

BAN CƠ YẾU CHÍNH PHỦ

HỌC VIỆN KỸ THUẬT MẬT MÃ

**----🙣🕮🙡---**

**BÀI TẬP LỚN**

MÔN: PHÒNG CHỐNG VÀ ĐIỀU TRA TỘI PHẠM   
MẠNG MÁY TÍN

Đề tài: Chương 6: Kết nối các điểm - Các sự kiện trên Logs

Giáo viên hướng dẫn: Thầy Lại Minh Tuấn

Lớp L02:

Các thành viên: Nguyễn Thị Hương Giang – AT12B

Trần Thị Phương Chi - AT12B

Nguyễn Thị Kim Anh – AT12E

***Hà Nội, tháng 11 năm 2019***

**Mục lục**

[**I.** **Giới thiệu:** 2](#_Toc25785746)

[**II.** **Các định dạng Logs** 3](#_Toc25785747)

[**III.** **Khám phá kết nối giữa logs và pháp y** 10](#_Toc25785748)

[**1.** **Security logs** 11](#_Toc25785749)

[**2.** **System logs** 13](#_Toc25785750)

[**3.** **Ứng dụng Log:** 13](#_Toc25785751)

[**IV.** **Thực hành Quản lý Log** 14](#_Toc25785752)

[**1.** **Cơ sở hạ tầng quản lý Log:** 16](#_Toc25785753)

[**2.** **Lập kế hoạch và chính sách quản lý Log** 18](#_Toc25785754)

[**V.** **Phân tích Log mạng bằng Splunk** 21](#_Toc25785755)

[**VI.** **Tổng kết** 30](#_Toc25785756)

**Chương 6: Kết nối các điểm - Các sự kiện trên Logs**

# **Giới thiệu:**

Chúng ta chỉ cần kết nối các điểm lại với nhau để xây dựng một bức tranh tổng thể, từ bức tranh của mạng lưới pháp y (NetWork Forensics), từ mạng lưới pháp y, chúng ta cần kết dính các tập tin với nhau để có bức tranh lớn về các hoạt động trên Network. Tất cả các thiết bị duy trì logs sự kiện là một tài nguyên tuyệt vời để theo dõi hoạt động của kẻ xâm phạm. Với vai trò là Mạng 007, chúng tôi sẽ sử dụng các logs này để thử và theo dõi từng bước của tuyến đường do kẻ đột nhập trong mạng của chúng tôi thực hiện.

Hãy bắt đâu bằng việc hiểu log là gì. Một log, như tên gợi ý, là bản ghi của thông tin được tạo ra để đáp ứng một sự kiện hoặc hoạt động cụ thể xảy ra trên hệ thống hoặc mạng. Một log nhằm mục đích nắm bắt ai, cái gì, khi nào và ở đâu của một sự kiện. Logs có thể bao gồm thông tin về ngày và thời gian của hoạt động; thiết bị hoặc ứng dụng nhật ký liên quan đến; người dùng hoặc tài khoản liên quan; loại nhật ký, chẳng hạn như lỗi, cảnh báo, thông tin và kiểm toán thành công hay thất bại, và tất nhiên, thông tin trừu tượng hoặc chi tiết (tùy thuộc vào cách logs được định cấu hình như thế nào) về chính sự kiện thực tế. Each specific event generates a log record. Mỗi sự kiện cụ thể đều tạo ra một bản ghi log. Nhiều bản ghi tạo thành một tập tin log, thường ở định dạng văn bản hoặc các bản ghi được lưu trữ trong cơ sở dữ liệu để dễ dàng quản lý và phân tích. Một trong những vấn đề điển hình mà chúng ta phải đối mặt với phân tích Nhật ký là do độ lớn của các bản ghi đã được tạo. Tùy thuộc vào số lượng thiết bị và mức độ log được bật, khối lượng có thể tăng lên. Để giải quyết vấn đề này, hầu hết các hệ thống ghi nhật ký được thiết kế để lưu trữ logs đến một khoảng thời gian cụ thể.

Do đó, đối với tường lửa, chúng tôi có thể giới hạn ở 10.000 logs của các sự kiện gần đây nhất hoặc 100 MB dữ liệu gần đây theo mặc định, trừ khi được định cấu hình khác. Tuy nhiên, quan điểm của người làm pháp y khác với quan điểm của một quản trị viên mạng bình thường. Mục đích của họ là phát hiện, theo dõi và xác định những kẻ xâm nhập và hoạt động bất hợp pháp trên mạng của chúng tôi. Do đó, việc truy cập vào các bản ghi cũ hơn có thể khá hữu ích. Việc so sánh các bản ghi có thể tạo ra kết quả thú vị. Một ví dụ đơn giản cho điều này là tăng gấp ba lần kích thước của tập tin log. Bản thân điều này có thể là cảnh báo về sự tăng lên của các hoạt động mạng, là một chỉ số có thể có để cảnh báo về phần mềm độc hại.

Một vấn đề lớn khác gặp phải do khối lượng các bản ghi được tạo ra là phân tích thủ công các bản ghi để tìm kiếm sự bất thường. Bất cứ nơi nào có hệ thống phân tích logs thủ công, trong một khoảng thời gian, chúng tôi thấy rằng logs có xu hướng bị bỏ qua. Điều này có thể được khắc phục bằng cách sử dụng các công cụ phân tích nhật ký tốt như Splunk và Logstash. Chúng tôi sẽ kiểm tra việc sử dụng Splunk để phân tích và hiển thị logs trong chương này.

Trong chương này, chúng tôi sẽ đề cập đến các chủ đề sau:

* Hiểu các định dạng logs.
* Khám phá kết nối giữa logs và pháp y
* Thực hành quản lý logs hợp lý
* Phân tích logs mạng bằng Splunk

# **Các định dạng Logs**

Ngay từ đầu chương này, chúng ta đã thảo luận cách logs ghi lại bốn Ws liên quan đến sự kiện. Đó là khi nào(When), ai (Who), ở đâu (Where ), và cái gì (What) về sự kiện. Chúng ta hãy hiểu cách mỗi thứ được thực hiện chi tiết hơn trong bảng sau. Chúng ta hãy tìm hiểu những điều này làm thế nào được thực hiện chi tiết hơn một chút trong bảng sau:

|  |  |  |
| --- | --- | --- |
|  | Thuộc tính | Ghi chú |
| Khi nào (When) | Ngày tháng và thời gian log.  Ngày và giờ của sự kiện | Ngày và thời gian log có thể khác với ngày của một sự kiện trong một số trường hợp, chẳng hạn như trong trường hợp dữ liệu sự kiện được thu thập từ xa vào các thời điểm không liên tục |
| Ở đâu (Where ) | Địa chỉ ứng dụng  Dịch vụ  Định vị địa lý  Cửa sổ / trang web / biểu mẫu  ID ứng dụng | Tên hệ thống hoặc máy chủ, địa chỉ IP, số cổng và ID thiết bị cục bộ  Tên và giao thức  Vĩ độ và kinh độ, nếu có thể  URL điểm nhập, URL điểm truy cập trang web, hộp thoại, v.v.  Tên và phiên bản ứng dụng |
| Ai(Who) | Mã vị trí | Đường dẫn và tên của mô-đun / tập lệnh mã |

|  |  |  |
| --- | --- | --- |
| What | Loại sự kiện  Mức độ nghiêm trọng của sự kiện  Cờ sự kiện  Mô tả sự kiện | Ví dụ: đăng nhập, đăng xuất, truy cập, xóa, v.v ..  Ví dụ: nghiêm trọng, lỗi, cảnh báo, thông tin, v.v. Các cờ liên quan đến bảo mật, ví dụ, đăng nhập thất bại  Mô tả dài dòng về một sự kiện, ví dụ: Một tài khoản không đăng nhập được, với tên tài khoản, ID bảo mật, tên miền và nhiều thông tin khác |

Khía cạnh quan trọng trong việc phân tích các bản ghi là khi nào (When). Mỗi mục nhật ký có một đầu vào ngày và thời gian, sẽ thiết lập thời điểm của một sự kiện cụ thể. Nếu ngày hoặc thời gian này không chính xác, tất cả các mối tương quan sẽ trở nên vô vọng và cuối cùng có thể làm rối tung phân tích.

Để minh họa điều này, hãy cùng xem một sự cố đã từng gặp trong một trong những cuộc điều tra .

**Ví dụ một vụ án đã xảy ra**

Một hãng hàng không đã nhận được một mối đe dọa bom điện tử cho một trong những chuyến bay đường dài đã được lên kế hoạch cho một ngày cụ thể. Mối đe dọa này đã được thông báo cho hãng hàng không thông qua trang hỗ trợ / phản hồi của khách hàng, được lưu trữ trên máy chủ sử dụng Internet-facing. Theo thông lệ, dữ liệu được thu thập bởi form của trang web và được gửi đến một trung tâm hỗ trợ và xử lý phản hồi ở nước ngoài. Nhà điều hành chuyên xử lý các khiếu nại đã xem xét mối đe dọa và thông báo cho người giám sát của cô của. Dựa trên đánh giá của người giám sát, mối đe dọa tiếp tục leo thang và cả hãng hàng không và cơ quan thực thi pháp luật đều được thông báo. Để hỗ trợ điều tra kỹ thuật vụ án, nhóm của chúng tôi đã được gọi đến.

Bước đầu tiên, chúng tôi đã yêu cầu quyền truy cập vào tất cả các bằng chứng kỹ thuật số cũng như nhật ký của máy chủ web, nơi form hỗ trợ được đăng. Nhưng không có hồ sơ về các địa chỉ IP được sử dụng để gửi e-mail như vậy. Đây là một trở ngại lớn.

Nhìn vào cách tiếp nhận được mà form hỗ trợ web đã gửi tin nhắn dưới dạng e-mail có cấu trúc đến trung tâm hỗ trợ. Các tiêu đề của e-mail đã được kiểm tra và điều đầu tiên chúng tôi nhận thấy là thời gian của e-mail để di chuyển từ máy chủ web đến máy chủ SMTP tiếp theo trong một chuỗi là hơn 27 phút. Điều này có vẻ khá kỳ lạ, do đó, chúng tôi đã cố gắng sao chép quy trình bằng cách gửi e-mail kiểm tra từ mẫu web và thấy rằng, trong trường hợp này, thời gian để vượt qua cùng một hop chỉ trong vòng một phút.

Đối mặt với sự bất thường này, chúng tôi đã lặp lại bài kiểm tra nhiều lần và đưa ra kết quả tương tự mỗi lần: dưới một phút. 27 phút được thực hiện bởi thư gốc là không giải thích được. Chúng tôi đã nói chuyện với quản trị viên về vấn đề này và được thông báo rằng tại thời điểm chúng tôi yêu cầu bằng chứng, họ đã quan sát thấy rằng thời gian máy tính không chính xác và không được đồng bộ hóa với thời gian hiện tại. Do đó, họ đã "sửa" điều này bằng cách thay đổi thời gian thêm "khoảng" 30 phút. Điều này thực sự đã đòn giáng mạng vào chúng tôi. Chúng tôi không còn có thể dựa vào độ chính xác của dữ liệu trong header liên quan đến thời gian. Để hiểu tác động của "sự khác biệt" này, chúng ta cần nhận ra rằng trên một trang web hàng không kinh doanh buôn bán nhiều, có thể có một ngàn khách truy cập / giờ. Do đó, thời gian bù khoảng 30 phút sẽ giúp tăng khoảng 500 khách truy cập.

Thời gian càng lúc càng ít! Ngày thảm họa đã gần kề. Một nỗ lực suốt ngày đêm đã được tiến hành. Tất cả các nhật ký web đã được thu thập và phân tích sâu các nhật ký đã được thực hiện. Thời gian bù gần đúng được kết hợp với thời gian email và tương quan với nhật ký lần truy cập trang web và được phân tích. Sáu mươi mốt địa chỉ IP nghi ngờ đã được xác định. Sự hỗ trợ của các nhà cung cấp dịch vụ Internet (ISP) địa phương đã được yêu cầu khẩn cấp. Mỗi địa chỉ IP được xác định và quy kết một cách tỉ mỉ. Hai trong số này thuộc về các quán cà phê mạng, hoạt động theo mô hình trả tiền khi bạn sử dụng. Chủ quán cà phê Cyber ​​đã hỗ trợ cảnh sát địa phương xác định các máy tính được sử dụng vào thời điểm đó. Hình ảnh forensics đã được thực hiện để phân tích. Video camera quan sát đã được kiểm tra và một số nghi phạm đã được xác định. Hình ảnh được thực hiện và lưu hành đến an ninh sân bay. Đồng thời, các chuyên gia pháp y của chúng tôi đã kiểm tra các hình ảnh ổ cứng và xác định một trong số chúng đã được sử dụng để truy cập trang web của hãng hàng không và gửi tin nhắn. Bất kỳ hoạt động ổ cứng bổ sung nào trong khoảng thời gian truy cập trang web đều được kiểm tra chi tiết. Chúng tôi thấy rằng nghi phạm cũng đã truy cập vào tài khoản email cá nhân của họ. Chúng tôi tóm được anh ta! Với một chút trợ giúp từ Nhà cung cấp dịch vụ thư (MSP), cơ quan thực thi pháp luật địa phương đã xác định được nghi phạm và đón anh ta về thẩm vấn. Hóa ra nghi phạm đã gửi thư là một sinh viên. Một người bạn của anh ta đang đi trên chuyến bay đó và anh ta đã đe dọa anh ta rằng anh ta sẽ không để bạn mình rời khỏi đất nước.

Như bạn có thể thấy, lần này cuộc điều tra có thể có thời gian điều tra ngắn hơn đáng kể nếu thời gian của máy chủ được đồng bộ hóa đúng cách. Thật vậy, máy chủ Network Time Protocol (NTP) được khuyến nghị nên được triển khai để duy trì thời gian trên một mạng để các trường hợp này không xảy ra trong quá tình điều tra pháp y trên Network.

Để hiểu rõ hơn về cấu trúc của một logs file, hãy xem tiêu chuẩn của World Wide Web Consortium (W3C). (<https://www.w3.org/>).

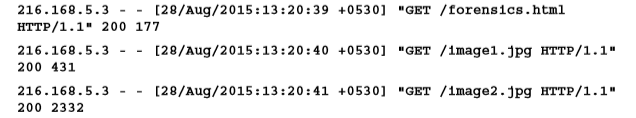
Sau đây là cấu trúc của Định dạng tệp dữ liệu chung:

  
Chúng ta hãy tìm hiểu từng cái một, như thể hiện trong phần sau:

|  |  |
| --- | --- |
| Các trường | Giải thích |
| remotehost | Đây thường là tên máy chủ DNS hoặc địa chỉ IP, trong trường hợp DNSLookup bị tắt. |
| rfc931 | Đây là giao thức Lực lượng đặc nhiệm kỹ thuật Internet (IETF) cung cấp phương tiện để xác định người dùng của một kết nối TCP cụ thể. Điều này được gọi là Giao thức nhận dạng. Điều này biểu thị tên đăng nhập từ xa của người dùng. |
| authuser | Đây là ID người dùng của người yêu cầu tài liệu. Điều này được xác định bởi tên người dùng được sử dụng để xác thực. |
| [date] | Điều này cho biết ngày và thời gian của yêu cầu |
| "request" | Đây là dòng yêu cầu từ khách hàng. Ví dụ: khách hàng có thể yêu cầu một trang web được gọi là forensics.html |
| status | Đây là mã trạng thái HTTP để đáp ứng yêu cầu của khách hàng. Ví dụ: mã trạng thái 404 có nghĩa là Không tìm thấy. |
| bytes | Đây là kích thước của tài liệu được trả về theo byte để đáp ứng yêu cầu. |

Giả sử rằng người ngoài muốn truy cập vào một trang web được gọi forensics.html trên máy chủ web của chúng tôi.

Các mục log sẽ được tạo sẽ tương tự như sau:



Nói một cách đơn giản, điều này có nghĩa là một trình duyệt web có địa chỉ IP 216.168.5.3 đưa ra yêu cầu cho trang forensics.html vào ngày 28 tháng 8 năm 2015, lúc 13:20:39 giờ IST, GMT +5.30 giờ. Máy chủ đã trả lời với một phản hồi thành công (200) và gửi tệp qua. Kích thước tệp là 177 byte.

Khi nhận được tệp này, trình duyệt web đã phát hiện ra rằng trang web forensics.html có hai hình ảnh nội tuyến là: image1.jpg và image2.jpg. Sau đó, trình duyệt đưa ra yêu cầu cho mỗi hình ảnh này, được máy chủ phản hồi thành công. Kích thước của hình ảnh được truyền đi lần lượt là 431 và 2332 byte.

# **Khám phá kết nối giữa logs và pháp y**

Trong phần trước, chúng ta đã hiểu rõ về nhật ký là gì và loại dữ liệu chứa trong đó. Tôi chắc chắn rằng giống như bất kỳ nhà điều tra giỏi nào, chúng tôi có cảm giác thực sự rằng những điều này khá quan trọng. Hãy khám phá chính xác lý do tại sao lại như vậy.

Như chúng ta đã thấy trong phần trước, một mục logs phản ánh một sự kiện xảy ra trong mạng của một tổ chức. Một nhóm các mục logs tạo một file logs. Nhiều tệp nhật ký như vậy có liên quan trực tiếp đến bảo mật, trong khi các tệp khác có thể có một số mục cụ thể cho các vấn đề liên quan đến bảo mật. Nhật ký liên quan đến bảo mật có thể được tạo bởi các công cụ chống vi-rút, tường lửa, hệ thống phát hiện và ngăn chặn xâm nhập (IDPS), hệ điều hành, thiết bị và ứng dụng mạng, v.v.

Các yếu tố chính cần hiểu là logs là một bản ghi độc lập của con người về hoạt động của hệ thống và người dùng trong mạng. Điều này làm cho chúng đặc biệt không thiên vị và cho phép tòa án chấp nhận làm bằng chứng, miễn là chúng được thu thập, xử lý và bảo quản một cách hợp pháp. Logs cung cấp “dấu vân tay” trong một sự cố. Chúng có thể cho chúng ta biết điều gì đã xảy ra, chuỗi sự kiện, hệ thống nào bị ảnh hưởng / liên quan, thông tin nào bị xâm phạm / điều gì không, người dùng đã gây ra hoặc phản hồi sự cố như thế nào và khung thời gian tất cả những điều này xảy ra.

Các loại Logs được biết đến rộng rãi là Logs Security, system logs, and application logs.

## **Security logs**

Bất kì tổ chức nào thành thạo trong vấn đề này đều có một số biện pháp an ninh tại chỗ. Tất cả các biện pháp đều tạo ra logs. Bao gồm những điều sau đây:

Phần mềm chống vi-rút / chống phần mềm độc hại: Phần mềm này ghi lại thông tin liên quan đến nhiều loại vi-rút, phần mềm độc hại, rootkit, ngày và thời gian phát hiện, các hệ thống được phát hiện đầu tiên, cố gắng kiểm dịch / xóa, cách ly, v.v.

Bộ định tuyến (Routers): Bộ định tuyến thường là tuyến phòng thủ đầu tiên trong nhiều mạng. Chúng được cấu hình để cho phép hoặc chặn lưu lượng truy cập mạng, cụ thể là dựa trên các chính sách được quản trị viên mạng triển khai. Một nghiên cứu về lưu lượng logs bị chặn khá có hữu ích từ góc độ pháp y.

Fiewalls: Giống như các bộ định tuyến, tường lửa cho phép hoặc chặn các hoạt động mạng dựa trên chính sách đã triển khai; tuy nhiên, các phương pháp được sử dụng để kiểm tra lưu lượng mạng tinh vi hơn nhiều. Điều này bao gồm: theo dõi tình trạng lưu lượng mạng và kiểm tra nội dung. Sự khác biệt chính giữa bộ định tuyến và tường lửa nằm ở các chính sách phức tạp hơn và hoạt động chi tiết hơn ghi logs của bộ định tuyến vis-à-vis.

Hệ thống phát hiện và ngăn chặn xâm nhập: Như bạn đã biết trong chương trước, các hệ thống phát hiện xâm nhập (IDS) và hệ thống ngăn chặn xâm nhập (IPS) xác định, ghi lại, phát hiện và ngăn chặn các hành vi và tấn công đáng ngờ. Vì các hệ thống IPS tự chủ động nên chúng có thể loại bỏ các gói độc hại. Việc ghi chép dài dòng các gói dữ liệu này rất hữu ích từ góc độ điều tra.. Nhật ký của các bài kiểm tra tính toàn vẹn được thực hiện bởi IDS cũng có thể chứa thông tin có giá trị, đặc biệt khi so sánh với các bài kiểm tra trước đó được thực hiện trước đây. Điều này có thể giúp chuyên gia pháp y thiết lập khung thời gian của vụ việc.

Phần mềm truy cập từ xa: Microsoft Remote Desk Protocol (RDP), LogMeIn, TeamViewer và RADmin là một số công cụ phần mềm truy cập từ xa có sẵn cho phép truy cập từ xa được bảo mật thông qua mạng riêng ảo (VPN). Trong số này có một vài loại hỗ trợ kiểm soát truy cập chi tiết (như Lớp cổng bảo mật hoặc SSL) và thường duy trì nhật ký chi tiết liên quan đến việc sử dụng tài nguyên bao gồm ngày và thời gian truy cập của người dùng, dữ liệu được truyền, cũng như chi tiết của tất cả những lần đăng nhập thành công và thất bại. Thông tin từ các nhật ký này có thể giúp xác định hành vi trộm cắp dữ liệu và hoạt động từ xa khác, bao gồm cả truy cập trái phép.

Proxy web: Một số tổ chức sử dụng proxy web để kích hoạt, giám sát, lọc và hạn chế quyền truy cập trang web cho người dùng của họ. Proxy web cũng lưu trữ các bản sao của các trang web được truy cập nhiều để mỗi lần truy cập hiệu quả hơn. Các proxy web được thiết kế để giữ một bản ghi của tất cả các URL truy cập thông qua chúng. Như chúng ta có thể thấy, các proxy web có thể giúp chúng ta xác định nơi người dùng đã chọn một ổ đĩa có thể gây rắc rối bằng cách tải xuống.

Phần mềm quản lý lỗ hổng: Hàng ngàn lỗ hổng được phát hiện mỗi năm. Điều này có nghĩa là mọi tổ chức phải liên tục cập nhật và vá môi trường của nó. Để làm điều này trên bất kỳ quy mô nào đòi hỏi phần mềm chuyên dụng cho công việc. Phần mềm quản lý lỗ hổng chuyên dụng này đóng vai trò kép gồm phần mềm quản lý bản vá và đánh giá lỗ hổng. Thông thường, nhật ký hệ thống quản lý lỗ hổng vá lịch sử cài đặt và trạng thái lỗ hổng của từng máy chủ. Điều này bao gồm các lỗ hổng đã biết / đã xác định trước đó trong các bản cập nhật phần mềm và có thể có thông tin về cấu hình của máy chủ lưu trữ. Phần mềm quản lý lỗ hổng thường được chạy trong chế độ hàng loạt và tạo ra một số lượng lớn các mục nhật ký.

Máy chủ xác thực: Bất cứ khi nào một thực thể (người dùng hoặc máy tính khác) yêu cầu quyền truy cập vào tài nguyên mạng, thông tin đăng nhập của nó cần được xác minh bởi máy chủ xác thực. Máy chủ xác thực bao gồm máy chủ thư mục, máy chủ đăng nhập một lần, điểm truy cập hoặc thậm chí là một công tắc. Chúng thường ghi lại mọi nỗ lực xác thực: nguồn gốc, tên người dùng, thành công hay thất bại, và tất nhiên, ngày và giờ.

## **System logs**

Máy chủ, hệ thống, thiết bị mạng như bộ định tuyến và bộ chuyển mạch có hệ điều hành riêng. Chúng đều ghi lại một loạt các thông tin liên quan đến bảo mật. Phổ biến nhất là các sự kiện hệ thống và hồ sơ kiểm toán, như được trình bày dưới đây:

Sự kiện hệ thống: Từ góc độ điều tra pháp y, các sự kiện hệ thống như bắt đầu dịch vụ, tắt hệ thống hoặc sự kiện không thành công có thể khá hữu ích. Tất cả các sự kiện như vậy thường được ghi lại theo ngày và thời gian, và có thể bao gồm trạng thái, mã lỗi, tên dịch vụ và tài khoản được liên kết với sự kiện.

Hồ sơ kiểm toán: Những thông tin bảo mật lưu trữ này, bao gồm các lần xác thực thất bại và thành công, tạo tệp, truy cập, xóa, thay đổi chính sách bảo mật, tạo và xóa tài khoản, leo thang đặc quyền, v.v. Quản trị viên có thể cài đặt mức độ chi tiết được ghi lại trong logs.

## **Ứng dụng Log:**

Các ứng dụng được sử dụng để tạo, lưu trữ, truy cập và thao tác dữ liệu. Hệ điều hành cung cấp nền tảng cho các ứng dụng này trong khi vai trò của các hệ thống bảo mật là bảo vệ dữ liệu này. Trong khi một số ứng dụng tạo nhật ký riêng, một số ứng dụng sử dụng hệ thống ghi nhật ký của HĐH để ghi nhật ký dữ liệu liên quan.

Một số dữ liệu có thể được tìm thấy trong Log của các ứng dụng liên quan như sau:

* **Yêu cầu và phản hồi của Máy khách / Máy chủ:** Chúng có thể rất quan trọng trong việc xây dựng lại chuỗi sự kiện. Nếu xác thực người dùng thành công được ghi lại, có thể xác định người dùng đã thực hiện yêu cầu. Máy chủ email ghi lại người gửi, người nhận, chủ đề e-mail và chi tiết đính kèm; hồ sơ máy chủ web yêu cầu các URL và phản hồi được cung cấp bởi máy chủ; và các ứng dụng tài chính ghi lại các hồ sơ được truy cập bởi mỗi người dùng. Tất cả thông tin này được tập hợp lại có thể là một kho tàng bằng chứng trong một cuộc điều tra.
* **Thông tin liên quan đến tài khoản:** Các file Log có thể chứa thông tin liên quan đến các lần xác thực thành công và thất bại, tạo và xóa tài khoản, leo thang đặc quyền tài khoản và sử dụng đặc quyền thực tế. Các Log này giúp xác định các sự kiện như thử mật khẩu bắt buộc cũng như xác định ứng dụng được sử dụng khi nào và bởi ai.
* **Thông tin liên quan đến sử dụng:** Có thể tìm thấy thông tin như số giờ giao dịch (hoặc phút) và kích thước cụ thể của từng giao dịch (như kích thước e-mail, kích thước tải lên tệp, v.v.) trong file Log. Cơ quan pháp lý có thể sử dụng điều này để xác định lưu lượng thư điện tử tăng đột ngột, có thể là do virus hoặc phần mềm độc hại hoặc tải xuống các tệp lớn (như phim) vi phạm chính sách của công ty.
* **Thông tin liên quan đến hành động quan trọng:** Điều này có thể bao gồm các hành động như khởi động ứng dụng, tắt ứng dụng, các lỗi trong ứng dụng, cũng như việc thay đổi cấu hình ứng dụng chính.

# **Thực hành Quản lý Log**

Nền móng của sự thành công trong bất kỳ loại điều tra pháp y chính quá trình chuẩn bị. Như chúng ta đã thấy, Log là nguồn thông tin mẹ và nếu không có chúng, pháp y mạng sẽ bị tê liệt nghiêm trọng. Tội phạm cũng nhận ra điều này. Khi một tội phạm đã có quyền truy cập vào mạng của chúng ta, một trong những điều đầu tiên họ cố gắng làm là che dấu vết. Bước đầu tiên trong quy trình này là loại bỏ các Log ghi lại hoạt động của họ và sau đó sẽ phá vỡ đi sự an toàn của mạng. Để chống lại rủi ro này, các quy trình quản lý Log hợp lý phải được thực hiện.

Trong mọi tổ chức, có vô số hệ điều hành, nhiều phần mềm bảo mật và một số lượng lớn các ứng dụng; mỗi trong số đó tạo ra các bản ghi. Tất cả điều này làm cho việc quản lý Log rất phức tạp. Một số vấn đề liên quan đến quản lý Log được trình bày như sau:

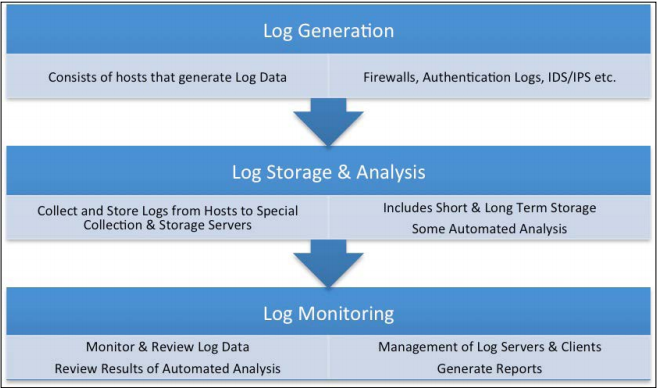
* **Nhiều nguồn nhật Log:** Log có thể được tạo trên nhiều máy chủ trong toàn tổ chức. Hơn nữa, một số ứng dụng có thể tạo ra nhiều Log, ví dụ, thông tin xác thực có thể đi vào một Log và việc sử dụng ứng dụng có thể được lưu trữ trong một Log khác.
* **Cấu trúc tệp Log không nhất quán:** Các nguồn khác nhau sử dụng các định dạng khác nhau cho các tệp Log của chúng. Log có thể được lưu trữ trong Giá trị phân tách bằng dấu phẩy (Comma Separated Values) (CSV), giá trị được phân tách bằng tab (tab-separated values) (TSV), cơ sở dữ liệu (databases), Ngôn ngữ đánh dấu mở rộng (Extensible Markup Language) (XML), sysLog, Giao thức đánh dấu mạng đơn giản (Simple Network Markup Protocol) (SNMP) và tệp nhị phân (binary)(như trong trường hợp của Windows).
* **Biểu diễn trường dữ liệu không nhất quán:** Các Log khác nhau thể hiện nội dung khác nhau. Một số Log sẽ xác định một máy chủ sử dụng địa chỉ IP của nó, trong khi một số khác sẽ sử dụng tên người dùng. Điều này có thể gây khó khăn cho việc tương quan giữa hai Log mặc dù chúng đại diện cho cùng một thực thể. Một vấn đề khác là làm thế nào các giá trị dữ liệu được thể hiện ra, ví dụ, ngày có thể được lưu trữ ở định dạng MMDDYYYY hoặc MM-DD-YYYY. Tương tự, một hệ thống có thể ghi Log sử dụng Giao thức truyền tệp (File Transfer Protocol) (FTP), trong khi một hệ thống khác chỉ có thể xác định nó bằng số cổng của nó 21.
* **Mốc thời gian không nhất quán:** Trong khi mỗi Log đều ghi lại các mốc thời gian, hệ thống đưa ra các mốc thời gian chuẩn để tham chiếu với mốc thời gian trong Log. Do đó, nếu thời gian của hệ thống không chính xác, thì mốc thời gian của Log cũng không chính xác. Trong trường hợp không thống nhất giữa nhiều bản ghi hệ thống bảo mật, một phân tích về chuỗi sự kiện trong vi phạm bảo mật có thể trở nên sai lầm nghiêm trọng.
* **Khối lượng dữ liệu lớn:** Bất kỳ tổ chức nào có kích thước hợp lý đều có thể tạo ra khối lượng lớn Log trong mạng của mình. Đa nguồn và giới hạn các lưu trữ cục bộ về các vấn đề phức tạp. Kích thước tệp Log thường được giới hạn bởi số lượng mục hoặc kích thước giới hạn của nó . Tuy nhiên, điều này có thể không phù hợp với cơ quan điều tra từ góc độ điều tra. Do đó, bắt buộc phải có cách thu thập và lưu trữ Log trong thời gian dài hơn ở các vị trí riêng biệt.

Các thành phần thiết yếu của hệ thống quản lý Log trong cơ quan pháp y bao gồm cơ sở hạ tầng quản lý Log và lập kế hoạch và chính sách quản lý Log.

## **Cơ sở hạ tầng quản lý Log:**

Các cuộc điều tra các vụ án về mạng dựa rất nhiều vào Log để làm sáng tỏ một vụ án. Đây thường là các chứng cứ mà tội phạm để lại sau mỗi lần tấn công. Nếu không có một cơ sở hạ tầng quản lý Log hiệu quả cũng như các Log cần thiết không có sẵn cuộc điều tra có thể bị đi lệch hướng hoặc ít nhất, bị cản trở nghiêm trọng.

Cơ sở hạ tầng quản lý Log thường có cấu trúc ba lớp như trong hình sau:



Cơ sở hạ tầng quản lý Log được yêu cầu thực hiện một số chức năng theo cách không thay đổi (hoặc ảnh hưởng đến tính toàn vẹn của) Log theo bất kỳ cách nào.

Một số chức năng cơ sở hạ tầng quản lý Log điển hình như sau:

|  |  |
| --- | --- |
| Chức năng | Mô tả |
| Phân tích Log | Điều này liên quan đến việc trích xuất dữ liệu từ Log và sau đó chuyển nó sang một quá trình Logging khác. |
| Lọc các sự kiện | Điều này liên quan đến việc lọc ra các sự kiện không có bất kỳ thông tin nào quan tâm. Điều này thường được thực hiện trong quá trình phân tích, báo cáo hoặc giai đoạn lưu trữ dài hạn để không làm thay đổi Log ban đầu. |
| Tổng hợp sự kiện | Điều này liên quan đến việc hợp nhất nhiều mục tương tự thành một mục, trong khi kết hợp số lần xuất hiện của sự kiện. |
| Luân chuyển Log | Điều này được thực hiện để giữ cho kích thước tệp Log có thể quản lý được và giữ được Log ban đầu. Điều này liên quan đến việc đóng tệp Log dựa trên một tiêu chí nhất định, chẳng hạn như kích thước tệp hoặc dữ liệu liên quan đến một khoảng thời gian cụ thể, chẳng hạn như một ngày, tuần, tháng, v.v. và sau đó bắt đầu một tiêu chí mới. |
| Lưu trữ Log | Điều này thường liên quan đến việc lưu trữ Log trong một khoảng thời gian dài. Lưu trữ Log có thể là một phần của việc duy trì Log hoặc bảo quản Log. Lưu giữ Log là một phần của quy trình vận hành tiêu chuẩn (standard operating procedures) (SOP) của một tổ chức; bảo quản Log là một phần của việc tuân thủ các yêu cầu quy định,và một cuộc điều tra diễn ra với mục tiêu bảo vệ được những dữ liệu quan trọng. |
| Nén Log | Điều này liên quan đến việc lưu trữ tệp Log ở dạng nén để giảm dung lượng lưu trữ cần thiết. Điều này được thực hiện nhằm ngăn chặn sửa đổi ý nghĩa của nội dung của tệp. |
| Giảm Log | Điều này liên quan đến việc loại bỏ các mục không cần thiết trong một tệp log để làm cho nó có dung lượng nhỏ hơn. |
| Chuyển đổi Log | Sau khi phân tích cú pháp của tệp Log, một tệp Log trải qua một sự thay đổi trong định dạng trước khi lưu trữ. Điều này được gọi là chuyển đổi Log. |
| Chuẩn hóa Log | Điều này liên quan đến việc chuyển đổi từng trường dữ liệu thành định dạng chuẩn để phân loại nhất quán. Ví dụ, chuẩn hóa vô số định dạng ngày và giờ thành một định dạng chuẩn hóa duy nhất. Quá trình này là rất cần thiết để có thể thực hiện các mối tương quan và các phân tích khác dưới mọi hình thức. Đây có thể là một nhiệm vụ khá khó khăn. |
| Kiểm tra tính toàn vẹn của tệp Log | Điều này liên quan đến việc tạo ra một giá trị băm (sử dụng các thuật toán mã hóa như MD5 và SHA) để thiết lập và duy trì tính toàn vẹn của các tệp Log trong suốt quá trình quản lý Log. |
| Tương quan sự kiện | Điều này liên quan đến việc tìm kiếm mối quan hệ giữa nhiều mục Log. Ví dụ: xác định tất cả lưu lượng truy cập liên quan đến một địa chỉ IP cụ thể hoặc lưu lượng truy cập Internet Control Message Protocol (ICMP) (ping) từ một loạt các máy trong mạng để theo dõi sự lây lan của phần mềm độc hại. |
| Xem Log | Điều này liên quan đến việc sử dụng các công cụ chuyên dụng để hiển thị Log ở định dạng có thể đọc được. |
| Báo cáo Log | Đây là kết quả hiển thị của việc phân tích. Báo cáo Log có thể liên quan đến đầu ra bảng tổng hợp cũng như các báo cáo điển hình. Ngày nay, các hệ thống cung cấp khả năng phân tích sâu để tăng mức độ chi tiết của báo cáo. |
| Loại bỏ Log | Log không thể được lưu trữ vô thời hạn. Dựa trên các tiêu chuẩn, chính sách duy trì và các yêu cầu pháp lý, các bản ghi phải được loại bỏ theo thời gian. |

## **Lập kế hoạch và chính sách quản lý Log**

Để bất kỳ một cuộc điều tra pháp y nào có ý nghĩa, tổ chức cần lên kế hoạch trước và xác định các yêu cầu và mục tiêu liên quan đến quản lý Log.

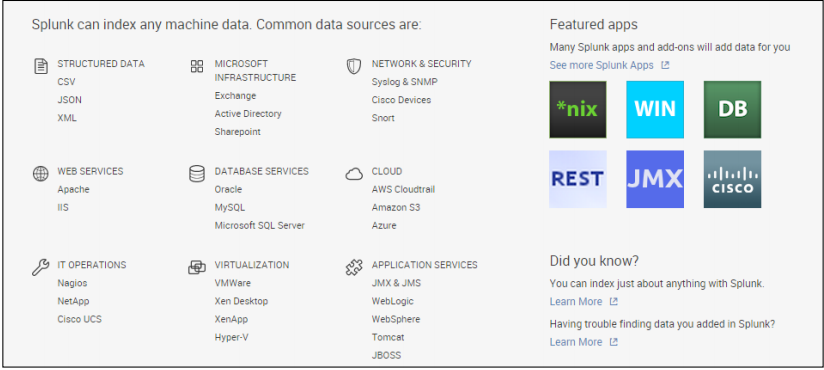
Những điều này có thể bị ảnh hưởng bởi môi trường pháp lý, chẳng hạn như Đạo luật về trách nhiệm giải trình và trách nhiệm bảo hiểm y tế (Health Insurance Portability and Accountability Act)(HIPAA), Tiêu chuẩn bảo mật dữ liệu thẻ thanh toán công nghiệp (Payment Card Industry Data Security Standard) (PCI DSS), v.v. Từ góc độ kinh doanh liên tục, một sự cân bằng phải được duy trì giữa thời gian và nguồn lực cần thiết và giảm hồ sơ rủi ro của tổ chức. Các khía cạnh của quản lý Log cần được xem xét khi xác định các chính sách và thủ tục như sau:

|  |  |
| --- | --- |
| Khía cạnh quản lý Log | Các vấn đề cần giải quyết |
| Vai trò và trách nhiệm | Vai trò và trách nhiệm của những điều sau đây cần được xác định rõ ràng:  • Quản trị viên hệ thống  • Quản trị viên mạng  • Quản trị viên bảo mật  • Đội ứng phó khắc phục sự cố  • Kỹ thuật viên, lập trình viên.  • Kiểm toán viên  • Nhân viên an ninh thông tin  • Giám đốc (CIO) |
| Thiết lập chính sách khai thác | Các yêu cầu bắt buộc và các yêu cầu cần được xác định cho các mục sau:   * Tạo Log * Những máy chủ và loại máy chủ sẽ đăng nhập? * Những thành phần máy chủ nào (HĐH / dịch vụ / ứng dụng) sẽ đăng nhập? * Những sự kiện nào để đăng nhập (sự kiện bảo mật, v.v.)? * Thông tin nào được ghi lại cho mỗi sự kiện (tên người dùng và IP nguồn cho đăng nhập)? * Tần suất đăng nhập sẽ là gì? (mọi lúc, 100 lần, v.v.) * Truyền Log * Những máy chủ nào sẽ chuyển đến máy chủ quản lý Log? * Dữ liệu / mục nào nên được chuyển đến cơ sở hạ tầng? * Khi nào và bao lâu thì nên làm điều này? * Những phương thức / giao thức nên được sử dụng để truyền? (Yêu cầu về băng thông cho việc truyền tải sẽ cần được xem xét.) * Tính bảo mật, tính toàn vẹn và tính sẵn sàng (CIA) sẽ được duy trì như thế nào trong quá trình này? * Lưu trữ và xử lý Log * Bao lâu Log sẽ được luân chuyển? * Bao lâu bản ghi sẽ được lưu trữ? * Làm thế nào các yêu cầu bảo quản pháp lý sẽ được xử lý? * Cần bao nhiêu dung lượng lưu trữ? * Những quy trình nào sẽ được theo sau để xóa / xử lý an toàn? * Tính bảo mật, tính toàn vẹn và tính sẵn sàng sẽ được duy trì như thế nào trong quá trình lưu trữ và xử lý? * Phân tích Log * Bao lâu thì mỗi loại Log sẽ được phân tích? * Các bước cần tuân thủ khi một hoạt động đáng ngờ được xác định là gì? * Ai sẽ được phép truy cập dữ liệu Log và ai sẽ làm theo các bước này? * Việc vô tình tiếp xúc với dữ liệu bí mật trong Log (như nội dung / mật khẩu email) sẽ được xử lý như thế nào? * Tính bảo mật, tính toàn vẹn và tính sẵn sàng của phân tích Log và báo cáo sẽ được bảo vệ như thế nào? |

Bây giờ chúng ta đã hiểu các nguyên tắc cơ bản đằng sau Log và cách quản lý chúng, hãy xem cách phân tích Log mạng.

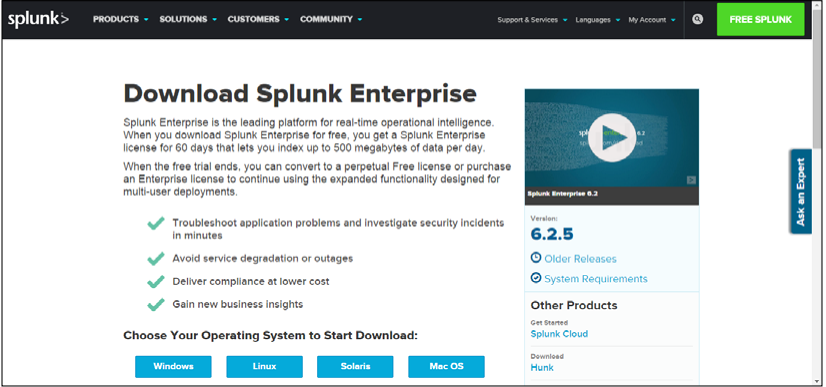
# **Phân tích Log mạng bằng Splunk**

Công cụ lựa chọn là Splunk. Đây là một công cụ rất linh hoạt (nó cũng có phiên bản miễn phí) cung cấp cho người dùng khả năng thu thập tệp Log từ nhiều nguồn, lập chỉ mục và chuẩn hóa dữ liệu bên trong, sau đó tiến hành phân tích chuyên sâu để tìm kiếm sự bất thường, chuẩn bị báo cáo, và hình dung kết quả. Ví dụ:



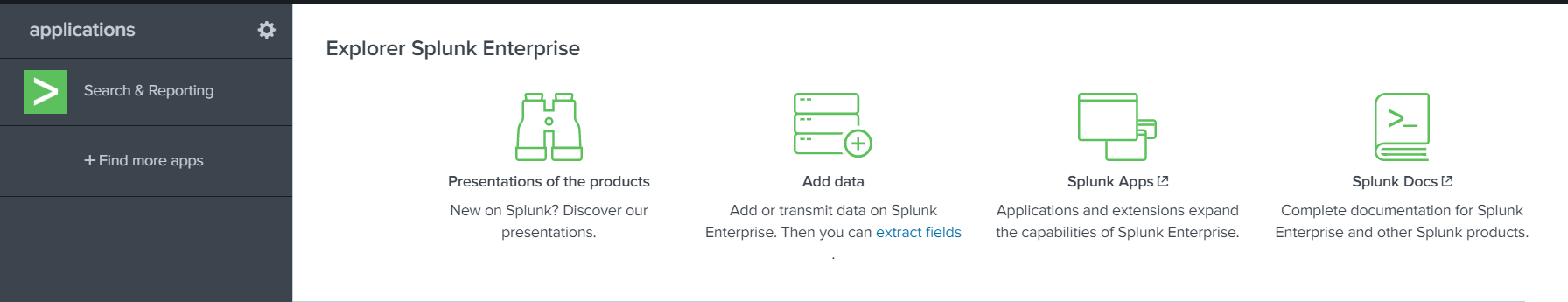
Splunk cung cấp phương tiện để nhập và lập chỉ mục dữ liệu theo vô số định dạng. Điều này bao gồm dữ liệu có cấu trúc, dịch vụ web, thiết bị mạng, máy chủ Microsoft, dịch vụ ứng dụng, đám mây và một loạt các dịch vụ khác. Sau đây là các bước để tải xuống và bắt đầu sử dụng Splunk:

Để bắt đầu, cần tải xuống Splunk từ http: //www.splunk. com /, như thể hiện trong ảnh chụp màn hình sau:



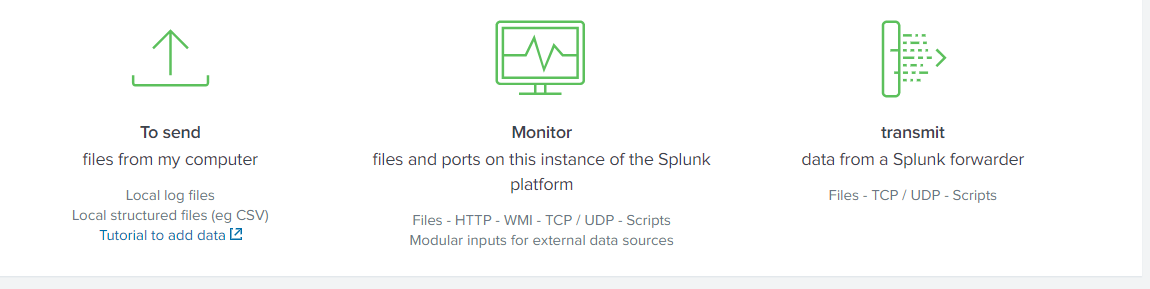
1. Tất cả những gì chúng ta cần để bắt đầu là tải xuống tệp nhị phân có liên quan dựa trên hệ điều hành mà chúng ta dự định sử dụng để phân tích. Như chúng ta có thể thấy trong ảnh chụp màn hình trước đó, Splunk hỗ trợ Windows, Linux, Solaris và Mac OS.
2. Trong trường hợp này, ta sẽ tải xuống tệp nhị phân Windows để có thể thu thập Log trực tiếp từ chính hệ thống và xem liệu có thể có được một số dữ liệu thú vị hay không.
3. Khi quá trình tải xuống hoàn tất, Splunk yêu cầu người dùng tạo tên người dùng và mật khẩu. Mật khẩu được tạo cho phép quản trị viên truy cập vào Spunk thông qua màn hình đăng nhập dựa trên trình duyệt.

Bước tiếp theo, chúng ta sẽ thêm dữ liệu vào Splunk:

****

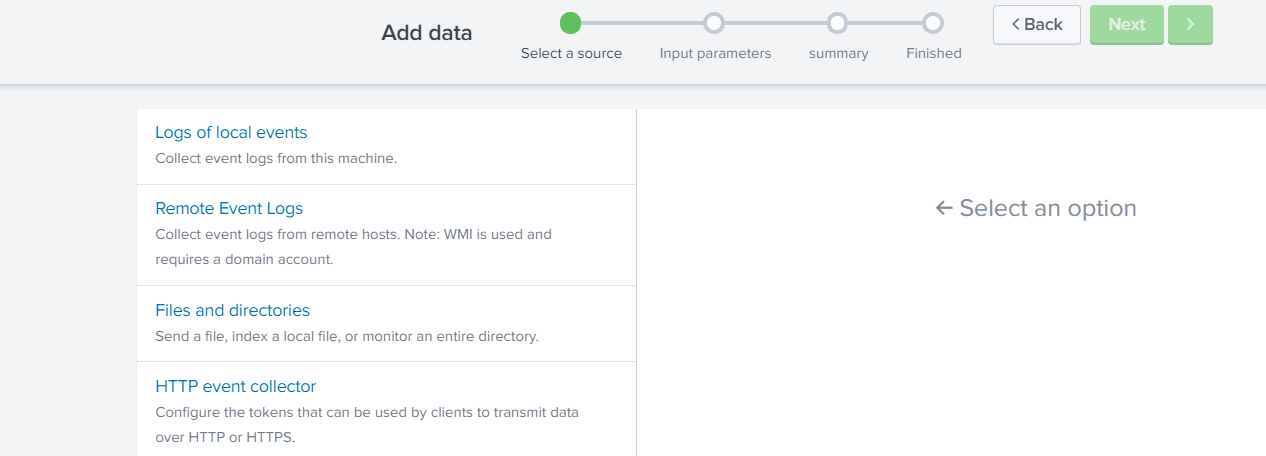
Để thực hiện việc này, hãy hoàn thành các bước sau:

1. Nhấp vào tùy chọn đầu tiên, **thêm dữ liệu**.
2. Sau đó, ta được cung cấp ba tùy chọn như trong ảnh dưới đây:

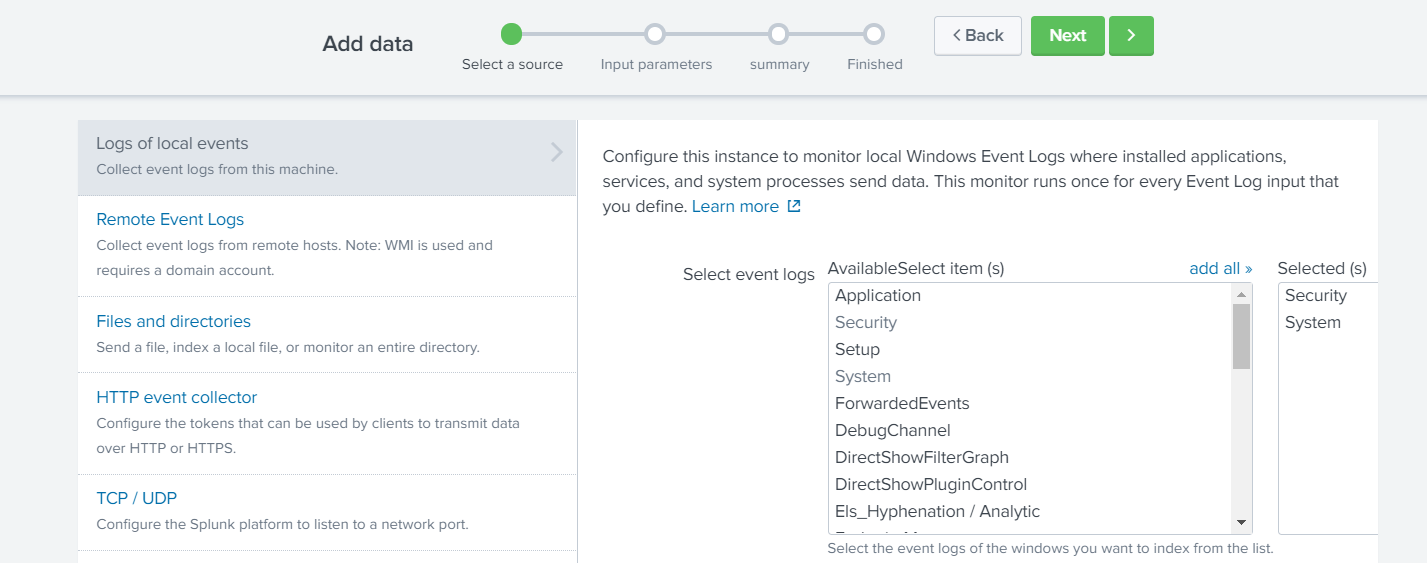


* Đầu tiên là tùy chọn tải lên: Tùy chọn này sẽ được chọn khi chúng ta muốn phân tích các tập tin logs mà chúng ta đã mang từ máy tính đang được điều tra hoặc có cục bộ trên ổ đĩa hoặc phương tiện lưu động. Trong trường hợp này, phân tích dữ liệu của các sự kiện đã xảy ra trong quá khứ và các tập tin logs không tự động thay đổi. Chúng ta có thể sử dụng cho các tập tin logs của hệ thống được tải xuống như là một phần của hướng dẫn trên trang web Splunk.
* Thứ hai là tùy chọn màn hình: Tùy chọn này, chúng ta có thể giám sát các cổng cũng như các tập tin đang thay đổi tự động trên hệ thống. Điều này khá hữu ích khi chúng ta muốn theo dõi logs hoạt động trong thời gian thực.
* Thứ ba là tùy chọn chuyển tiếp: Tùy chọn này cho phép nhiều trường hợp của Splunk chạy trên các máy khác nhau. Tại một đầu, Splunk đóng vai trò như một người thu gom và ở bên kia, nó đóng vai trò như một người nhận. Điều này giúp trong việc thu thập và củng cố và hợp nhất các bản ghi qua mạng.

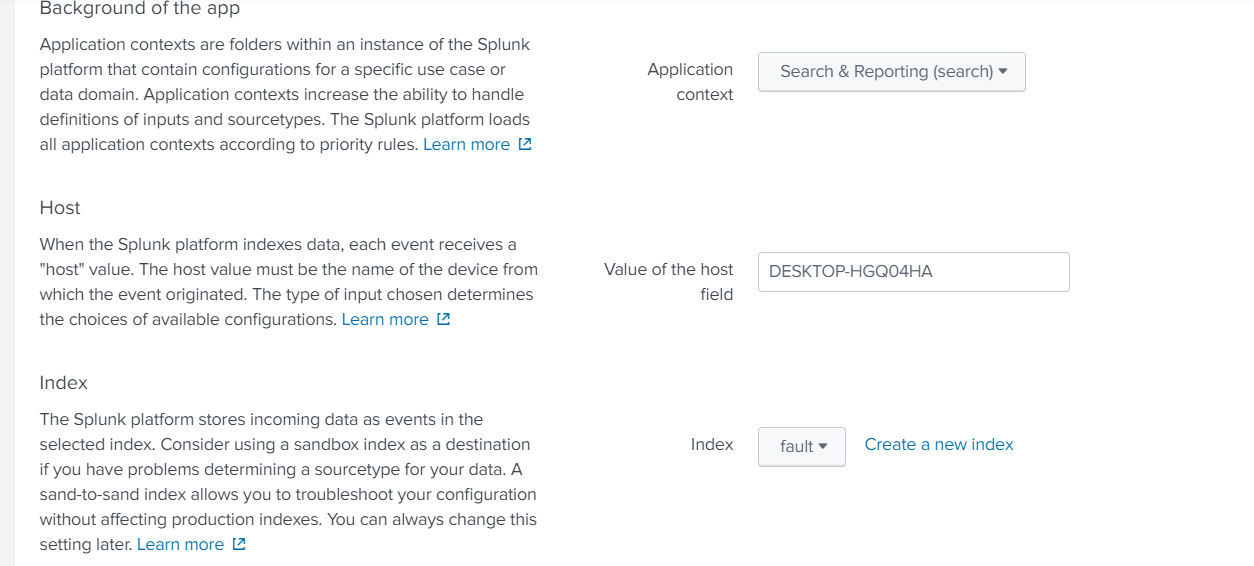
1. Trong trường hợp này, hãy bắt đầu bằng cách phân tích các bản ghi logs hệ thống của chính chúng ta một cách tự động. Bước đầu tiên là chọn nguồn dữ liệu. Trong trường hợp của chúng ta, chúng ta sẽ chọn Local Event Logs (Nhật ký sự kiện cục bộ), như được hiển thị trong hình dưới đây:



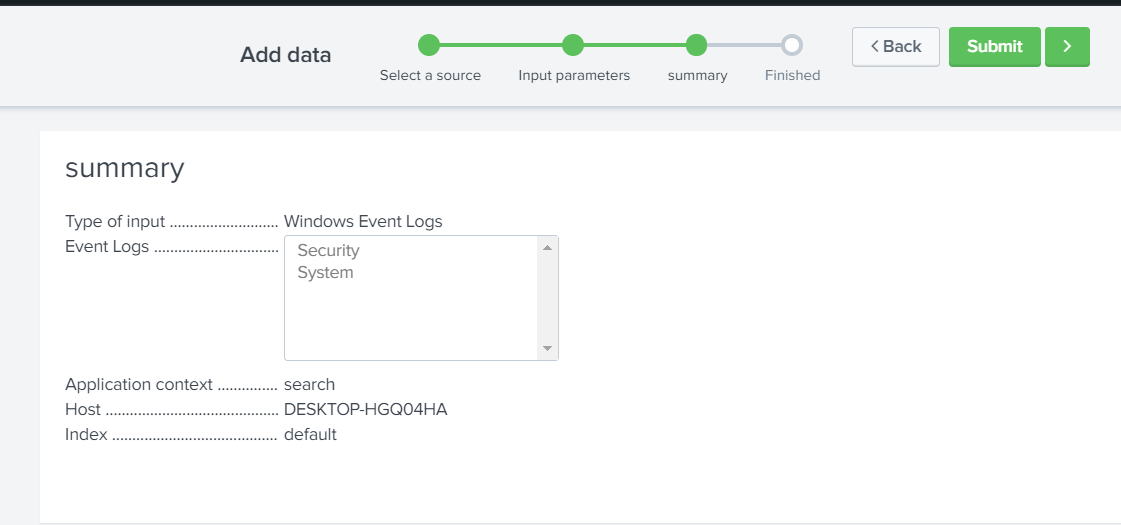
1. Khi chúng ta thực hiện như trên, Splunk tìm kiếm tất cả các loại logs trên hệ thống cục bộ. Khi chúng được xác định (chỉ mất vài giây), chúng ta sẽ thấy một cửa sổ hiển thị các bản ghi mà chúng ta muốn nhập vào Splunk và sau đó lập chỉ số để phân tích sau này, hiển thị như trong ảnh dưới đây:



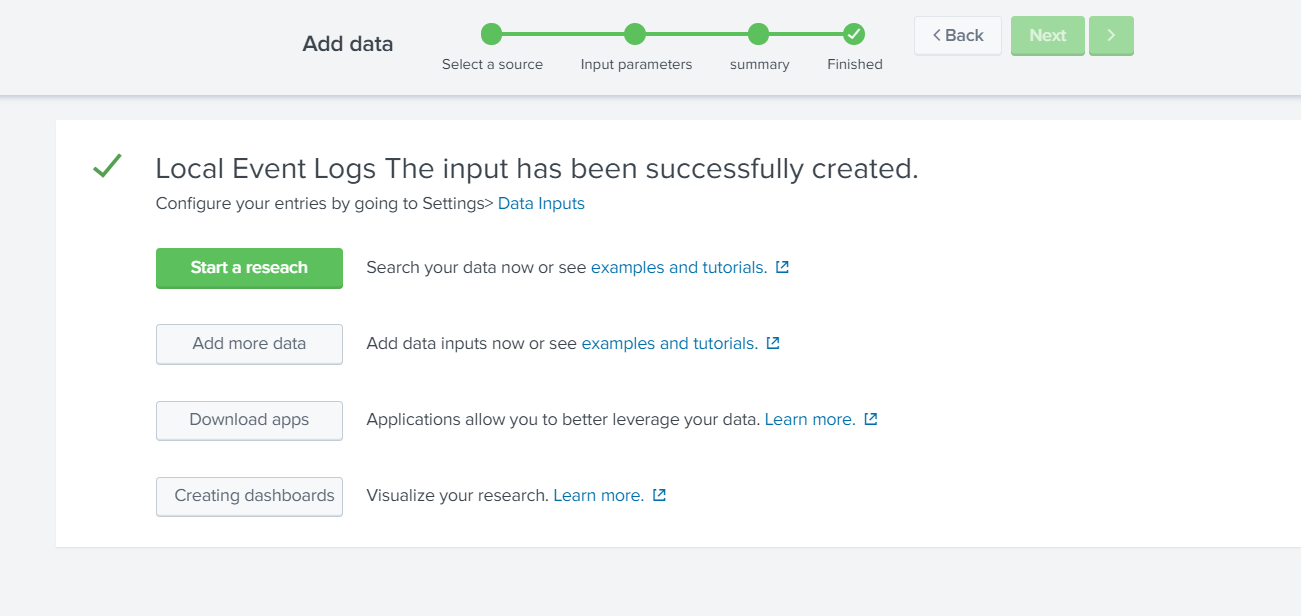
1. Chúng ta có thể chọn một logs hoặc thêm tất cả vào cùng một lúc. Sau khi thực hiện xong, chúng ta chuyển sang bước cài đặt đầu vào, như hình sau:



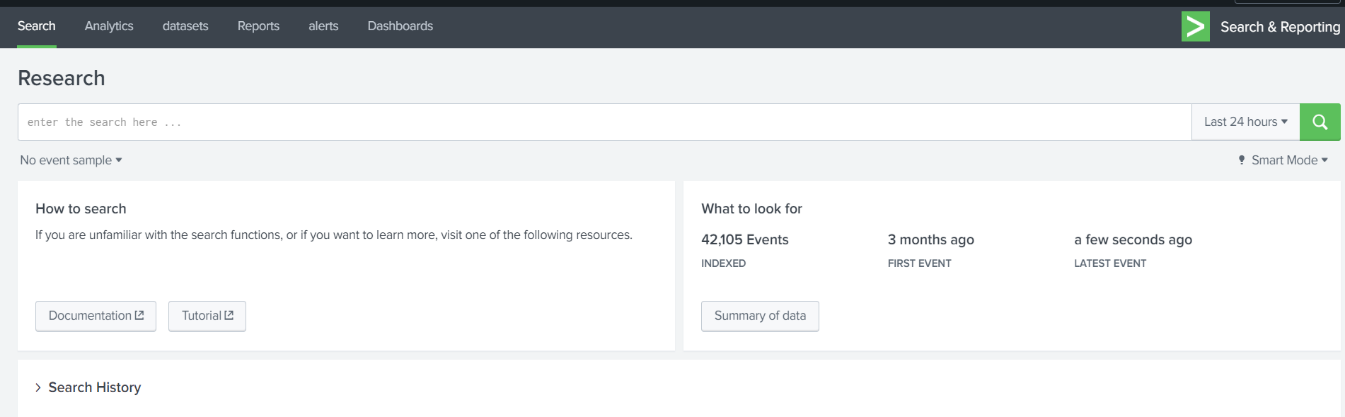
1. Giai đoạn này là tùy chọn. Ở đây, các tham số đầu vào bổ sung cho dữ liệu được xác định. Trong trường hợp này, chúng ta chỉ chấp nhận mặc định và chuyển sang giai đoạn tổng hợp, như thể hiện trong hình sau:



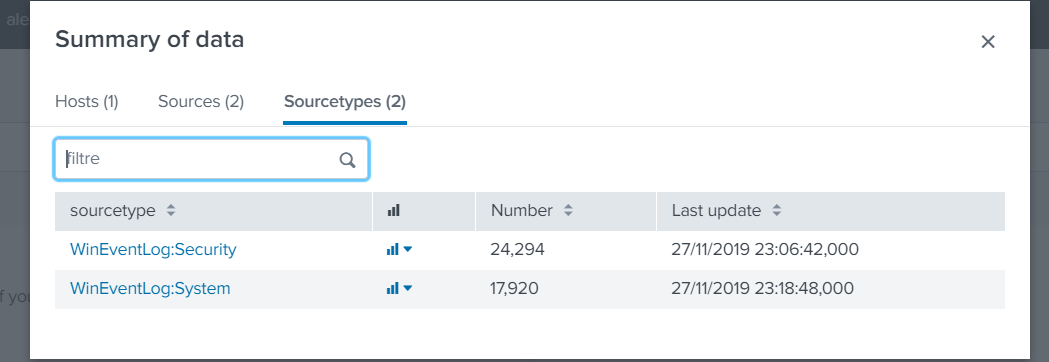
1. Chúng ta xem lại bản tổng hợp của chúng ta và xác nhận rằng tất cả mọi thứ đều phải như vậy. Sau khi hoàn tất, việc nhập dữ liệu của chúng ta vào Splunk được thực hiện. Trong nền, Splunk di chuyển sang chế độ lập chỉ số, nơi nó xây dựng các chỉ số chuyên sâu đầy đủ từ các logs để truy xuất tốc độ cao, như được hiển thị trong ảnh chụp màn hình sau:



Đây là giai đoạn chờ đợi của chúng ta đã kết thúc và chúng ta có thể tận hưởng tính năng 007. Bây giờ chúng ta có thể bắt đầu tìm kiếm tất cả các hoạt động khả nghi. Trang tìm kiếm Splunk được hiển thị trong hình sau:



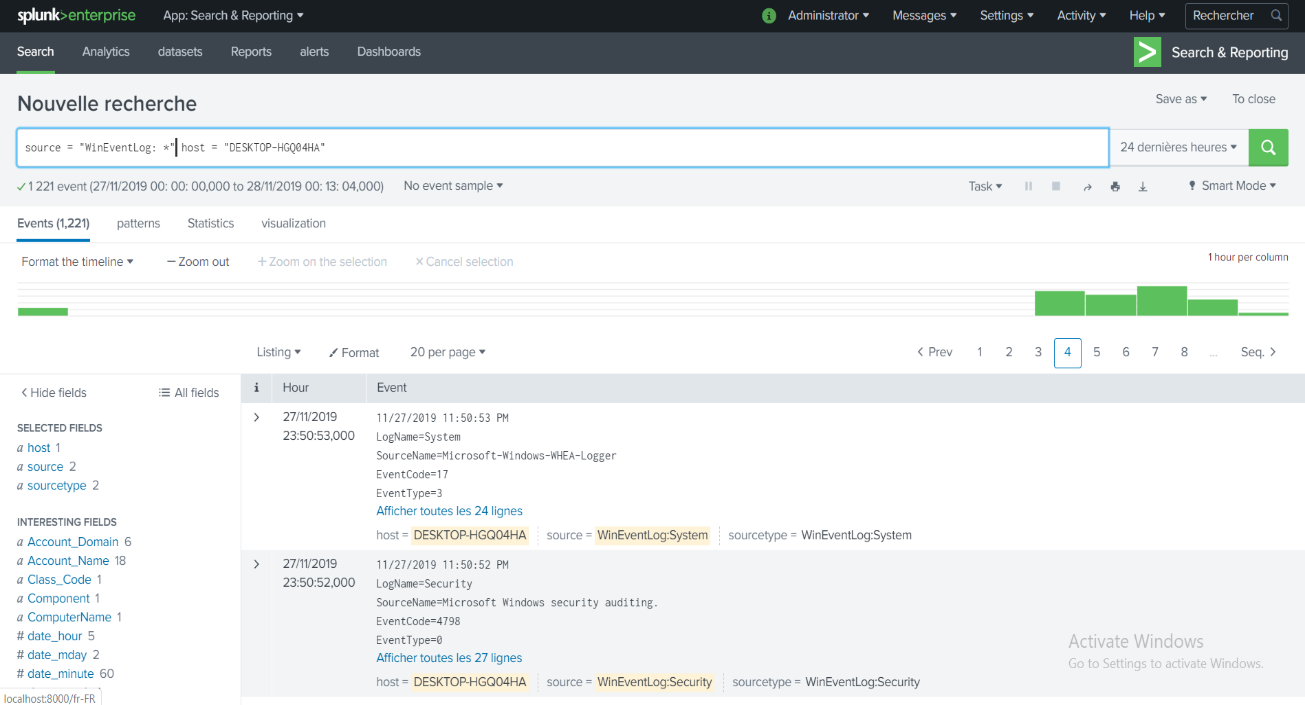
Bên dưới, Splunk liệt kê số lượng events được lập chỉ số cũng như event trước đó và sau đó.Nhấp chuột vào Data Summary hiển thị kết quả sau:



Có ba tab. Tab Hosts đầu tiên cho thấy rằng dữ liệu được Splunk lập chỉ số liên quan đến một máy chủ duy nhất. Hai tab tiếp theo cho chúng ta thấy rằng dữ liệu liên quan đến bốn nguồn và loại nguồn khác nhau. Chúng ta hãy chọn một máy chủ duy nhất và xem liệu có bất kỳ sự kiện bảo mật nào có giá trị hay không.

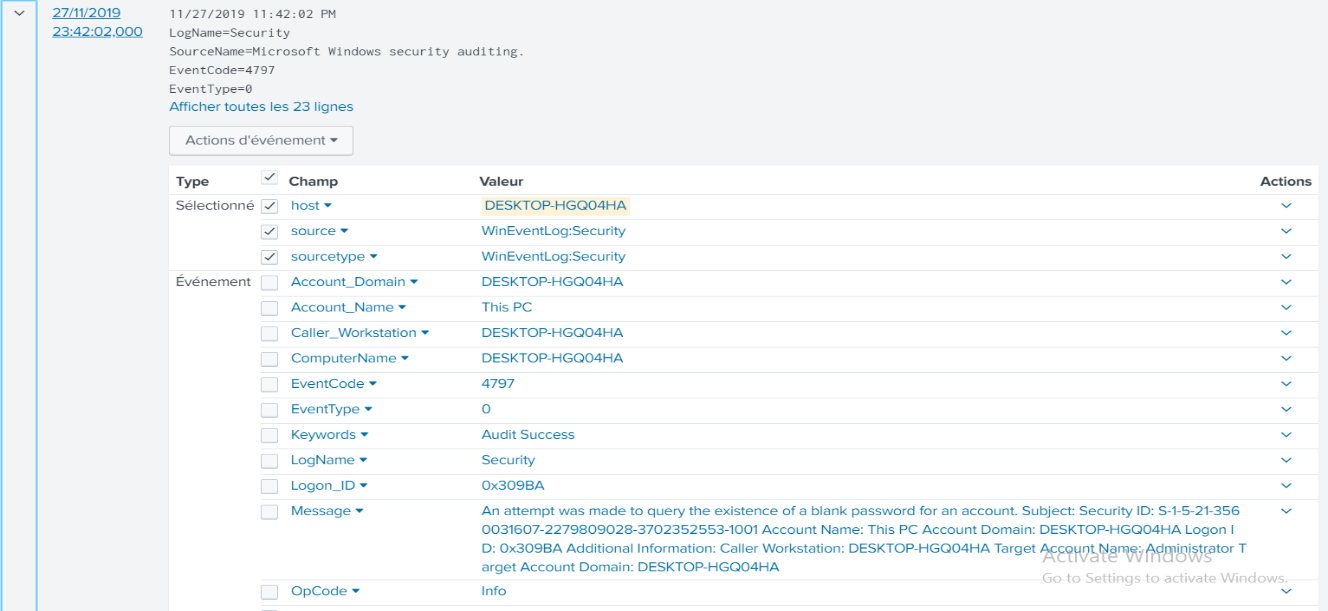
Điểm hay của Splunk là nó có giao diện tìm kiếm kiểu Google. Chúng ta chỉ cần gõ vào văn bản mà chúng ta muốn tìm kiếm và nó cho chúng ta thấy kết quả. Trong trường hợp này, hãy tìm các log có từ đăng nhập, cũng như một từ bắt đầu bằng pass.

Thẻ wild \* biểu thị rằng đầu ra sẽ bao gồm mọi thứ bắt đầu bằng pass bao gồm cả pass và mật khẩu (đó là những gì chúng ta đang tìm kiếm). Nó được hiển thị trong hình sau:



Nhìn vào giao diện chúng ta thấy rằng có 1221 events đáp ứng tiêu chí của chúng ta vào tháng 11 năm 2019. Đây là điều mà chúng ta chắc chắn muốn tìm hiểu chi tiết hơn.

Hãy mở rộng các mục này bằng cách nhấp vào dấu “>” ở phía bên trái của mục, như được hiển thị trong phần sau:



Mọi thứ đã trở nên rất rõ ràng, chi tiết. Nếu chúng ta nhìn kỹ, chúng ta sẽ thấy rằng ngày 27 tháng 11 năm 2015, vào khoảng 23 giờ 42 phút, một xâm nhập đã cố gắng để đăng nhập vào PC Account Domain: DESKTOP-HGQ04HA với một mật khẩu trống. Đây là một ví dụ về những điều thú vị mà chúng ta đang tìm hiểu.

Trong khi điều này phản ánh một sự kiện duy nhất, nó sẽ không tốt nếu chúng ta có thể nhìn thấy tất cả các sự kiện tương tự một cách trực quan. May mắn cho chúng ta, đây là những gì Splunk làm rất tốt.

Loại thông tin này rất có giá trị vì nó cho chúng ta một cái nhìn tổng quan ngay lập tức về loại tấn công đang được đưa vào nguồn thông tin của chúng ta và cho chúng ta điểm khởi đầu lý tưởng để bắt đầu cuộc điều tra.

Với một chút thực hành và sử dụng các hướng dẫn trực tuyến tốt có sẵn với Splunk, nó là khá dễ dàng để có được các kiến thức chuyên môn cần thiết để bắt đầu quá trình phân tích logs và nâng cao sự hiểu biết của bạn về mạng pháp y ở một khía cạnh khác.

# **Tổng kết**

Trong chương này, chúng ta đã tìm hiểu được thêm về một tài nguyên có giá trị khác trong nỗ lực tìm hiểu mạng pháp y. Tầm quan trọng của việc ghi logs và các tham số khác nhau trong quản lý logs đã được nhấn mạnh và thảo luận. Chúng ta hiểu những câu hỏi cần trả lời khi chúng ta thiết lập một hệ thống ghi logs để hỗ trợ chúng ta tại thời điểm điều tra mạng pháp y. Chúng ta đã thấy tầm quan trọng của độ chính xác của đồng hồ, mối tương quan, thu thập, lưu trữ và một loạt các yếu tố khác trong việc ghi lại. Chúng ta đã thấy làm thế nào để có được nhiều logs khác nhau vào một hệ thống quản lý logs duy nhất có thể giúp chúng ta kết nối các dấu chấm và cung cấp cho chúng ta cái nhìn sâu sắc và toàn diện về hoạt động của những kẻ xâm nhập không tham gia vào mạng của chúng ta.