

BASIC SPLUNK SPL QUERIES

What is Splunk? Splunk is a Security Information and Event Management (SIEM) tool that helps cybersecurity teams monitor, detect, and respond to threats in real-time. It centralizes log data from diverse sources like firewalls, servers, applications, endpoints and analyzes it for suspicious activities or vulnerabilities.



How Splunk Supports Cybersecurity Teams

Real-Time Threat Detection

Splunk identifies potential security incidents by analyzing patterns in data.

For example, it can flag unusual login attempts, unexpected file transfers, or communication with known malicious IP addresses.

Incident Response

Splunk streamlines response workflows by correlating logs, enabling analysts to understand the "who, what, when, where, and how" of an attack.

Compliance and Reporting

With built-in dashboards and reporting features, Splunk helps organizations meet compliance requirements like GDPR, HIPAA, and PCI-DSS.

Forensic Investigations

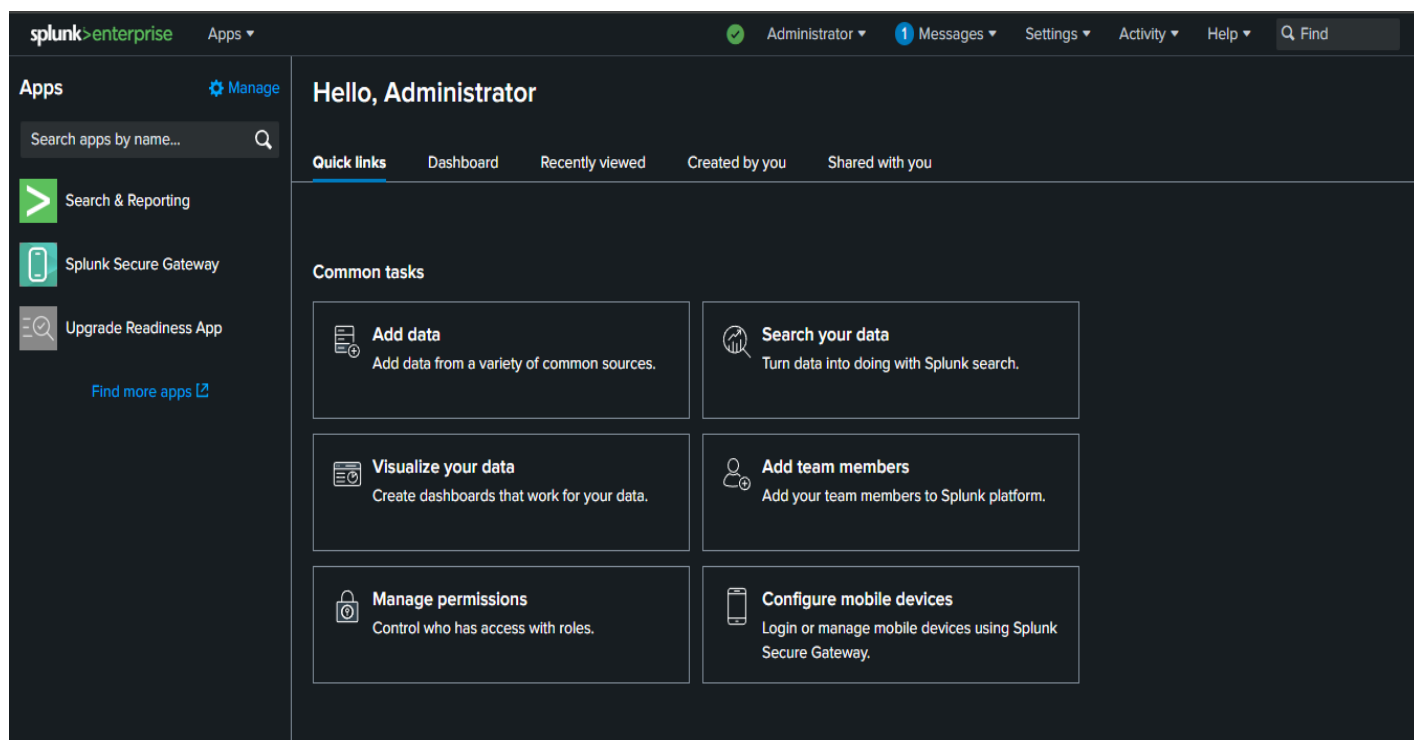
By retaining and indexing historical data, Splunk allows SOC teams to investigate breaches, uncover root causes, and implement preventive measures.

How SOC Analysts Use Splunk

SOC analysts rely on Splunk to:

- Correlate logs and detect advanced persistent threats (APTs).
- Investigate incidents, such as detecting C2 traffic or URL-based attacks.
- Monitor for Indicators of Compromise (IoC's) like suspicious IPs or malicious file hashes.

Adding Data to Splunk



Quick Links Tab

On the Splunk home dashboard or the main page of the Search & Reporting app, find the Quick Links panel on the screen.

Click "Add Data"

Within the Quick Links tab, there is a prominent "Add Data" option. Clicking this link opens the same "Add Data" wizard you would access through the Settings menu.

Follow the Data Input Wizard: The steps to add data are the same as in the Settings menu:

Or get data in with the following methods



Upload

files from my computer

Local log files
Local structured files (e.g. CSV)
[Tutorial for adding data](#)



Monitor

files and ports on this Splunk platform instance

Files - HTTP - WMI - TCP/UDP - Scripts
Modular inputs for external data sources



Forward

data from a Splunk forwarder

Files - TCP/UDP - Scripts

Upload a File: Select and upload a file from your system.

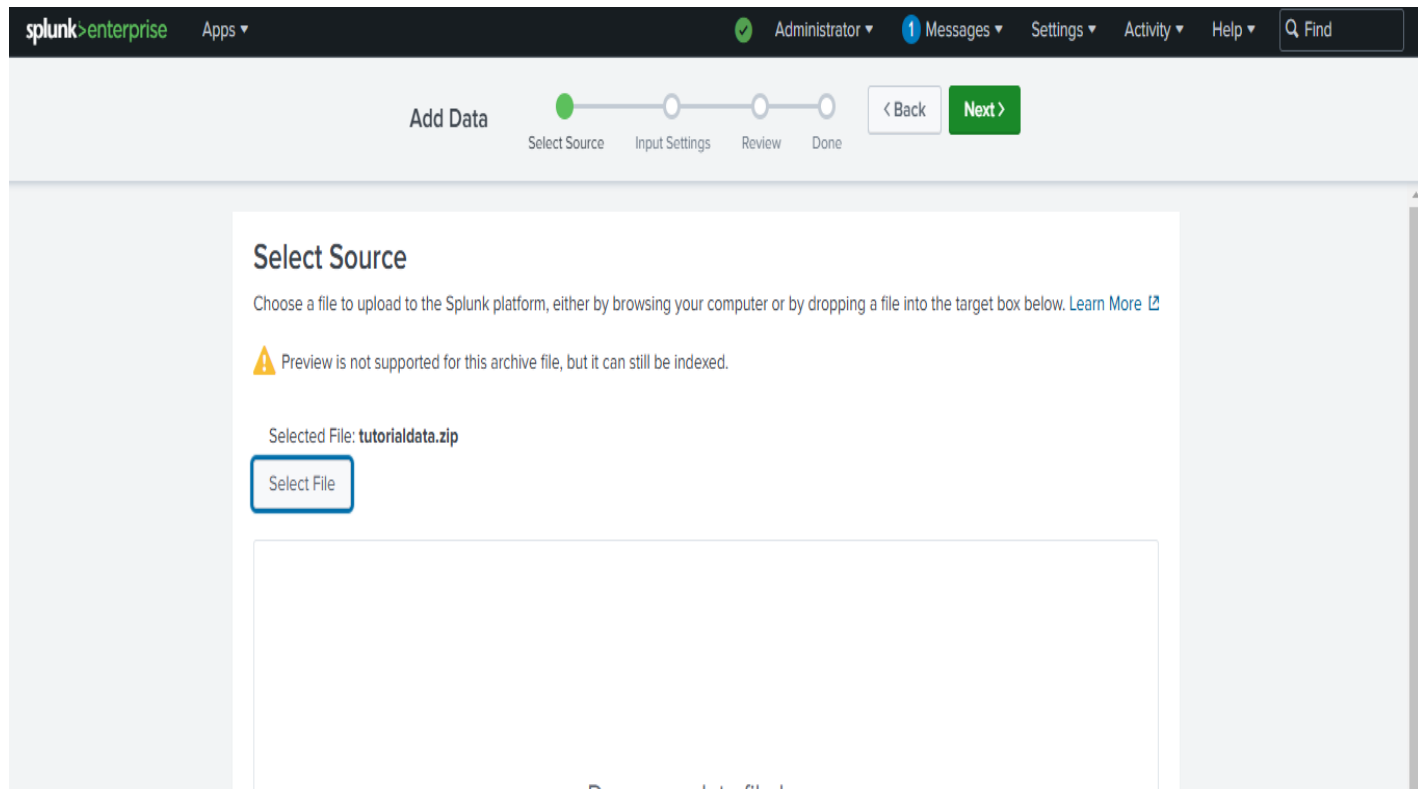
Monitor a Source: Specify a directory, file path, or network source to monitor.

Forward Data: Configure Splunk forwarders for continuous data streaming.

Configure Source Type and Index

Splunk attempts to auto-detect the data's source type. Adjust manually if necessary.

Choose an index for storing the data.



The screenshot shows the Splunk Enterprise web interface. At the top, the navigation bar includes the Splunk logo, 'enterprise' branding, and links for 'Apps', 'Administrator', 'Messages', 'Settings', 'Activity', and 'Help'. A search bar is on the right. Below the navigation bar, a progress bar indicates the current step is 'Select Source' (highlighted with a green dot). The main content area is titled 'Select Source' and contains instructions: 'Choose a file to upload to the Splunk platform, either by browsing your computer or by dropping a file into the target box below. [Learn More](#)'. A warning icon and text state: 'Preview is not supported for this archive file, but it can still be indexed.' Below this, it says 'Selected File: tutorialdata.zip' and features a 'Select File' button. A large empty box for dropping files is at the bottom.

Input Settings

In this step, you define how Splunk processes the ingested data.

Source Type

Splunk automatically detects the data source type (e.g., syslog, Apache logs).

Manually set the source type if auto-detection is incorrect:

Use categories like csv, json, or access_combined.

splunk>enterprise Apps ▾

Administrator ▾ 1 Messages ▾ Settings ▾ Activity ▾ Help ▾ Find

Add Data

Select Source Input Settings Review Done

< Back Review >

Source type

The source type is one of the default fields that the Splunk platform assigns to all incoming data. It tells the Splunk platform what kind of data you've got, so that the Splunk platform can format the data intelligently during indexing. And it's a way to categorize your data, so that you can search it easily.

Automatic Select New

Host

When the Splunk platform indexes data, each event receives a "host" value. The host value should be the name of the machine from which the event originates. The type of input you choose determines the available configuration options. [Learn More](#)

☒ Constant value
☐ Regular expression on path
☐ Segment in path

Host field value: DESKTOP-E24DCPL

Index

The Splunk platform stores incoming data as events in the selected index. Consider using a "sandbox" index as a destination if you have problems determining a source type for your data. A sandbox index lets you troubleshoot your configuration without impacting production indexes. You can always change this setting later. [Learn More](#)

Index: test1 ▾ Create a new index

Optionally preview the parsed data for verification.

Host Assignment

Assign a host value for the data

IP Address: Use the IP address of the machine sending the data.

Hostname: Use the hostname of the machine sending the data.

Custom Value: Manually enter a value (e.g., web_server_1).

Set Index

Choose the index where the data will be stored.

Default index is usually main.

Use a custom index for specific types of data for better organization.

Click Next to proceed.

The screenshot shows the Splunk Enterprise interface. At the top is a dark navigation bar with the 'splunk>enterprise' logo, an 'Apps' dropdown, and user/notifications links: 'Administrator' (with a green checkmark), '1 Messages' (with a blue notification icon), 'Settings', 'Activity', and 'Help'. A search bar with a magnifying glass icon and the text 'Find' is on the right. Below the navigation bar is a light gray header area. On the left, it says 'Add Data'. To its right is a progress indicator with four steps: 'Select Source', 'Input Settings', 'Review', and 'Done'. The first three steps are marked with green circles, and they are connected by a green line. The 'Review' step is the current active step. To the right of the progress indicator are two buttons: a white button with a left arrow and the text '< Back', and a green button with the text 'Submit >'. Below the header is a large white rectangular area containing the 'Review' section. The title 'Review' is at the top left of this area. Below the title are five lines of configuration details, each with a label followed by a dotted line and a value: 'Input Type Uploaded File', 'File Name tutorialdata.zip', 'Source Type Automatic', 'Host DESKTOP-E24DCPL', and 'Index test1'.

splunk>enterprise Apps ▾ Administrator ▾ 1 Messages ▾ Settings ▾ Activity ▾ Help ▾ Find

Add Data

Select Source Input Settings Review Done

< Back Submit >

Review

Input Type Uploaded File
File Name tutorialdata.zip
Source Type Automatic
Host DESKTOP-E24DCPL
Index test1

Review

The final step before ingestion.

Above is a sample query in Splunk

***index="test1" sourcetype=access_combined_wcookie
clientip="87.194.216.51" productId="SC-MG-G10"***

Here's how I'm breaking it down:

1. Specify **index="test1"** to search within the test1 index where the data is stored.
2. Narrow down the search by using **sourcetype=access_combined_wcookie** to target events formatted as web server logs with cookies.
3. Filter events by applying **clientip="87.194.216.51"** to focus on logs from this specific IP address.
4. Include **productId="SC-MG-G10"** to isolate events related to the specified product ID.

This approach helps to efficiently identify and analyze relevant events within the dataset.

Splunk Search Processing Language Syntax Structure

<search criteria> | <command> <arguments>

Search criteria: Specifies where to retrieve the data from (e.g., index, sourcetype).

Pipe (|): Passes the results from one command to another.

Commands: Operate on the data (e.g., stats, eval, table).

Example Query

index="test1" sourcetype="access_combined" | stats count by clientip

- Retrieves events from test1 with the specified sourcetype.
- Counts occurrences of events grouped by clientip.

SPL is central to Splunk's functionality, making it a versatile and powerful tool for analyzing machine-generated data.

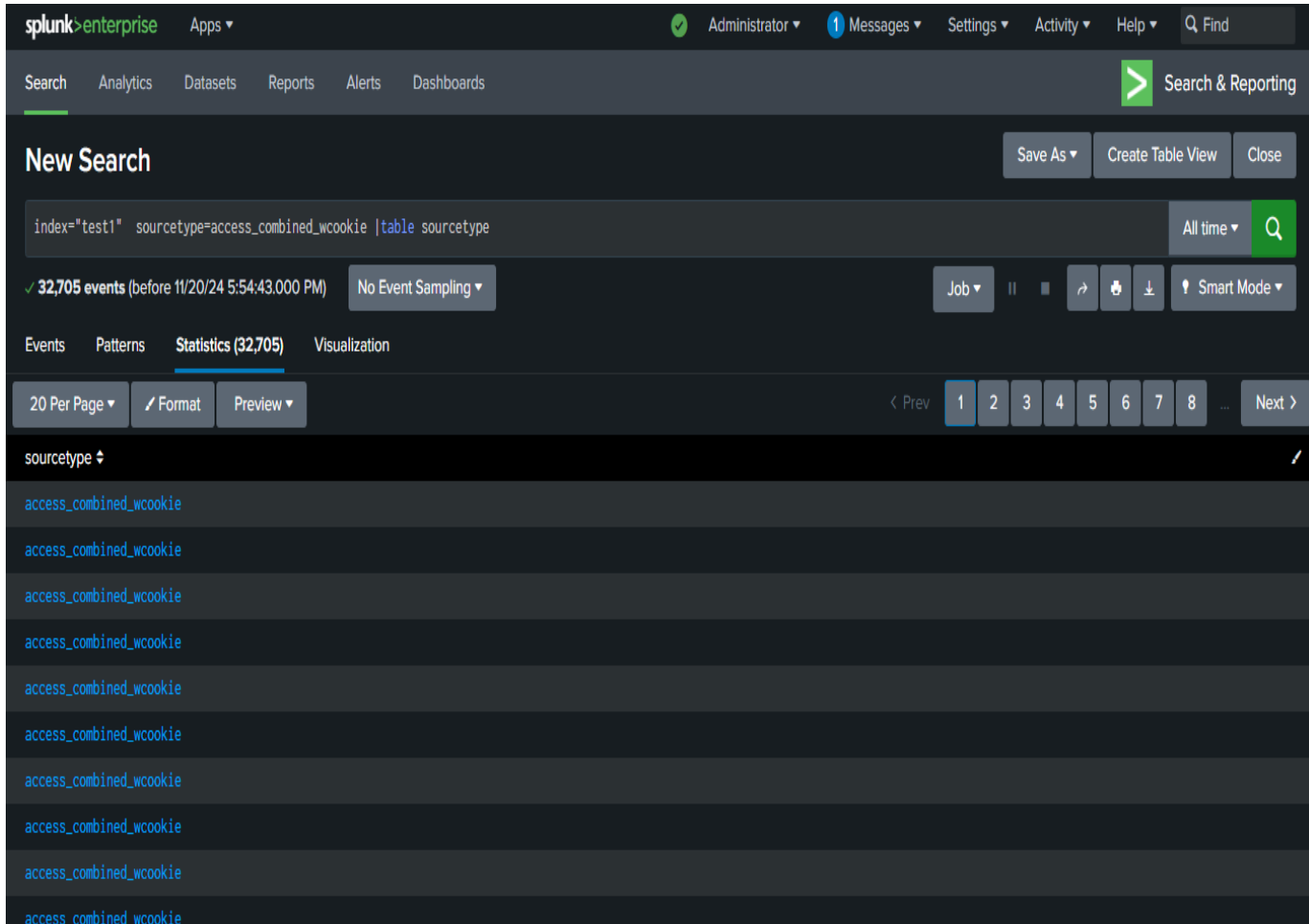
Basic Query Commands in Splunk

1.Table- Displays search results in a structured, tabular format with specified fields. It's useful for organizing data and making it easier to analyze or share.

[illegible]

2. dedup- Removes duplicate events based on specified fields. It ensures that only unique entries are displayed in the search results.

For example, below we can see that there are multiple entries for sourcetype “access_combined_wcookie”



In this case, dedup sourcetype doesn't change much, because all events in your search already share the same sourcetype value (access_combined_wcookie).

It will keep the first event and remove any duplicates based on the sourcetype field as shown below.

New Search

Save As Create Table View Close

index="test1" sourcetype=access_combined_wcookie |table|sourcetype|dedup sourcetype All time Q

32,705 events (before 11/20/24 5:53:02.000 PM) No Event Sampling Job Pause Stop Refresh Download Smart Mode

Events Patterns Statistics (1) Visualization

20 Per Page Format Preview

sourcetype
access_combined_wcookie

3. **Field**- Specifies which fields to include or exclude in the results. It helps reduce the volume of data and makes the results more focused.

You can use fields to list the fields you want to keep (or exclude), improving performance and clarity in search results.

New Search Save As ▾ Create Table View Close

index="test1" sourcetype=access_combined_wcookie clientip="87.194.216.51"
productId="SC-MG-G10" | fields clientip productId All time ▾ Q

✓ 61 events (before 11/20/24 5:22:35.000 PM) No Event Sampling ▾ Job ▾ || ▮ ↶ ▴ ⬇ Smart Mode ▾

Events (61) Patterns Statistics Visualization

Format Timeline ▾ – Zoom Out + Zoom to Selection x Deselect 1 hour per column

List ▾ Format 20 Per Page ▾ < Prev 1 2 3 4 Next >

< Hide Fields All Fields

SELECTED FIELDS
a clientip 1
a productId 1
+ Extract New Fields

i	Time	Event
>	9/8/22 8:59:36.000 AM	87.194.216.51 - - [08/Sep/2022:08:59:36] "GET /category.screen?categoryId=SIMULATION&JSESSIONID=SD9SL4FF10ADFF50456 HTTP 1.1" 200 872 "http://www.buttercupgames.com/cart.do?action=view&itemId=EST-18&productId=SC-MG-G10" "Mozilla/5.0 (iPad; U; CPU OS 4_3_5 like Mac OS X; en-us) AppleWebKit/533.17.9 (KHTML, like Gecko) Version/5.0.2 Mobile/8L1 Safari/6533.18.5" 407 clientip = 87.194.216.51 productId = SC-MG-G10
>	9/8/22 6:49:30.000 AM	87.194.216.51 - - [08/Sep/2022:06:49:30] "GET /cart.do?action=addtocart&itemId=EST-21&productId=SC-MG-G10&JSESSIONID=SD5SL8FF9ADFF50030 HTTP 1.1" 200 326 "http://www.google.com" "Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.2.28) Gecko/20120306 YFF3 Firefox/3.6.28 (.NET CLR 3.5.30729; .NET4.0C)" 665 clientip = 87.194.216.51 productId = SC-MG-G10
>	9/8/22 4:33:34.000 AM	87.194.216.51 - - [08/Sep/2022:04:33:34] "GET /product.screen?productId=SC-MG-G10&JSESSIONID=SD3SL7FF4ADFF49495 HTTP 1.1" 200 2440 "http://www.yahoo.com" "Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.1; WOW64; Trident/5.0; BOIE9;ENUS)" 315 clientip = 87.194.216.51 productId = SC-MG-G10

For Including Specific fields. This query will return only the clientip and productId fields from the events in the test1 index that match the access_combined_wcookie sourcetype.

4 **Stats**- The stats command in Splunk is used to calculate statistics over your data, such as counts, sums, averages, etc.

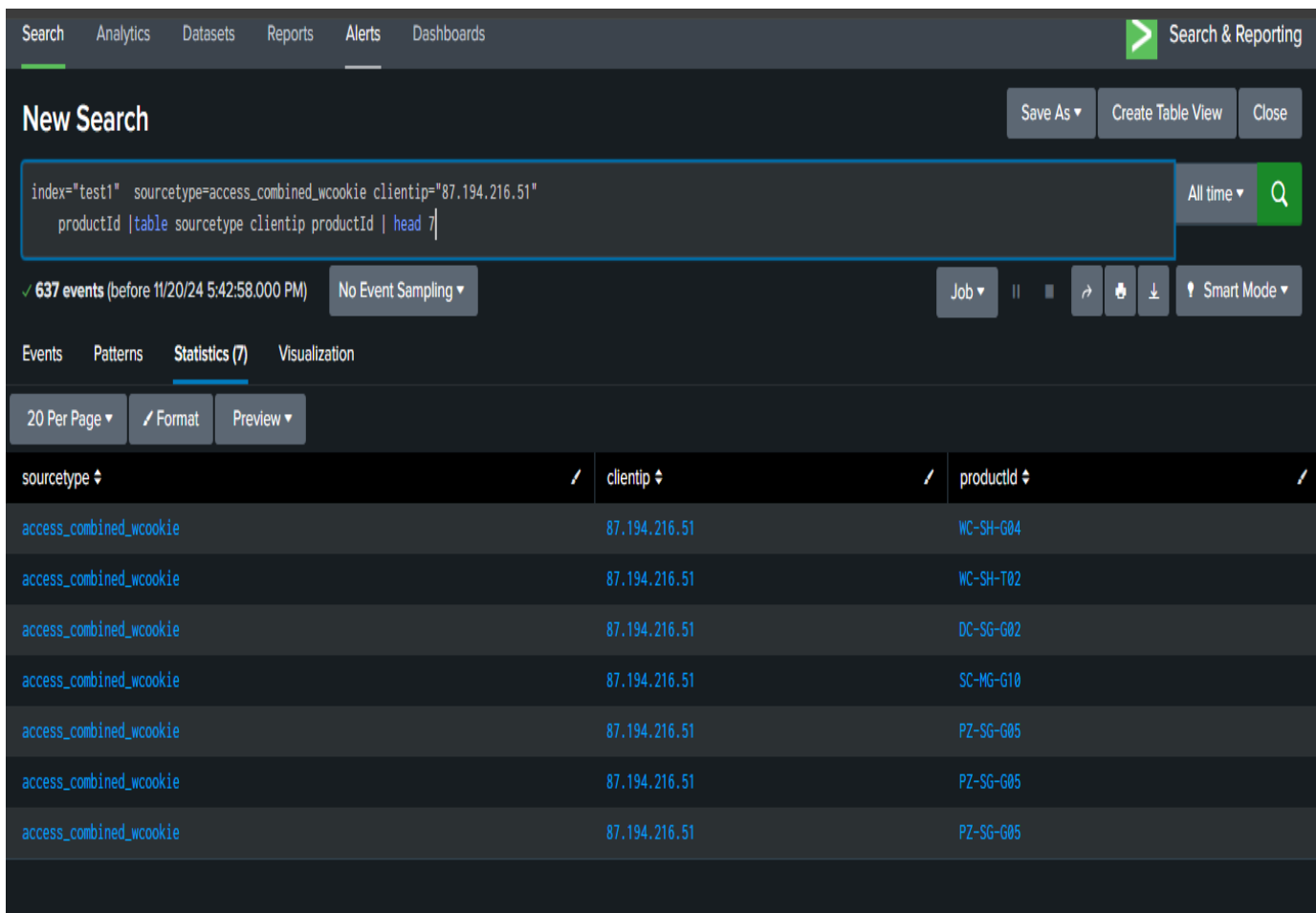
When you use the count function in the stats command, it provides a count of distinct (unique) values for a field.

The screenshot shows the Splunk Search interface. At the top, there's a 'New Search' header with buttons for 'Save As', 'Create Table View', and 'Close'. Below this is a search bar containing the query: `index="test1" sourcetype=access_combined_wcookie clientip="87.194.216.51" productId="SC-MG-G10" | stats count by clientip`. To the right of the search bar are buttons for 'All time' and a search icon. Below the search bar, it shows '✓ 61 events (before 11/20/24 5:28:58.000 PM)' and a 'No Event Sampling' button. There are also buttons for 'Job', 'Pause', 'Refresh', 'Download', and 'Smart Mode'. Below this is a navigation bar with 'Events', 'Patterns', 'Statistics (1)', and 'Visualization'. Under 'Statistics (1)', there are buttons for '20 Per Page', 'Format', and 'Preview'. The results table has two columns: 'clientip' and 'count'. The first row shows '87.194.216.51' and '61'.

clientip	count
87.194.216.51	61

5.**head**- Retrieves the first N results from your search. This command is useful when you want to quickly examine the top records in your data.

By default, head returns the first 10 events, but you can specify any number of results.



The screenshot displays the Splunk Search & Reporting interface. At the top, there's a navigation bar with tabs for Search, Analytics, Datasets, Reports, Alerts, and Dashboards. The 'Search' tab is active. Below the navigation bar, the 'New Search' section is visible. A search bar contains the query: `index="test1" sourcetype=access_combined_wcookie clientip="87.194.216.51" productid |table sourcetype clientip productid | head 7`. To the right of the search bar, there are buttons for 'Save As', 'Create Table View', and 'Close'. Below the search bar, a status bar indicates '637 events (before 11/20/24 5:42:58.000 PM)' and 'No Event Sampling'. There are also buttons for 'Job', 'Pause', 'Refresh', 'Download', and 'Smart Mode'. Below the status bar, there are tabs for 'Events', 'Patterns', 'Statistics (7)', and 'Visualization'. The 'Statistics (7)' tab is selected. Below the tabs, there are buttons for '20 Per Page', 'Format', and 'Preview'. The main content area shows a table with 7 rows of data. The table has three columns: 'sourcetype', 'clientip', and 'productid'. The data is as follows:

sourcetype	clientip	productid
access_combined_wcookie	87.194.216.51	WC-SH-G04
access_combined_wcookie	87.194.216.51	WC-SH-T02
access_combined_wcookie	87.194.216.51	DC-SG-G02
access_combined_wcookie	87.194.216.51	SC-MG-G10
access_combined_wcookie	87.194.216.51	PZ-SG-G05
access_combined_wcookie	87.194.216.51	PZ-SG-G05
access_combined_wcookie	87.194.216.51	PZ-SG-G05

6 **Tail**- Retrieves the last N results from your search. It's commonly used to inspect the most recent data or see the tail end of a dataset.

By default, tail returns the last 10 events, but you can adjust the number of results.

New Search

Save As

Create Table View

Close

index="test1" sourcetype=access_combined_wcookie clientip="87.194.216.51"
productid |table sourcetype clientip productid | tail 5

All time

✓ 637 events (before 11/20/24 5:44:18.000 PM)

No Event Sampling

Job

Smart Mode

Events

Patterns

Statistics (5)

Visualization

20 Per Page

Format

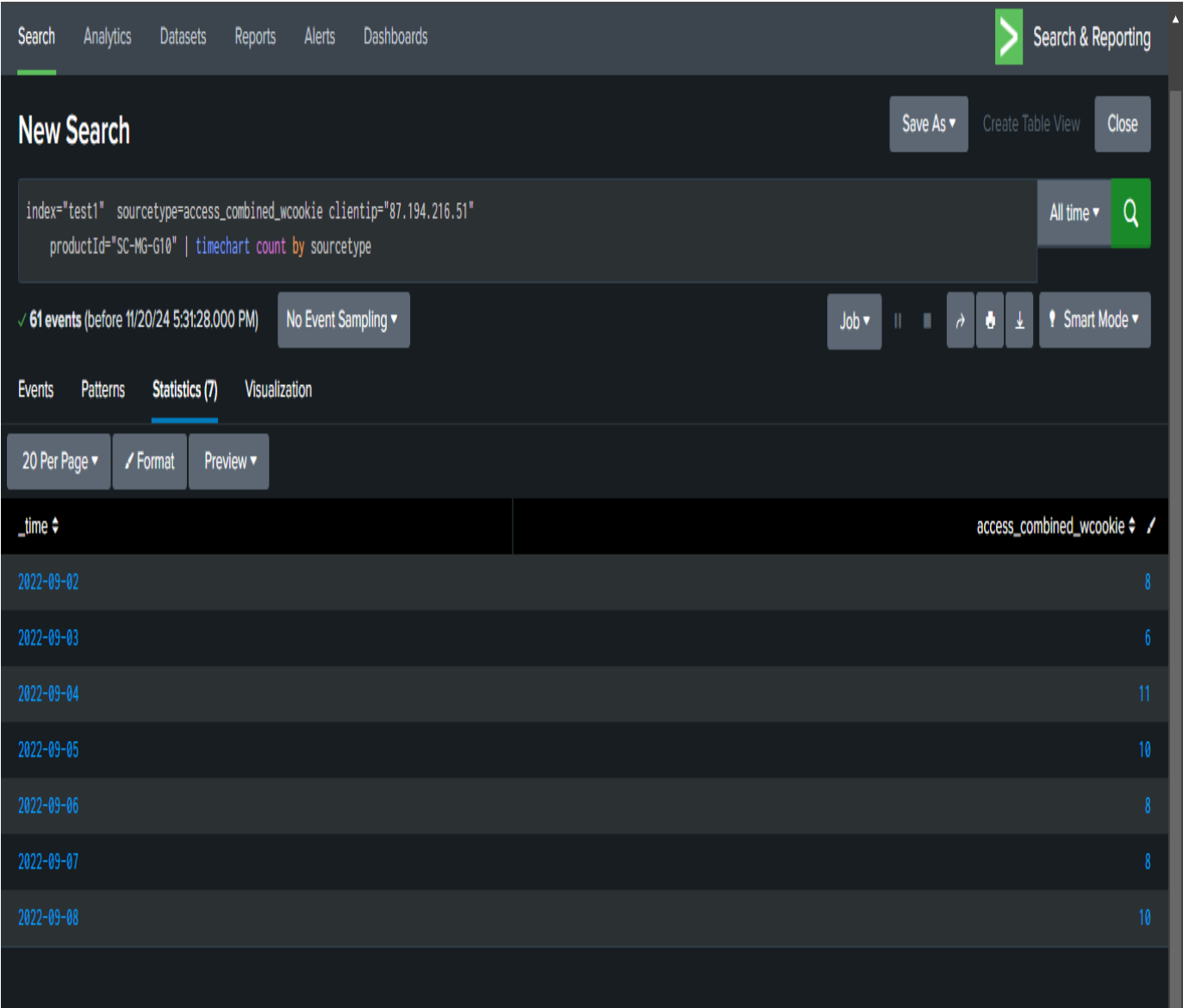
Preview

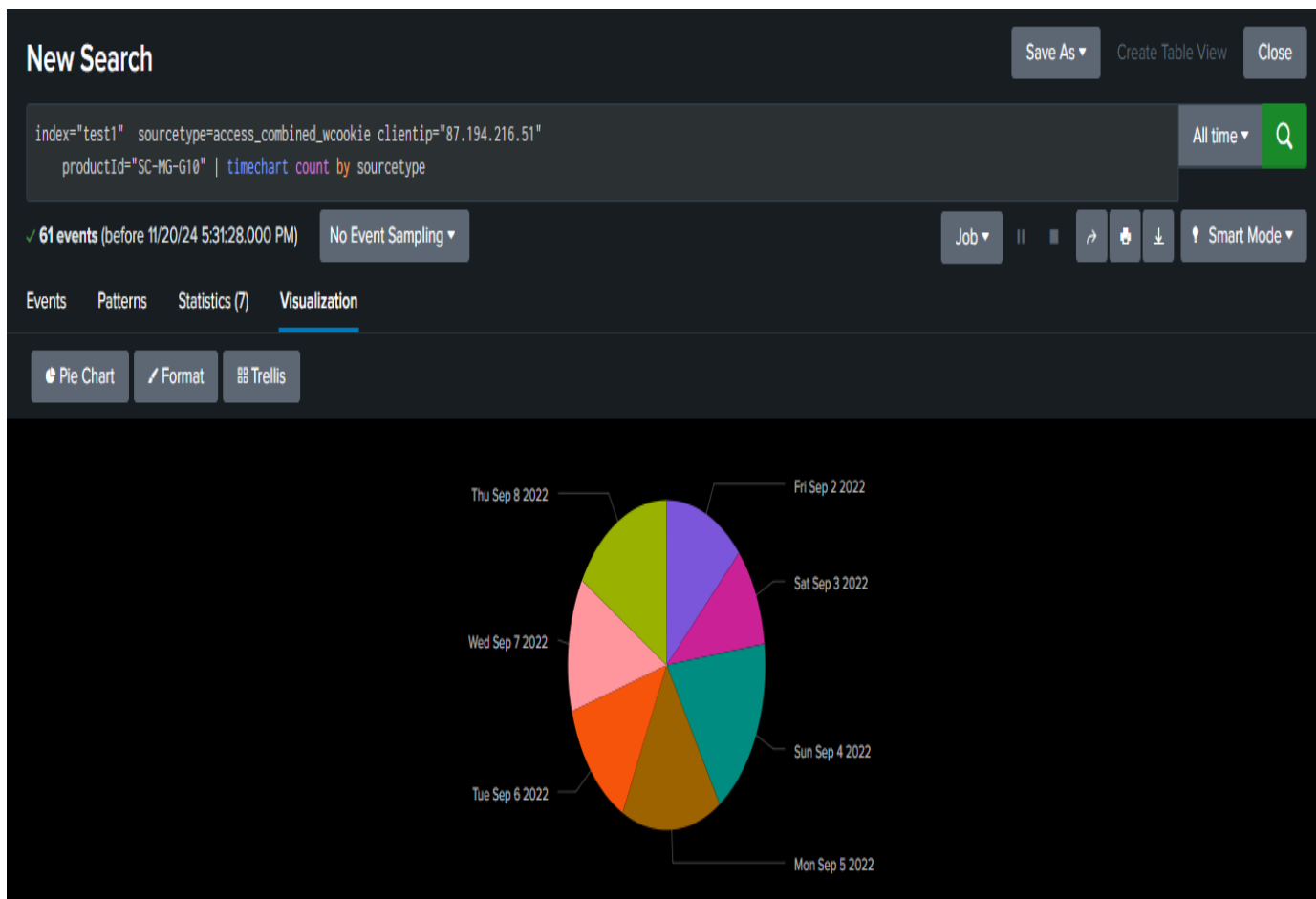
sourcetype	clientip	productid
access_combined_wcookie	87.194.216.51	DB-SG-G01
access_combined_wcookie	87.194.216.51	BS-AG-G09
access_combined_wcookie	87.194.216.51	WC-SH-A01
access_combined_wcookie	87.194.216.51	CU-PG-G06
access_combined_wcookie	87.194.216.51	WC-SH-T02

7 **timechart**- command in Splunk is used to create time-based visualizations from your event data.

It helps in analyzing trends and patterns over time by summarizing data points at specified time intervals.

This is particularly useful for monitoring logs, detecting anomalies, and visualizing metrics like event counts, averages, and sums.





In the above we can see a timechart visualization in form of a piechart.

Splunk's Search Processing Language (SPL) is a powerful tool for security analysts, allowing them to perform deep searches, analyze logs, and generate reports from large datasets. Here's a summary of its importance:

1. **Efficient Log Search and Filtering**

SPL allows security analysts to search through massive amounts of log data efficiently. Commands like `search` help filter relevant events, making it easier to identify suspicious activities or potential security incidents.

2. **Data Extraction and Analysis**

SPL enables analysts to extract meaningful information from raw log data using commands like *`stats`*, *`eval`*, and *`rex`*. For instance, *`stats`* aggregates data to summarize key metrics, and *`eval`* lets analysts create custom fields for better insights.

3. **Correlation and Alerts**

Analysts can use SPL to correlate events from multiple sources. For example, using commands like *`join`* or *`transaction`*, security analysts can link data from different logs

to detect patterns of attack, such as those seen in a multi-stage cyberattack.

4. **Real-Time Monitoring**

SPL supports real-time search and monitoring, helping analysts to detect threats as they occur.

Commands like *tail* and *streamstats* let analysts watch live data flows, crucial for identifying ongoing security incidents.

5. **Data Enrichment and Customization**

SPL allows analysts to enrich raw data with additional context, using commands such as *lookup* to map IP addresses to geolocation data or hostnames.

This enrichment enhances the analyst's ability to understand the severity and impact of an incident.

6. **Dashboards and Visualizations**

SPL helps create custom dashboards and visualizations for presenting security data, allowing security teams to monitor trends and identify anomalies visually.

Commands like *timechart* and *chart* are vital for creating time-based visualizations, which are key in tracking the evolution of attacks or detecting patterns over time.

7. **Automation of Response**

SPL can be integrated into automation workflows, enabling analysts to set up alerts and automated responses for predefined security thresholds.

This helps analysts react quickly to potential threats, improving overall incident response times.

8. **Tailored Investigations**

By combining SPL commands, security analysts can perform deep investigations into suspicious activities.

Whether it's looking at specific IP addresses, domains, or user behavior, SPL offers the flexibility to customize searches based on the unique needs of each investigation.

In summary, SPL is essential for security analysts because it helps them effectively manage, search, and analyze vast amounts of data, detect potential security issues, and respond to incidents quickly and efficiently. It enhances the ability to correlate data, create actionable insights, and automate responses, which is key in any modern security operation.