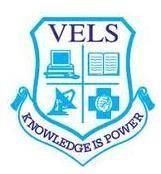


**SCHOOL OF COMPUTING SCIENCES**

**DEPARTMENT OF INFORMATION TECHNOLOGY**



**SECURITY INFORMATION AND EVENT MANAGEMENT (SIEM) SOLUTION**

***Submitted by***

BHUVANESH.G (20139105)

SHRIRAM.M (20139135)

MUKUND VIGNESH.D (20139113)

**Under the guidance of**

**Mrs. Elavarasi**

**A PROJECT REPORT**

**2022- 2023**

**BONAFIDE CERTIFICATE**

Certified that this project report **SECURITY INFORMATION AND EVENT MANAGEMENT** “**SIEM” SOLUTION PROJECT** is the Bonafide work of **BHUVANESH.G(20319105), SHRIRAM.M(20139135),** **MUKUND VIGNESH.D(20139113)** who carried out the project work under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form part of any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

*SIGNATURE SIGNATURE*

**Mrs. Elavarasi**

**Dr. T. KAMALAKANNAN**

Assistant Professor HEAD OF THE DEPARTMENT

Department of Computer Application Department of Computer Application

VISTAS, Chennai VISTAS, Chennai

**ACKNOWLEDGEMENT**

We thank the Almighty GOD for the abundant blessings showered on us. We extend our deepest love and gratitude to our dear parents who built up our career and backed us up in life.

We feel thankful to the Head of the Department **Dr. T. Kamalakannan,** Department of Information Technology, VISTAS, for all his encouragement, which has sustained our efforts.

We express our deepest gratitude to the internal guide **Mrs. Elavarasi** Professor, Department of Information Technology, VISTAS, for his valuable guidance, ideas and support.

We extend our sincere thanks to the project coordinator **Prof. Dhanaraj** **& Mr Stalin** for is guidance and support.

We would like to thank all other faculty members of Department of Information Technology for their help and advice throughout our life in this campus.

Finally, we are thankful to all our friends and all others who encouraged us and helped us in doing this project.

**Table of Contents**

[Abstract 5](#_Toc133966306)

[Introduction 6](#_Toc133966307)

[Scope 8](#_Toc133966308)

[SIEM Solution 9](#_Toc133966309)

[Log management 10](#_Toc133966310)

[Log Aggregation: 10](#_Toc133966311)

[Log Parsing 12](#_Toc133966312)

[Normalization 12](#_Toc133966313)

[Event Categorization 13](#_Toc133966314)

[Enrichment 14](#_Toc133966315)

[Log Indexing 14](#_Toc133966316)

[Storage 16](#_Toc133966317)

[Basic Technical Architecture 16](#_Toc133966318)

[Components and Capabilities of standardized SIEM Solution 17](#_Toc133966319)

[Open Source SIEM Platform – WAZUH 18](#_Toc133966320)

[Wazuh Solution 19](#_Toc133966321)

[Technical Components Used 20](#_Toc133966322)

[Application Architecture Components 21](#_Toc133966323)

[Major Implementation Steps 21](#_Toc133966324)

[Benefits of SIEM Implementation 23](#_Toc133966325)

[Results and Conclusions 24](#_Toc133966326)

# Abstract

Security Information and Event Management (SIEM) systems have been widely deployed as a powerful tool to prevent, detect, and react against cyber-attacks. SIEM solutions have evolved to become comprehensive systems that provide a wide visibility to identify areas of high risks and proactively focus on mitigation strategies aiming at reducing costs and time for incident response. Currently, SIEM systems and related solutions are slowly converging with big data analytics tools.

As part of the project, WAZUH is a free and open-source security platform were used in the pilot project have got multiple features to detect and alert security events reported either from desktops, laptops, servers, network and security devices either operating in on premise or in cloud platforms. The solution aggregates, normalizes it, parses the data and analyzes the data using pre-defined rule sets and presents the security events in visual dashboard for the security analyst to take appropriate timely actions.

The technical solution implemented on AWS Cloud platform in a single tier architecture with two different log sources in agent / agentless mode to simulate the security events from different log sources.

Various learnings imparted from the project towards the purpose of SIEM solution, effective deployment of open-source solution to detect intrusions, threats and behavior anomalies. As cyber threats are becoming more sophisticated, real-time monitoring and security analysis are needed for fast threat detection and remediation.

# Introduction

In today's increasingly digital world, the importance of data security cannot be overstated. Organizations of all sizes face an ever-growing number of cyber threats, including malware, phishing attacks, and data breaches. In order to protect their networks and sensitive information, organizations need effective security measures in place. One such measure is Security Information and Event Management (SIEM).

Security information and event management (SIEM) is a security solution that helps organizations recognize potential security events before they have a chance to disrupt business operations. It identifies the user and systemic behavior anomalies and uses use cases, signature-based threat detection and machine learning algorithms to detect security threats from across enterprise networks.

Security information and event management (SIEM) tools collect and aggregate log and event data to help identify and track security breaches. The solution provides cyber soc team members an insights into what's happening in an IT environment and a track record of relevant security events that have happened in the past.

The SIEM solution provides real-time analysis of security alerts generated by network hardware, and applications. By analyzing data from various sources, such as firewalls, routers, intrusion detection / prevention systems, and servers etc. SIEM systems can identify security incidents and provide a comprehensive view of security events across an organization's entire network.

The purpose of this report is to explore the concept of SIEM and its benefits for organizations. The report will discuss the key components of a SIEM system, including data collection, event correlation, and incident response.

Furthermore, the report will analyze the challenges and limitations of implementing SIEM in an organization, such as the high cost of deployment and maintenance, the need for skilled personnel to manage and monitor the system, and the complexity of integrating SIEM with existing security infrastructure.

Finally, the report will conclude with recommendations for organizations considering the implementation of SIEM, including the importance of selecting the right solution, investing in adequate training for personnel, and regularly evaluating and updating the SIEM system to ensure its effectiveness.

Overall, this report aims to provide a comprehensive overview of SIEM, and its technical open source Implementation “WAZUH” steps and its results and conclusions to provide organizations seeking to enhance their data security.

# Scope

The Scope of the Project to provide an overview of security information event management solution. Design and deploy an open source free ware “WAZUH” technical solution in a cloud platform and demonstrate some of the use cases to detect and alert potential anomalies / security events or behaviors and that will help cyber security team members to monitor and take further steps for mitigation.

# SIEM Solution

SIEM (Security Information and Event Management) is a software solution that provides real-time monitoring and analysis of security-related events occurring on an organization's IT systems. A SIEM solution collects security event data from various sources, such as network devices, servers, and applications, and correlates this data to detect security threats and incidents.

The SIEM solution continuously captures various logs from all kinds of data sources and match with the use cases and trigger security alerts to the cyber security team for further monitoring and mitigations. Some of the following security alerts are triggered but not limited to.

* Login authentication success / Failures
* Changes in user privileges
* Abnormal network traffic failures
* File integrity changes

The primary function of a SIEM solution is to provide visibility into an organization's security posture by aggregating and analyzing security event data from multiple sources. The solution typically includes several components as follows.

Diagram

Description automatically generated

Figure 1 - SIEM Components

## Log management

Log Management is the process of collecting, storing, and managing log data from various sources across an organization's IT infrastructure. This data includes information about user activity, network traffic, and system events. This information is then used to monitor system activity, identify security threats, and audit compliance.

The system collects logs from firewalls, intrusion detection/prevention systems, web proxies, and more. The collected data is then stored in a central repository where it's indexed and searched. This allows administrators to quickly identify and investigate suspicious activity.

There is a wealth of information in log files that can help identify problems and patterns in production systems. Log monitoring involves scanning log files, searching for patterns, rules, or inferred behavior that indicates important events, and triggering an alert sent to operations.

Log monitoring can help identify problems before they are experienced by users. It can uncover suspicious behavior that might represent an attack on organizational systems. It can also help record baseline behavior of devices, systems, or users, in order to identify anomalies that require investigation.

Additionally, by analyzing log data over time, administrators can develop better detection strategies and improve their overall security posture. While SIEM log management can be a time-consuming and complex process, it is an essential part of any effective security program. Log management is a part of the existing SIEM solution to collect the logs from multiple sources for effective monitoring.

## Log Aggregation:

SIEM platforms collect data from thousands of different sources because these events provide the data we need to analyze the health and security of our environment. In order to get a broad end-to-end view, we need to consolidate what we collect onto a single platform. Aggregation is the process of moving data and log files from disparate sources into a common repository. Collected data is placed into a homogenous data store – typically purpose-built flat file repositories or relational databases – where analysis, reporting, and forensics occur; and archival policies are applied.

The process of aggregation – compiling these dissimilar event feeds into a common repository – is fundamental to Log Management and in most SIEM platforms. Data aggregation can be performed by sending data directly into the SIEM/LM platform (which may be deployed in multiple tiers), or an intermediary host can collect log data from the source and periodically move it into the SIEM system. Aggregation is critical because we need to manage data in a consistent fashion: security, retention, and archive policies must be systematically applied. Perhaps most importantly, having all the data on a common platform allows for event correlation and data analysis, which are key to addressing the use cases we have described.

There are some downsides to aggregating data onto a common platform. The first is scale: analysis becomes exponentially harder as the data set grows. Centralized collection means huge data stores, greatly increasing the computational burden on the SIEM/LM platform. Technical architectures can help scale, but ultimately these systems require significant horsepower to handle an enterprise’s data. Systems that utilize central filtering and retention policies require all data to be moved and stored – typically multiple times – increasing the burden on the network.

Some systems scale using distributed processing, where filtering and analysis occur outside the central repository, typically at the distributed data collection point. This reduces the compute burden on the central server and allows processing to occur on smaller, more manageable data sets. It does require that policies, along with the code to process them, be distributed and kept current throughout the network. Distributed agent processes are a handy way to “divide and conquer”, but increase IT administration requirements. This strategy also adds a computational burden on the data collection points, degrading their performance and potentially slowing enough to drop incoming data.

## Log Parsing

Logs need to be parsed before they can be used to derive meaningful insights. Parsing is the process of extracting key pieces of information from each logged event and putting them into a common format. These values are then stored for later analysis. Logs can be quite large and contain lots of useless data. Parsing extracts only the relevant pieces of data while discarding the rest.

One example of parsing is mapping original timestamps to the values of a single time zone. Timestamps are critical metadata related to an event, and you can have different timestamps in your logs depending on your log sources.

A parser can extract other important pieces of information, such as usernames, source and destination IP addresses, the network protocol used, and the actual message of the log. For example, parsing can also filter out data to keep only ERROR and WARNING type events, while excluding anything less severe.

## Normalization

If the process of aggregation is to merge dissimilar events feeds into one common platform, normalization takes it one step further by reducing the records to just common event attributes. As we mentioned in the data collection post, most data sources collect the same base event attributes: time, user, operation, network address, and so on. Facilities like syslog not only group the common attributes, but provide means to collect supplementary information that does not fit the basic template. Normalization is where known data attributes are fed into a generic template, and anything that doesn’t fit is simply omitted from the normalized event log.

Depending upon the SIEM or Log Management solution, the original non-normalized records may be kept in a separate repository for forensics purposes prior to later archival or deletion, or they may simply be discarded. In practice, discarding original data is a bad idea, since the full records are required for any kind of legal enforcement. Thus, most products keep the raw event logs for a user-specified period prior to archival. In some cases, the SIEM platform keeps a link to the original event in the normalized event log which provides ‘drill-down’ capability to easily reference extra information collected from the device.

Normalization allows for predicable and consistent storage for all records, and indexes these records for fast searching and sorting, which is key when battling the clock in investigating an incident. Additionally, normalization allows for basic and consistent reporting and analysis to be performed on every event regardless of the data source.

Technically normalization is no longer a requirement on current platforms. Normalization was a necessity in the early days of SIEM, when storage and computer power were expensive commodities, and SIEM platforms used relational database management systems for back-end data management. Advances in indexing and searching unstructured data repositories now make it feasible to store full source data, retaining original data, and eliminating normalization overhead.

## Event Categorization

Categorizing collected events is one of its most important tasks. Categorization provides useful insights into an event and accurate categorization enables a security analyst to take appropriate actions for blocking or eliminating cybersecurity threats. Traditionally categorization is done manually. It requires a substantial amount of domain knowledge, vulnerability information, and most importantly experience. It is a known fact that there exists a resource shortage in the cybersecurity space. As more and more devices are being integrated with SIEM and the number of new events increases, it is becoming a daunting task for any SIEM provider to perform this essential work manually. Current siem solutions perform automatic event categorization using machine learning alogirthms for analyzing an event and recommending an event category.

## Enrichment

In reality, we are seeing a number of platforms doing data enrichment, adding supplemental information (like geo-location, transaction numbers, application data, etc.) to logs and events to enhance analysis and reporting. Enabled by cheap storage and Moore’s Law, and driven by ever-increasing demand to collect more information to support security and compliance efforts, we expect more platforms to increase enrichment. Data enrichment requires a highly scalable technical architecture, purpose-built for multi-factor analysis and scale, making tomorrow’s SIEM/LM platforms look very similar to current business intelligence platforms.

But that just scratches the surface in terms of enrichment, because data from the analysis can also be added to the records. Examples include identity matching across multiple services or devices, behavioral detection, transaction IDs, and even rudimentary content analysis. It is somewhat like having the system take notes and extrapolate additional meaning from the raw data, making the original record completer and more useful.

## Log Indexing

To speed up log searching, we create indexes from log files and data. It’s a critical part of log management to ensure that large data silos can still be accessible and searched. If you have terabytes of data from the last few weeks and need to go back and find a specific event, indexing log files will expedite the process.

The way the logs are indexed depends on the way you configure the index. Administrators can configure indexes to ensure that they are optimized based on standard search parameters. Generally, logs are indexed based on dates and times, but we can index logs in other ways, such as categories, usernames, log IDs, and event types.

Log indexing speeds up searches, but it also organizes the data within your archives. It helps you avoid duplicates, which can be expensive when you need storage for terabytes of data, primarily if we’re storing duplicate data. Sorting and organizing logs enhance search performance, but it is also beneficial for log management tools.

Indexing tools create keys that sort log files based on configurations (e.g., by date or type), so when you search these files, the search tool returns results quickly since it can execute your query on sorted data. Of course, there is more technical background involved in indexes and organizing data, but indexing ensures that your search queries perform faster than simply having files stored randomly.

When archiving log data, rotation is an important concept to understand because it’s the point at which current log data passes to an archive on the system’s local storage, and we create a new log. The frequency of log rotation depends on the system, configurations, and the size of each log file. For example, if you only have a few log entries a day in a log, it might make sense to rotate logs weekly instead of daily.

The main difficulty in log rotation is moving data to an older archive file, then creating a new current file without losing any data. This process is not complicated when you have a system with low traffic, but a high-traffic system can receive several events every second. In the latter case, the system must switch from an older log file to a current one without losing data. Your system resources and log tools play a significant role in ensuring that data is not lost during this process.

Another aspect of log rotation is file compression. Most systems compress files into a .gz format so that they take up less storage space. Compression saves space, but we must extract the file data before we can read it. However, compression saves on cost and storage for unneeded files in production. It’s common for organizations to compress log files as they move from a production environment to an archive location.

Every organization that handles network infrastructure and logs that track events should have an archiving solution in place. They should optimize this archiving solution for storage, compliance, speed, and reliability. In addition to optimizing for speed and costs, an archive solution should also have cybersecurity in mind. Attackers who access or archive could get plenty of information to launch attacks and potentially steal sensitive data. The process of log archiving and indexing is a critical component in log management.

## Storage

Logs collected from multiple sources are stored in a respective format. Organizations need to maintain the logs as part of the regulatory or contractual requirements for a defined period of time.

# Basic Technical Architecture

The SIEM solution collects logs from multiple sources but not limited to, parse, analyze and match the logs to the use case defined in the SIEM solution. The logs are collected either through agent based and agent less model for further processing and analyzing the sameDiagram

Description automatically generated

Figure 2 - SIEM Basic Technical Architecture

## Components and Capabilities of standardized SIEM Solution

Enclosed are the standard components involved in security information and event management solution with its capabilities.

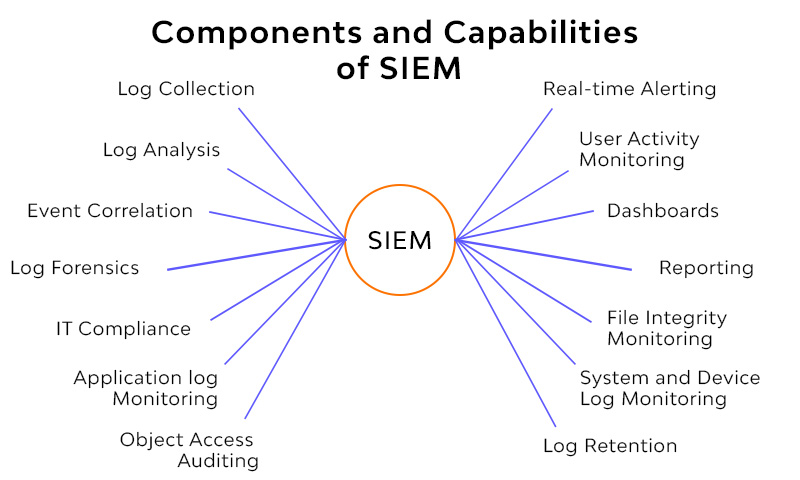


Figure 3 - Standard SIEM Capabilities

# Open Source SIEM Platform – WAZUH

“Wazuh” is an open-source SIEM solution that provides real-time threat detection, incident response, and compliance management capabilities. Here are some of the key features:

* Host-Based Intrusion Detection: Wazuh provides host-based intrusion detection capabilities, allowing organizations to monitor activity on individual servers and workstations to detect and respond to security incidents.
* Log Analysis: Wazuh can collect and analyze log data from various sources, including servers, network devices, and applications. It can also parse and normalize log data, making it easier to analyze and correlate security events.
* File Integrity Monitoring: Wazuh can monitor file system activity on servers and workstations, detecting changes to files and directories that may indicate a security threat.
* Threat Intelligence Integration: Wazuh integrates with popular threat intelligence feeds, allowing organizations to detect and respond to emerging threats.
* Compliance Management: Wazuh provides features such as audit trails, policy enforcement, and compliance reporting, helping organizations meet regulatory compliance requirements.
* Real-time Alerts: Wazuh can generate real-time alerts based on predefined rules or anomalies detected in log data. Alerts can be sent to security teams via email, SMS, or other notification methods.
* Incident Response Workflows: Wazuh provides incident response workflows that guide security teams through the process of investigating and responding to security incidents.
* Open Source: Wazuh is an open-source solution, meaning that organizations can download and use the software for free. This can help organizations reduce their security costs while still maintaining a high level of security.

Overall, Wazuh provides organizations with a comprehensive set of security capabilities, allowing them to detect and respond to security threats in real-time, meet regulatory compliance requirements, and improve their overall security posture.

## Wazuh Solution

The Wazuh platform provides XDR and SIEM features to continuously monitor and protect your cloud, container, and server workloads. These include log data analysis from multiple sources of logs, intrusion and malware detection, file integrity monitoring, configuration assessment, vulnerability detection, and support for regulatory compliance.

The Wazuh solution is based on the Wazuh agent, which is deployed on the monitored endpoints, and on three central components: the Wazuh server, the Wazuh indexer, and the Wazuh dashboard.

* The [Wazuh indexer](https://documentation.wazuh.com/current/getting-started/components/wazuh-indexer.html) is a highly scalable, full-text search and analytics engine. This central component indexes and stores alerts generated by the Wazuh server.
* The [Wazuh server](https://documentation.wazuh.com/current/getting-started/components/wazuh-server.html) analyzes data received from the agents. It processes it through decoders and rules, using threat intelligence to look for well-known indicators of compromise (IOCs). A single server can analyze data from hundreds or thousands of agents, and scale horizontally when set up as a cluster. This central component is also used to manage the agents, configuring and upgrading them remotely when necessary.
* The [Wazuh dashboard](https://documentation.wazuh.com/current/getting-started/components/wazuh-dashboard.html) is the web user interface for data visualization and analysis. It includes out-of-the-box dashboards for security events, regulatory compliance (e.g., PCI DSS, GDPR, CIS, HIPAA, NIST 800-53), detected vulnerable applications, file integrity monitoring data, configuration assessment results, cloud infrastructure monitoring events, and others. It is also used to manage Wazuh configuration and to monitor its status
* [Wazuh agents](https://documentation.wazuh.com/current/getting-started/components/wazuh-agent.html) are installed on endpoints such as laptops, desktops, servers, cloud instances, or virtual machines. They provide threat prevention, detection, and response capabilities. They run on operating systems such as Linux, Windows, macOS, Solaris, AIX, and HP-UX. The Wazuh agent is a single and lightweight monitoring software. It is a multi-platform component that can be deployed to laptops, desktops, servers, cloud instances, containers, or virtual machines. It provides visibility into the endpoint's security by collecting critical system and application records, inventory data, and detecting anomalies.
* In addition to agent-based monitoring capabilities, the Wazuh platform can monitor agent-less devices such as firewalls, switches, routers, or network IDS, among others.

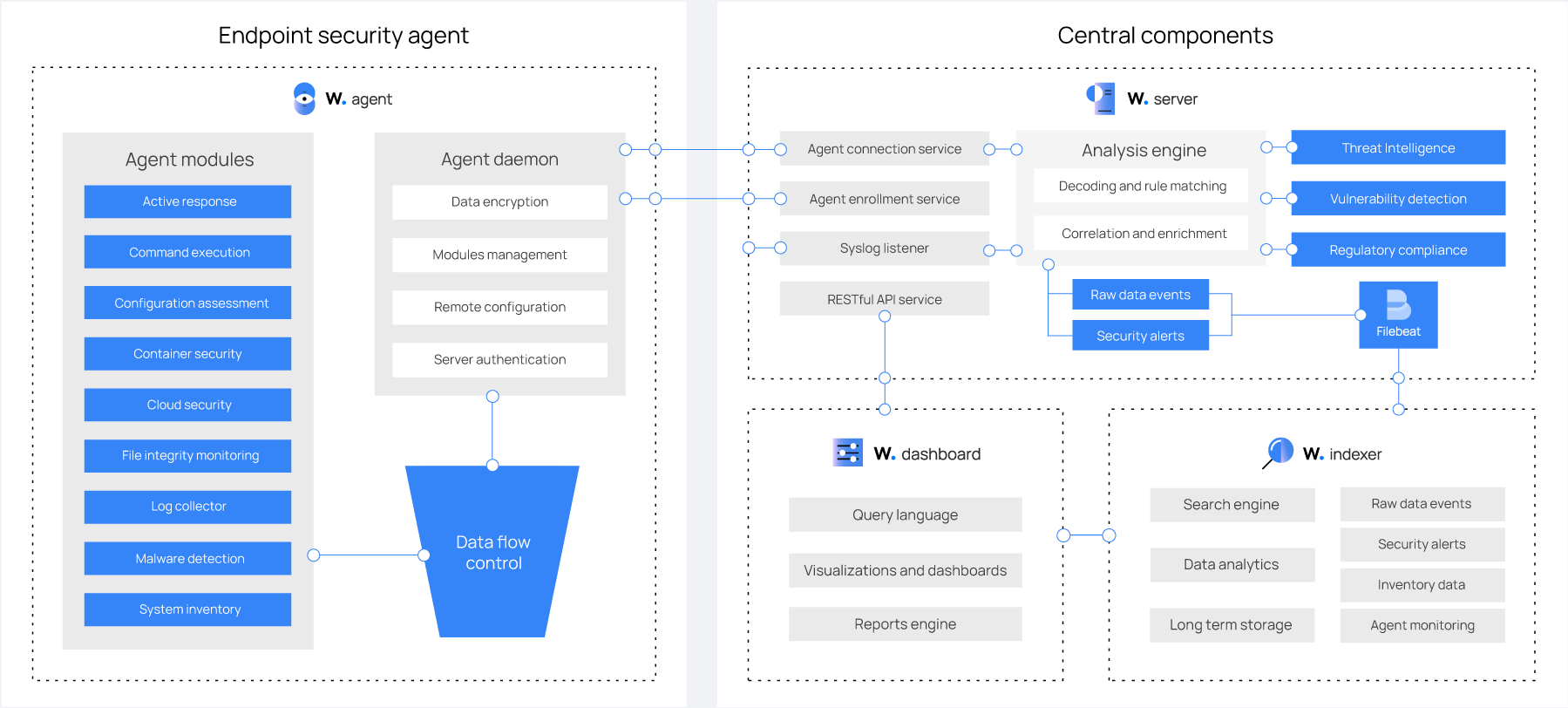


Figure 4 - WAZUH Open Source SIEM Components

## Technical Components Used

* Wazuh free Open-Source Software solution
* 2 Virtual Machines / EC2 Instances on the Cloud
* Wazuh Agents
* Windows Operating Systems

# Application Architecture Components

The following architecture diagram represents the standard application architecture supported by the open-source solution.

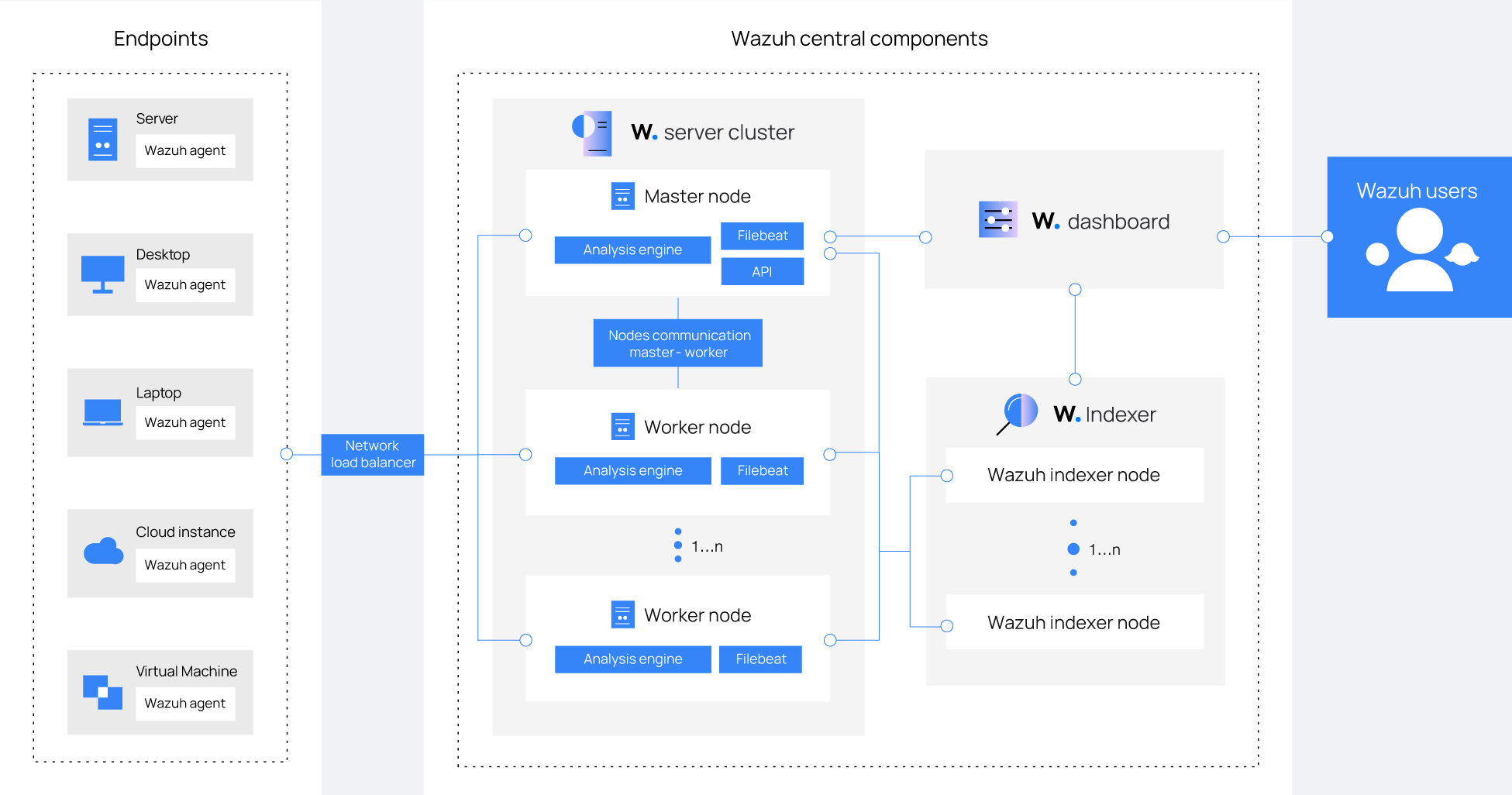


Figure 5 - Wazuh Application components

# Major Implementation Steps

Note: SIEM solution can be setup in Single / distributed architecture. However, for the project, the plan is to setup in a single tier architecture.



Figure 6 - Wazuh application components

* Deploy EC2 instances with Windows OS
* Download and Install Wazuh Open source solution from the trusted website in one of the EC2 system.
* Install the Wazuh Indexer, server on a single host as part of this project setup.
* Install the Wazuh agent in an another windows host to trigger logs to Wazuh Server
* Configure the Wazhu with indexer nodes installation and cluster intialization
* Enable the services to view on the dashboard.
* Define use cases for monitoring the alerts

The solution shall be setup in single tier architecture where all the server components (the Wazuh server, the Wazuh indexer, and the Wazuh dashboard) shall be installed on single VM to enable the solution. Similarly, to generate log sources, sample one Virtual desktops shall be setup to forward the system / security events of the logs to the WAZU Server. Additionally, one more VM Desktop shall be deployed with WAZUH agent to detect other endpoint security related events.

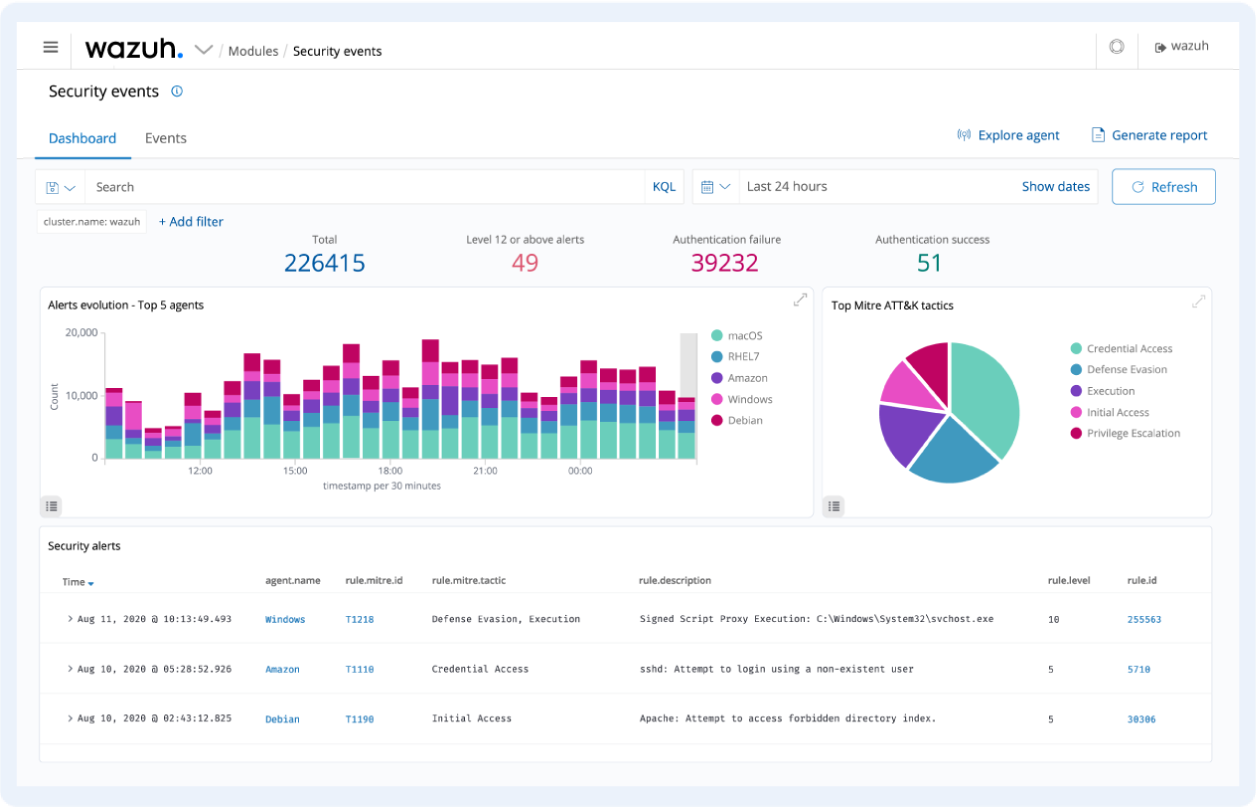


Figure 7 - Wazuh Sample Security Monitoring Dashboard

# Benefits of SIEM Implementation

SIEM (Security Information and Event Management) solutions offer several benefits to organizations. Here are some of the key benefits of implementing a SIEM solution:

* Log collection, Aggregation, correlation, alerting, analysis, reporting, log retention in one single platform.
* Reduce manual efforts for monitoring the individual logs from multiple resources.
* Improved Threat Detection: SIEM solutions can help organizations detect security threats in real-time by analyzing data from multiple sources and correlating events to identify patterns and anomalies that may indicate a security breach
* Rapid Incident Response: With a SIEM solution in place, organizations can respond quickly to security incidents by providing security teams with real-time alerts and incident response workflows. This can help minimize the impact of security incidents and reduce the time to resolution.
* Increased Visibility: SIEM solutions provide organizations with a centralized view of their security posture, allowing them to see security events and incidents across their entire IT infrastructure. This can help organizations identify vulnerabilities and prioritize security efforts.
* Compliance Management: SIEM solutions can help organizations meet regulatory compliance requirements by providing features such as audit trails, policy enforcement, and compliance reporting.
* Operational Efficiency: SIEM solutions can help organizations improve operational efficiency by automating security processes such as log collection, analysis, and reporting. This can help reduce the workload on security teams and enable them to focus on more strategic security initiatives.
* Cost Savings: By detecting and responding to security threats more quickly, organizations can reduce the impact of security incidents and avoid costly remediation efforts. SIEM solutions can also help organizations reduce the risk of data breaches, which can result in significant financial losses and damage to the organization's reputation.

Overall, SIEM solutions offer organizations a comprehensive approach to managing security events and incidents, enabling them to improve their security posture and protect against cyber threats.

# Results and Conclusions

* The SIEM solution provides continuous security monitoring solution using open-source solution deployed on the cloud platform to detect, alert and mitigate security alerts on a timely manner by the security team members.
* Solution supports deployment of large-scale implementation for various organizations who have got systems and applications connected to the internet which has continuous cyber threats today.
* The Open source SIEM solution provides multiple features such as Log analysis, vulnerability detection, file integrity monitoring, integrate with multiple threat intelligence sources, and ensure compliance.