|  |
| --- |
| BAN CƠ YẾU CHÍNH PHỦ  **HỌC VIỆN KỸ THUẬT MẬT MÃ**  ¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯ |
| THU THẬP VÀ PHÂN TÍCH THÔNG TIN AN NINH MẠNG  **TÌM HIỂU CÁC CÔNG CỤ PHÂN TÍCH NHẬT KÝ ỨNG DỤNG TRONG PHÁT HIỆN TẤN CÔNG WEB** |
| Khoa: Công nghệ thông tin  Chuyên ngành: An toàn thông tin |
| TP.HCM, 2020 |

|  |
| --- |
| BAN CƠ YẾU CHÍNH PHỦ  **HỌC VIỆN KỸ THUẬT MẬT MÃ**  ¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯ |
| THU THẬP VÀ PHÂN TÍCH THÔNG TIN AN NINH MẠNG  **TÌM HIỂU CÁC CÔNG CỤ PHÂN TÍCH NHẬT KÝ ỨNG DỤNG TRONG PHÁT HIỆN TẤN CÔNG WEB** |
| Khoa: Công nghệ thông tin  Chuyên ngành: An toàn thông tin  Giáo viên hướng dẫn: **ThS Phạm Tấn Toàn**  Sinh viên thực hiện:   * **Nguyễn Hoàng Lân** * **Phan Hoàng Trung** * **Lê Thanh Lâm** * **Trần Thị Ngọc**   Lớp: **AT13N01** |
| TP.HCM, 2020 |

BAN CƠ YẾU CHÍNH PHỦ

**HỌC VIỆN KỸ THUẬT MẬT MÃ**

¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯

BÁO CÁO BÀI TẬP LỚN

THU THẬP VÀ PHÂN TÍCH THÔNG TIN AN NINH MẠNG

ĐỀ TÀI:

**TÌM HIỂU CÁC CÔNG CỤ PHÂN TÍCH NHẬT KÝ ỨNG DỤNG TRONG PHÁT HIỆN TẤN CÔNG WEB**

Nhận xét của cán bộ hướng dẫn:..................................................................................

.......................................................................................................................................

.......................................................................................................................................

.......................................................................................................................................

.......................................................................................................................................

Điểm chuyên cần:.........................................................................................................

Điểm báo cáo:...............................................................................................................

**Xác nhận của cán bộ hướng dẫn**

MỤC LỤC

[LỜI NÓI ĐẦU 1](#_Toc47296533)

[CHƯƠNG 1. TÌM HIỂU CHUNG VỀ PHÂN TÍCH NHẬT KÝ AN NINH MẠNG 2](#_Toc47296534)

[1.1. Khái niệm? 2](#_Toc47296535)

[1.2. Tầm quan trọng của việc phân tích nhật ký trang. 2](#_Toc47296536)

[1.2.1. Giảm thời gian chẩn đoán và giải quyết vấn đề 2](#_Toc47296537)

[1.2.2. Cải thiện sử dụng tài nguyên & chi phí cơ sở hạ tầng sản xuất 3](#_Toc47296538)

[CHƯƠNG 2. CÁC CÔNG CỤ PHÂN TÍCH NHẬT KÝ 4](#_Toc47296539)

[2.1. Logstas 4](#_Toc47296540)

[2.1.1. Khái niệm và thành phần 4](#_Toc47296541)

[2.1.2. Cách thức hoạt động 4](#_Toc47296542)

[2.2. Solarwind Loggly 7](#_Toc47296543)

[2.2.1. Solarwind Log Analyze 8](#_Toc47296544)

[2.3. GrayLog 9](#_Toc47296545)

[2.4. Splunk 11](#_Toc47296546)

[2.4.1. Khái niệm 11](#_Toc47296547)

[2.4.2 Tính năng 11](#_Toc47296548)

[2.4.3 Splunk Architecture 12](#_Toc47296549)

[2.5. Datadog 14](#_Toc47296550)

[2.6. Logentries 14](#_Toc47296551)

[2.7. Stackify 15](#_Toc47296552)

[2.8. ManageEngine EventLog Analyzer 16](#_Toc47296553)

[2.9. Logalyze 17](#_Toc47296554)

[2.10. LogDNA 18](#_Toc47296555)

[2.11. XpoLog 19](#_Toc47296556)

[2.12. Papertrail 20](#_Toc47296557)

[2.13. EvenTracker 21](#_Toc47296558)

[TÀI LIỆU THAM KHẢO 23](#_Toc47296559)

# LỜI NÓI ĐẦU

Hiện nay, có rất nhiều các thiết bị, công cụ hỗ trợ điều tra và phân tích tấn công một cách dễ dàng, ví dụ như các hệ thống: IDS/IPS, honey pot, honey net,... Tuy nhiên trong bài viết này sẽ đưa ra hai phương pháp chính hỗ trợ điều tra, phân tích tấn công web phía máy chủ, với trường hợp máy chủ Linux Apache và không hỗ trợ các hệ thống phát hiện xâm nhập hay phân tích dữ liệu hiện đại, chủ yếu dựa trên các công cụ mã nguồn mở miễn phí.

Đề tài này sẽ cung cấp cho các bạn sinh viên một cái nhìn tổng quan về công cụ phân tích nhật ký ứng dụng trong phát hiện tấn công web.

Nội dung đề tài gồm 2 chương:

Chương 1. Tìm hiểu chung về phân tích nhật ký an ninh mạng.

Chương 2. Các công cụ phân tích nhật ký.

Qua quá trình tìm hiểu và nghiên cứu nhóm đã đạt được những mục tiêu đề ra, tuy nhiên thời gian còn hạn chế việc chọn lọc và trình bày các vấn đề trong quá trình nghiên cứu không tránh khỏi thiếu sót. Rất mong nhận được các ý kiến đóng góp của thầy, cô, cũng như các bạn học viên để nghiên cứu này được hoàn thiện hơn.

# CHƯƠNG 1. TÌM HIỂU CHUNG VỀ PHÂN TÍCH NHẬT KÝ AN NINH MẠNG

## 1.1. Khái niệm?

Phân tích nhật ký là quá trình hiểu ý nghĩa của thông điệp tường trình do máy tính tạo ra, còn được gọi là sự kiện nhật ký, hồ sơ theo dõi kiểm toán hoặc đơn giản là nhật ký. Phân tích nhật ký cung cấp các số liệu hữu ích vẽ nên một bức tranh rõ ràng về những gì đã xảy ra trên cơ sở hạ tầng. Bạn có thể sử dụng dữ liệu này để cải thiện hoặc giải quyết các vấn đề về hiệu suất trong một ứng dụng hoặc cơ sở hạ tầng. Nhìn vào bức tranh lớn hơn, các công ty phân tích nhật ký để chủ động và phản ứng giảm thiểu rủi ro, tuân thủ các chính sách bảo mật, kiểm toán và quy định và hiểu hành vi của người dùng trực tuyến.

## 1.2. **Tầm quan trọng của việc phân tích nhật ký trang.**

Hầu hết các doanh nghiệp được yêu cầu thực hiện lưu trữ nhật ký và phân tích nhật ký như là một phần của quy định tuân thủ của họ. Họ phải thường xuyên theo dõi và phân tích nhật ký hệ thống để tìm kiếm lỗi, sự bất thường hoặc hoạt động đáng ngờ hoặc trái phép đi lệch khỏi định mức. Điều này cho phép họ tạo lại chuỗi sự kiện dẫn đến sự cố và khắc phục sự cố một cách hiệu quả.

Hơn nữa, mặc dù thoạt nhìn, phân tích nhật ký dữ liệu dường như chỉ ảnh hưởng đến khía cạnh CNTT của doanh nghiệp của bạn, nhưng thực tế nó ảnh hưởng đến tất cả các khía cạnh của nó, từ pháp lý, đến tài chính, bán hàng và tiếp thị, nhân lực, bảo mật và hoạt động. Khi tận dụng phân tích nhật ký, bạn có thể phát hiện các vấn đề trước hoặc khi chúng xảy ra và tránh lãng phí thời gian, sự chậm trễ không cần thiết và chi phí bổ sung, như chúng tôi sẽ giải thích ngay sau đây.

### ****1.2.1.**** Giảm thời gian chẩn đoán và giải quyết vấn đề

Tìm kiếm các vấn đề có thể là một nhiệm vụ tẻ nhạt và lãng phí thời gian, đặc biệt là khi nó không rõ ràng nếu vấn đề nằm ở lớp ứng dụng hoặc cơ sở hạ tầng.

Cho dù đó là cái trước, cái sau hay sự kết hợp của cả hai, chúng sẽ kéo dài thời gian ứng dụng của bạn mang lại trải nghiệm người dùng kém. Phân tích tệp nhật ký cho phép bạn thực hiện một cách tiếp cận chủ động bằng cách chỉ ra các vấn đề và nguyên nhân gốc rễ của chúng trước hoặc khi chúng xảy ra. Điều này tránh mất thời gian và giảm MTTR.

Ngoài ra, DevOps có thể can thiệp và giải quyết vấn đề nhanh hơn, do đó cho phép họ tập trung hơn vào việc cải thiện các chức năng hiện có và thêm các sản phẩm và dịch vụ mà họ đang tạo thay vì mất thời gian khắc phục sự cố. Điều này, đến lượt nó, làm tăng giá trị của phần mềm họ đang xây dựng, dẫn đến việc phát hành thường xuyên hơn và tăng giá trị tổng thể cho doanh nghiệp.

### ****1.2.2.**** Cải thiện sử dụng tài nguyên & chi phí cơ sở hạ tầng sản xuất

Một trong những khía cạnh thách thức nhất của bất kỳ tổ chức nào là quản lý tài nguyên - từ băng thông mạng đến chu kỳ CPU hoặc dung lượng lưu trữ và hơn thế nữa.

Bạn có thể suy đoán kích thước tài nguyên, nhưng bạn kết thúc với không đủ tài nguyên - dẫn đến hiệu suất kém, khách hàng thất vọng và cuối cùng, mất doanh số - hoặc quá nhiều - làm tăng chi phí, do đó ảnh hưởng đến lợi nhuận của bạn. Thay vì đoán các yêu cầu tài nguyên của bạn, phân tích tệp nhật ký - cùng với việc sử dụng và lập kế hoạch tài nguyên dựa trên số liệu - cho phép bạn hiểu dễ dàng và chính xác hơn việc sử dụng tài nguyên hiện tại và các yêu cầu tài nguyên trong tương lai của bạn.

Ngoài ra, khi nói đến hiệu năng hệ thống, thường thì không phải là phần mềm bị lỗi mà là người dùng yêu cầu quá tải hệ thống của bạn đến mức gặp khó khăn khi xử lý nhu cầu. Phân tích nhật ký cho phép bạn theo dõi việc sử dụng tài nguyên và phát hiện nơi hệ thống của bạn đang gặp khó khăn để bạn có thể bổ sung thêm dung lượng.

Mặt khác, bạn cũng có thể thấy các tài sản bị sử dụng không đúng mức hoặc đã chết để bạn có thể cơ cấu lại và tối ưu hóa cơ sở hạ tầng của mình để cải thiện năng suất và sự thành thạo. Bạn cũng có thể sử dụng dữ liệu mở rộng máy chủ để tối ưu hóa chi phí cơ sở hạ tầng tại chỗ hoặc trên nền tảng đám mây.

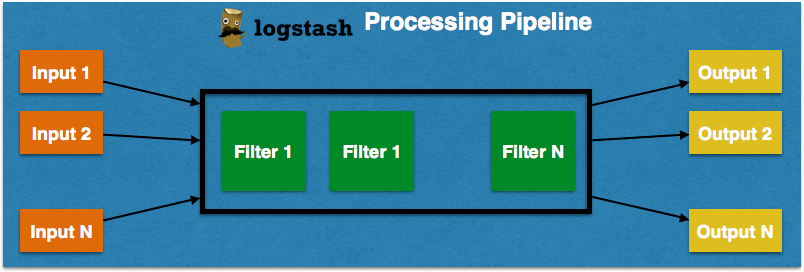
# CHƯƠNG 2. CÁC CÔNG CỤ PHÂN TÍCH NHẬT KÝ

## ****2.1.**** Logstas

### 2.1.1. Khái niệm và thành phần

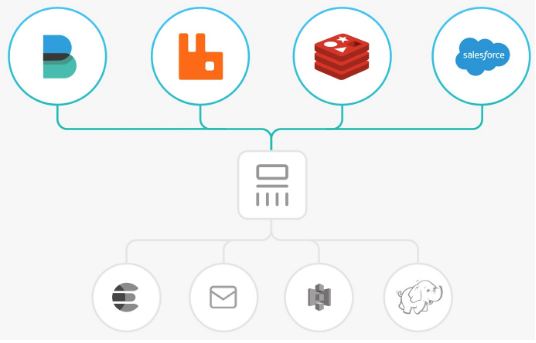
* Logstash là một công cụ mã nguồn mở thu thập dữ liệu có khả năng liên hợp theo thời gian thực. Logstash có thể hợp nhất dữ liệu từ các nguồn khác nhau và chuẩn hóa dữ liệu ở phần xử lý tiếp theo. Loại bỏ và đồng hóa tất cả dữ liệu đó trong một số use case cần phân tích và thể hiện trên biểu đồ.
* Logstash có 3 thành phần chính cũng chính là 3 bước xử lý chính của logstash đó là:
  + INPUT: Nó có thể lấy đầu vào từ TCP/UDP, các file, từ syslog, Microsoft Windows EventLogs, STDIN và từ nhiều nguồn khác. Chúng ta có thể lấy log từ các ứng dụng trên môi trường của chúng ta rồi đẩy chúng tới Logstash.
  + FILTER: Khi những log này tới Server Logstash, có một số lượng lớn các bộ lọc mà cho phép ta có thể chỉnh sửa và chuyển đổi những event này. Ta có thể lấy ra các thông tin mà ta cần từ những event log.
  + OUTPUT: Khi xuất dữ liệu ra, Logstash hỗ trợ rất nhiều các đích tới bao gồm TCP/UDP, email, các file, HTTP, Nagios và số lượng lớn các dịch vụ mạng. Ta có thể tích hợp Logstash với các công cụ tính toán số liệu (metric), các công cụ cảnh báo, các dạng biểu đồ, các công nghệ lưu trữ hay ta có thể xây dựng một công cụ trong môi trường làm việc của chúng ta.

### 2.1.2. Cách thức hoạt động



Đường ống xử lý sự kiện của Logstash có ba giai đoạn: input → filter → output. Các đầu vào tạo ra các sự kiện, bộ lọc sửa đổi chúng và các đầu ra sẽ chuyển chúng tới nơi khác. Đầu vào và đầu ra hỗ trợ codec cho phép bạn mã hóa hoặc giải mã dữ liệu khi nó vào hoặc thoát khỏi đường dẫn mà không phải sử dụng bộ lọc riêng biệt.

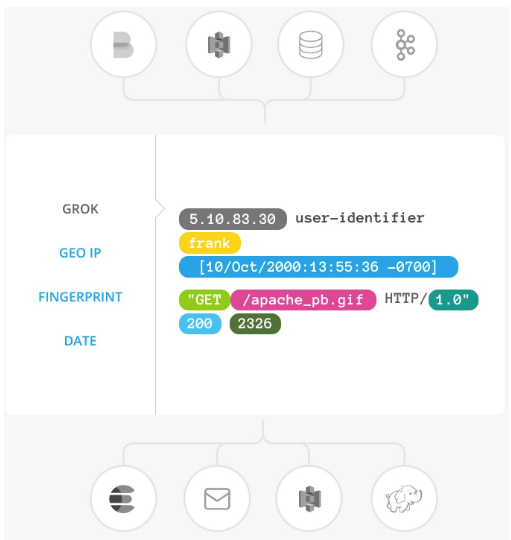
**INPUT**



Chúng ta sử dụng Input để lấy dữ liệu vào Logstash. Một số đầu vào thường được sử dụng là :

* File : đọc từ một tệp trên hệ thống, giống như lệnh UNIX tail -0F
* Syslog : nghe trên cổng 514 nổi tiếng cho các thông báo nhật ký hệ thống và phân tích cú pháp theo định dạng RFC3164.
* Redis : đọc từ máy chủ redis, sử dụng cả kênh redis và danh sách redis. Redis thường được sử dụng như một “broker” trong một mô hình Logstash tập trung, có hàng đợi các sự kiện Logstash từ các “shippers” từ xa.
* Beats : xử lý các sự kiện do beats gửi.

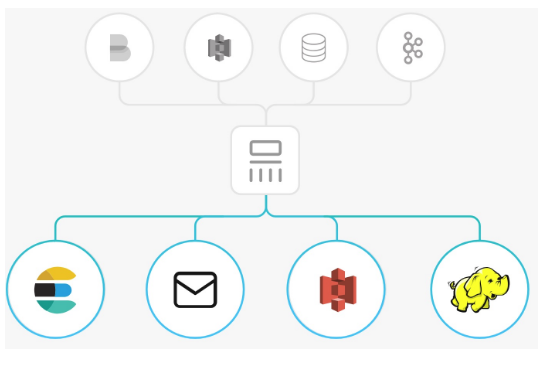
**FILTER**



Filter là thiết bị xử lý trung gian trong đường dẫn Logstash. Chúng ta có thể kết hợp các bộ lọc với các điều kiện để thực hiện một hành động trên một sự kiện nếu nó đáp ứng các tiêu chí nhất định. Một số bộ lọc hữu ích bao gồm :

* Grok : phân tích cú pháp và cấu trúc văn bản tùy ý - chỉnh sửa định dạng log từ client gửi về. Grok hiện là cách tốt nhất trong Logstash để phân tích cú pháp dữ liệu nhật ký không được cấu trúc thành một thứ có cấu trúc và có thể truy vấn được. Với 120 mẫu được tích hợp sẵn trong Logstash, nhiều khả năng chúng ta sẽ tìm thấy một mẫu đáp ứng nhu cầu của mình.
* Mutate : thực hiện các phép biến đổi chung trên các trường sự kiện. Bạn có thể đổi tên, xóa, thay thế và sửa đổi các trường trong sự kiện của mình.
* Drop : xóa hoàn toàn sự kiện, ví dụ: debug events.
* Clone : tạo bản sao của sự kiện, có thể thêm hoặc xóa các trường.
* Geoip : thêm thông tin về vị trí địa lý của địa chỉ IP (cũng hiển thị biểu đồ tuyệt vời trong Kibana).

**OUTPUTS**



Các đầu ra là pha cuối cùng của đường ống Logstash. Một sự kiện có thể đi qua nhiều đầu ra, nhưng một khi tất cả xử lý đầu ra đã hoàn tất, sự kiện đã hoàn tất việc thực thi của nó. Một số đầu ra thường được sử dụng bao gồm :

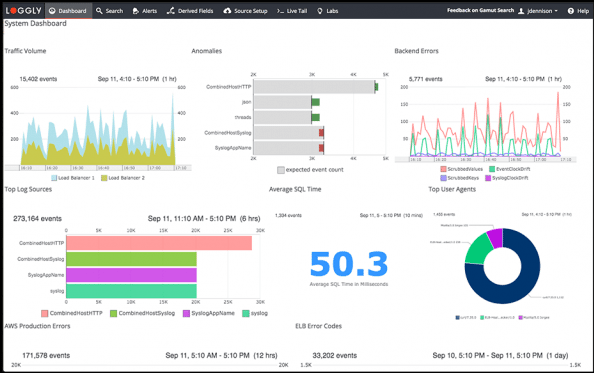
* Elasticsearch : gửi dữ liệu sự kiện tới Elasticsearch. Nếu chúng ta đang có kế hoạch để lưu dữ liệu trong một định dạng hiệu quả, thuận tiện, và dễ dàng truy vấn … Elasticsearch là con đường để đi.
* File : ghi dữ liệu sự kiện vào file trên bộ nhớ.
* Graphite : gửi dữ liệu sự kiện tới graphite, một công cụ nguồn mở phổ biến để lưu trữ và vẽ đồ thị số liệu.
* Statsd : gửi dữ liệu sự kiện đến statsd, một dịch vụ lắng nghe và thống kê.

## 2.2. Solarwind Loggly

Loggly là một ưu đãi SaaS cho phép bạn xem hiệu suất ứng dụng và phát hiện hoạt động bất thường và theo dõi hành vi tổng thể của hệ thống. Bạn có thể phân tích và giám sát các ứng dụng, cho dù trên cơ sở hạ tầng Amazon AWS hoặc Microsoft Azure, môi trường đám mây lai, thiết lập IoT hoặc sắp xếp microservice.

Với Loggly, bạn có thể dễ dàng xác định nguyên nhân gốc của các sự cố với sự trợ giúp của dữ liệu nhật ký đến từ các ngăn xếp và từ các dịch vụ được kết nối của bên thứ ba. Nó cũng bao gồm các công cụ DevOps và có thể tích hợp với Slack, HipChat, GitHub, Jira và PagerDuty.

Loggly có trực quan hóa tốt để giúp bạn theo dõi tuân thủ SLA, xem xu hướng hiệu suất và theo dõi và báo cáo về KPI. Phân tích nhật ký và thông tin chi tiết của bạn cũng có thể được chia sẻ dễ dàng bằng cách sử dụng các tích hợp công cụ được đề cập ở trên.



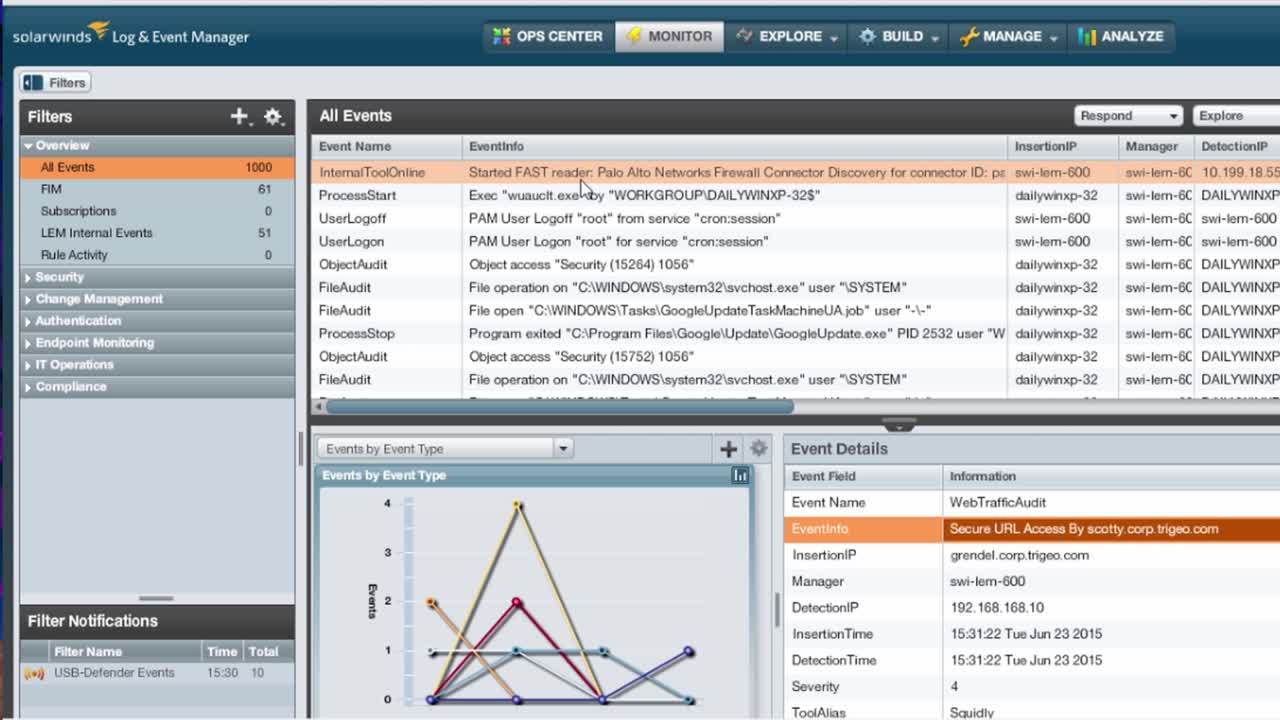
### 2.2.1. Solarwind Log Analyze

Trình phân tích nhật ký được thiết kế để cung cấp phân tích và bộ sưu tập cho syslog, bẫy và các sự kiện Windows và VMware.

Trình phân tích nhật ký cho phép bạn theo dõi thông tin theo thời gian thực về các sự cố phần cứng và phần mềm và nhật ký mạng. Với tìm kiếm dữ liệu linh hoạt, bạn cũng có thể lọc dữ liệu nhật ký được theo dõi. Công cụ này bao gồm các bộ lọc bên ngoài để giúp bạn tiết kiệm thời gian và được thiết kế để cung cấp phân tích nhật ký chuyên sâu thông qua trực quan hóa.

Trình phân tích nhật ký bao gồm trực quan hóa khối lượng nhật ký và kết quả tìm kiếm và biểu đồ tương tác hiển thị các khung thời gian trong đó các nhật ký đã được đối chiếu. Các trực quan hóa cung cấp thông tin được mã hóa màu và làm nổi bật nhật ký nào bạn cần kiểm tra chi tiết hơn, để bạn có thể đi đến gốc rễ của vấn đề và hiểu rõ hơn về hiệu suất cơ sở hạ tầng.

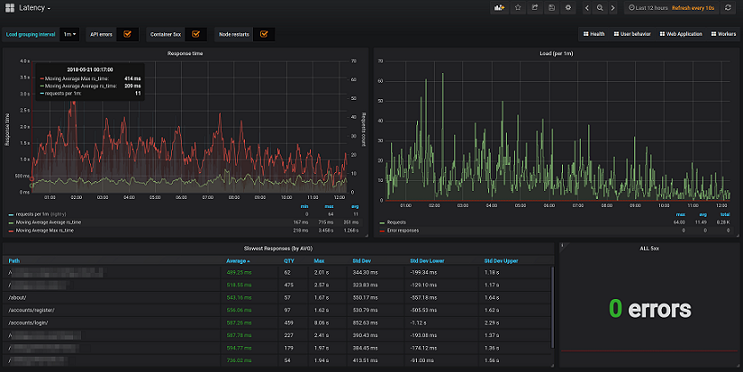
Một trong những tính năng đáng chú ý của công cụ này, khiến nó khác biệt với các tùy chọn SolarWinds khác, là nó tích hợp hoàn toàn với Nền tảng Orion® của SolarWinds. Điều này cho phép bạn kết hợp nó với các công cụ SolarWinds khác để tạo ra một giải pháp toàn diện với một bảng điều khiển duy nhất. Các cảnh báo được cung cấp thông qua tích hợp Orion cho bạn biết khi nào có vấn đề về bảo mật hoặc hiệu suất, do đó bạn có thể bắt đầu khắc phục sự cố.



## 2.3. GrayLog

Graylog là một nền tảng mạnh mẽ cho phép quản lý nhật ký dễ dàng cả dữ liệu có cấu trúc và không cấu trúc cùng với các ứng dụng gỡ lỗi. Nó dựa trên Elaticsearch, MongoDB và Scala. Graylog có một máy chủ chính, nhận dữ liệu từ các máy khách của nó được cài đặt trên các máy chủ khác nhau và giao diện web, trực quan hóa dữ liệu và cho phép làm việc với các bản ghi được tổng hợp bởi máy chủ chính.

Chúng tôi sử dụng Graylog chủ yếu làm bản ghi cho nhật ký của các ứng dụng web mà chúng tôi xây dựng. Tuy nhiên, nó cũng hiệu quả khi làm việc với các chuỗi thô (tức là syslog): công cụ phân tích nó thành dữ liệu có cấu trúc mà chúng ta cần. Nó cũng cho phép tìm kiếm tùy chỉnh nâng cao trong nhật ký bằng các truy vấn có cấu trúc. Nói cách khác, khi được tích hợp đúng cách với một ứng dụng web, Graylog giúp các kỹ sư phân tích hành vi hệ thống trên hầu hết các dòng mã.



Ưu điểm chính của Graylog là nó cung cấp một thể hiện hoàn hảo của bộ sưu tập nhật ký cho toàn hệ thống. Điều này có ích nếu cơ sở hạ tầng hệ thống lớn và phức tạp. Nó có thể được phân phối xung quanh nhiều nơi và không phải tất cả các thành viên trong nhóm có thể truy cập ngay vào tất cả các thành phần của nó

Graylog lưu trữ nhật ký có cấu trúc của các ứng dụng, do đó dễ dàng điều hướng trong đó và tìm thấy bất kỳ thông tin nào mà nhà phát triển có thể quan tâm, cho phép theo dõi trạng thái hệ thống trong mỗi thời điểm cụ thể

**Lợi thế của Graylog**

Graylog là một phần mềm mã nguồn mở miễn phí.

Các hành động hoặc thông báo kích hoạt của nó ngay lập tức thông báo cho chúng ta khi có điều gì đó cần chú ý, vì vậy chúng ta liên tục bám sát hiệu suất hệ thống.

Với dấu vết ngăn xếp lỗi nhận được từ Graylog, các kỹ sư hiểu ngữ cảnh của bất kỳ vấn đề nào trong mã nguồn. Điều này giúp tiết kiệm thời gian và nỗ lực để gỡ lỗi / khắc phục sự cố và sửa lỗi.

Công cụ này có cú pháp tìm kiếm mạnh mẽ, vì vậy thật dễ dàng để tìm thấy chính xác những gì bạn đang tìm kiếm, ngay cả khi bạn có terabyte dữ liệu nhật ký. Các truy vấn tìm kiếm có thể được lưu.

Graylog cung cấp chức năng lưu trữ, vì vậy mọi thứ cũ hơn 30 ngày có thể được lưu trữ trên bộ lưu trữ chậm và được nhập lại vào Graylog khi nhu cầu đó xuất hiện (ví dụ: khi nhóm nhà phát triển cần điều tra một sự kiện nào đó từ quá khứ).

Các ứng dụng Python có thể dễ dàng kết nối với Graylog vì có một thư viện sẵn có cho việc này.

## 2.4. Splunk

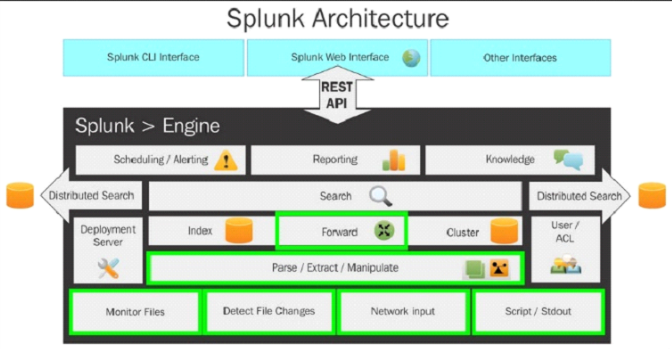
### 2.4.1. Khái niệm

Splunk là một phần mềm giám sát mạng dựa trên sức mạnh của việc phân tích Log. Splunk thực hiện các công việc tìm kiếm, giám sát và phân tích các dữ liệu lớn được sinh ra từ các ứng dụng, các hệ thống và các thiết bị hạ tầng mạng. Nó có thể thao tác tốt với nhiều loại dịnh dạng dữ liệu khác nhau (Syslog, csv, apache-log, access\_combined…). Splunk được xây dựng dựa trên nền tảng Lucene and MongoDB.

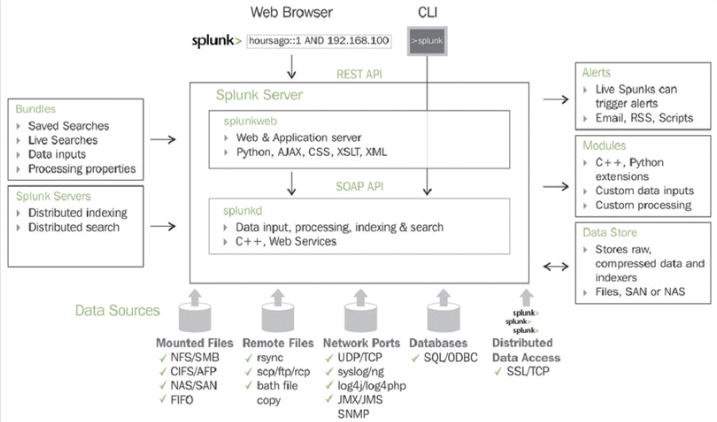
### 2.4.2 Tính năng

* **Định dạng Log:** Hỗ trợ hầu như tất cả các loại log của hệ thống, thiết bị hạ tầng mạng, phần mềm, Firewall, IDS/IPS, Log Event, Register của các máy trạm ….
* **Các hình thức thu thập dữ liệu:** Splunk có thể thực hiện việc thu thập log từ rất nhiều nguồn khác nhau. Từ một file hoặc thư mục (kể cả file nén) trên server, qua các kết nối UDP, TCP từ các Splunk Server khác trong mô hình Splunk phân tán, từ các Event Logs, Registry của Windows …Splunk kết hợp rất tốt với các công cụ thu thập log khác.
* **Cập nhật dữ liệu:** Splunk cập nhật dữ liệu liên tục khi có thay đổi trong thời gian thực. Giúp cho việc phát hiện và cảnh báo trong thời gian thực.
* **Đánh chỉ mục dữ liệu:** Splunk có thể đánh chỉ mục dữ liệu với một khối lượng dữ liệu rất lớn trong một khoảng thời gian ngắn. Giúp việc tìm kiếm diễn ra nhanh chóng và thuận tiện.
* **Tìm kiếm thông tin:** Splunk làm việc rất tốt với dữ liệu lớn và cập nhật liên tục. Nó cung cấp cơ chế tìm kiếm với một “Splunk Language” cực kỳ thông minh bao gồm các từ khóa, các hàm và cấu trúc tìm kiếm giúp người sử dụng có thể truy xuất mọi thứ, theo rất nhiều tiêu chí từ tập dữ liệu rất lớn. Những nhà quản trị mạng cao cấp và chuyên nghiệp thường gọi Splunk với cái tên “Splunk toàn năng” hay “Splunk as Google for Log files” để nói lên sức mạnh của Splunk.
* **Giám sát và cảnh báo:** Splunk cung cấp cho người dùng một cơ chế cảnh báo dựa trên việc tìm kiếm các thông tin do chính người sử dụng đặt ra. Khi có vấn đề liên quan tới hệ thống phù hợp với các tiêu chí mà người dùng đã đặt ra thì hệ thống sẽ cảnh báo ngay tới người dùng (cảnh bảo trực tiếp qua giao diện, giử Email).
* Khắc phục sự cố: Splunk còn cung câp một cơ chế tự động khắc phục với các vấn đề xảy ra bằng việc tự động chạy các file Script mà người dùng tự tạo (Ví dụ như: Chặn IP, đòng Port …) khi có các cảnh báo xảy ra.
* **Hiển thị thông tin:** Splunk cung cấp một cơ chế hiển thị rất trực quan giúp người sử dụng có thể dễ dàng hình dung về tình trạng của hệ thống, đưa ra các đánh giá về hệ thống. Splunk còn từ động kết xuất ra các báo cáo với nhiều loại định dạng một cách rất chuyên nghiệp.
* **Phát triển:** Cung cấp các API hỗ trợ việc tạo các ứng dụng trên Splunk của người dùng. Một số bộ API điển hình như Splunk SDK (cung cấp các SDK trên nền tảng Python, Java, JS, PHP), Shep (Splunk Hadoop Intergration - đây là sự kết hợp giữa Splunk và Hadoop), Shuttl (là một sản phẩm hỗ trợ việc sao lưu dữ liệu trong Splunk), Splunkgit (Giúp bạn hình dung dữ liệu tốt hơn), Splunk power shell resource Kit (Bộ công cụ hỗ trợ việc mở rộng và quản lý hệ thống).

### 2.4.3 Splunk Architecture



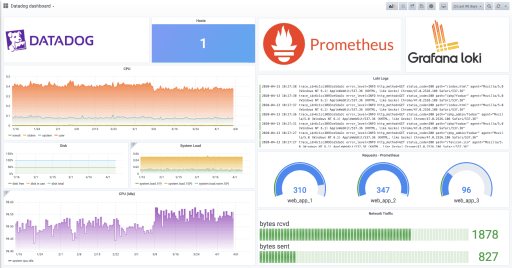
* Mức thấp nhất của kiến trúc Splunk mô tả các phương thức nhập liệu khác nhau được hỗ trợ bởi Splunk. Những phương thức nhập này có thể được cấu hình để gửi dữ liệu trên các bộ phân loại Splunk.
* Trước khi dữ liệu đến được các bộ phân loại Splunk, nó có thể được phân tích cú pháp hoặc thao tác, có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



* Splunkd cho phép người dùng tìm kiếm, điều hướng và quản lý dữ liệu trên Splunk Enterprise thông qua giao diện web được gọi là Splunk Web. Nó là một máy chủ ứng dụng web dựa trên Python cung cấp một giao diện web để sử dụng Splunk. Trong phiên bản trước của Splunk: splunkd và SplunkTeb là hai quy trình riêng biệt, nhưng từ Splunk 6, cả hai quy trình đã được tích hợp là 1. Nó cho phép người dùng tìm kiếm, phân tích và hình dung dữ liệu bằng cách sử dụng giao diện web. Giao diện Splunk Web có thể được truy cập bằng cách sử dụng cổng Web Splunk, và Splunk cũng cho thấy REST API để truyền thông thông qua cổng quản lý chia sẻ.
* Một trong những thành phần quan trọng của kiến ​​trúc của Splunk là kho dữ liệu. Nó có trách nhiệm nén và lưu trữ dữ liệu ban đầu (nguyên vẹn). Dữ liệu được lưu trữ trong các tệp Time Series Index (T SIDX). Một kho dữ liệu cũng bao gồm lưu trữ và lưu trữ dựa trên chính sách lưu giữ có thể cấu hình.
* Các triển khai của Splunk Enterprise có thể bao gồm từ việc triển khai các máy chủ đơn (có chỉ số vài gigabyte dữ liệu mỗi ngày và được truy cập bởi một vài người dùng đang tìm kiếm, phân tích và hình dung dữ liệu) tới các triển khai lớn của doanh nghiệp ở nhiều trung tâm dữ liệu, lập chỉ mục hàng trăm terabytes dữ liệu và tìm kiếm được thực hiện bởi hàng trăm người dùng. Splunk hỗ trợ truyền thông với một cá thể khác của một máy chủ Splunk thông qua TCP để chuyển tiếp dữ liệu từ một máy chủ Splunk sang một máy khác để lưu trữ dữ liệu và các yêu cầu phân phối và phân phối dữ liệu khác thông qua giao tiếp TCP Splunk-to-Splunk.
* Bundles là các thành phần của kiến trúc Splunk lưu trữ cấu hình dữ liệu đầu vào, tài khoản người dùng, ứng dụng Splunk, tiện ích và môi trường khác.
* Các mô-đun là những thành phần của kiến trúc Splunk được sử dụng để thêm các tính năng mới bằng cách sửa đổi hoặc tạo bộ xử lý và đường ống. Các mô-đun chỉ là các kịch bản tùy chỉnh và các phương pháp nhập dữ liệu hoặc phần mở rộng có thể thêm một tính năng mới hoặc sửa đổi các tính năng hiện có của Splunk.

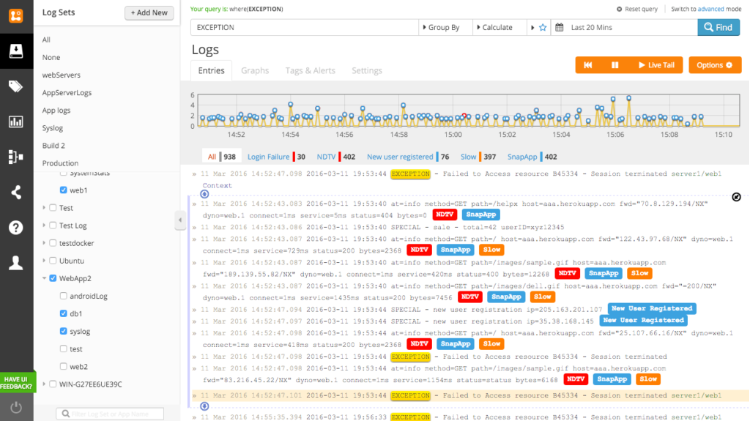
## 2.5. Datadog

Datadog là một dịch vụ giám sát, tập hợp số liệu và sự kiện từ các máy chủ, cơ sở dữ liệu, các ứng dụng, các công cụ và dịch vụ để trình bày một quan điểm thống nhất của các cơ sở hạ tầng. Những khả năng này được cung cấp trên một nền tảng phân tích dữ liệu SaaS dựa trên phép Dev và Ops đội làm việc cộng tác trên cơ sở hạ tầng để tránh thời gian chết, giải quyết vấn đề hiệu suất và đảm bảo rằng các chu kỳ phát triển và triển khai hoàn thành đúng thời hạn.



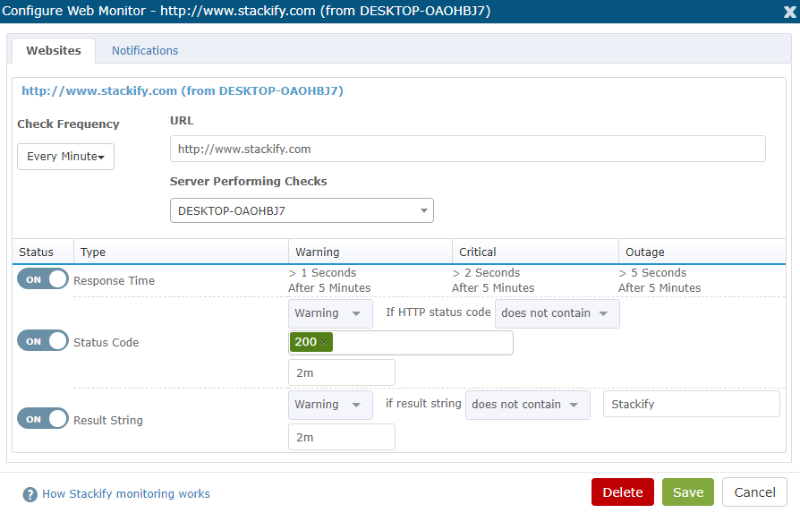
## 2.6. Logentries

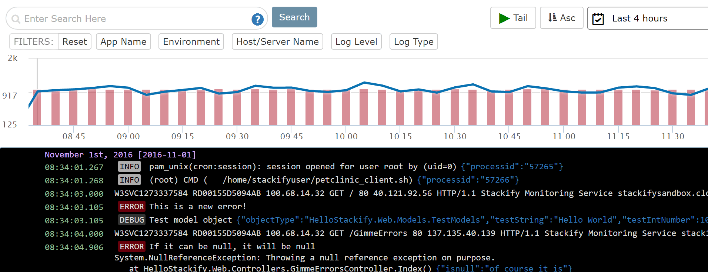
Logentries là một công cụ giám sát nhật ký cơ bản cho phép theo dõi và tìm kiếm nhật ký theo thời gian thực. Nó cũng bao gồm các công cụ phân tích và trực quan dễ sử dụng và dễ nhìn, do đó bạn có thể theo dõi các xu hướng dài hạn hơn và xem các sự kiện có tương quan như thế nào trên các hệ thống của bạn. Và với cảnh báo theo thời gian thực, mọi vấn đề sẽ được gắn cờ, vì vậy bạn có thể giải quyết chúng trước khi bạn kết thúc với vi phạm bảo mật hoặc trải nghiệm người dùng cuối bị gián đoạn.



## 2.7. Stackify

Stackify chủ yếu là một công cụ phân tích dành cho các nhà phát triển và do đó, nó là một lựa chọn tốt để xem xét nếu bạn có một nhóm phát triển cần phần mềm phân tích nhật ký. Nó kết hợp các công cụ phân tích nhật ký, quản lý hiệu suất ứng dụng, lỗi và nhật ký tích hợp và cấu hình mã trong một gói. Điều này làm cho nó trở thành một lựa chọn linh hoạt để phân tích nhật ký như là một phần của quá trình phát triển lớn hơn.



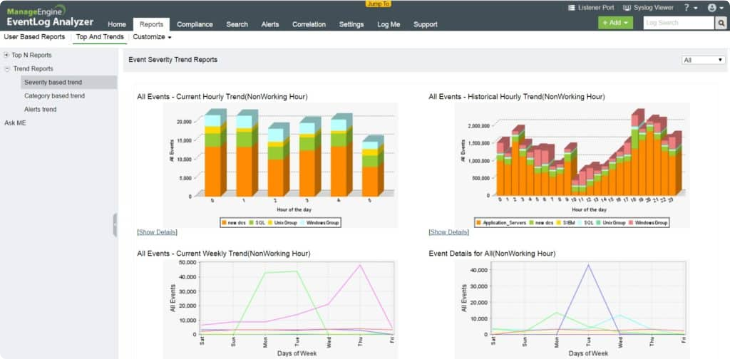


## 2.8. ****ManageEngine EventLog Analyzer****

ManageEngine EventLog Analyzer là một công cụ phân tích nhật ký với trải nghiệm người dùng hợp lý. ManageEngine EventLog Analyzer thu thập nhật ký từ các nền tảng cơ sở dữ liệu, máy chủ web, bộ định tuyến, thiết bị chuyển mạch, bộ ảo hóa, máy quét lỗ hổng, hệ thống Linux, hệ thống Unix, tường lửa và Giải pháp bảo mật điểm cuối.

Để giúp bạn điều hướng dữ liệu nhật ký, ManageEngine EventLog Analyzer sử dụng hệ thống cảnh báo. Cảnh báo có thể tùy chỉnh và cảnh báo bạn trong thời gian thực qua email hoặc SMS nếu chương trình phát hiện điều gì đó cần sự chú ý của bạn. Cảnh báo được phân loại là mức độ ưu tiên cao, trung bình hoặc thấp để giúp bạn phản hồi phù hợp với thông báo.

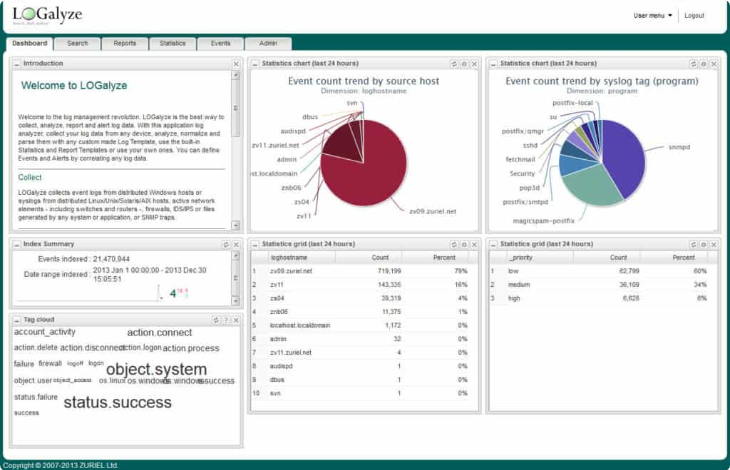
Phần mềm này tuân thủ quy định đối với một số chính sách bao gồm HIPAA, PCI DSS, ISO 27001, GLBA, SOX, FISMA, v.v. Báo cáo tuân thủ giúp đảm bảo rằng bạn có tất cả các tài liệu cần thiết để giữ cho tổ chức của bạn không bị băng đỏ. Ví dụ, tuân thủ HIPAA báo cáo việc xử lý các đối tượng, đăng nhập / đăng xuất thành công và nhật ký hệ thống để đảm bảo có một bản ghi rõ ràng về hoạt động của người dùng.



## 2.9. Logalyze

Logalyze là một bộ phân tích nhật ký mã nguồn mở và giám sát mạng cho người dùng doanh nghiệp. Sản phẩm hỗ trợ các thiết bị, máy chủ windows và máy chủ Linux / Unix với khả năng phát hiện sự kiện theo thời gian thực. Khi dữ liệu nhật ký đã được thu thập, bạn có thể sử dụng tính năng tìm kiếm chương trình để tìm thông tin bạn cần.

Người dùng cũng có thể xác định cảnh báo của riêng họ. Khi một cảnh báo đã được đưa ra thì vé có thể được tạo để ghi lại vấn đề cho đến khi nó được giải quyết. Ngoài ra còn có thêm tài liệu có sẵn dưới dạng các báo cáo được lên lịch mà bạn có thể sử dụng để xem các cập nhật thường xuyên về trạng thái của mạng. Các báo cáo tuân thủ PCI-DSS, SOX, v.v.

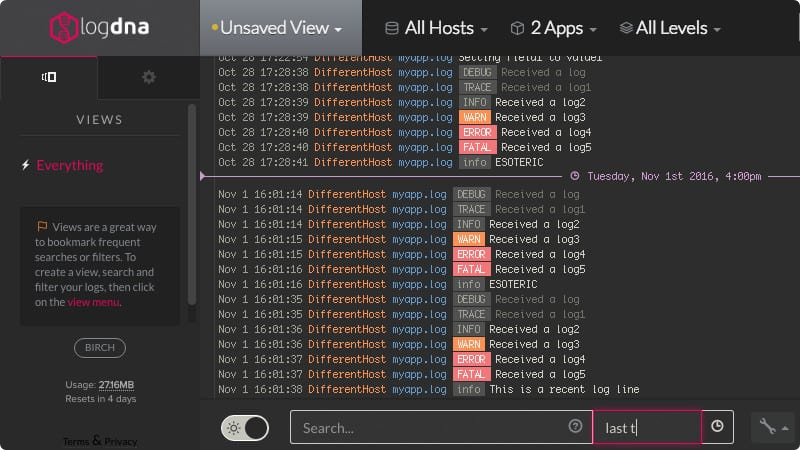


## 2.10. LogDNA

LogDNA là một nền tảng phần mềm quản lý nhật ký có thể theo dõi dữ liệu nhật ký theo thời gian thực. Công cụ này dựa trên đám mây và được cấu hình trong vòng chưa đầy hai phút để thu thập nhật ký từ AWS, Heroku, Elastic, Docker và các nhà cung cấp khác. Công cụ này ngay lập tức tổng hợp nhật ký từ các ứng dụng và máy chủ trên mạng của bạn với băng thông để xử lý một triệu sự kiện nhật ký mỗi giây.

Một trong những điều thú vị về LogDNA là tác nhân LogDNA và giao diện CLI là nguồn mở. Trong thực tế, điều này cho phép bạn tùy chỉnh kinh nghiệm quản lý nhật ký của bạn. Tuy nhiên, nếu bạn không muốn làm điều đó, giao diện người dùng chuẩn có quá nhiều tính năng để giúp bạn theo dõi nhật ký hệ thống một cách hiệu quả.

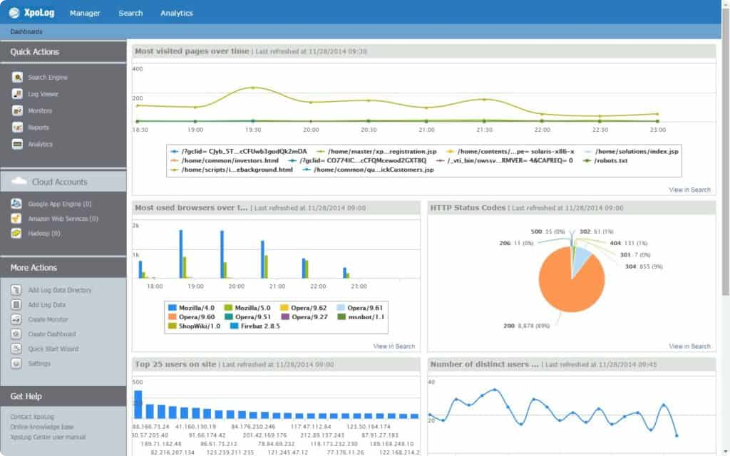
LogDNA là phải có cho các tổ chức cần một giải pháp quản lý nhật ký dựa trên đám mây và có thể mở rộng. LogDNA có sẵn dưới dạng giải pháp dựa trên đám mây hoặc gói tại chỗ / tự lưu trữ. Có bốn tùy chọn giá có sẵn: Miễn phí, Birch, Maple và Oak. Phiên bản miễn phí hỗ trợ một người dùng.



## 2.11. XpoLog

XpoLog là một nền tảng giám sát nhật ký có thể thu thập và phân tích nhật ký từ các thiết bị trên mạng. XpoLog theo dõi nhật ký trong thời gian thực để khám phá các vấn đề về hiệu suất và tạo cảnh báo. Người dùng có thể xác định các quy tắc để cảnh báo và thực hiện các quy tắc lọc của riêng họ.

Một trong những tính năng giúp XpoLog nổi bật là phát hiện lỗi do AI cung cấp. AI có thể phát hiện ra lỗi, rủi ro bảo mật và phân biệt các mẫu nhật ký cho thấy hiệu suất kém. Phát hiện lỗi phục vụ để tự động hóa quản lý nhật ký và đảm bảo bạn không bỏ lỡ bất kỳ hoạt động có vấn đề nào. Tuy nhiên, nếu bạn muốn xem xét kỹ hơn, bạn có thể sử dụng tính năng tìm kiếm nhật ký tự động để xem trí thông minh của máy khi bạn chạy tìm kiếm thủ công.

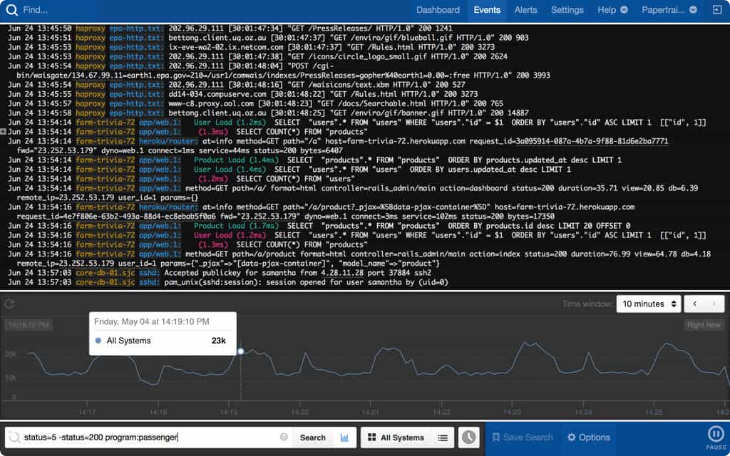


## 2.12. Papertrail

Papertrail là một công cụ phân tích nhật ký cho Windows, tự động quét qua dữ liệu nhật ký của bạn. Khi quét dữ liệu nhật ký, bạn có thể chọn thông tin bạn muốn kết quả quét hiển thị. Chẳng hạn, bạn có thể chọn quét có chứa địa chỉ IP, địa chỉ email, GUID / UUID, URL (s) URL, tên miền, máy chủ lưu trữ, tên tệp và văn bản được trích dẫn hay không.

Một trọng tâm của Papertrail là giải quyết sự kiện. Để giúp bạn tìm ra nguyên nhân của các sự kiện bảo mật mạng nhanh hơn, bạn có thể lọc các sự kiện nhật ký theo thời gian, nguồn gốc hoặc trường tùy chỉnh mà bạn chọn. Lọc nhật ký theo cách này cho phép bạn loại bỏ dữ liệu không liên quan và tập trung vào dữ liệu quan trọng nhất.

Một tùy chọn lọc tương tự khác được cung cấp bởi Papertrail cho phép bạn phát hiện các xu hướng trong dữ liệu nhật ký. Bạn có thể lọc các sự kiện theo nguồn, dữ liệu, mức độ nghiêm trọng, cơ sở hoặc nội dung tin nhắn. Khi tìm kiếm được lọc hoàn tất, bạn sẽ có thể xem biểu đồ kết quả ở cuối màn hình.

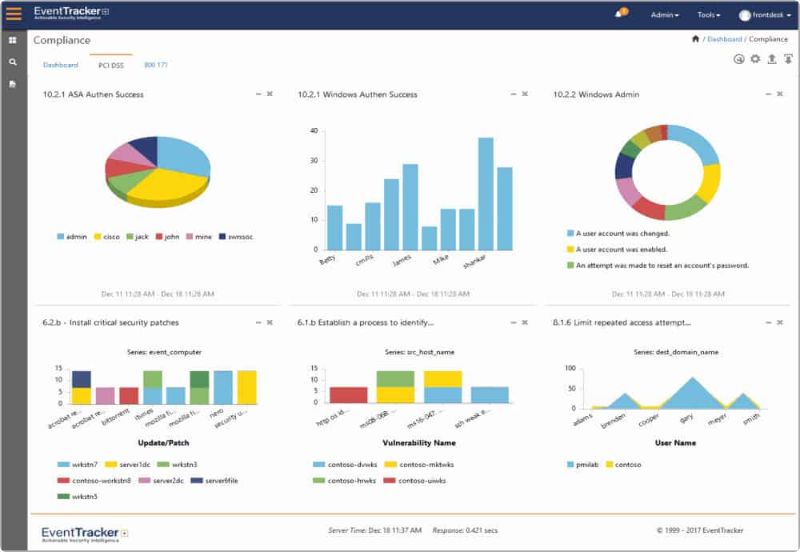


## 2.13. EvenTracker

Tiến trình (trước đây là IpSwitch), nhóm đằng sau công cụ giám sát mạng phổ biến WhatsUp Gold cũng có một giải pháp quản lý nhật ký có tên là EventTracker. EventTracker có thể thu thập và phân tích dữ liệu nhật ký trên các tệp nhật ký Sự kiện Windows, Syslog và W3C / IIS. Chương trình có thể phát hiện các sự kiện bảo mật trong thời gian thực. Các khả năng phân tích nhật ký thời gian thực của EventTracker làm cho nó trở thành một lựa chọn tốt để phát hiện và phản hồi sự kiện.

Cảnh báo sự kiện trong thời gian thực cung cấp thêm một lớp khả năng hiển thị vào hoạt động nhật ký. Có hàng trăm cảnh báo khác nhau được cung cấp ngoài hộp với EventTracker (mặc dù bạn cũng có thể tự tạo thông báo). Cảnh báo bao gồm phân tích pháp y để bạn có thêm dữ liệu cần rút ra khi khắc phục sự cố cho giải pháp cho sự kiện bảo mật.

Để giữ cho nhóm của bạn trong vòng lặp phát triển nhật ký sự kiện, EventTracker tự động phân phối báo cáo cho nhân viên chính, người quản lý và các bên liên quan. Các báo cáo được tuân thủ với HIPAA, Sarbanes, OXLEY, PCI DSS, NISPOM, MiFID và FISMA. Có hơn 1.500 báo cáo khác nhau để bạn lựa chọn. Sử dụng thông tin từ các báo cáo này giúp xác định xem có bất kỳ lỗ hổng nào trong mạng của bạn cần được xử lý hay không.



# TÀI LIỆU THAM KHẢO

[1] <https://viblo.asia/p/phat-hien-xam-nhap-voi-splunk-OeVKB8MdlkW?fbclid=IwAR0zTgcsVrFcsYUjjcCSjV7P3I9fcPepPlC3LfancPCuHKeFHDw7E3STYwg>

[2] <https://sematext.com/blog/log-analysis/?fbclid=IwAR0mojfZd-Mwv4fgn4Ce8R5jkam5SKRqJAxPysrLdFSsf1iHRxKfUZExnps>

[3] <https://sematext.com/blog/log-analysis/?fbclid=IwAR0mojfZd-Mwv4fgn4Ce8R5jkam5SKRqJAxPysrLdFSsf1iHRxKfUZExnps>

[4] <https://www.keycdn.com/blog/log-analysis-tools?fbclid=IwAR0o8E-FikwieJ6SO71jAxCVonf79gsDO3hsyPQd13fNiD9ABHgrssfCzoQ>

[5] <https://www.dnsstuff.com/log-analysis-tools?fbclid=IwAR19denahnr_Fddn1W14EzTcemcmIJRjHb01oSOBhKrIgpVsiE2FZTaETv0#Graylog>

[6] <https://dzone.com/articles/how-to-use-graylog-for-technical-monitoring-in-sof>