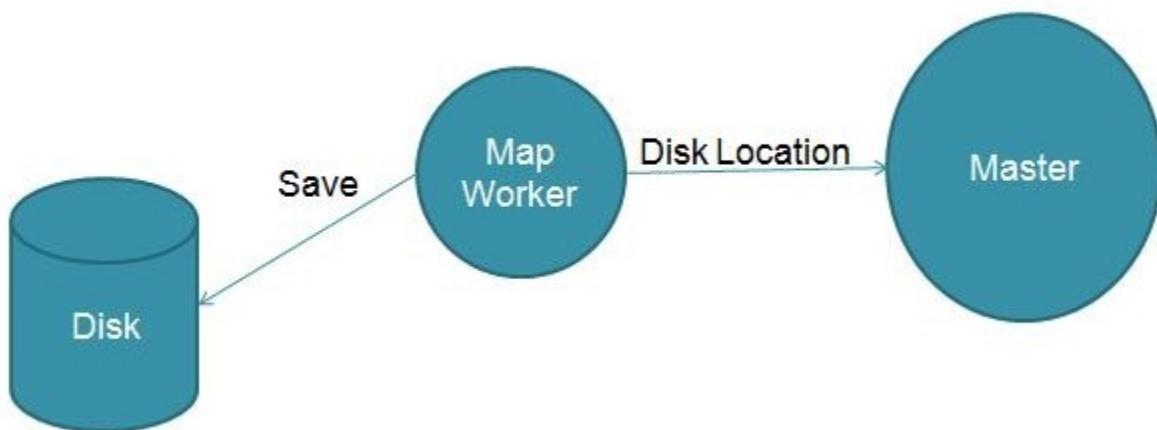
-Thực hiện hàm Map

Máy master sẽ phân phối các tác vụ Map và Reduce vào các worker đang rảnh rỗi. Các tác vụ này được Master phân phối cho các máy dựa trên vị trí của dữ liệu liên quan trong hệ thống. Máy Worker khi nhận được tác vụ Map sẽ đọc dữ liệu mà nó được nhận từ phân vùng dữ liệu đã gán cho nó và thực hiện hàm Map. Kết quả đầu ra là các cặp <key,value> trung gian. Các cặp này được lưu tạm trên bộ nhớ đệm của các máy.



Hình 8 : Thực hiện hàm Map cho ra kết quả <key,value>.

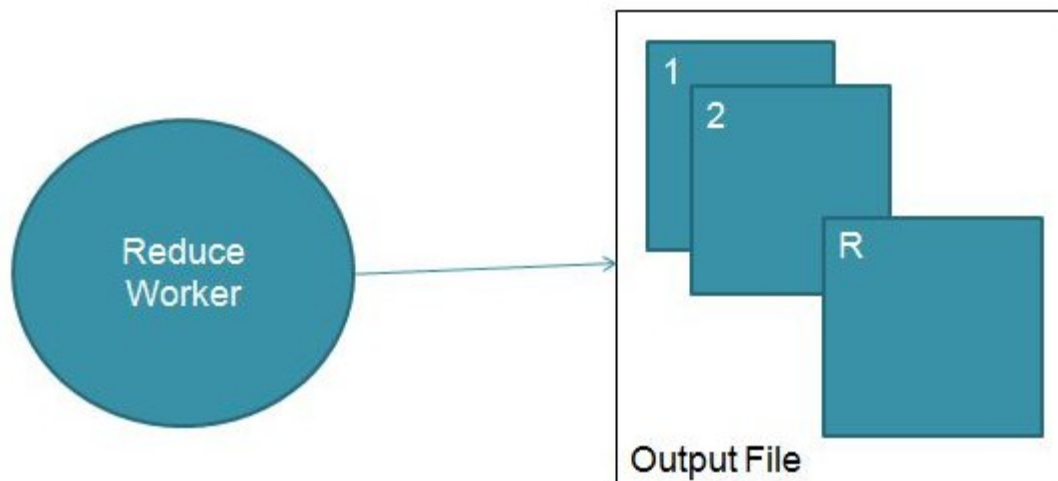
-Sau khi thực hiện xong công việc Map . Các máy Worker làm nhiệm vụ chia các giá trị trung gian thành R vùng (tương ứng với R tác vụ Reduce) lưu xuống đĩa và thông báo kết quả , vị trí lưu cho máy Master



Hình 9 : Thực hiện hàm Reduce và thông báo kết quả cho Master.

-Thực thi tác vụ Reduce

Master sẽ gán các giá trị trung gian và vị trí của các dữ liệu đó cho các máy thực hiện công việc Reduce. Các máy reducer làm nhiệm vụ xử lý sắp xếp các key, thực hiện hàm reduce và đưa ra kết quả cuối.



Hình 10 : Thông báo chương trình mapreduce hoàn tất và kết quả được lưu trữ trên R tập tin.

-Thông báo kết quả.

Master sẽ kích hoạt thông báo cho chương trình người dùng quá trình mapreduce đã hoàn tất. Kết quả đầu ra được lưu trữ trên R tập tin.

4.2 Hướng dẫn tìm kiếm và sử dụng Splunk hiệu quả

Chìa khóa để tạo một câu lệnh tìm kiếm hiệu quả đó chính là tận dụng lợi thế của index. Index của Splunk là một kho từ lớn và nhân tố ảnh hưởng tới kết quả tìm kiếm, đó là có bao nhiêu event được lấy ra từ disk.

4.2.1 Một số điều cần lưu ý khi tìm kiếm dữ liệu trong Splunk:

- Splunk không phân biệt chữ hoa, thường. Các từ ngữ tìm kiếm như error, ErRoR, ERROR đều trả về kết quả tìm kiếm như nhau.
- Splunk truy vấn dữ liệu tại một thời gian cụ thể.
- Có thể kết hợp các từ khóa tìm kiếm với Boolean (AND , OR ,NOT..) hoặc nhóm các điều kiện với nhau để tìm kiếm hiệu quả hơn. Boolean khi sử dụng phải viết hoa.
- Từ khóa tìm kiếm phải nguyên 1 từ, không phải 1 phần của từ. Tìm kiếm từ khóa “foo” sẽ không khớp với kết quả “foobar”.
- Từ khóa là những từ được bao quanh bởi khoảng cách hoặc dấu chấm câu. Ví dụ 1 đoạn log 2014-02-14 Hello world thì từ khóa được index là 2014,02,14,Hello,world.
- Con số chưa phải là định dạng số cho tới khi nó được phân tích tại thời điểm tìm kiếm.
- Tên của các field phải viết thường. Ví dụ: host=hoasen sẽ hoạt động, Host=hoasen sẽ không hoạt động.

4.2.2 Tìm hiểu về Boolean và nhóm các điều kiện

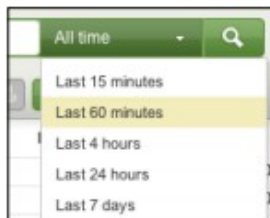
- AND kết quả tìm kiếm phải thỏa cả hai giá trị. Ví dụ : error AND mary.
- OR kết quả tìm kiếm chỉ cần thỏa 1 hay cả hai giá trị. Ví dụ : error OR mary
- NOT áp dụng cho điều kiện tìm kiếm tiếp theo đó. Ví dụ: error NOT mary. Kết quả sẽ tìm kiếm các event có từ error và không chứa từ mary.
- " " dùng để tìm một câu theo đúng thứ tự. Ví dụ: "Out of order" . Kết quả tìm kiếm sẽ trả về câu theo đúng thứ tự. Nếu không dùng " " với câu khi tìm kiếm, kết quả tìm kiếm có thể sẽ không đúng theo thứ tự các từ trong câu.
- () dùng để nhóm các điều kiện. ví dụ (bob AND (error OR mary)) AND NOT debug
- = được dành riêng để xác định các fields
- [] dùng để thực hiện subsearch(tìm kiếm con)

4.2.3 Sử dụng * để tìm kiếm 1 cách hiệu quả

- Mặc dù index dựa vào từ để tìm kiếm nhưng ta có thể dùng * khi ta không biết chính xác từ đó.
 - Nên sử dụng * sau cùng, sau khi đã áp dụng các từ khóa tìm kiếm trước mà vẫn không tìm được kết quả như ý.
- Ví dụ: Bob* sẽ cho kết quả tìm kiếm Bobby

4.2.4 Tìm kiếm các sự kiện bằng thời gian

- Có thể tùy chỉnh thời gian để xem trong khoảng thời gian bao nhiêu phút đã có bao nhiêu sự kiện xảy ra

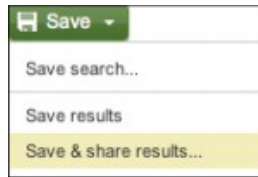


Hình 11 : Ví dụ tìm các sự kiện xảy ra trong 60 phút trước.

- Có thể dùng lệnh tìm kiếm bằng thời gian trên thanh search
 - + Để tìm kiếm error ảnh hưởng user bob trong 60 phút vừa qua, sử dụng earliest = -60m bob error
 - + Để tìm kiếm error ảnh hưởng user bob trong 3 giờ trước, sử dụng earliest = -3h@h bob error
 - + Để tìm kiếm error ảnh hưởng user bob ngày hôm qua, sử dụng earliest = -1d@d latest = -0d@d bob error
 - + Để tìm kiếm errors ảnh hưởng user bob từ thứ hai lúc nửa đêm, sử dụng earliest = -0@w1 bob error

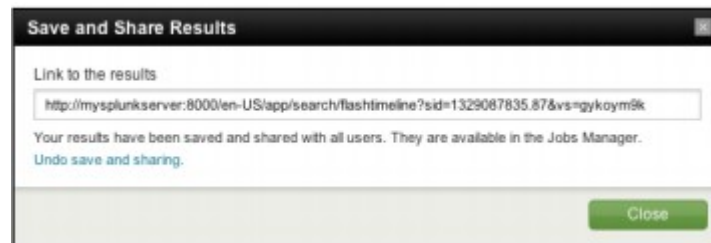
4.2.5 Chia sẻ kết quả tìm kiếm với người khác

- Sau khi tìm được các kết quả mong muốn ta có thể nhấn chọn Save& share result từ menu Save



Hình 12 : Lưu và chia sẻ kết quả tìm kiếm.

-Nó sẽ mở ra panel Save and Share Results

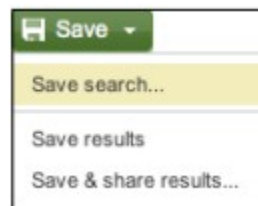


Hình 13 : Kết quả có thể được chia sẻ dưới dạng link.

-Phía dưới dòng Link the results là link URL đến kết quả tìm kiếm mà ta muốn chia sẻ . Chỉ cần copy link URL và gửi cho người ta cần chia sẻ.

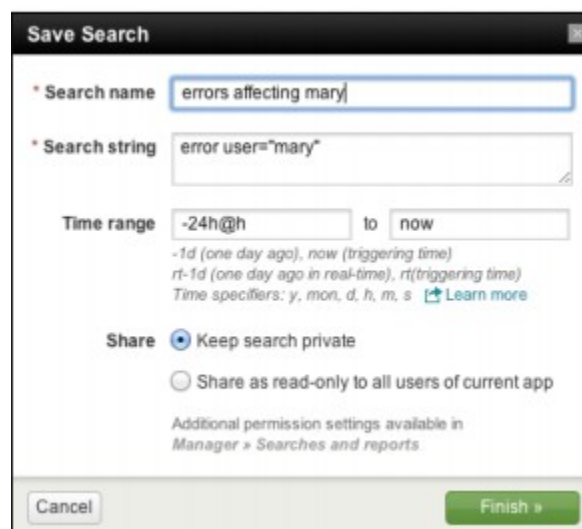
4.2.6 Lưu kết quả tìm kiếm để sử dụng lại

-Chọn Save search từ menu Save



Hình 14 : Lưu kết quả tìm kiếm

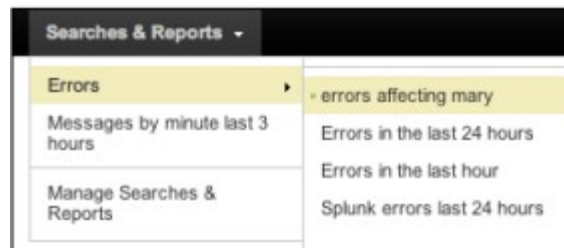
-Cửa sổ Save Search xuất hiện:



Hình 15 : Những kết quả phải thỏa những điều kiện được thiết lập mới được lưu.

-Nhập vào giá trị cho Search name, trong hình là ,errors affecting mary. Thời gian là từ 24h trước. Có thể tùy chọn private hoặc chia sẻ cho user khác.

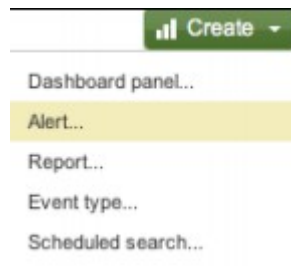
-Kết quả search sẽ xuất hiện trong menu Searches & Report dưới Error



Hình 16 : Kết quả tìm kiếm sẽ xuất hiện trong menu Search & Report.

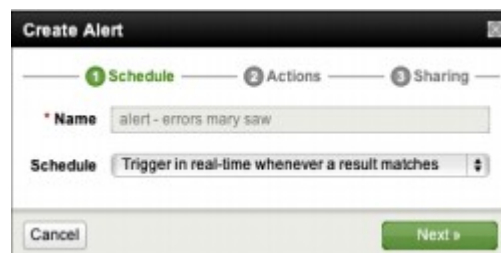
4.2.7 Tạo alerts từ kết quả tìm kiếm

-Từ menu Create chọn alert



Hình 17 : Tạo một alert.

-Menu Create Alert xuất hiện



Hình 18 : Đặt tên alert và điều kiện để kích hoạt alert.

Hình 19 : Chạy kết quả tìm kiếm event mỗi giờ, khởi động arlet khi kết quả tìm kiếm lớn hơn 0.

+Option Trigger in real-time whenever result matches có nghĩa là kết quả tìm kiếm sẽ chạy theo thời gian thực và sẽ tự động cảnh báo khi tìm thấy event.

Create Alert

1 Schedule 2 Actions 3 Sharing

* Name alert - errors mary saw

Schedule Run on a schedule once every...

Hour

Search will run over selected schedule interval.

Trigger If Number of results

Is greater than 0

Cancel Next >

Hình 19 : Chạy kết quả tìm kiếm event mỗi giờ, khởi động alert khi kết quả tìm kiếm lớn hơn 0.

+Option Run on a schedule once every... : làm xuất hiện nhiều option khác

Create Alert

1 Schedule 2 Actions 3 Sharing

* Name alert - errors mary saw

Schedule Monitor in real-time over a rolling window of...

5 minute

Trigger If Number of results

Is greater than 5

Cancel Next >

Hình 20 : Nếu số lượng event tìm được trong 5 phút bé hơn 5 thì kích hoạt alert.

+Monitor in real-time over a rolling window of...: rất hữu ích trong việc tạo cảnh báo. Ví dụ nếu số lượng event diễn ra trong 1 phút dưới 100 thì gửi cảnh báo.

-Sau khi tùy chỉnh xong, nhấn Next để qua phần Action

Hình 21 : Các option trong Alert.

-Bảng action giúp chúng ta quyết định sẽ làm gì đối với kết quả của alert. Một số option:

+Send mail:gửi mail dựa trên danh sách e-mail đã nhập.

+Run a script: chạy script với kết quả của quá trình tìm kiếm.

+Show triggered alerts in Alert manager: Liệt kê các alerts phổ biến trong Saved search

-Sau khi xong các tùy chọn, có thể nhấn Next và thực hiện chức năng Sharing nếu có nhu cầu.

4.3 Table , chart trong Splunk

4.3.1 Giới thiệu một số hàm cơ bản trong việc tạo table

-Hàm pipe (|) trong splunk dùng để đưa kết quả output của 1 tiến trình thành input cho 1 tiến trình khác.

-Một số hàm để tạo fields :eval, rex

-Hàm lọc event: head, where

-Hàm thay thế event với report : top, stats

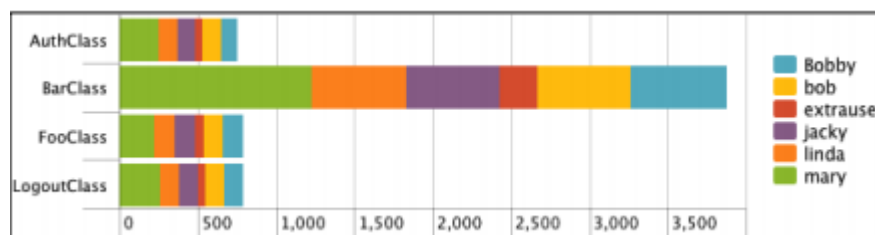
4.3.2 Ví dụ về một table cụ thể:

-Sử dụng câu lệnh search: source="impl_splunk_gen" error | top logger. Kết quả tìm kiếm trả về là một table

	logger ↕	count ↕	percent ↕
1	BarClass	242	63.185379
2	FooClass	49	12.793734
3	AuthClass	47	12.271540
4	LogoutClass	45	11.749347

Hình 22 : Một table dạng số.

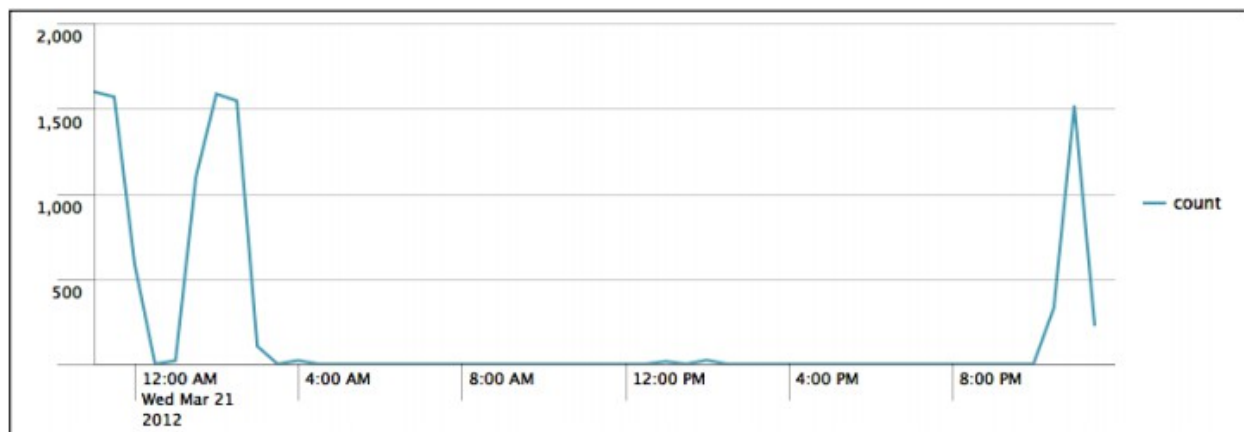
-Sau đó có thể nhấn vào icon chart phía trên table để chuyển đổi table thành chart



Hình 23 : Một table dạng chart.

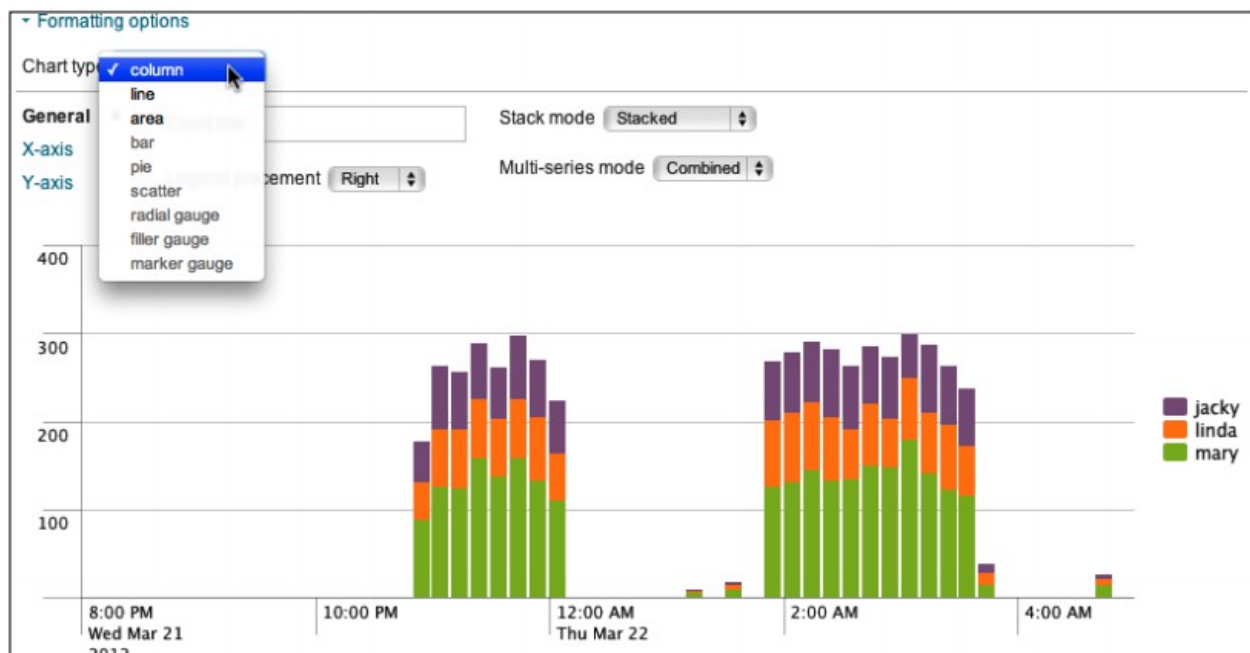
-Một dạng khác của chart là timechart, dùng để biểu diễn dữ liệu số theo thời gian

+Gõ câu lệnh sourcetype= "impl_splunk_gen" error |timechart count



Hình 24 : Biểu đồ chart dữ liệu nhận được trong một khoảng thời gian.

-Formating options phía trên chart cho ta nhiều lựa chọn tùy biến

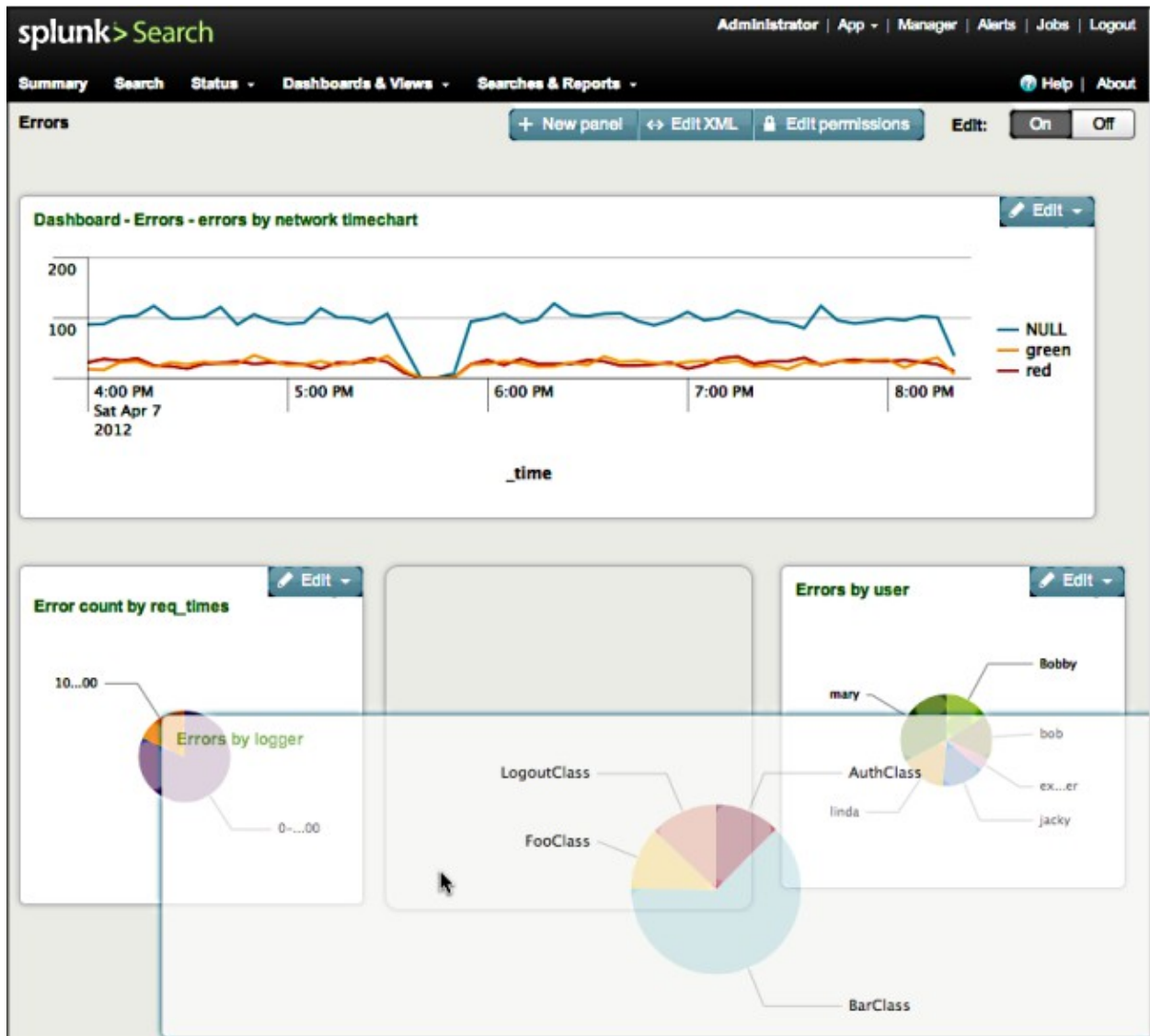


Hình 25 : Các tùy chọn formatting của chart.

4.4 Dashboard

Dashboard là công cụ giúp chúng ta nắm bắt, nhóm và tùy chỉnh các bảng, biểu đồ một cách hiệu quả. Nó chứa nhiều bảng thông tin, mỗi bảng chạy một truy vấn khác nhau. Mỗi dashboard có 1 link URL riêng biệt, dễ dàng trong việc chia sẻ. Dashboard có thể tùy biến, tùy chỉnh hiển thị các giá trị cần thiết, thanh tìm kiếm trong Dashboard được loại bỏ. Nhiều công ty sử dụng dashboard trên máy chiếu của họ để đưa 1 cái nhìn lướt qua về môi trường của công ty tới khách hàng. Dashboard còn có thể lập lịch để send file pdf bằng email

Ví dụ về một dashboard cơ bản:



Hình 26 : Ví dụ về một dashboard cơ bản.

4.4.1 SEARCH LANGUAGE trong Splunk

Lệnh	Mô tả	Xem thêm
abstract	Đưa ra các bản tóm tắt cho mỗi kết quả tìm kiếm.	Highlight
accum	Giữ hoạt động của 1 số trường số cụ thể.	Delta, streamstats, trendline
Addcoltotals	Tính toán sự kiện chứa các trường số cho sự kiện trước đó.	Stats
Addinfo	Thêm 1 trường chứa các thông tin về các lệnh tìm kiếm thông thường của lệnh tìm kiếm hiện tại.	Search
Addtotals	Tính tổng các trường số cho mỗi kết quả.	Stats
Append	Thêm các kết quả của subsearch cho kết quả hiện tại	appendcols, appendcsv, appendlookup, join, set
appendcols	Thêm vào trường của kết quả subsearch vào kết quả hiện tại.	Append, join, set, appendcsv
Audit	Trả lại những thông tin được chứa trong audix index.	
chart	Trả lại kết quả trong 1 bản, dữ liệu đầu ra là dạng biểu đồ.	bucket, sichart, timechart
Cluster	Gom, tổng hợp những sự kiện tương tự.	anomalies, anomalousvalue, cluster, kmeans, outlier
Collect, stash	Đem những kết quả tìm kiếm vào index tóm tắt.	overlap
concurrency	Dùng những trường tồn tại để kiểm soát sự kiện đồng thời của từng sự kiện.	timechart
convert	Chuyển đổi trường giá trị sang giá trị số.	eval
crawl	Thu thập file hệ thống làm tài nguyên cho index mới.	
Dbinspect	Trả lại thông tin cho 1 index cụ thể nào đó.	
dedup	Xóa các chuỗi kết quả ứng với các tiêu chí cụ thể.	uniq
Delete	Xóa các sự kiện cụ thể hoặc tìm kiếm kết quả	
Diff	Trả về sự khác nhau giữa 2 kết quả tìm kiếm.	
erex	Cho phép chỉ định ví dụ hoặc đếm giá trị ví dụ để tự động xuất ra những trường có giá trị tương đương.	extract, kvform, multikv, regex, rex, xmlkv

4.4.2 Định nghĩa chức năng một số hàm tìm kiếm

Các lệnh tìm kiếm

Lệnh	Mô tả	Xem thêm
------	-------	----------

Eval	Tính toán các hàm và đẩy giá trị vào 1 trường.	where
Eventcount	Trả về số sự kiện trong index.	Dbinspect
Extract, kv	Xuất ra trường giá trị từ kết quả tìm kiếm.	kvform, multikv, xmlkv, rex
Eventstats	Chèn tóm tắt vào tất cả các giá trị tìm kiếm.	stats
filldown	Thay thế giá trị rỗng với giá trị cuối cùng không phải là rỗng.	Fillnull
fillnull	Thay thế giá trị rỗng với 1 giá trị cụ thể.	
findtypes	Tạo ra 1 danh sách đề nghị các loại sự kiện.	typer
format	Lấy kết quả của Subsearch và định dạng của nó vào 1 kết quả riêng.	
Genttimes	Khởi tạo thời gian tìm kiếm kết quả.	
Head	Trả về kết quả đầu tiên của 1 kết quả tìm kiếm.	Reverse, tail
history	Trả về lịch sử tìm kiếm, định dạng như là 1 danh sách sự kiện hoặc như là 1 bảng.	search
Input	Thêm dữ liệu vào splunk hoặc làm vô hiệu hóa các nguồn từ splunk.	
Multisearch	Thực hiện 1 lúc nhiều quá trình tìm kiếm.	Append,join
overlap	Tìm sự kiện trong index tóm tắt mà bị trùng thời gian hoặc bị mất.	collect
rangemap	Thiết lập trường khoảng các tên	
Rare	Hiển thị các giá trị ít nhất trong 1 trường.	si rare, stats, top
Replace	Thay thế giá trị 1 trường cụ thể với 1 giá trị mới cụ thể.	
return	Chỉ ra giá trị để trả về từ 1 subsearch	format, search
run	Hiển thị script	
sort	Sắp xếp kết quả tìm kiếm bởi 1 trường cụ thể.	reverse
table	Tạo ra 1 bảng sử dụng các trường cụ thể	fields
tail	Trả về giá trị cuối cùng .	Head, reverse
uniq	Xóa các tìm kiếm có trùng với kết quả trước đó.	dedup

Các định dạng biến ngày tháng, giờ.

Biến thời gian

Biến	Mô tả
%Ez	Splunk chỉ ra vùng thời gian trong phút.
%H	Giờ (định dạng 24h) là số decimal gồm từ 00 tới 23
%I	Giờ (định dạng 12) bao gồm số từ 01-12
%k	Giống %H nhưng số 0 ở đầu bị thay thế bằng khoảng trắng (0 tới 23)
%M	Phút, là số decimal (00 tới 59)
%p	AM hoặc PM
%S	Giây , là số decimal (00 tới 60)
%T	Thời gian trong 24 giờ , định dạng (%H:%M:%S)

Biến dữ liệu

Biến	Mô tả
%F	Định dạng %Y-%m-%d (theo chuẩn ISO 8601 định dạng ngày tháng)
%A	Cả tuần (chủ nhật tới thứ 2)
%d	Ngày trong tháng, là số decimal gồm các số từ 01 tới 31
%e	Như %d nhưng số 0 đầu tiên thay bằng khoảng trắng (từ 1 tới 31)
%j	Số ngày trong năm , là số decimal gồm các số từ 001 tới 366
%w	Thứ trong tuần bằng số decimal (Sunday=0,..... Satuday =6)

Biến tháng

Biến	Mô tả
%b	Tên viết tắt tên tháng (Jan, Feb, etc.)
%B	Tên đầy đủ của tháng . (January, February, etc.)
%m	Tháng đặt theo số decimal (01 – 12)

Biến năm

Biến	Mô tả
%y	Số năm theo dạng decimal (00-99)
%Y	Số năm theo dạng đầy đủ (2014)

4.4.3 Một số cú pháp search language trong splunk:

Chú thích:

*(...):đặt ở đầu câu lệnh search nhằm báo rằng đã có tác vụ tìm kiếm nào đó trước khi đưa vào pipe

* |: đặt ở đầu câu lệnh search nhằm ngăn không cho thêm vào trước câu lệnh tìm kiếm.

+administrative

Xem thông tin của index “audit”	Index=_audit audit
Thu thập thông tin root và thư mục gốc sau đó add kết quả tìm được vào file inputs.conf	crawl root="/;/Users/" input add
Hiển thị biểu đồ trong khoảng thời gian một ngày	dbinspect index=_internal span=1d
Trả về giá trị “host” cho các sự kiện trong index “_internal”	metadata type=hosts index=_internal
Trả về thông tin typehead cho sources trong index “_internal”	typehead prefix=source count=10 index=_internal

+alerting

Gửi kết quả tìm kiếm tới một địa chỉ mail cụ thể	... sendmail to="tuan@splunk.com"
--	-------------------------------------

+add

Lưu lại số lần xuất hiện của “total_count”	... accum count AS total_count
Thêm thông tin về tìm kiếm cho mỗi event	... addinfo
Tìm kiếm các event “404” và thêm các fields trong mỗi sự kiện vào các kết quả tìm kiếm trước.	... appendcols [search 404]
So sánh biến ‘count’ với giá trị trước đó của nó và lưu kết quả vào ‘countdiff’	... delta count AS countdiff
Trích xuất giá trị “7/01” và đưa vào thuộc tính ngày tháng	... erex monthday examples="7/01"
Thiết lập tốc độ về dạng distance/time	... eval velocity=distance/time
Trích xuất giá trị và thiết lập lại quá trình trích xuất field từ ổ đĩa	... extract reload=true
Trích xuất giá trị giới hạn bởi “ ;” và “=:”.	... extract pairdelim=" ;", kvdelim="=:", auto=f
Thêm thông tin về địa chỉ ip	... iplocation
Trích xuất giá trị từ “eventtype” nếu file đó tồn tại	... kvform field=eventtype
Đặt range là “green” nếu giá trị date_second từ 1-30; “blue” nếu từ 31-39, “red” nếu từ 40-59 và “gray” là các giá trị còn lại.	... rangemap field=date_second green=1-30 blue=31-39 red=40-59 default=gray
Tính toán sự liên quan của phép tính tìm kiếm và sắp xếp kết quả theo thứ tự giảm dần	Disk error relevancy sort -relevancy
Trích field dữ liệu “author” từ định dạng XML hoặc JSON (áp dụng cho sách)	... spath output=author path=book{@author}
Thêm field “comboIP”. Giá trị của nó = “”sourceIP” + “/” + “destIP””	... strcat sourceIP “/” destIP comboIP

+convert

Chuyển đổi giá trị của tất cả field thành giá trị số ngoại trừ giá trị của field “foo”	... convert auto(*) none(foo)
--	---------------------------------

Thay đổi giá trị memory trong field “virt” thành Kilobytes.	... convert memk(virt)
Thay đổi định dạng đơn vị của syslog(D+HH:MM:SS) thành giây	... convert dur2sec(delay)
Chia giá trị “foo” thành nhiều giá trị	... makemv delim=”:” allowempty=t foo
Kết hợp giá trị của field gửi thành một giá trị và hiển thị 10 giá trị đầu tiên(Dùng trong hoạt động sendmail)	Eventtype=”sendmail” nomv senders top senders
+filter	
Giữ field “host” và “ip” và hiển thị theo thứ tự “host”, “ip”	... fields + host, ip
Xóa field “host” và “ip”	... fields – host, ip
+modify	
Xây dựng biểu đồ thời gian các sự kiện web của host và điền các fields trống = NULL	Sourcetype=”web” timechart count by host fillnull value=NULL
Thay đổi field “_ip” thành “IPAddress”.	... rename _ip as IPAddress
Thay đổi các host có giá trị kết thúc là localhost thành localhost	... replace *localhost with localhost in host
+formatting	
Hiển thị bảng tóm tắt 5 dòng cho mỗi kết quả tìm kiếm	... abstract maxlines=5
So sánh giá trị “ip” của kết quả tìm kiếm thứ nhất và thứ ba	... diff pos1=1 ps2=3 attribute=ip
Làm nổi bật các từ “login” và “logout”	... highlight login,logout
+delete	
Xóa events có từ “invalid” trong index “imap”	Index=imap invalid delete
+summary	
Đặt events “download” trong index tên là “downloadcount”	Eventtype=”download” collect index=downloadcount
Tìm events trùng lặp trong “summary”	Index=summary overlap
+reporting	
Tính tổng các fields số của mỗi kết quả và để vào fields “sum”	... addtotals fieldname=sum
Phân tích fields số để dự đoán giá trị “is_activated”	... af classfield=is_activated
Trả về số lượng events trong index “_internal”	eventcount index=_internal
Loại bỏ các giá trị trùng lặp cùng giá trị “host” và trả về tổng số lần trùng lặp	... stats dc(host)
Tìm log truy cập và trả về 100 giá trị đầu tiên của “referrer domain”	Sourcetype=access_combined top limit=100 referer_domain stats sum(count)
Tính toán giá trị trung bình của “CPU” mỗi phút của từng “host”	... timechart span=1m avg(CPU) by host
Tính toán trung bình “CPU” và “MEM” mỗi phút trên mỗi “host”	... timechart span=1m eval(avg(CPU) * avg(MEM)) by host
Định dạng lại kết quả tìm kiếm	... timechart avg(delay) by host untable _time

	host avg_delay
+results	
Trả về những events bất thường	... anomalies
Xóa kết quả trùng cùng giá trị host	... dedup host
Join kết quả của nó với field “id”	... selfjoin id
Tìm từ ngày 25/10 đến nay	gentimes start=10/25/14
Tìm events được tạo ra bởi job với id=123.2	loadjob 123.2 events=t
Trở về 20 kết quả đầu tiên	... head 20
Trở về 20 kết quả cuối cùng	... tail 20
Hiển thị events từ file “messages.1” nếu events được indexed vào splunk	inputcsv all.csv search error outputcsv errors.csv
Xuất kết quả tìm kiếm ra file csv “mysearch.csv”	... outputcsv mysearch
+search	
Giữ kết quả tìm kiếm có giá trị “src” và “dst” định trước	Src=”10.9.165.*” OR dst=”10.9.165.8”
Tìm giá trị “URL” chứa chuỗi “404” hoặc “303” nhưng không phải cả hai	set diff [search 404 fields url] [search 303 fields url]

Tham khảo: <https://sites.google.com/site/chapterhut/hoc-tap/mon-hoc/map-reduce>

4.5 Splunk Forwarder

Nhiệm vụ của Splunk Forwarder là forward dữ liệu về Splunk server để index.

4.5.1 Các loại Forwarder:

Universal forwarder là một lightweight forwarder mới của splunk. Nó có chức năng thu thập dữ liệu từ nhiều input và forward dữ liệu tới Splunk server để index(chứa) và tìm kiếm

Light forwarder là một phiên bản nhỏ của forwarder, được lược bỏ hầu hết các tính năng của Splunk full nhằm phục vụ cho mục đích tối ưu, nó không phân tích mà chỉ forward dữ liệu tới hệ thống Splunk Enterprise hoặc hệ thống của bên thứ ba(third-party). Light forwarder nhẹ và cấu hình đơn giản. Nó ít được sử dụng ở phiên bản splunk 6.0.

Heavy forwarder là phiên bản Splunk full, với một vài tính năng được lược bỏ để tối ưu hóa. Nó là một loại forwarder, có thể phân tích dữ liệu và forward dữ liệu tới Hệ thống Splunk Enterprise khác hoặc hệ thống third-party khác. Nó không có khả năng thực hiện tìm kiếm phân phối. Nhiều chức năng mặc định của nó như splunk web có thể bị disable để tối ưu hơn. Nó cũng có thể index(chứa) dữ liệu nội bộ trong khi forward dữ liệu tới một Splunk index khác. Nó chiếm gấp đôi dung lượng bộ nhớ, CPU so với Light Forwarder và cấu hình phức tạp hơn.

4.5.2 So sánh universal forwarder với Splunk full:

Mục đích duy nhất của universal forwarder là forward dữ liệu. Nó không thể index dữ liệu hoặc tìm kiếm dữ liệu. Universal forwarder có một số hạn chế:

+Không có tính năng tìm kiếm , index(chứa dữ liệu), hay tính năng cảnh báo.

- +Không phân tích dữ liệu
- +Không đẩy dữ liệu ra ngoài dưới dạng syslog
- +Không giống như Splunk full, nó không có hỗ trợ Python.

Universal forwarder được tối ưu chỉ bao gồm các thành phần cần thiết để forward dữ liệu tới Splunk indexers. Universal forwarder có thể nói là 1 công cụ tốt nhất để forward dữ liệu tới indexer.

4.5.3 So sánh universal forwarder với light forwarder:

- + Universal forwarder sử dụng ít CPU, chiếm ít bộ nhớ và không gian ổ đĩa.
- + Universal forwarder có tốc độ truyền dữ liệu mặc định là 256Kbps
- + Universal forwarder không hỗ trợ Python
- + Universal forwarder chỉ làm nhiệm vụ forward, không thể chuyển đổi thành Splunk full.

4.6 Một số khái niệm về các file Splunk.conf

Props.conf: Định nghĩa các sự kiện nào theo tên host, source và sourcetype

Input.conf: Điều khiển dữ liệu vào Splunk, có chức năng blacklist và whitelist, ngăn chặn hoặc cho phép loại dữ liệu nào vào splunk, tùy chọn bỏ qua không index các dữ liệu cũ, input dữ liệu bằng cách lắng nghe trên port, có thể input dữ liệu bằng scripts

Transforms.conf: nơi chuyển đổi và tra cứu các events, có thể được tham chiếu theo tên trong file props.conf, tạo ra các field

Fields.conf: nơi để add dữ liệu

Outputs.conf: là file cấu hình để splunk forward event ra ngoài.

Indexes.conf: là file quyết định nơi lưu trữ dữ liệu trên ổ đĩa, lưu giữ bao nhiêu, và trong bao lâu. Index thực chất là tên của thư mục có cấu trúc đặc biệt. Bên trong chứa thư mục con gọi là bucket và dữ liệu index.

Authorize.conf: Lưu thông tin định nghĩa vai trò và các roles. Nó ảnh hưởng đến quá trình tìm kiếm và giao diện web.

Savedsearches.conf: Nơi lưu trữ kết quả của các quá trình tìm kiếm

Time.conf: định nghĩa thời gian xuất hiện trong bảng chọn thời gian.

Commands.conf: gồm những lệnh đặc biệt cung cấp bởi app.

Web.conf: thay đổi port cho web server, chứng chỉ SSL.

4.7 Hướng dẫn cấu hình input log từ syslog server vào splunk server

+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 : # service iptables stop

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

Sửa dòng lệnh từ

```
# This file controls the state of SELinux on the system.
# SELINUX= can take one of these three values:
#     enforcing - SELinux security policy is enforced.
#     permissive - SELinux prints warnings instead of enforcing.
#     disabled - No SELinux policy is loaded.
SELINUX=enforcing
```

Hình 27 : Cấu hình mặc định của Selinux.

Thành

```
# This file controls the state of SELinux on the system.
# SELINUX= can take one of these three values:
#     enforcing - SELinux security policy is enforced.
#     permissive - SELinux prints warnings instead of enforcing.
#     disabled - No SELinux policy is loaded.
SELINUX=disabled
```

Hình 28 : Tắt selinux.

Trên máy Splunk server:

Tải gói splunk rpm 64 bit từ trang web www.splunk.com và cài đặt dưới quyền admin

Gói cài đặt để trong thư mục opt

Trên terminal, cd vào thư mục opt, gõ rpm -ivh splunk-6.0.3-204106-linux-2.6-x86_64.rpm để tiến hành cài đặt

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. Mật khẩu truy cập như mặc định sẽ là admin/changeme, chúng ta phải thay đổi mật khẩu mặc định.

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

```
# Provides UDP syslog reception
#$ModLoad imudp
#$UDPServerRun 514

# Provides TCP syslog reception
#$ModLoad imtcp
#$InputTCPServerRun 514
```

Hình 29 : Cấu hình mặc định trong file rsyslog.conf.

Thành

```
# Provides UDP syslog reception
$ModLoad imudp
$UDPServerRun 514

# Provides TCP syslog reception
$ModLoad imtcp
$InputTCPServerRun 514
```

Hình 30 : Cấu hình để mở port 514 cho syslog.

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

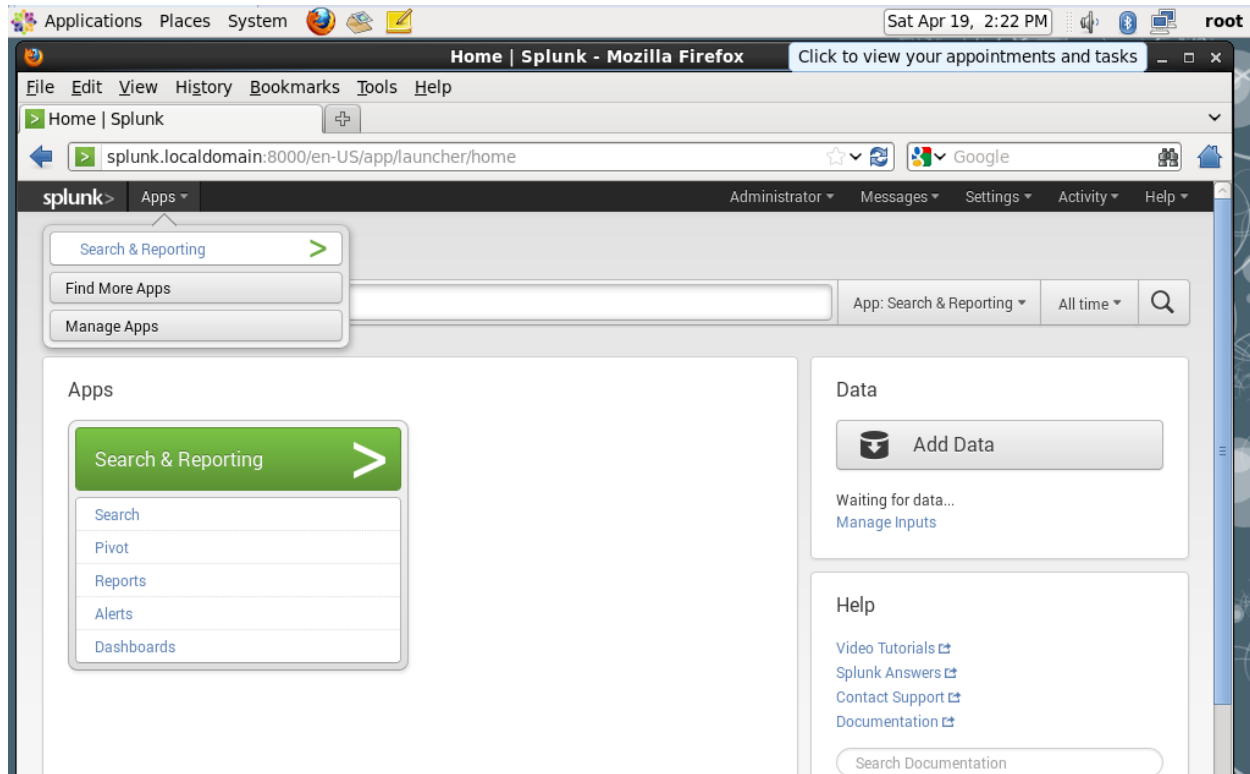
Sau đó thử switch user trên syslog server để sản sinh log.Sau đó qua kiểm tra bên splunk server xem log đó đã được send qua hay chưa.

4.8 Hướng dẫn cấu hình input log Window vào splunk server

Cài đặt splunk

Trên Splunk Server, tiến hành cài đặt app for windows trên giao diện web

Kết nối vào giao diện web serv của splunk, chọn apps ở góc trái màn hình -> manage apps

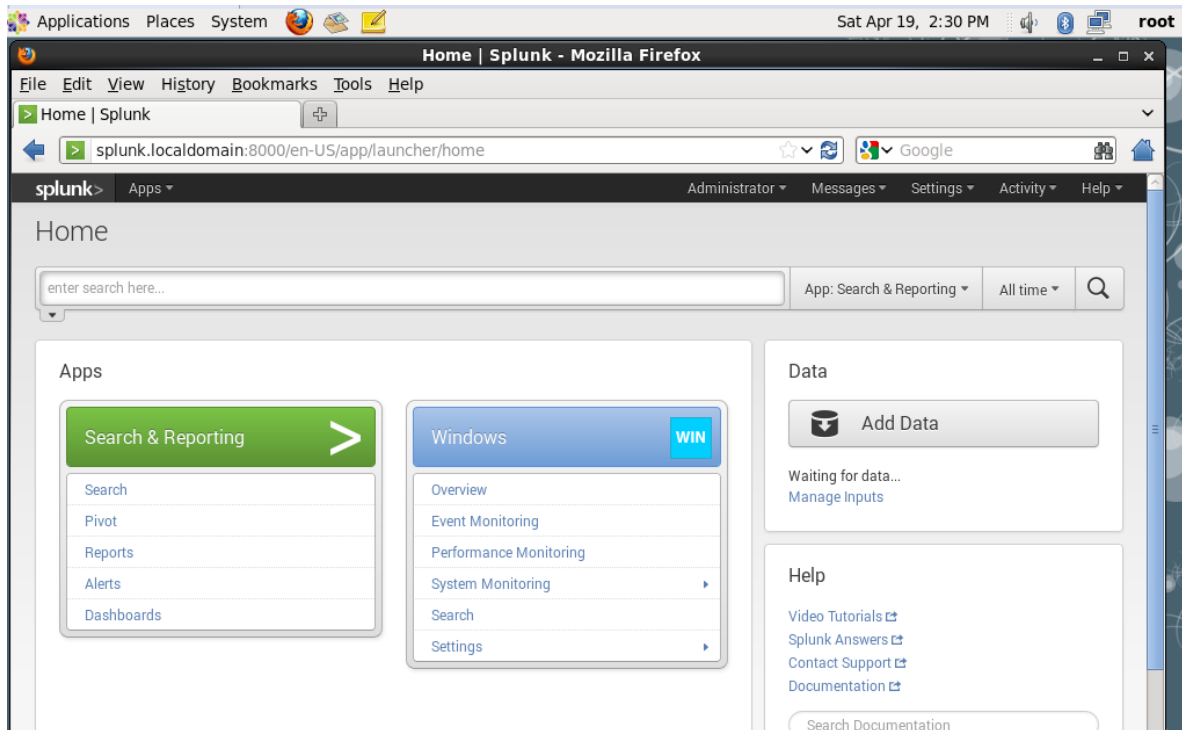


Hình 31 : Giao diện web của Splunk.

Ở đây ta tiến hành cài đặt app vào splunk từ source đã chuẩn bị.

Chọn Browse tìm đường dẫn thư mục chứa file cài đặt sau đó chọn install app from file.

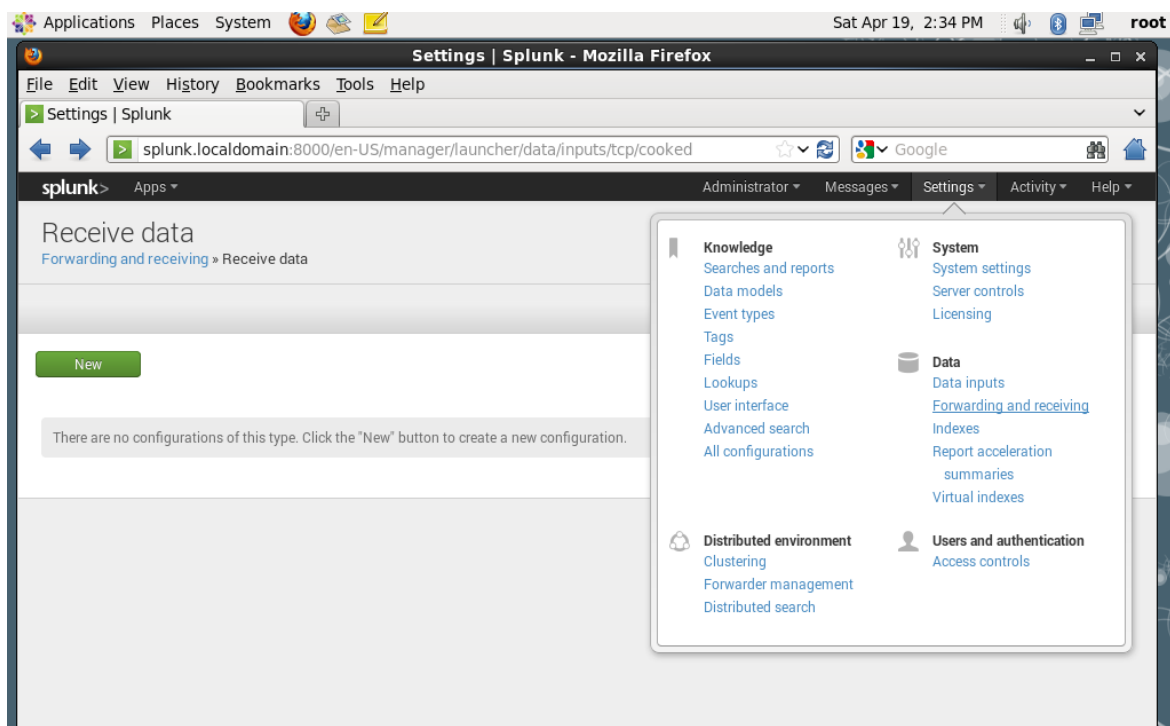
Sau khi install app restart lại splunk server để hoàn tất, apps nhận trên server. Ta có thể thấy apps cài ở homepage.



Hình 32 : Giao diện splunk đã có thêm add-on Windows.

Gán port cho splunk nhận dữ liệu từ forwarder

Ở góc phải chọn Setting chọn tiếp ở mục Data (Forwarding – Receiving)



Hình 33 : Cấu hình Forwarding and Receiving.

Chọn tiếp **Configure Receiving > Add new** để tiến hành gắn port vào. Chọn **save** để hoàn tất quá trình.

Cài đặt Splunk forwarder trên 1 máy khác để gửi log vào splunk server.

Chuẩn bị gói cài đặt:

Gói Universal Forwarder trên trang chủ của Splunk, ở đây ta sử dụng gói cho windows `splunkforwarder-6.0.2-196940-x86-release.msi`.

Sau đó tiến hành cài đặt, chọn nơi cài. Ở mục **Receiving Indexer** : gõ IP của Splunk Server, port giống như port đã tạo ở trên.

Chọn **Local Data only** để lấy log trên máy , bấm **Next**, sau đó chọn loại log mà ta cần: Ở đây ta lấy **Windows Event Logs** và **Performance Logs**.



Hình 34 : Tùy chọn các loại log mà universalforwarder sẽ gửi.

Chọn **next** , để kết thúc cài đặt.

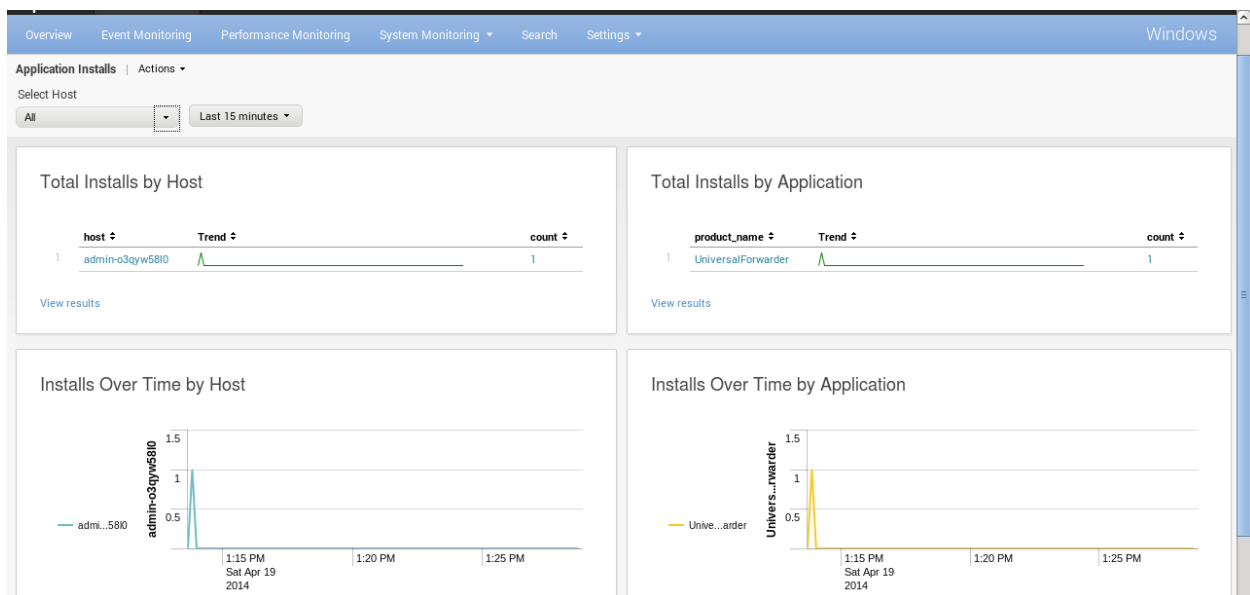
Sau đó cài đặt **Splunk Technology Add-on** trên máy forwarder. Vào trang chủ splunk down gói `Splunk_TA_windows`.

Bung gói đó ra ta sẽ được 1 thư mục cùng tên. Vào thư mục `Splunk_TA_windows\default` chép file `input.conf` vào thư mục `Splunk_TA_windows\local`

Sau đó chép thư mục `Splunk_TA_windows` vào đường dẫn cài `forwarder\etc\apps`

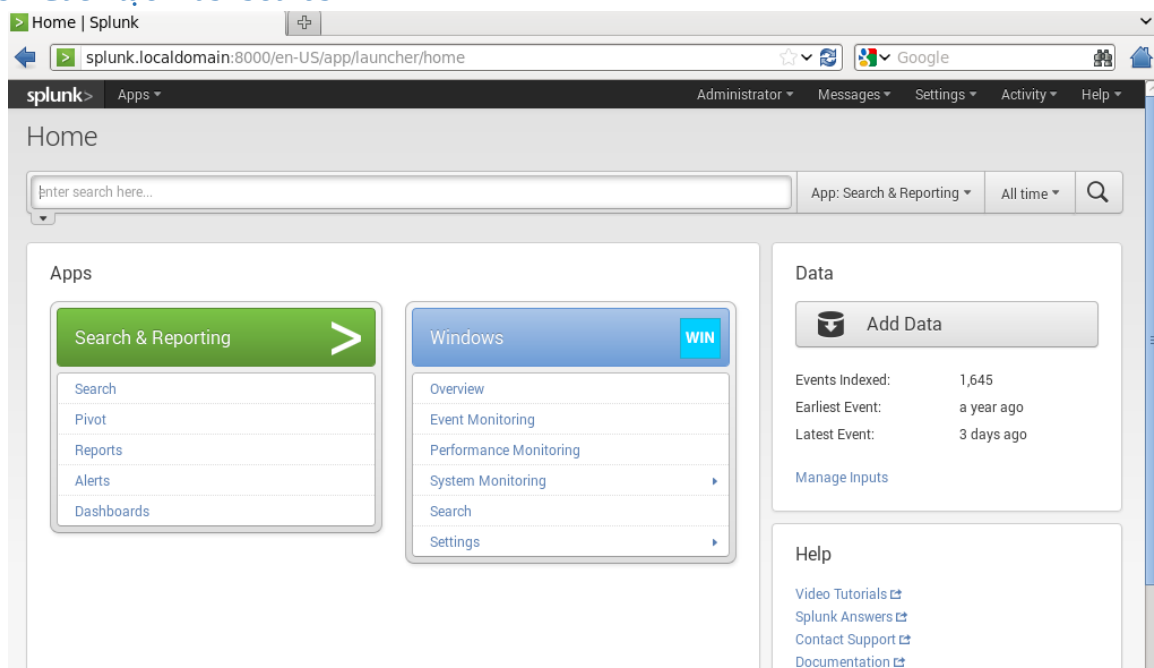
Khởi động lại máy để hoàn tất quá trình.

Vào máy chủ Splunk chọn Apps for Win để kiểm tra



Hình 35 : Splunk đã nhận được log của Window.

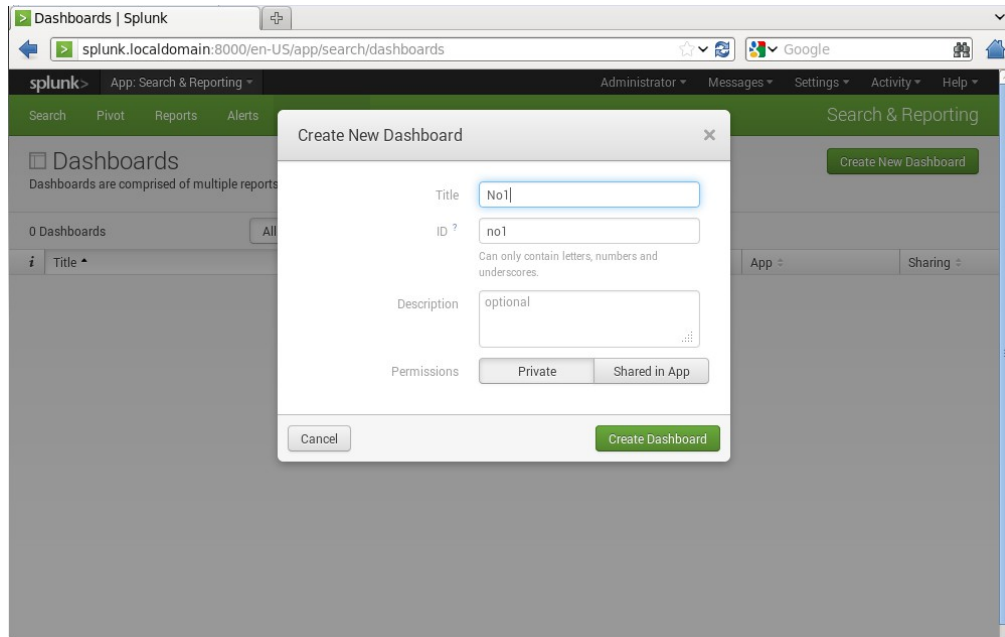
4.9 Cách tạo Dashboards



Hình 36 : Menu chính của Splunk.

Ở giao diện chính màn hình “ Search & Reporting” chọn Dashboards để tạo

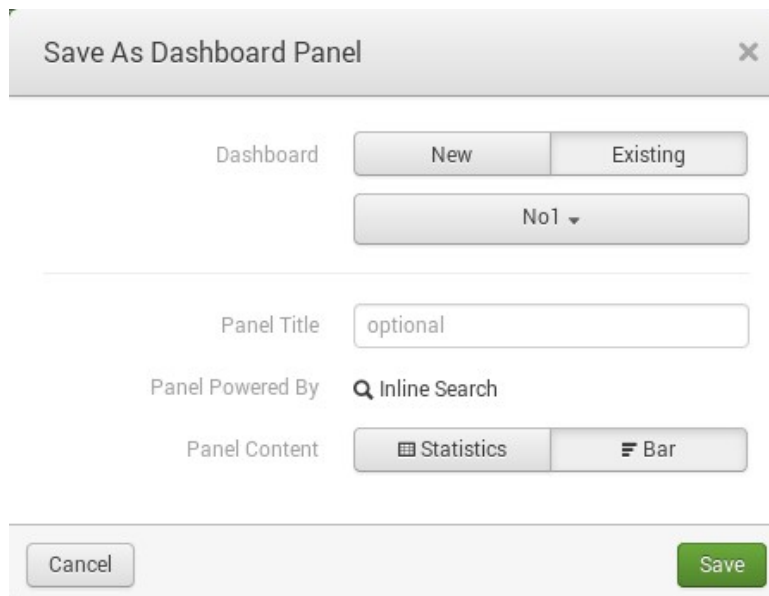
Đặt tên: No1 , và chú thích lại nếu muốn, chọn quyền cho Dashboard (Private / Share). Bấm *Create Dashboard* để khởi tạo.



Hình 37 : Tạo một Dashboard mới.

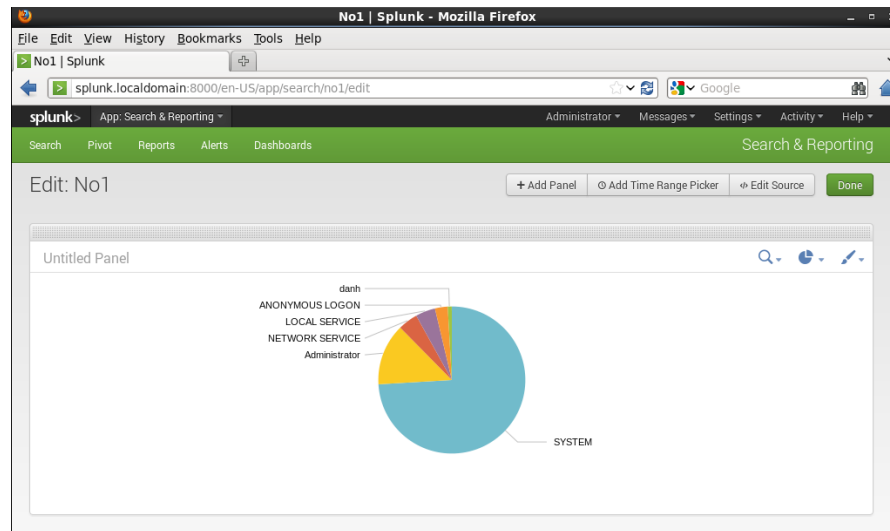
Sau khi tạo xong Dashboard, tiếp theo sẽ tạo panel để đưa lên Dashboard. Ta thử search ở ô lệnh lấy log. Chọn *Save as > Dashboard Panel* , chọn Dashboard đã tạo khi này (No1), bấm *Save* . Ta có thể tạo mới 1 Dashboard ở đây hoặc sử dụng Dashboard đã tạo.

Ở mục Panel Content: cho phép chọn kiểu biểu đồ xuất ra.

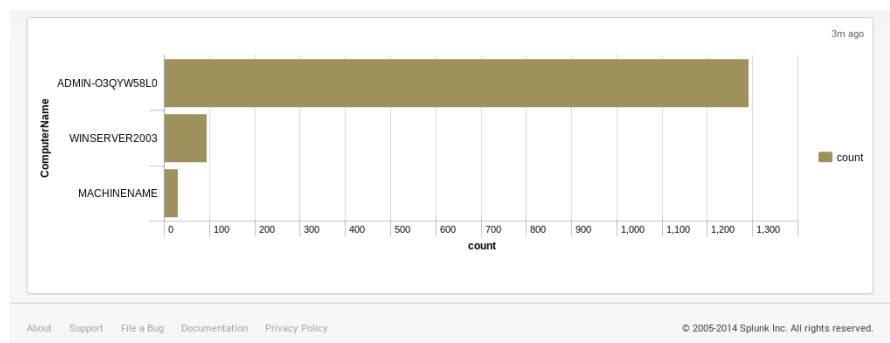


Hình 38 : Tùy chỉnh kiểu Dashboard sẽ xuất ra.

Tạo tương tự vậy với 2 biểu mẫu.

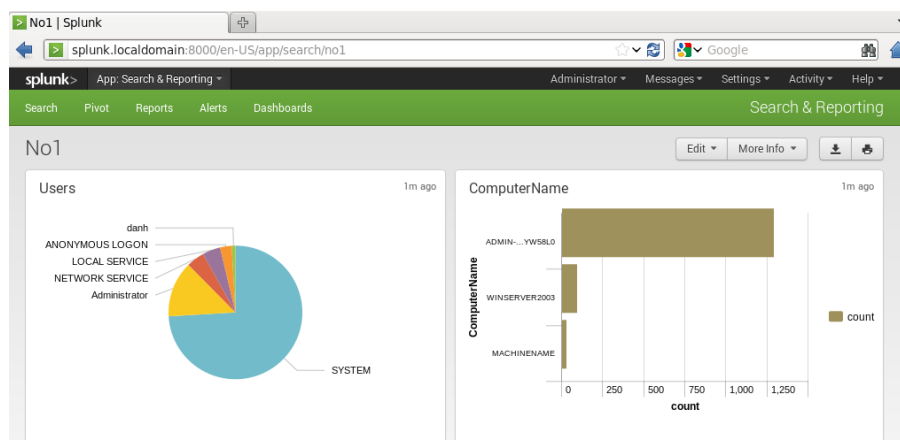


Hình 39 : Biểu đồ biểu diễn log hệ thống Window dạng pie.



Hình 40 : Biểu đồ biểu diễn log hệ thống Window dạng cột.

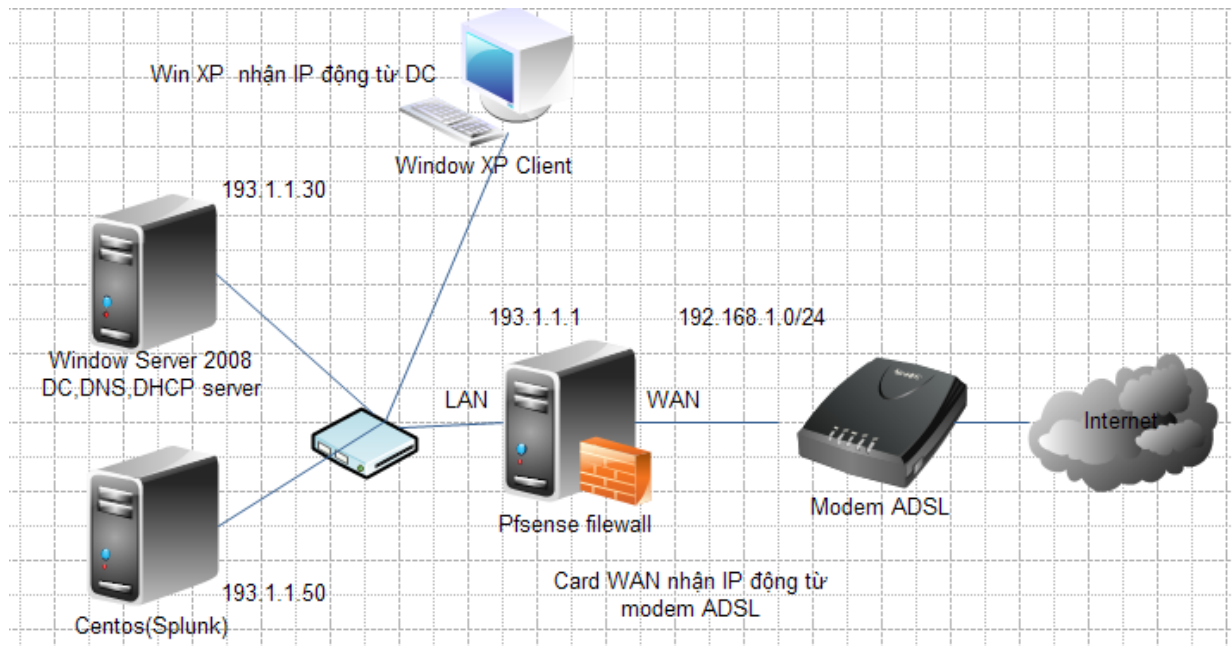
Sau khi add xong panel vào, ta chọn Done để hoàn tất việc tạo Dashboard và chèn panel. Ở đây ta có thể tùy chỉnh biểu đồ theo ý muốn mục *Edit > Edit panel*, ở trên góc phải mỗi panel ta có thể tùy chỉnh loại biểu đồ. Bấm *Done* để hoàn tất.



Hình 41 : Thêm ghép nhiều biểu đồ sẽ trở thành một dashboard.

5 Demo Lab lấy log từ hệ thống mạng nhỏ

5.1 Mô hình:

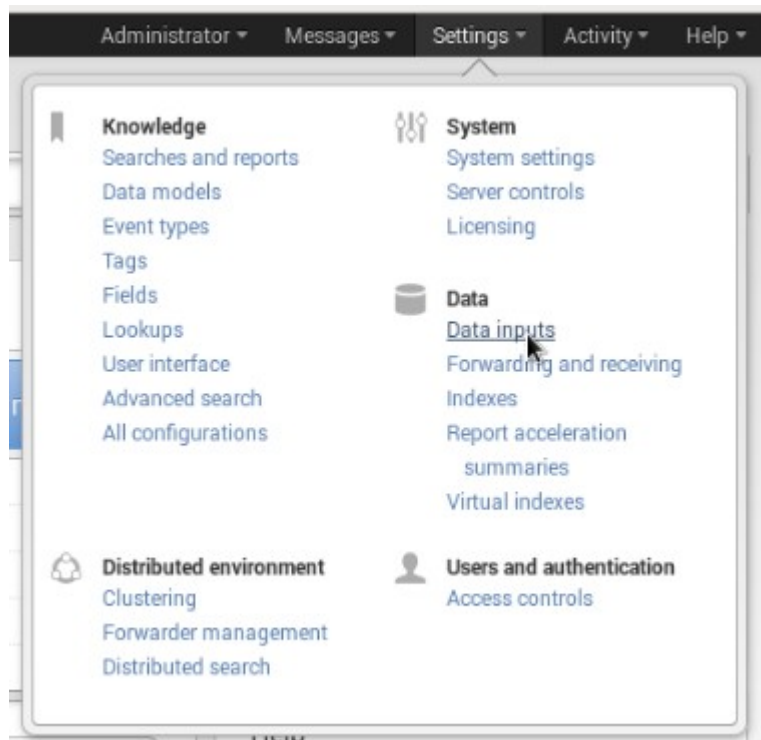


Chuẩn bị:

- + 1 máy Window Server 2008, lên Domain Controller, Cài đặt DNS và DHCP server với IP 193.1.1.30
- + 1 máy CentOS cài sẵn Splunk để thu log với IP 193.1.1.50
- + 1 máy PfSense firewall với 2 card mạng LAN và WAN, card LAN có IP là 193.1.1.1
- + 1 máy XP Client kết nối vào PfSense để cấu hình trên giao diện Web

5.1.1 Bước 1: Lấy log từ Pfsense vào Splunk

+Trên máy Centos đã cài sẵn Splunk:



Chọn Setting =>Data inputs

splunk> Apps ▾ Administrator ▾ Messages ▾ Settings ▾ Activity ▾

Data inputs

Set up data inputs from files and directories, network ports, and scripted inputs. If you want to set up forwarding and receiving between two Splunk instances, go to [Forwarding and receiving](#).

[Add data](#)

Type	Inputs	Actions
Files & directories <i>Upload a file, index a local file, or monitor an entire directory.</i>	6	Add new
TCP <i>Listen on a TCP port for incoming data, e.g. syslog.</i>	0	Add new
UDP <i>Listen on a UDP port for incoming data, e.g. syslog.</i>	0	Add new
Scripts <i>Run custom scripts to collect or generate more data.</i>	1	Add new

Chọn Add new UDP

splunk> Apps Administrator Messages Settings Activity Help

Add new

Data inputs » UDP » Add new

Source

UDP port *

514

Source name override

If set, overrides the default source value for your UDP entry (host:port).

Source type

Set sourcetype field for all events from this source.

Set sourcetype *

Manual

Source type *

*

☐ More settings

Cancel Save

Chọn port nhận các gói tin UDP từ client là 514, Đặt sourcetype là Manual, chọn Save

UDP

Data inputs » UDP

Successfully saved "514".

New

Showing 1-1 of 1 item

Results per page 25

UDP port	Source type	Status	Actions
514	*	Enabled Disable	Clone Delete

Splunk sẽ nhận các gói tin UDP từ port 514

+ Trên máy PfSense:

Vào Status=>System Logs

Status: System logs: General

System Firewall DHCP Portal Auth IPsec PPP VPN Load Balancer OpenVPN NTP **Settings**

General Gateways Routing Resolver Wireless

Last 50 system log entries

Jun 14 13:11:54	check_reload_status: Reloading filter
Jun 14 13:11:58	php: rc.newwanip: Resyncing OpenVPN instances for interface WAN.
Jun 14 13:11:58	php: rc.newwanip: Creating red undata script

Chọn tab Setting

IP Protocol IPv4 ▼
This option is only used when a non-default address is chosen as the source above. This option only expresses a preference; If an IP address of the selected type is not found on the chosen interface, the other type will be tried.

Enable Remote Logging ☒ **Send log messages to remote syslog server**

Remote Syslog Servers

Server 1

Server 2

Server 3

IP addresses of remote syslog servers, or an IP port.

Remote Syslog Contents

☐ Everything

☒ System events

☒ Firewall events

☐ DHCP service events

☒ Portal Auth events

☐ VPN (PPTP, IPsec, OpenVPN) events

☒ Gateway Monitor events

☐ Server Load Balancer events

☐ Wireless events

Save

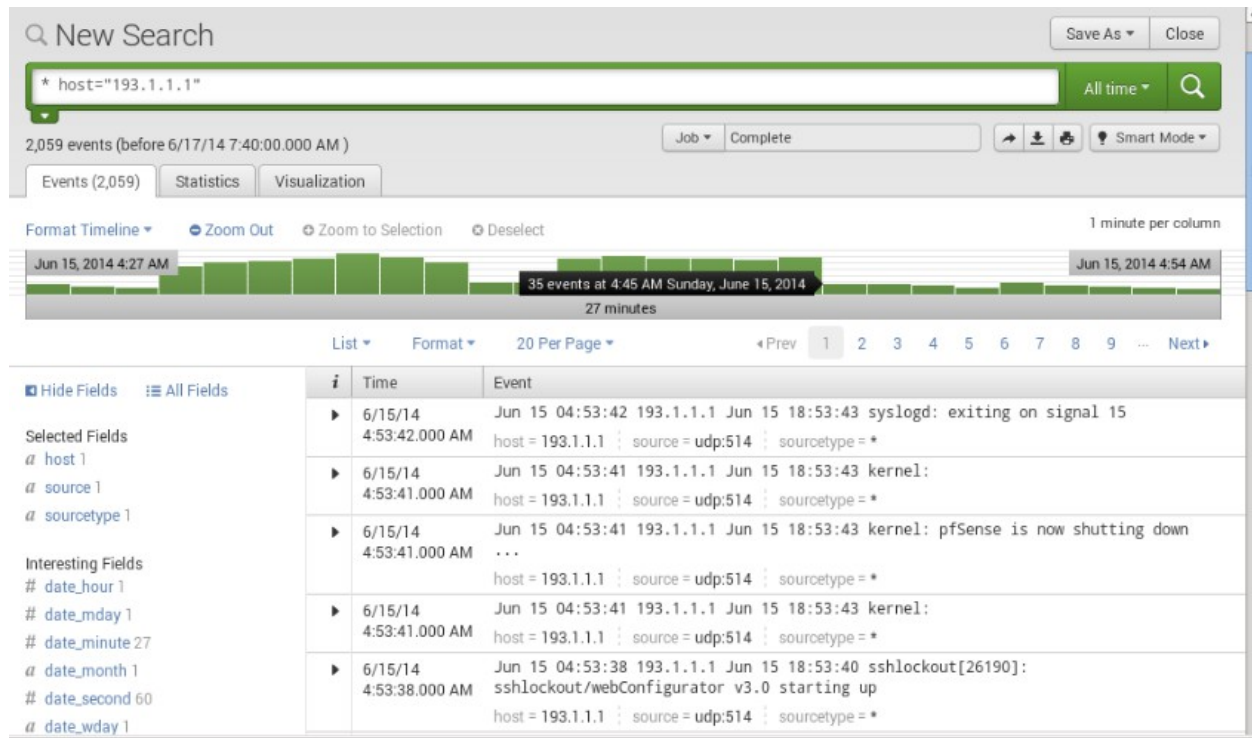
Notes:
syslog sends UDP datagrams to port 514 on the specified remote syslog server, unless another port is specified. Be sure to set syslogd on the remote server to accept syslog messages from pSense.

Tích vào ô Send log Messages to remote syslog server, IP remote server nhận log là 193.1.1.50(máy Splunk), tích tùy chọn các log muốn gửi qua Splunk, Chọn Save.

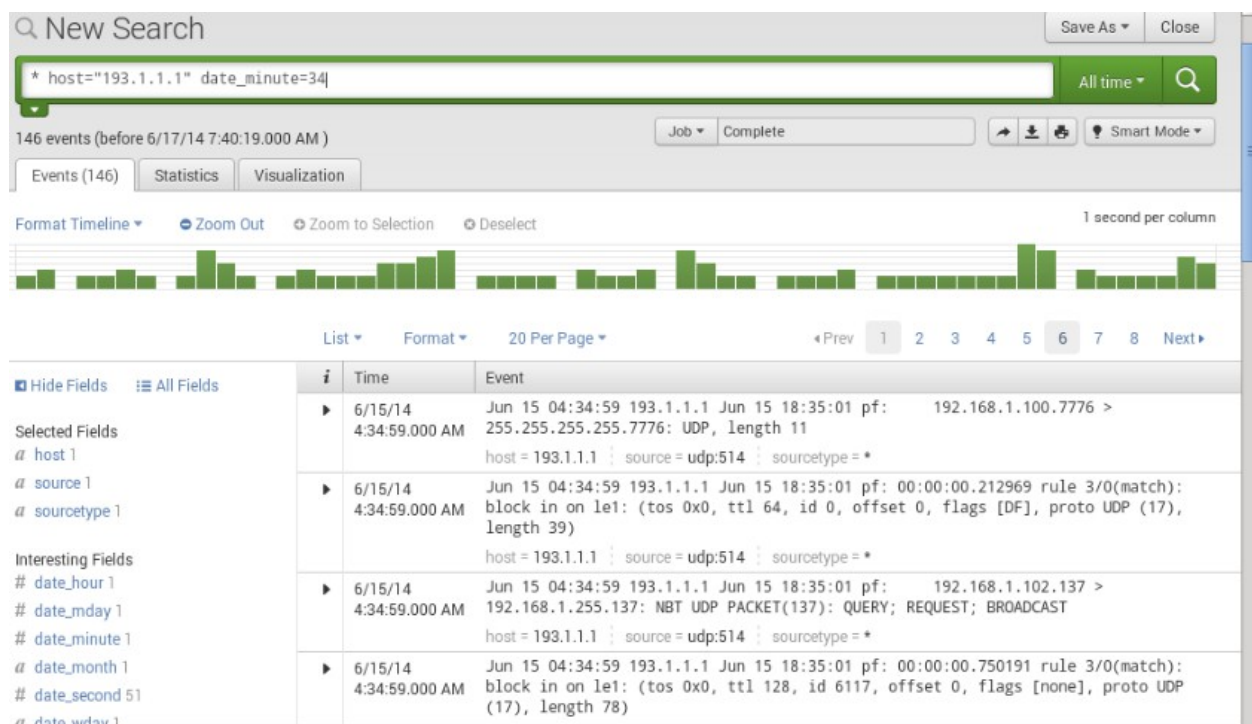
Lưu ý là Pfsense chỉ gửi log bằng giao thức UDP.

Kết quả:

Splunk đã nhận được log từ Pfsense



Nhấn thanh search tìm địa chỉ IP 193.1.1.1 của Pfsense



* host="193.1.1.1" | top limit=20 date_minute

All time



2,059 events (before 6/17/14 7:40:38.000 AM)

Job Complete



Smart Mode

Events

Statistics (20)

Visualization

20 Per Page

Format

Preview

date_minute	count	percent
34	146	7.090821
40	137	6.653715
35	134	6.508014
44	132	6.410879
39	130	6.313745
33	129	6.265177
42	126	6.119475
41	126	6.119475
43	122	5.925206
32	119	5.779505
36	116	5.633803
31	115	5.585236
30	99	4.808159
38	40	1.942691

Events

Statistics (54)

Visualization

20 Per Page

Format

Preview

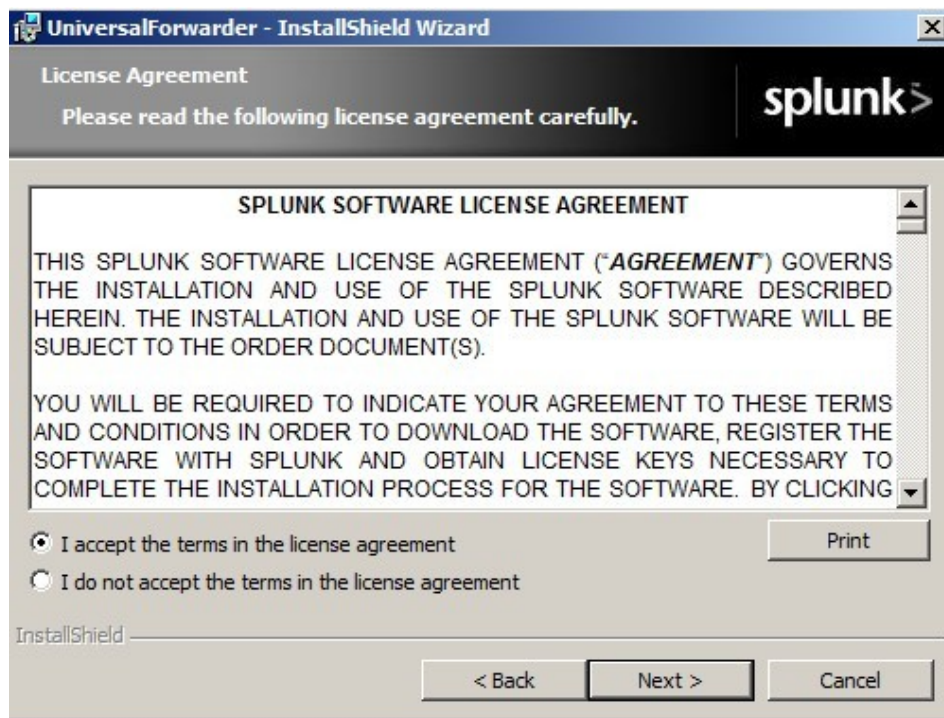
Prev 1 2 3 Next

_time	avg(date_second)
2014-06-15 04:27:00	16.529412
2014-06-15 04:27:30	47.200000
2014-06-15 04:28:00	13.125000
2014-06-15 04:28:30	48.538462
2014-06-15 04:29:00	18.600000
2014-06-15 04:29:30	48.545455
2014-06-15 04:30:00	14.242424
2014-06-15 04:30:30	44.500000
2014-06-15 04:31:00	15.034483
2014-06-15 04:31:30	44.964912
2014-06-15 04:32:00	14.836066
2014-06-15 04:32:30	45.051724
2014-06-15 04:33:00	14.800000
2014-06-15 04:33:30	46.028986
2014-06-15 04:34:00	14.816901
2014-06-15 04:34:30	45.533333
2014-06-15 04:35:00	14.597015

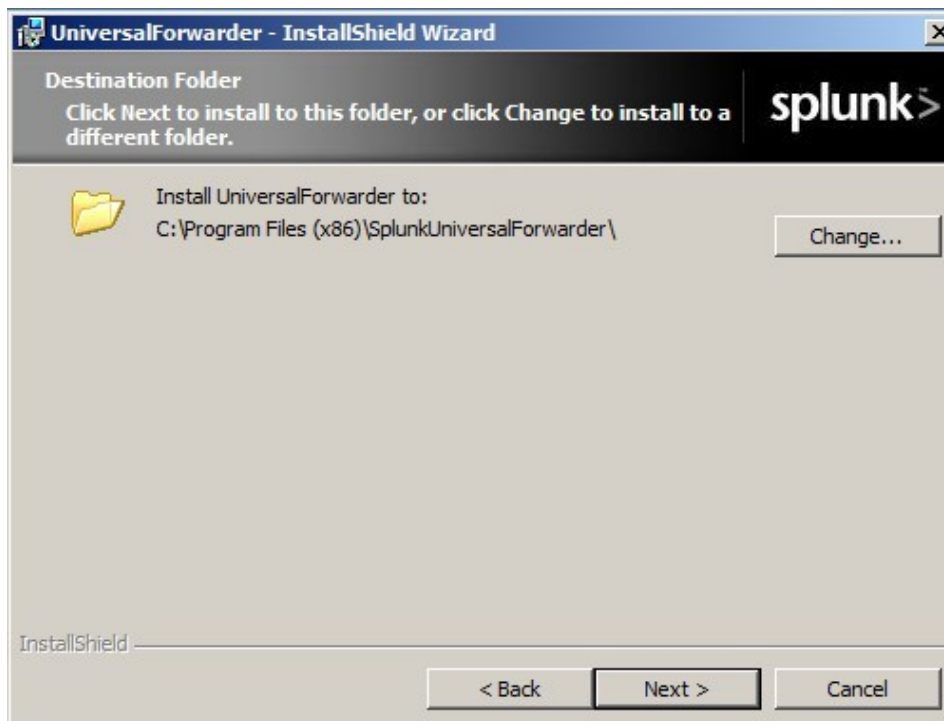
5.1.2 Bước 2: Lấy log từ Window Server 2k8 DC vào Splunk

+Trên máy Window Server 2008:

Cài đặt Splunk Forwarder



Chọn chấp nhận các điều khoản của splunk sau đó bấm Next



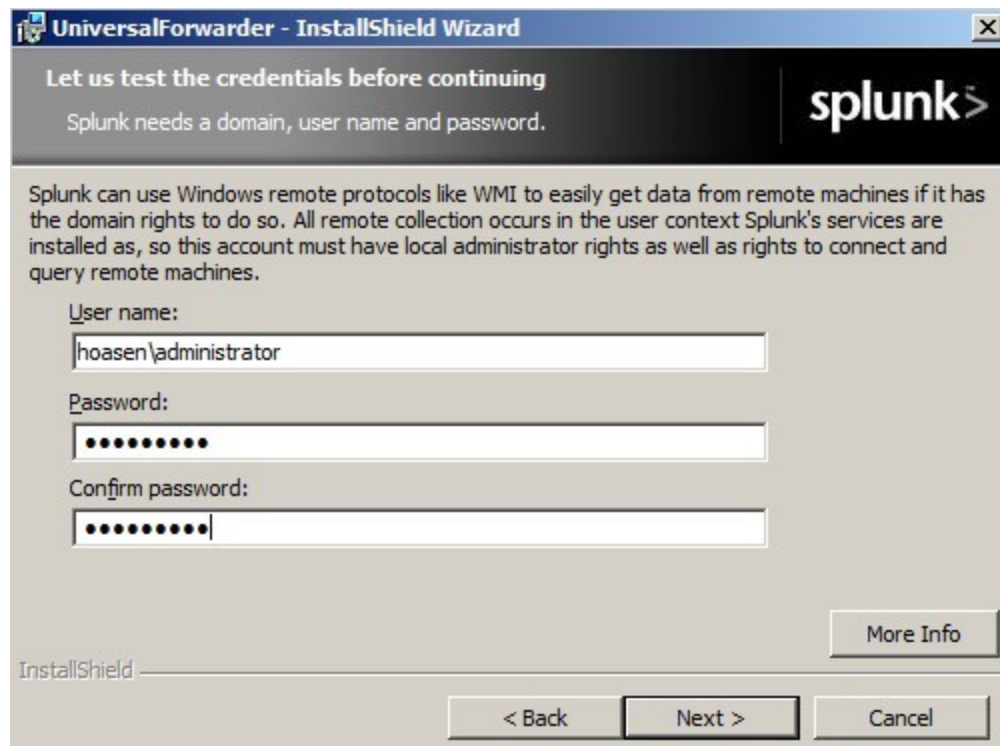
Chọn nơi cài đặt cho Splunk



Nhập vào địa chỉ của máy chủ splunk và port. Lưu ý: ta chọn port trùng với port sẽ cấu hình trên Splunk (Setting > Forwarding and Receive data) để dữ liệu có thể gửi qua Splunk.



Chọn mục Remote Windows Data để gửi thông tin các log, event, và performance của DomainController



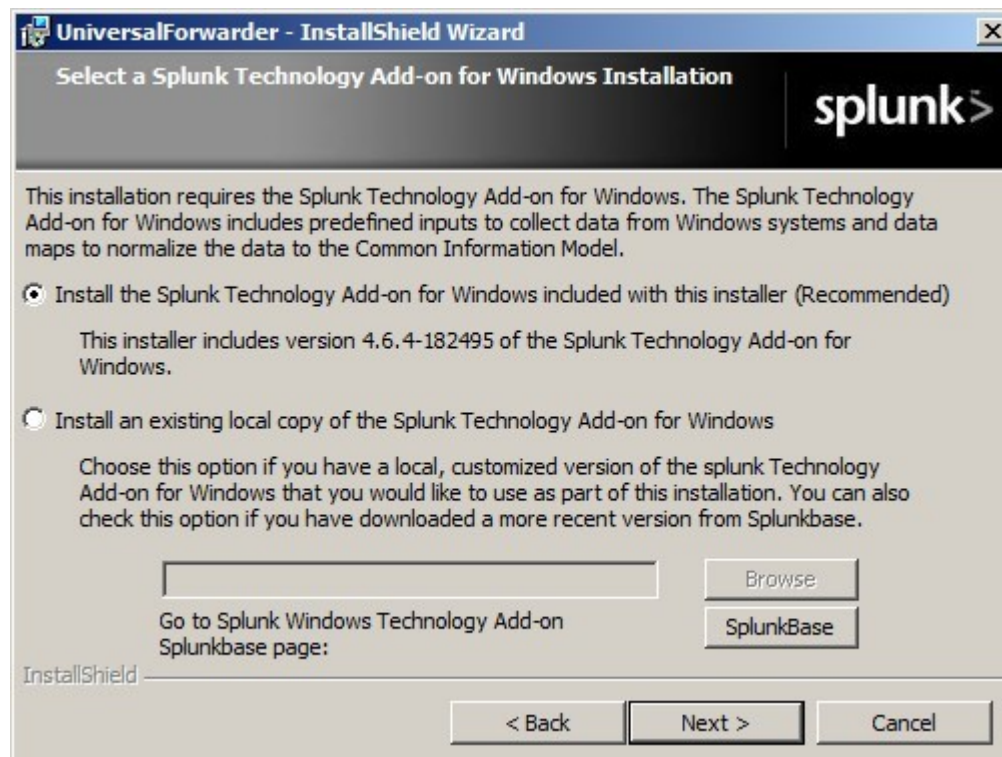
The screenshot shows the 'UniversalForwarder - InstallShield Wizard' window. The title bar includes the Splunk logo. The main heading is 'Let us test the credentials before continuing', followed by the instruction 'Splunk needs a domain, user name and password.' Below this, a paragraph explains that Splunk can use Windows remote protocols like WMI to get data from remote machines, and that the user context must have local administrator rights and remote machine access rights. The form contains three input fields: 'User name:' with the text 'hoasen\administrator', 'Password:' with masked characters, and 'Confirm password:' also with masked characters. At the bottom right is a 'More Info' button. At the bottom center are '< Back' and 'Next >' buttons, and at the bottom right is a 'Cancel' button. The 'InstallShield' logo is in the bottom left corner.

Nhập tên tài khoản administrator của DomainController.



The screenshot shows the 'UniversalForwarder - InstallShield Wizard' window at the 'Enable Windows Inputs' step. The title bar includes the Splunk logo. The main heading is 'Enable Windows Inputs', followed by the instruction 'Optionally select some basic Windows inputs to enable'. The form is divided into two columns of checkboxes. The left column, titled 'Windows Event Logs', includes 'Application Log', 'Security Log', 'System Log', 'Forwarded Events Log', and 'Setup Log'. The right column, titled 'PerfMon', includes 'CPU Load', 'Memory', 'Disk Space', and 'Network Stats'. Below these is a section titled 'Active Directory Monitoring' with a checkbox for 'Enable AD monitoring'. At the bottom, there is a 'Path to monitor' label, an empty text box, and 'File...' and 'Directory...' buttons. At the bottom right are '< Back', 'Next >', and 'Cancel' buttons. The 'InstallShield' logo is in the bottom left corner.

Chọn loại log mà ta cần giám sát, ta có thể tùy chỉnh lại ở file sau khi cài đặt.



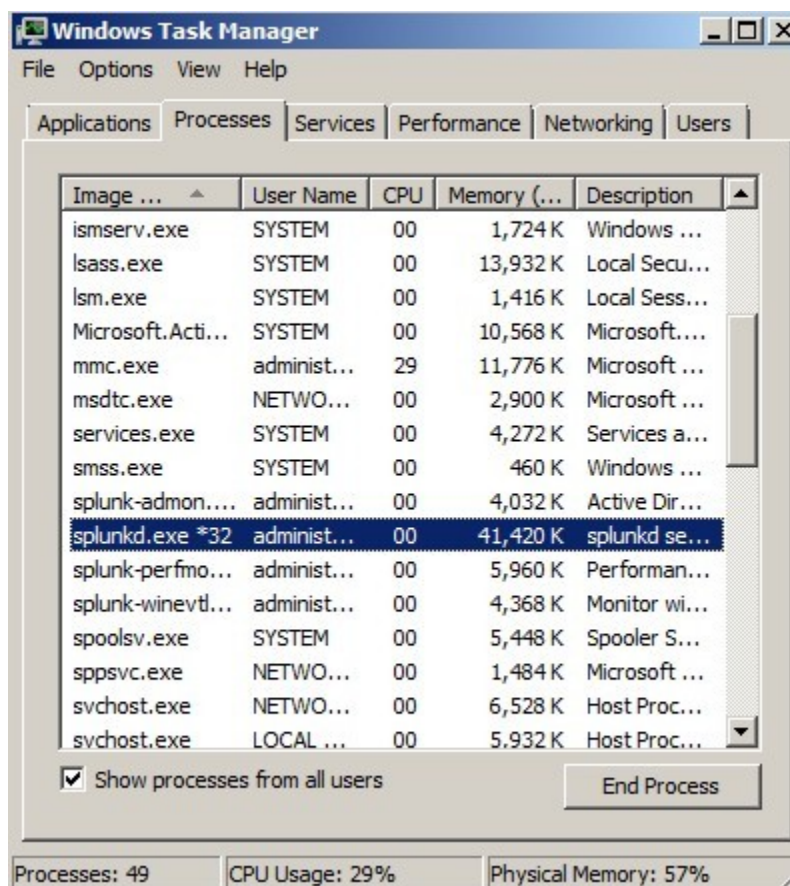
Ta chọn cài đặt luôn Splunk Add-on for Windows.



Finish để kết thúc quá trình cài đặt.

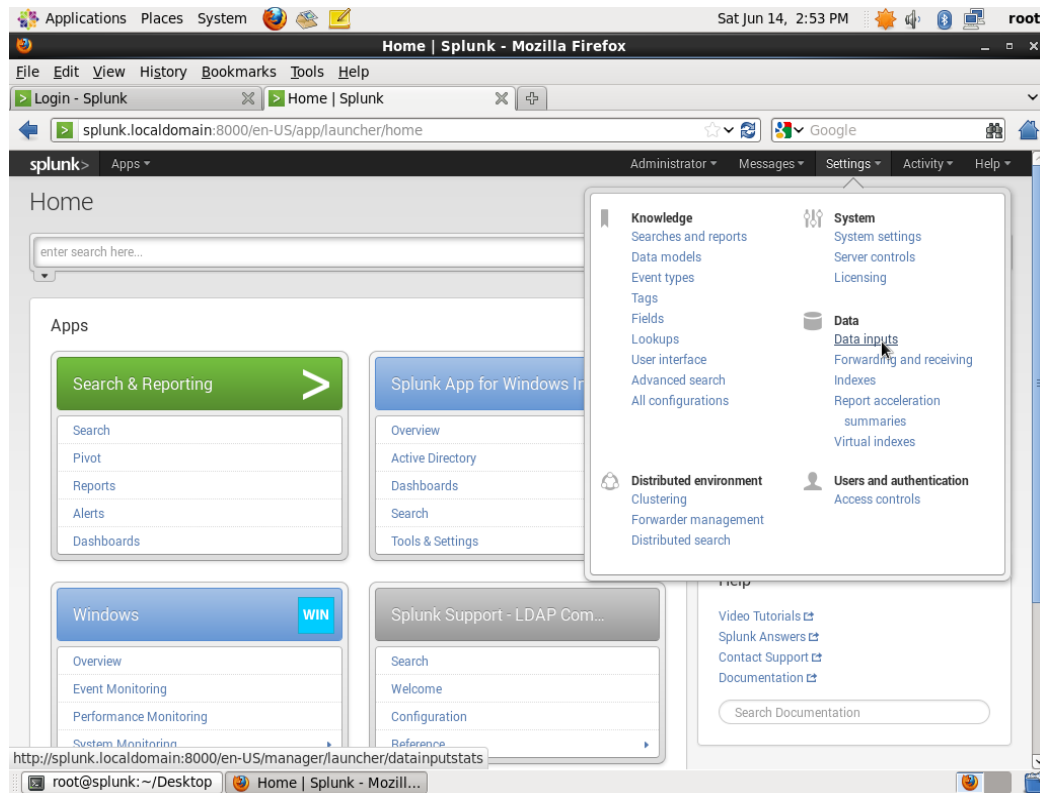
Name ^	Date modified	Type	Size
learned	6/17/2014 11:29 PM	File folder	
search	6/17/2014 11:28 PM	File folder	
Splunk_TA_windows	6/17/2014 11:29 PM	File folder	
SplunkUniversalForwarder	6/17/2014 11:29 PM	File folder	
TA-DNSServer-NT6	6/17/2014 11:32 PM	File folder	
TA-DomainController-NT6	6/17/2014 11:32 PM	File folder	

Copy 2 thư mục TA-DNSServer-NT6 và TA-DomainController-NT6 vào thư mục cấu hình của Splunk để có thể gửi các thông tin của DC. Sau đó ta restart server để splunk có thể hoạt động.

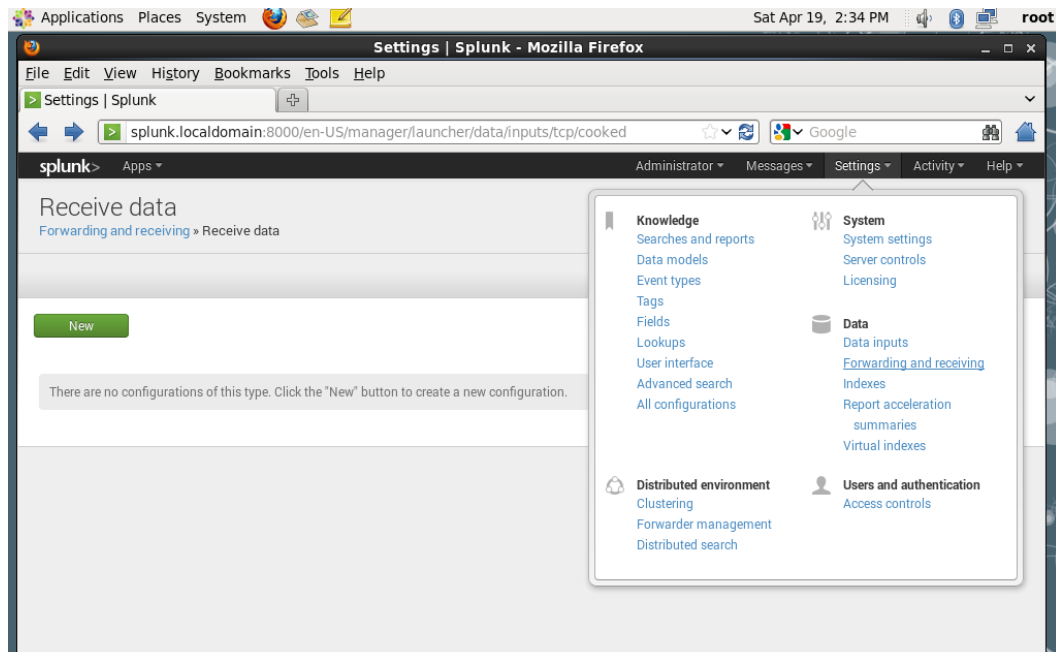


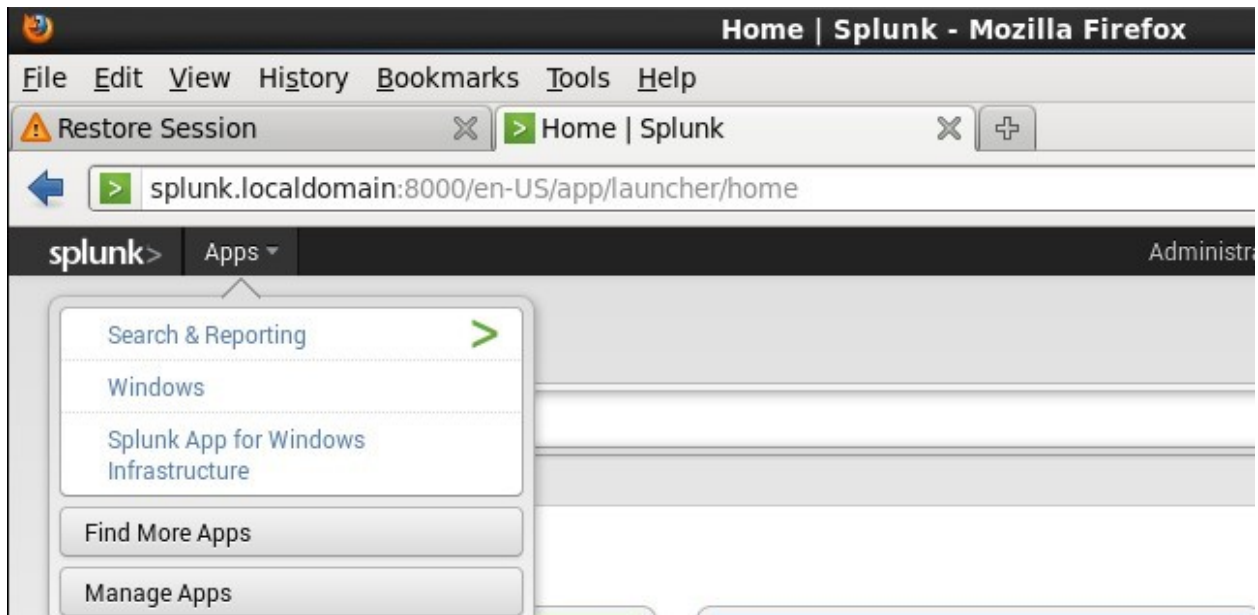
Sau khi restart lại DC, kiểm tra thấy tiến trình Splunk đã hoạt động.

+Cấu hình trên Splunk

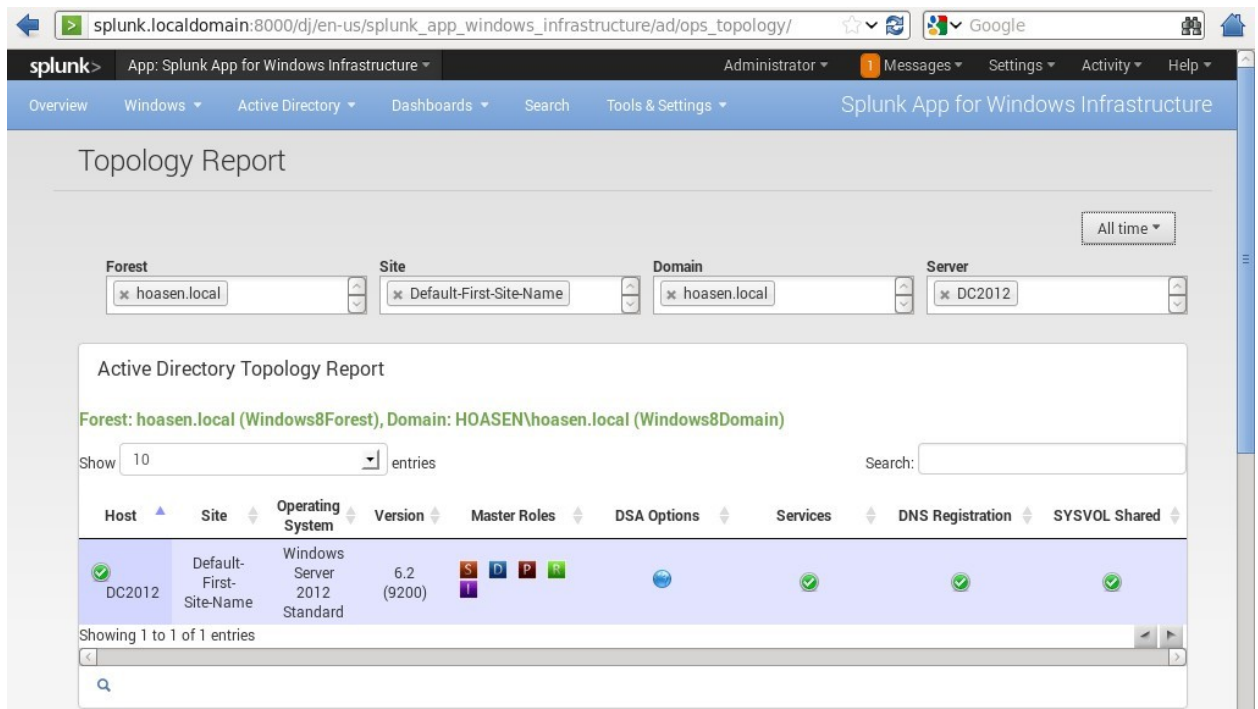


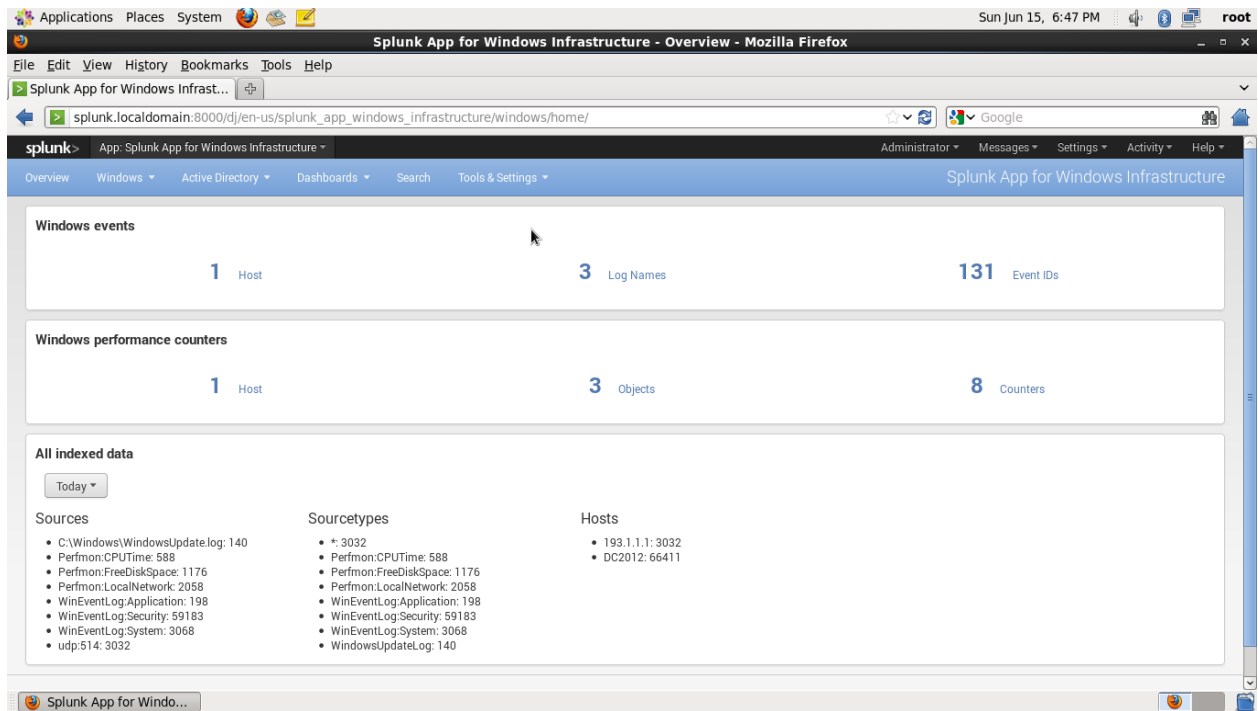
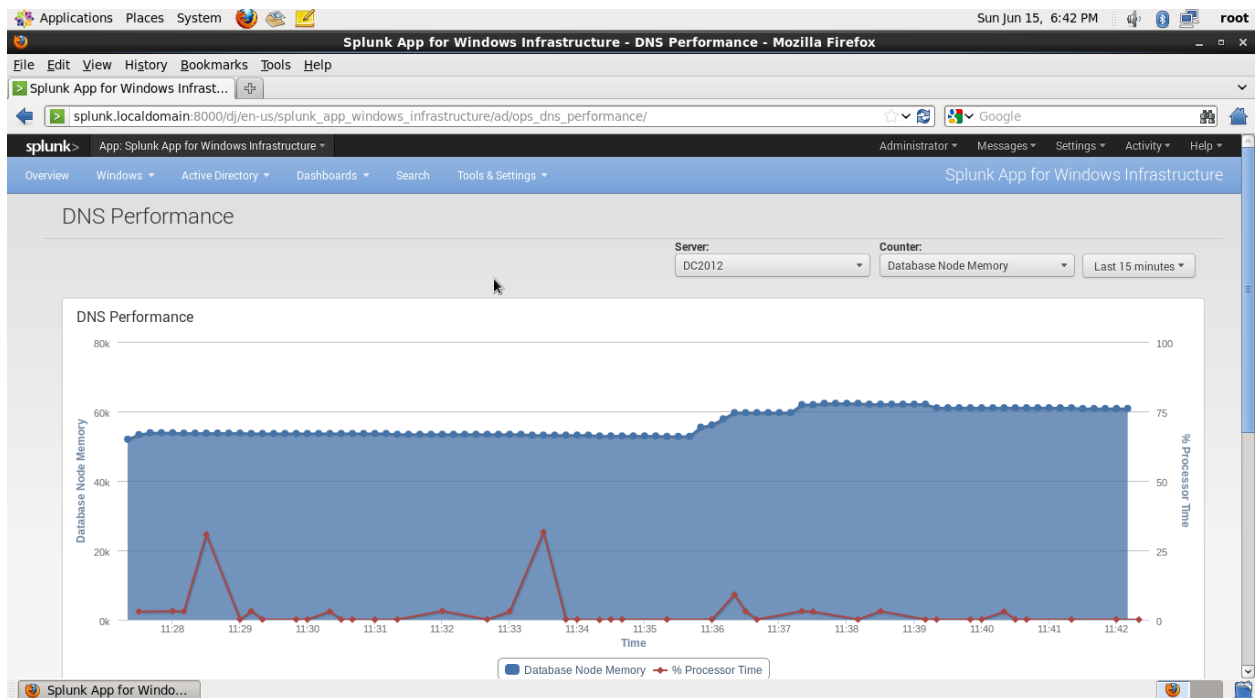
Vào Forwarding and Receiving add thêm port 10000 trùng với port lúc cài đặt ở DC.





Sau khi gán port ta kiểm tra dữ liệu đã được gửi qua cho Splunk, ở mục Splunk App for Windows Infrastructure





Splunk App for Windows Infrastructure - DC Status - Mozilla Firefox

splunk.localdomain:8000/dj/en-us/splunk_app_windows_infrastructure/ad/ops_dc_status/

Domain Controller: DC2012 Last 15 minutes

Domain Controller Status

Server	HOASEN \ DC2012
Domain	HOASEN \ hoasen.local
Site	Default-First-Site-Name
Forest	hoasen.local
Operating System	Windows Server 2012 Standard
Service Pack	
OS Version	6.2 (9200)
DSA Options	
Master Roles	
Highest USN	49184
Schema Version	56 (Windows Server 2012)
Services	Distributed File Replication ✓ Intersite Messaging ✓ Kerberos ✓ Distribution ✓ Network Logon ✓ NT File Replication ✗ Windows Time ✓

Directory Services Performance

Performance Counter	Average	Average Value
LDAP Client Sessions		6.000000
LDAP Searches/sec		11.885110
LDAP Successful Binds/sec		7.923407
Negotiated Binds/sec		7.923407

Replication Performance

No results found.

http://splunk.localdomain:8000/dj/en-us/splunk_app_windows_infrastructure/ad/ops_site_status

Splunk App for Windows Infrastructure - DNS Zone Information - Mozilla Firefox

splunk.localdomain:8000/dj/en-us/splunk_app_windows_infrastructure/ad/ops_dns_zoneinfo/

DNS Zone Information

DNS Zone: 1.1.193.in-addr.arpa All time

Zone Settings

column	row 1
Zone	1.1.193.in-addr.arpa
Aging	False
AllowUpdate	1
AutoCreated	False
AvailForScavengeTime	0
Caption	
DsIntegrated	True
ForwarderSlave	False
ForwarderTimeout	0
NoRefreshInterval	168
RefreshInterval	168
Paused	False
Reverse	True
Shutdown	False
Status	

DNS Servers - Zone

host	A	AAAA	CNAME	HINFO	MX	NS	SOA	SRV	TXT	TotalRecords
DC2012	0	0	0	0	0	1	1	0	0	

Splunk App for Windows Infrastructure - DNS Server Status - Mozilla Firefox

splunk.localdomain:8000/dj/en-us/splunk_app_windows_infrastructure/ad/ops_dns_server_status/

splunk> App: Splunk App for Windows Infrastructure Administrator Messages Settings Activity Help

Overview Windows Active Directory Dashboards Search Tools & Settings Splunk App for Windows Infrastructure

DNS Server Status

DNS Server: DC2012 All time

DNS Server Status

Server	DC2012
DNS Name	DC2012.hoasen.local
Operating System	Windows Server 2012 Standard
Service Pack	
OS Version	6.2 (9200)
Directory Available	✓
Auto Reverse Zones	✓
Auto Cache Update	✗
Recursion	✓
Round Robin	✓
Local Net Priority	✓
Strict File Parsing	✗
Loose Wildcards	✗
Bind Secondaries	✗
Write Authoritative NS	✗
Secure Responses	✓

Query Performance

Performance Counter	Average	Average Value
Total Query Received/sec		19.081199
Total Response Sent/sec		27.010080
UDP Query Received/sec		19.081199
UDP Response Sent/sec		27.010080

Recursion Performance

Performance Counter	Average	Average Value
Recursive Queries/sec		5.921338
Recursive Query Failure/sec		3.962974
Recursive Timeout/sec		4.468467

Splunk App for Windows Infrastructure - DNS Zone Information - Mozilla Firefox

splunk.localdomain:8000/dj/en-us/splunk_app_windows_infrastructure/ad/ops_dns_zoneinfo/

splunk> App: Splunk App for Windows Infrastructure Administrator Messages Settings Activity Help

Overview Windows Active Directory Dashboards Search Tools & Settings Splunk App for Windows Infrastructure

DNS Zone Information

DNS Zone: 1.1.193.in-addr.arpa All time

Zone Settings

column	row 1
Zone	1.1.193.in-addr.arpa
Aging	False
AllowUpdate	1
AutoCreated	False
AvailForScavengeTime	0
Caption	
DnsIntegrated	True
ForwarderSlave	False
ForwarderTimeout	0
NoRefreshInterval	168
RefreshInterval	168
Paused	False
Reverse	True
Shutdown	False
Status	

DNS Servers - Zone

host	A	AAAA	CNAME	HINFO	MX	NS	SOA	SRV	TXT	TotalRecords
DC2012	0	0	0	0	0	1	1	0	0	

6 Kết luận và hướng phát triển đề tài

6.1 Kết luận

Với mục tiêu đã đề ra, nhóm chúng tôi đã hoàn thành công việc tìm hiểu, nghiên cứu cũng như triển khai áp dụng được Splunk vào mô hình mạng thực tế. Qua đó, nhóm chúng tôi đã kiểm chứng những yếu tố sau của splunk:

- Nguyên lý hoạt động trong môi trường bigdata
- Một số tính năng cơ bản và nâng cao trong việc xử dụng splunk.
- Sức mạnh trong việc truy vết xự cố phát sinh trong hệ thống.
- Điều quan cốt lõi là chúng tôi đã thu được lượng kiến thức xoay quanh vấn đề syslog cũng như hiểu thêm về tầm quan trọng của công tác an ninh, bảo mật, phục hồi mạng.

Kết quả đạt được giúp chúng tôi hiểu sâu hơn về các tính năng mà splunk cung cấp. Tuy nhiên, do hạn chế về tài liệu, phí sử dụng bản quyền cũng như kinh nghiệm của nhóm nên bài báo cáo còn nhiều thiếu sót. Nhưng nhóm chúng tôi sẽ cố gắng tiếp tục tìm hiểu sâu hơn nữa, kể cả sau khi kết thúc bài báo cáo nghiên cứu khoa học này.

6.2 Hướng phát triển

Thế giới công nghệ thông tin nói chung và môi trường mạng nói riêng đang ngày càng phát triển vượt bậc. Song song với đó ngày càng có nhiều lỗ hổng mạng được khai thác tạo điều kiện thuận lợi cho hacker xâm nhập gây ảnh hưởng tiêu cực đến hệ thống. Ngoài ra, đối với vị trí IT system administrator chúng ta phải luôn lắng nghe tất cả thông điệp được phát đi từ hệ thống.

Qua đó cho thấy Splunk hoàn toàn phù hợp, cần thiết và đầy đủ khả năng để đáp ứng những yêu cầu đặt ra đối với một chương trình quản lý, giám sát, cảnh báo tất cả sự kiện đang âm thầm diễn ra trong hệ thống. Thực tế, Splunk là một trình dịch viên cao cấp giúp người quản trị giao tiếp một cách trực quan nhất đối với hệ thống. Qua cái nhìn trực quan ấy mà người quản trị xác định chính xác đâu là nguyên nhân dẫn đến sự cố để có thể khắc phục hiệu quả nhất cũng như xây dựng phương hướng phát triển hệ thống. Tất cả những điều trên chứng minh rằng, Splunk có đầy đủ khả năng hoạt động trong môi trường datacenter, mạng doanh nghiệp, mạng dịch vụ, hạ tầng mạng,... cũng như hỗ trợ các thiết bị cisco, IBM,...

Tổng hợp các điều trên chỉ ra, bất kỳ nơi nào có hệ thống mạng tồn tại, nơi nào có hệ thống log tồn tại là nơi đó có thể ứng dụng Splunk.