Capstone Report

Introduction:

Logs play a key role in management of IT infrastructure and their analysis is crucial when problems arise and as a part of the resolution process. This process usually involves collection of different audit trail records on a centralized log server and forwarding them to a log analyzer for analysis purposes. A common logging conundrum is which components should be logged and how to prioritize the amount of log data to be collected. Too much data creates noise and makes it challenging to identify the needle in the haystack. So, availability of right data to right people at the right time becomes essential in ensuring the security of an organization.

Technologies used:

For this whitepaper on network monitoring tools, we are using AWS to create multiple instances which can act like GNS3 server, Vector, and other test instances. Graphical Network Simulator-3 (GNS3) is used to simulate the functionality of Cisco IOS devices. The events generated in these devices will be used for our analysis. We will be using Vector a lightweight tool as our Splunk forwarder to forward all the events to other AWS instances each with log analyzer tools installed. We will be using Splunk, Datadog and SolarWinds as network monitoring tools. In this whitepaper, we will be analyzing only system logs (syslogs).

GNS3:

The complexity level for installation and configuration of GNS3 is moderate. GNS3 consists of two components: GNS3 GUI aka client and GNS3 VM aka server. To simulate the functionalities of network devices we must install GNS3 server along with client. The GNS3 server can be installed in an AWS EC2 instance along with OpenVPN. Two simple command line scripts can grab and install both GNS3 server and OpenVPN on our EC2 instance. After OpenVPN installation is complete, we can now download the OpenVPN client configuration file. Then, we must download and install OpenVPN and GNS3 clients for our workstation and connect the VPN via the configuration file. The tunnel IP address is to be noted and used to connect to the GNS3 server via the client. Some of the issues I faced with GNS3 are:

1. The GNS3 client and server needs to be of the same version and does not work otherwise due to compatibility issues.
2. I installed the GNS3 server on my workstation prior to installing it on the EC2 instance. When I tried to connect to the server on EC2 instance after using it with the server on local machine it failed to connect. Only when I reinstalled the GNS3 client, it connected to the remote GNS3 server.

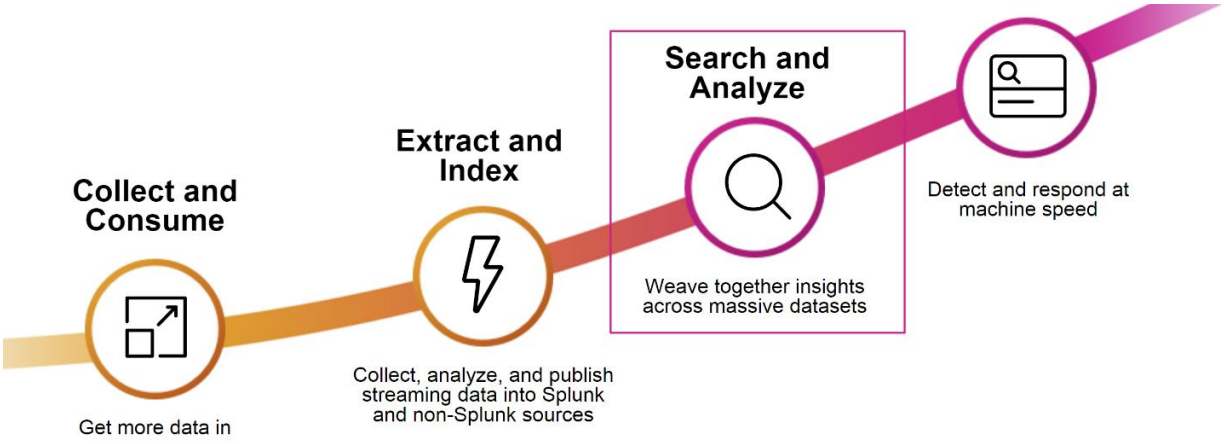
Vector:

The complexity level for installation and configuration of vector is moderate, although its well documented. Vector can be installed via a command line script which performs both platform detection and installation. All the configurations exist in the vector.toml file. Vector acts like a forwarder and forwards all the events to other AWS instances when the elastic IP addresses of the instances are added to the vector.toml file. All the IP addresses associated with the instances are dynamic in nature. Hence, we need a static IP (Elastic IP) to be added to the configuration file. Elastic IP can be easily obtained through AWS. The issue I faced with the vector is it randomly dies after a few days of inactivity. Also, it does not start properly when start command is issued and fail some health checks. However, it starts working after a few attempts.

Product 1: Splunk

Why Splunk?

Splunk is arguably one of the most widely used integrated solutions for big data analysis. Splunk provides continuous near real-time monitoring and can be used to correlate, identify trends in the events or even compare it against current activity while providing valuable insights. With Splunk we can turn raw data into digestible and actionable data. Also, we can search any data source to visualize relationships among data to spot any outliers and create alerts when new issues or threats are discovered.



Components:

One of the most useful functionalities of Splunk is the ability to search huge amounts of data. In Spunk's search and reporting app, the user can pass the queries by using Splunk's Search Processing Language (SPL) and look at the desired events as well as visualize them. The basic components of using search are:

Search assistant: The search assistant helps by giving out suggestions on how to complete the string.

Time range: Time range is one of the important components as the results can be tuned to the user's preferences. The time range can be selected from a wide range of presets or can be customized which will influence the search speed and efficiency.

Saved jobs: All search results can be saved and can be accessed later. These search jobs can also be shared and allow multiple users to work on them at the same time.

Fields: Splunk automatically discovers several of these interesting key value pairs called fields which help users to get valuables insights on the data.

Installation:

The installation of Splunk in an Amazon EC2 instance is extremely easy. First, we must create a Splunk account which is straightforward. A free Splunk account would suffice for our case study as it provides 500 MB of Daily Indexing Volume. A simple 'wget' command then grabs and installs Splunk on the target machine. A quick tip is to pass the '--accept-license' as argument to the start command when running Splunk for the first time can avoid us going through the entire license. Then we are prompted to create the Splunk admin account. Finally, the Splunk instance starts running and we can access it from our local machine.

Configuration:

The complexity level for configuration of Splunk is moderate. It is best practice NOT to run Splunk as super user for *nix OR as administrator for Windows. Splunk should also be able to access data sources like /var/log. Also, we must ensure time synchronization between Splunk indexer and the local instance to ensure correct timestamping. Finally, we must add the rule to open port 8000 in the AWS security group, which allows us to access the Splunk Web UI.

Usage:

For monitoring and analysis of data, it is required to have a basic understanding of Splunk Processing Language (SPL). The data summary section can tell us about all the events Splunk has indexed. Although it can be overwhelming to look at all the data upfront, with the robust search functionality we can get rich insightful data. Again, the time range functionality is important and dictates the speed and efficiency of the searches. There are three modes of search namely smart, fast, and verbose mode which have visible and varying impacts on large datasets. These search results can either be saved in form of reports, can be used to create alerts which trigger other actions or create visualizations. Reports and visualizations can really be useful when dealing with large datasets. There is a plethora of information which can be inferred from these logs. Each event consists of following information: Host information like source IP and Port number, source type, message which gives information about the event like interface down or login failed and timestamps.

Issues:

User needs to always be mindful of the storage space left in the partition. Can miss crucial events if overlooked. The dashboard warning message could be as follows:

“The index processor has paused data flow. Current free disk space on partition '/' has fallen to 2670MB, below the minimum of 5000MB. Data writes to index path '/opt/splunk/var/lib/splunk/audit/db' cannot safely proceed. Increase free disk space on partition '/' by removing or relocating data.”

The instance storage can be increased from default to resolve this issue.

Product 2: Datadog

Datadog is a purely SaaS alternative to other network monitoring tools. It is one of the fastest growing security monitoring tools on the market today.

Installation

The installation of Datadog is extremely easy and neatly documented. Datadog also requires creation of a user account. A free license can be obtained through GitHub's student developer program. A free trial of 1 week is available to others. A Datadog agent is available for almost all the operating systems. Installation of Datadog on Amazon Linux is done through a simple script.

Startup

The startup of Datadog is extremely easy. There is no startup required for Datadog. Since, it is SaaS based it is always up and running.

Configuration / Data Connection

The configuration of Datadog is extremely easy. The API key is easy to locate and should be added to log forwarder. There are over 250 vendor supported integrations that Datadog supports out-of-box.

Validation and Usage

It is easy to use Datadog. The responsive sidebar menu can provide us with all the functionalities we need like events, dashboards, integrations etc. The event stream can help us look at most recent events as well as provide us with valuable insights.

Issues

I did not face any issues with Datadog.

Conclusion:

After the testing of our two tools, it was made clear that while Splunk is the preferred tool for a more technical person, if it's possible to use a SaaS solution, Datadog has a more streamlined setup and integration. If an information system can support a SaaS solution (not a closed environment) the easier solution would be Datadog.

However, if your environment would like more defined, advanced searches, the dashboard building function of Splunk provides the ability to build specific search functions and display the data in a very easy to decipher format, if you have somebody with the skills to establish the dashboards.