Overview of SIEM technology

Previously, you learned about the SIEM process. In this reading, you'll explore more about this process and why SIEM tools are an important part of incident detection and response. As a refresher, a **security information and event management** (**SIEM**) tool is an application that collects and analyzes log data to monitor critical activities in an organization. You might recall that SIEM tools help security analysts perform **log analysis** which is the process of examining logs to identify events of interest.

SIEM advantages

SIEM tools collect and manage security-relevant data that can be used during investigations. This is important because SIEM tools provide awareness about the activity that occurs between devices on a network. The information SIEM tools provide can help security teams quickly investigate and respond to security incidents. SIEM tools have many advantages that can help security teams effectively respond to and manage incidents. Some of the advantages are:

- **Access to event data:** SIEM tools provide access to the event and activity data that happens on a network, including real-time activity. Networks can be connected to hundreds of different systems and devices. SIEM tools have the ability to ingest all of this data so that it can be accessed.
- **Monitoring, detecting, and alerting:** SIEM tools continuously monitor systems and networks in real-time. They then analyze the collected data using detection rules to detect malicious activity. If an activity matches the rule, an alert is generated and sent out for security teams to assess.
- **Log storage:** SIEM tools can act as a system for data retention, which can provide access to historical data. Data can be kept or deleted after a period depending on an organization's requirements.

The SIEM process

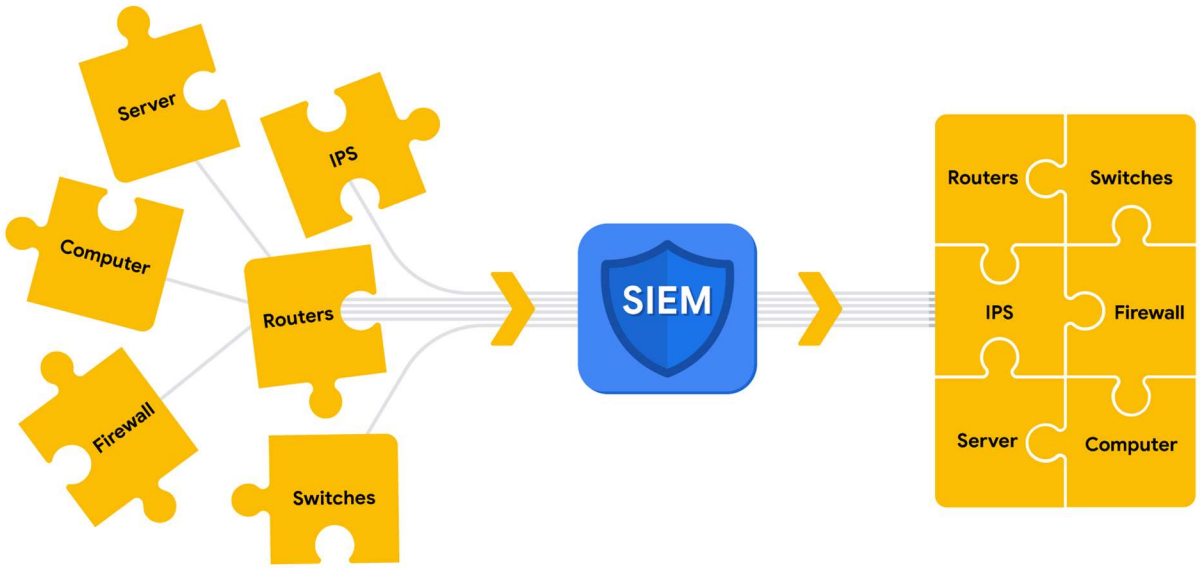
The SIEM process consists of three critical steps:

1. **Collect and aggregate data**
2. **Normalize data**
3. **Analyze data**

By understanding these steps, organizations can utilize the power of SIEM tools to gather, organize, and analyze security event data from different sources. Organizations can later use this information to improve their ability to identify and mitigate potential threats.

Collect and aggregate data

SIEM tools require data for them to be effectively used. During the first step, the SIEM collects event data from various sources like firewalls, servers, routers, and more. This data, also known as logs, contains event details like timestamps, IP addresses, and more. **Logs** are a record of events that occur within an organization’s systems. After all of this log data is collected, it gets aggregated in one location. Aggregation refers to the process of consolidating log data into a centralized place. Through collection and aggregation, SIEM tools eliminate the need for manually reviewing and analyzing event data by accessing individual data sources. Instead, all event data is accessible in one location—the SIEM.



Parsing can occur during the first step of the SIEM process when data is collected and aggregated. *Parsing* maps data according to their fields and their corresponding values. For example, the following log example contains fields with values. At first, it might be difficult to interpret information from this log based on its format:

`April 3 11:01:21 server sshd[1088]: Failed password for user nuhara from 218.124.14.105 port 5023`

In a parsed format, the fields and values are extracted and paired making them easier to read and interpret:

- host = `server`
- process = `sshd`
- source_user = `nuhara`
- source ip = `218.124.14.105`
- source port = `5023`

Normalize data

SIEM tools collect data from many different sources. This data must be transformed into a single format so that it can be easily processed by the SIEM. However, each data source is different and data can be formatted in many different ways. For example, a firewall log can be formatted differently than a server log.

Collected event data should go through the process of normalization. *Normalization* converts data into a standard, structured format that is easily searchable.

Analyze data

After log data has been collected, aggregated, and normalized, the SIEM must do something useful with all of the data to enable security teams to investigate threats. During this final step in the process, SIEM tools analyze the data. Analysis can be done with some type of detection logic such as a set of rules and conditions. SIEM tools then apply these rules to the data, and if any of the log activity matches a rule, alerts are sent out to cybersecurity teams.

Note: A part of the analysis process includes correlation. *Correlation* involves the comparison of multiple log events to identify common patterns that indicate potential security threats.

Scenario

Review the following scenario. Then complete the step-by-step instructions.

You are a security analyst working at the e-commerce store Buttercup Games. You've been tasked with identifying whether there are any possible security issues with the mail server. To do so, you must explore any failed SSH logins for the root account.

Note: *Use the incident handler's journal you started in a previous activity to take notes during the activity and keep track of your findings.*

Step-By-Step Instructions

Follow the instructions and answer the following questions to complete the activity.

Step 1: Access supporting materials

The following supporting materials will help you complete this activity. The data contains log and event information from Buttercup Games' mail servers and web accounts. This includes information like access and authentication logs, email logs, and more.

To download this data, click the link then click the download icon.

Link to supporting materials: tutorialdata.zip file

OR

If you don't have a Google account, you can download the supporting materials directly from the following attachment.

tutorialdata
ZIP File

Step 2: Create a Splunk Cloud account

To use Splunk Cloud, you must create an account. Follow *Part 1 - Create a Splunk Cloud account* and *Part 2 - Verify your email* in the Follow-along guide for Splunk sign-up to create an account.

Step 3: Sign up for a free Splunk Cloud trial

After you've created your Splunk account, you'll need to sign up for a free Splunk Cloud trial. Follow *Part 3 - Activate a Splunk Cloud trial* in the Follow-along guide for Splunk sign-up.

Note*: If you experience any issues activating your Splunk Cloud trial please check out the Splunk cloud tutorial video*.

Step 4: Upload data into Splunk

To operate effectively, it's essential that SIEM tools ingest and index data. SIEM tools collect and process data so that it becomes searchable events that can be queried, viewed, and analyzed.

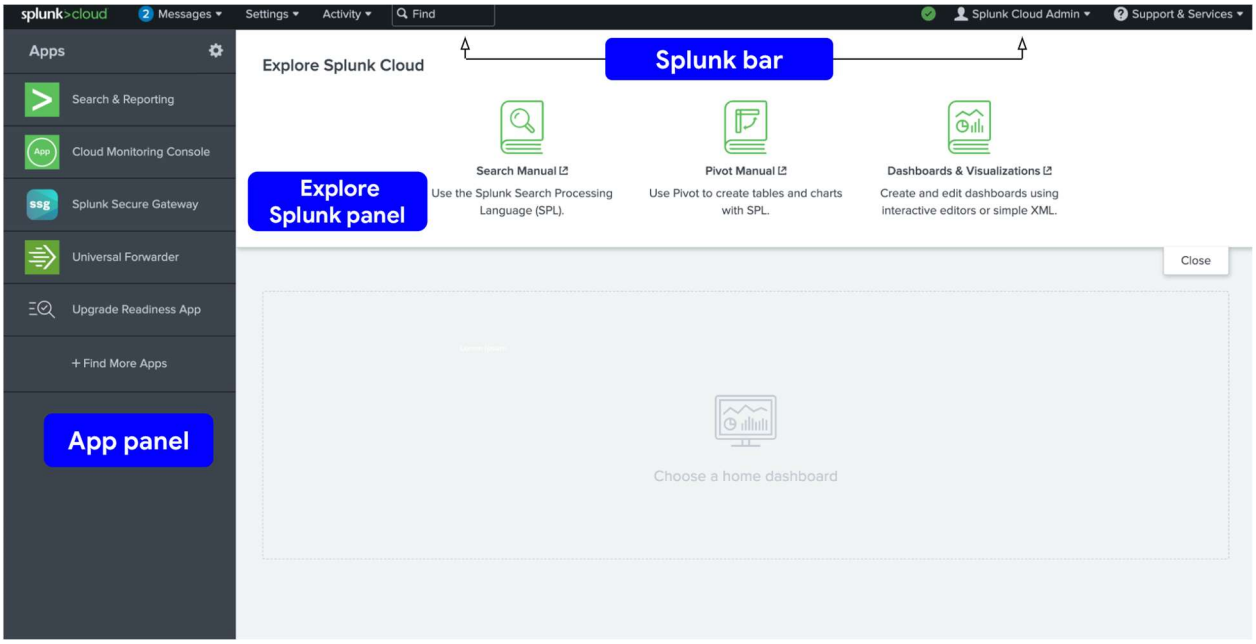
So far, you've created a Splunk account and activated and accessed the Splunk Cloud free trial, but your Splunk Cloud instance does not contain any data. Next, you'll need to upload data into Splunk to start querying. Complete the following steps to upload data into Splunk:

4. If you haven't already, download the data file from Step 1: tutorialdata.zip. Click the link then click the download icon. Do not uncompress the file.
5. Navigate to Splunk Home from your Splunk Cloud free trial instance. You might need to log in again using your credentials from Step 3.
6. On the Splunk bar, click **Settings.** Then click the **Add Data** icon.
7. Click **Upload**.
8. Click the **Select File** button.
9. Upload the `tutorialdata.zip` file, and click **Open**.
10. Click the **Next** button to continue to **Input Settings**.
11. By the **Host** section, select **Segment in path** and enter **1** as the segment number.
12. Click the **Review** button and review the details of the upload before you submit. The details should be as follows: Input Type: Uploaded File File Name: tutorialdata.zip Source Type: Automatic Host: Source path segment number: 1 Index: Default
13. Click **Submit**. Once Splunk has ingested the data, you will receive confirmation that the file was successfully uploaded.

***Note**: If you are experiencing issues uploading data into Splunk, refer to the Splunk Search Tutorial guide for help.*

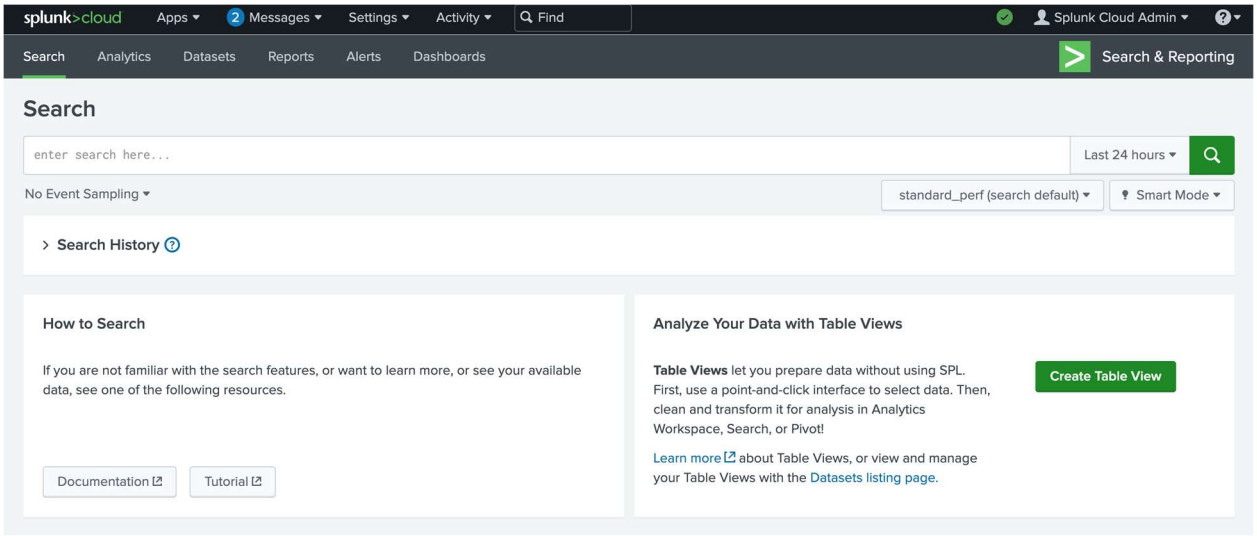
Step 5: Perform a basic search

Take a moment to examine the Splunk Cloud interface by locating the app panel, the Explore Splunk panel, and the Splunk bar.



Now that you've uploaded the data into Splunk, perform your first query to confirm that the data has been ingested, indexed, and is searchable. Follow these steps to perform a query:

14. Navigate to Splunk Home. (To return to Splunk Home, click the Splunk Cloud logo on the Splunk Cloud page.)
15. Click **Search & Reporting**. You may close any pop ups that appear.
16. In the search bar, enter your search query: **`index=main`** This search term specifies the index. An **index** is a repository for data. Here, the index is a single dataset containing events from an index named main.
17. Select **All Time** from the time range dropdown to search for all the events across all time.
18. Click the search button. Note that the search button is represented by the magnifying glass icon. Your search should retrieve thousands of events.

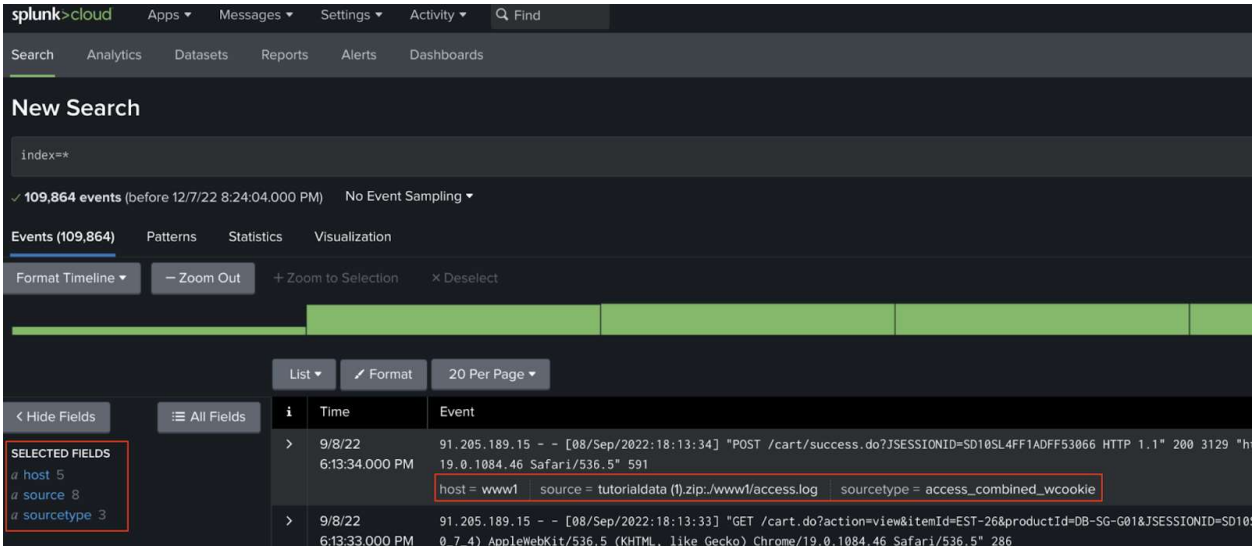


***Pro tip**: It's a best practice to use short time ranges in your searches because a shorter time range returns results faster and uses fewer resources. Adjust the time using the time range dropdown or by using time modifiers in your search.*

Step 6: Evaluate the fields

When Splunk indexes data, it attaches fields to each event. These fields become part of the searchable index event data. This helps security analysts easily search for and find the specific data they need. Now that you've run your first query, examine the search results and the fields.

For each event the fields are **`host`**, **`source`**, and **`sourcetype`**. Under **SELECTED FIELDS**, examine the same fields.



Examine the field values by clicking on the field under **SELECTED FIELDS**. You should observe the following:

- **host**: The host field specifies the name of the network host from which the event originated. In this search there are five hosts:
- **`mailsv`** - Buttercup Games' mail server. Examine events generated from this host.
- **`www1`** - This is one of Buttercup Games' web applications.
- **`www2`** - This is one of Buttercup Games' web applications.
- **`www3`** - This is one of Buttercup Games' web applications.
- **`vendor_sales`** - Information about Buttercup Games' retail sales.
- **source**: The source field indicates the file name from which the event originates. You should identify eight sources. Notice **`/mailsv/secure.log`**, which is a log file that contains information related to authentication and authorization attempts on the mail server.
- **sourcetype**: The sourcetype determines how data is formatted. You should observe three sourcetypes. Examine **`secure-2`**.

Step 7: Narrow your search

Because you've been tasked with exploring any failed SSH logins for the root account on the mail server, you'll need to narrow the search results for events from the mail server.

Under **SELECTED FIELDS**, click **host** and click **mailsv**.

Notice that a new term has been added to the search bar: **`index=main host=mailsv`**. The search results have narrowed to over 9000 events that are generated by the mail server.

Step 8: Search for a failed login for root

Now that you've narrowed your search results to events generated by the mail server, continue to narrow the search to locate any failed SSH logins for the root account.

19. Clear the search bar.
20. Enter **`index=main host=mailsv fail* root`** into the search bar. This search expands on the search from the previous task and searches for the keyword **`fail*`**. The wildcard tells Splunk to

expand the search term to find other terms that contain the word *fail* such as *failure*, *failed*, etc. Lastly, the keyword `root` searches for any event that contains the term root.

21. Click **search**.

Step 9: Evaluate the search results

Your search from the previous task should have retrieved search results for over 300 events. Navigate to other pages of the search results to observe the events not listed on the first page of results.

***Pro tip**: Splunk highlights search terms in search results to make it easier to identify where the search terms appear in the data.*

Step 10: Answer questions about the search results

1.

Question 1

How many events are contained in the main index across all time?

1 point

10,000
100-1,000
Over 100,000
10-99

2.

Question 2

Which field identifies the name of a network device or system from which an event originates?

1 point

source
sourcetype
host
index
3.

Question 3

Which of the following hosts used by Buttercup Games contains log information relevant to financial transactions?

1 point

www3
www2
vendor_sales
www1
4.

Question 4

How many failed SSH logins are there for the root account on the mail server?

1 point

None
One
100
More than 100

Key takeaways

In this activity, you used Splunk Cloud to perform a search and investigation. Using Splunk Cloud, you were able to:

- Upload sample log data
- Search through indexed data
- Evaluate search results
- Identify different data sources
- Locate failed SSH login(s) for the root account

If you would like to challenge yourself and explore more simulated incident investigations using Splunk, log in to Splunk and visit Splunk Boss of the SOC.