Acta Scientific COMPUTER SCIENCES

Volume 4 Issue 10 October 2022

Research Article

Elastic Stack: A Reliable Tool for Highly-Demanding Data Insight

Mohammed Daffalla Elradi*

Communication Systems Engineering Department, University of Science and Technology, Khartoum, Sudan

***Corresponding Author:** Mohammed Daffalla Elradi, Communication Systems Engineering Department, University of Science and Technology, Khartoum, Sudan.

Received: July 12, 2022

Published: September 02, 2022

Abstract

Cyber threats and attacks are evolving and been conducted more frequently and in various manners. Security analysts are challenged to detect, respond to, remediate and prevent those attacks as they impose immense risks. Logs are crucial in providing the tendency to have information about what happened, what is happening, and even predicting what will happen. Hence, having a tool that can centrally manage these logs, process them and even visualize them can be of great importance. There are many Data Analytics tools among which Elasticsearch Stack is standing out and getting more popular, formed from three main tools named Elasticsearch used for data storage and indexing, Logstash act as a data pipeline with input filter and output. In addition to Kibana that is used for data visualization and creating dashboards.

In this paper the Elasticsearch Stack was used to provide concise yet detailed dashboards to identify system metrics, network performance measurements as well as user management events and security events that give security analysts a thorough insight about security events occurring to help them promptly investigate on demand. It was found to be a reliable tool in such a highly-demanding environment.

Keywords: Elasticsearch Stack; Elasticsearch; Logstash; Kibana; Elastic Agent

Introduction

Data Analytics is the science that deals with the process of extracting information from raw data to derive insights. It encompasses many stages ranging from data set establishment, preparation for data processing to yield better visibility, applying models, identifying key findings and preparing reports accordingly [1].

Data is turning out to be a crucial prerequisite for any business, ranging from indicating the best performance requirements, deliver services and products that meets customers' expectations by identifying their habits and preferences, as well as targeting prospective customers by considering trends. Data analytics also plays a pivotal role when it comes to security and evaluating risks associated with the business itself [2].

Corporates need to have a transparent view about their business, which greatly influences the act of decision making. To achieve this clear vision, corporates should be well prepared for this by providing the appropriate structures, policies, skilled individuals and the required technology for the purpose of acquisition, identifying and utilizing the data available in order to derive results along with data visualization to help making decisions and implement observations accordingly.

In the recent years, cyber-attacks have been conducted frequently targeting wide sectors. Having the tendency to defend against these attacks is not an option anymore but a necessity [3]. Having a data analytic platform that gives a thorough view about security incidents will greatly impact the act of responding in a timely manner or even avoid been targeted by a cyber-attack or any malicious activities.

In the domain of IT and Information security, monitoring is critical as it represents the first line of defence for security teams especially in highly dynamic and continuously growing environments to maintain high availability, avoid being compromised as well as keeping up with regulatory standards. Hence, logs play a master role in the foundations of security as a whole. As events occur logs are shaped to give an insight of what happened, what is happening and even predicting what will happen [4].

In fact, there are many tools that are used to store, maintain, process and manage logs centrally and seamlessly. Among these tools stands out Elastic Stack formerly known as ELK Stack, which stands for Elasticsearch, Logstash and Kibana. These tools collectively form a powerful analytical tool that spans for diverse use cases ranging from maintaining, searching and analysing data on-time. In addition to a wide range of modules and integrations with many data sources in multiple forms to be viewed in rich visualizations to highlight every single aspect of your demanding data-driven analysis [4]. ELK stack is rich of features.

Elasticsearch

Elasticsearch is considered as the heart of the Elastic Stack. It has started to steadily get popular over the last several years as it is an open-source distributed search engine that can search and index data in miscellaneous formats either structured or unstructured and stores them as JSON documents, which are grouped together with indices. It also involves robust aggregation capabilities that makes it easy to identify patterns and sort out trends in a sophisticated manner [5].

Figure 1: Elasticsearch Indices.

Elasticsearch comprises too many features, the most prominent features are highlighted below.

Scalability and resiliency

Elasticsearch as the name indicates grows and extends as per your needs. It adapts with extension in data by creating additional nodes along with replica shards automatically.

High availability

Replication is a key feature for clusters to guarantee availability by implementing replica shards in case a node goes down.

Events correlation

Elasticsearch offers powerful data correlation capabilities that makes it easy to figure out anomalies.

Elasticsearch is an optimum option for various use cases that involve but not limited to log analysis, websites and application search, business analysis and security monitoring.

Logstash

Logstash is a server-side data processing pipeline that centrally collect logs from multiple sources, processes them and ingests them into Elasticsearch. It performs some main tasks as parsing, extracting, managing and structuring the data and derives structures from unstructured data.

It uses various plugins for inputs, data filters and outputs to send events and logs data to specific destination.

Figure 2: Logstash working mechanism.

Beats

Beats are agents installed on the server or end device intended to send events and logs data to the Elasticsearch, they are considered as lightweight data shippers used for a single purpose such as Filebeats for dealing with log files, Metricbeat for collecting system-related metrics such as system performance, CPU, memory

and disk usage. Similarly acts the Heartbeat, which indicates what services are accessible [6]. Also, Packetbeat for collecting network performance data, Auditbeat is used to collect audit events especially changes to files and configuration binaries. For monitoring Windows events logs, Winlogbeat is used.

As been depicted in figure 3, data from Beats are sent to the pipeline of Logstash for the purpose of being parsed and transformed or directly into Elasticsearch.

Figure 3: Ingesting Beats to Elasticsearch via Logstash.

Elastic agent

An agent that is used for the collection of metrics, logs, security events as well as preventing threats is referred to as elastic agent, which can be used to provide most of the functionalities of Beats collectively [11]. Elastic agents in different devices can be centrally managed using something called fleet server or can be run in standalone mode. Fleet server is a device proposed to act as a middle way device and collect data from miscellaneous elastic agents. It is simple to deploy because it is managed from the GUI in Elasticsearch.

The difference between elastic agent and Beats is that elastic agent provides central management, easier to deploy and configure and above all is endpoint security capabilities.

Kibana

Is the GUI (Graphical User Interface) used to visualize data and create intuitive dashboards composed of charts, maps, graphs and more to give a simple yet informative insight of data analysed by Elasticsearch. As soon as data reaches Elasticsearch from different sources via Logstash, index patterns are used to avail data to Kibana.

Kibana is easy to use by utilizing its rich features and using the broad modules along with preconfigured dashboards. An example of Kibana dashboard is shown in figure 4 below.

Figure 4: Kibana dashboard view.

The previously mentioned components form the architecture of the Elastic Stack illustrated in figure 5.

Figure 5: Elastic Stack architecture.

Literature Review

Actually, not too many studies have been conducted regarding Elastic Stack due to the fact that it is still growing and getting popular. Most of the researches were focused in the functionality of the Elastic Stack in general and few highlighted some specific use cases.

In [7], the ELK Stack was used to provide decision makers in Materials and Life Science Experimental Facility (MLF) at the Japan Proton Accelerator Research Complex (J-PARC) with analysis capabilities to maintain high performance for neutron instruments that encompass huge data and logs generated from the systems these instruments operate on. It was hard to utilize these logs due to the fact that they had different structures as they were generated from diverse systems.

As the ELK stack can handle unstructured data it was adopted to collect real-time logs from the various devices and systems installed on the neutron instruments to easily analyse logs such as surveillance logs, measurement logs and operation logs for the purpose of providing detailed analytical information to proactively monitor instrument performance metrics, maintenance and preventing failures of devices and systems due to depreciation.

Meteorological data is of high importance and having precise analytical tools is of great importance. To overcome the difficulties that emerge in dealing with such data. The ELK stack in [8] was considered to collect, process and visualize data collected from a group of sensors that measure some meteorological parameters about environmental variables like temperature, humidity, precipitation, … etc).

The ELK Stack provided the necessary toolset to gather data and metadata with different formats usually ASCII format and at various frequencies. The collected data was visualized in specific time series for relative humidity.

Another use case discussed in [9] where ELK Stack provided real-time visualizations for web and application servers on cloud platform to give a holistic view for system admins for any server outage.

Methods

This paper is intended to highlight the features that make Elastic Stack a reliable tool for giving a holistic view for security-related events needed by security analysts to identify anomalies. This approach is crucial as cyber threats are spreading rapidly in a sophisticated manner that makes it hard to identify and remediate.

For achieving such a goal, Elastic Stack composed of the three previously-mentioned tools Elasticsearch, Logstash and Kibana was implemented in a server with the specifications detailed in table 1 as below. The system was implemented in VMware® Workstation 15 Pro.

Specification	Details
System	VMware Virtual Platform
Processor	Intel(R) Core (TM) i7-7500U CPU @ 2.70GHz
Hard Disk	80 GB
RAM	8 GB
Operating System	Alma Linux 8.5

Table 1: Server specifications.

The Elastic Stack was installed following the official guide on the server mentioned in table 1. Regarding the client side, where logs are to be collected, it was implemented in a device described in table 2.

Specification	Details
System	HP Pavilion Notebook
Processor	Intel(R) Core (TM) i7-7500U CPU @ 2.70GHz
Hard Disk	512 GB
RAM	8 GB
Operating System	Microsoft Windows 10 Pro

Table 2: Client specifications.

The process of installing fleet server and elastic agent is straightforward and simple. Firstly, the Elastic Agent has to be installed in a device to act as a Fleet Server, where it acts a control plane for updating and controlling Elastic Agents. For Elastic Agents to join a fleet server, an Agent policy is needed to indicate what type of logs have to be shipped from devices where Elastic Agents are installed. Also, an enrolment token is generated to initiate the communication and interact with Elasticsearch, it is a simple process as it is being carried out from the user interface in Kibana.

After successfully adding the fleet server, data will start to flow from endpoints, which allows the tendency to search and filter, in order to visualize data in a concise way according to what is required.

Also, Kibana provides Dark mode which is getting popular and been involved in many websites, apps, smartphones and other interfaces. Although it is a subject of debate, it is proved to provide visual benefits that helps being focused and attentive [12,13]. The dark mode was used for figures shown in the next sections.

The rest of this paper will highlight the importance of the data collected from the elastic agent in giving security analysts informative interpretation about events occurring.

Results

This section describes the results obtained after implementing Elastic Stack. As been mentioned above, cyber-attacks are spreading drastically and been conducted in advanced ways that make it hard to detect unless having a reliable monitoring system that provides a holistic view about events happening, which is crucial for security analysts to help them identify threats beforehand and mitigate effects that might lead to serious outages.

The ELK Stack adopted in this paper aims at providing security analysts with such data insights to promptly cope up with incidents and minimize detection and investigation time.

Kibana provides an intuitive user interface to simply interact and deal with data collected and indexed to create rich dashboards. What really matters is the tendency to ensure that data is being collected from miscellaneous sources on time. For that purpose, Kibana provides a Fleet section for the purpose of centrally managing Elastic Agents and providing details about their current status as healthy, unhealthy, updating and offline. In addition, it also provides the last time the Elastic Agent was active. Figure 6 shows the section of Agents in Fleet.

Figure 6: The status of Elastic Agents in Fleet section.

Having real-time information about system metrics is important to figure many performance-related measurements that help recognize system load, CPU usage, memory usage, swap usage, disk usage, current processes and how much load they impose. In addition to network-related metrics that show inbound and outbound traffic on different interfaces as well as packet loss, which is of great importance to monitor network traffic and easily figure spikes. This can be easily fetched and visualized in Kibana as shown in figure 7 below.

Figure 7: System metrics dashboard.

Monitoring logon events plays a pivotal role in figuring out security breaches. Most analysts focus in figuring out failed logons which gives information about unauthorized access or wrong passwords being used. Also, failed logons within short interval might indicate a brute force attack or an attacker conducting lateral movement. In the other hand successful logons are important as well because it does not only show users and privileged administrators' activity as well as service accounts but also gives implications of suspicious activities like unusual logon times, logon from unexpected devices or many resources access for the same user.

Also, an important activity to monitor is the RDP (Remote Desktop Protocol) that allows remote desktop connections to an endpoint. It should be monitored to observe remote connections attempts from unknown users or unknown IP addresses. Figure 8 demonstrates the dashboard that shows logon information in addition to RDP connections and sources.

Figure 8: Logons dashboard.

User and account management is intended to track changes to users' accounts like users created, users changed, users renamed, users enabled, users disabled, users locked out, users unlocked, passwords changed or reset and users deleted. This information is essential to detect attacks and malware behaviour. The dashboard in figure 9 depicts the user management events.

Figure 9: User management events dashboard.

The Elastic Security feature allows for better security-related events observations and anomalies detection. It provides detailed information about the creation, manipulation and deletion of different events ranging from files, network and other events. It also involves uncommon processes, which indicate potential issues like malware actions. In addition, maps give a concise view of DNS queries, source and destination IP Geolocation. Figure 10 depicts the Security events dashboard.

Figure 10: Security events dashboard.

There are many other rich features that can be utilized to display more data according to demands.

Discussion

The logs collected from different sources are key factor in security detections and observations. Hence, having a platform that can adapt to such massive amount of data while maintaining a high level of events filtering and visualizing to help security analysts to deal with threats in a timely manner is of high priority.

The Elastic Stack providing such capabilities, along with central management for log sources makes it a great data analytics tool as well as SIEM (Security Information and Event Management) for monitoring your infrastructure either on-premises or on cloud. It can be considered as comprehensive by having many integrations and modules that enables collecting logs and events from diverse sources like Apache sever, AWS, Azure, Google Cloud, MSSQL, MySQL, Nginx, tomcat, Squid, Cisco, Juniper, Zoom and many others. This variation immensely empowers Elastic Stack and makes it a perfect fit for highly-demanding data insights.

Conclusion

In this paper, the Elastic Stack including Elasticsearch, Logstash and Kibana was adopted to provide detailed information about events and logs collected from different sources via what is called Elastic Agent that sends logs to a centrally management Fleet Server that controls the enrolled Elastic Agents. As soon as logs reach Elasticsearch, Kibana is used to visualize data to give detailed information about system and performance metrics, in addition to network monitoring measurements, source and destination IP addresses with geolocation and map locations. Also,

user management events involving successful and failed logons, account related processes in order to give an intuition about events occurring.

The results show that the Elastic Stack is very powerful and full of features that makes it a reliable tool to be used by security analysts to cope up with highly-demanding data insights for prompt action-taking regarding monitoring security events and responding to incidents in a timely manner which is crucial.

It is recommended to conduct further studies to highlight some other features that even empower the Elastic Stack more and more like Machine Learning.

Bibliography

1. Foster Provost and Tom Fawcett. "Data Science for Business: What you need to know about data mining and data-analytic thinking".
2. Harvard Business Review. "A Beginner Guide to Data Analytics".
3. Mohammed Daffalla Elradi., *et al*. "Cyber Security Professionals' Challenges: A Proposed Integrated Platform Solution". *Electrical Science and Engineering* 3.2 (2021).
4. Elastic. "Elastic Stack: Elasticsearch, Kibana, Beats and Logstash". (2022).
5. Elastic. "Elasticsearch: The Official Distributed Search and Analytics Engine | Elastic" (2022).
6. SRIVASTAVA ANURAG. "Kibana 7 Quick Start Guide: Visualize your Elasticsearch data with ease". PACKT Publishing Limited (2019).
7. Moriyama K., *et al*. "Development of Status Analysis System Based on ELK Stack at J-PARC MLF" (2018).
8. Almeida E., *et al*. "Exploratory study of the Elk Stack for Meteorological Observation System Data Analysis". *Journal of Computational Interdisciplinary Sciences* 8.3 (2017).
9. Sunny Advani., *et al*. "Log analytics using ELK stack on Cloud platform". *IJARCCE* (2016).
10. Yang C., *et al*. "Cyberattacks detection and analysis in a network log system using XGBoost with ELK stack". *Soft Computing* 26.11 (2022): 5143-5157.
11. "Beats and elastic agent capabilitiesedit". Elastic. (n.d.) (2022).
12. Austin Erickson., *et al*. "An Extended Analysis on the Benefits of Dark Mode User Interfaces in Optical See-Through Head-Mounted Displays". *ACM Transactions on Applied Perception* 18.3 (2021): 22.
13. Apple. "How to Use Dark Mode on Your Mac". *Apple Support*, 7 Oct. (2019).