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

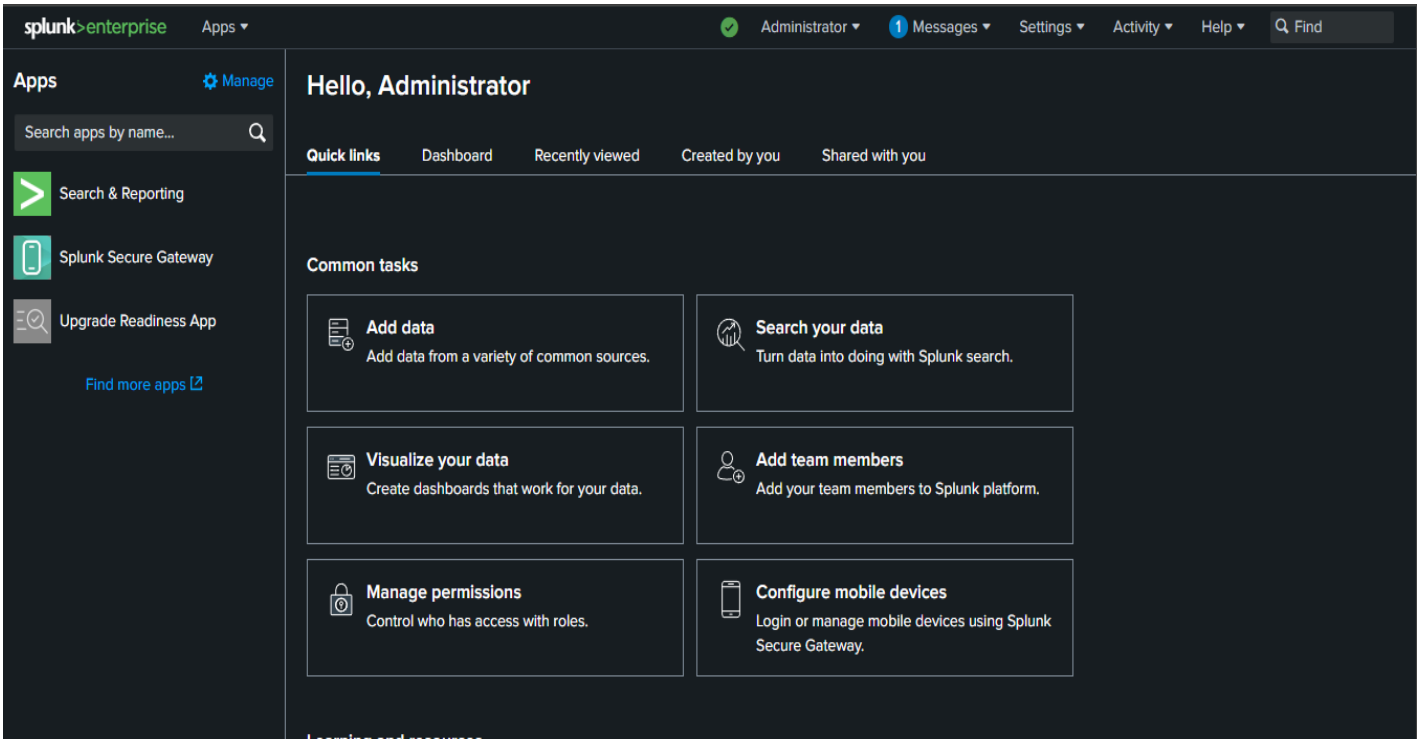
By retaining and indexing historical data, Splunk allows SOC teams to investigate breaches, uncover root causes, and implement preventive measures.

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



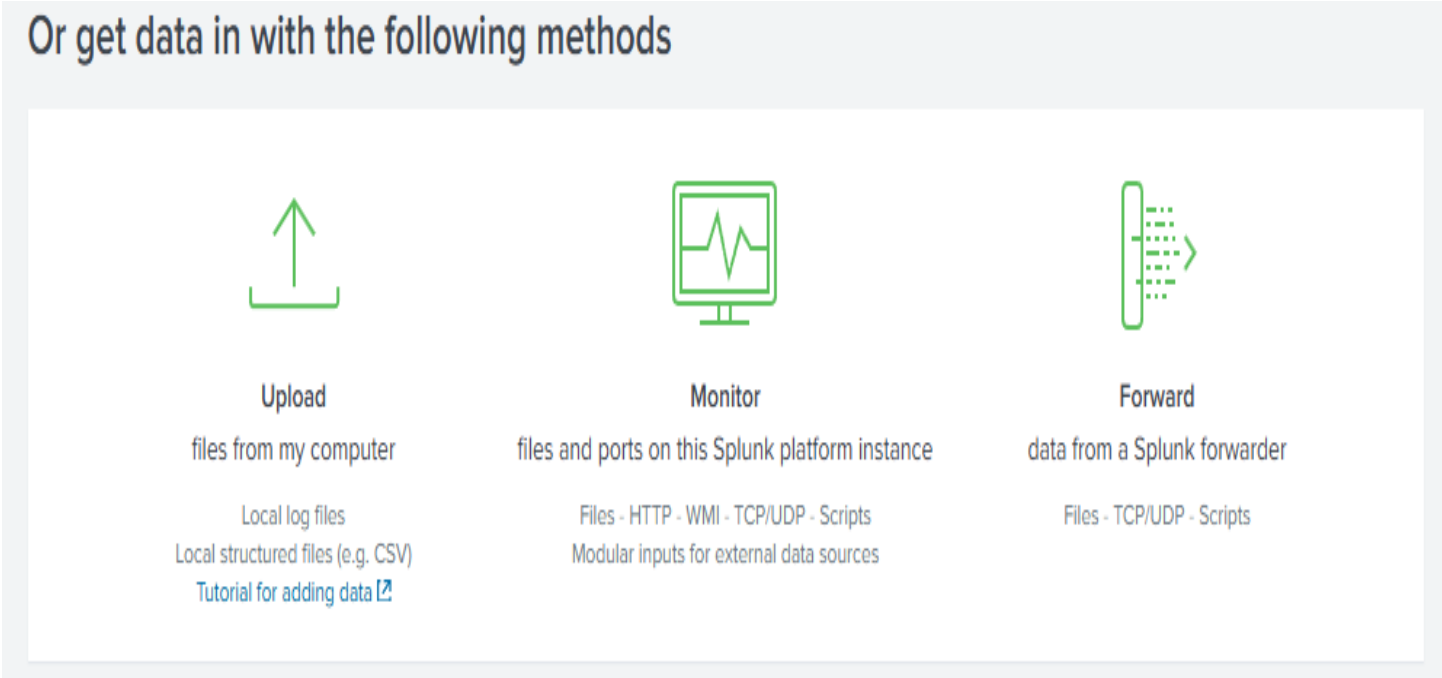
Quick Links Tab

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:



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

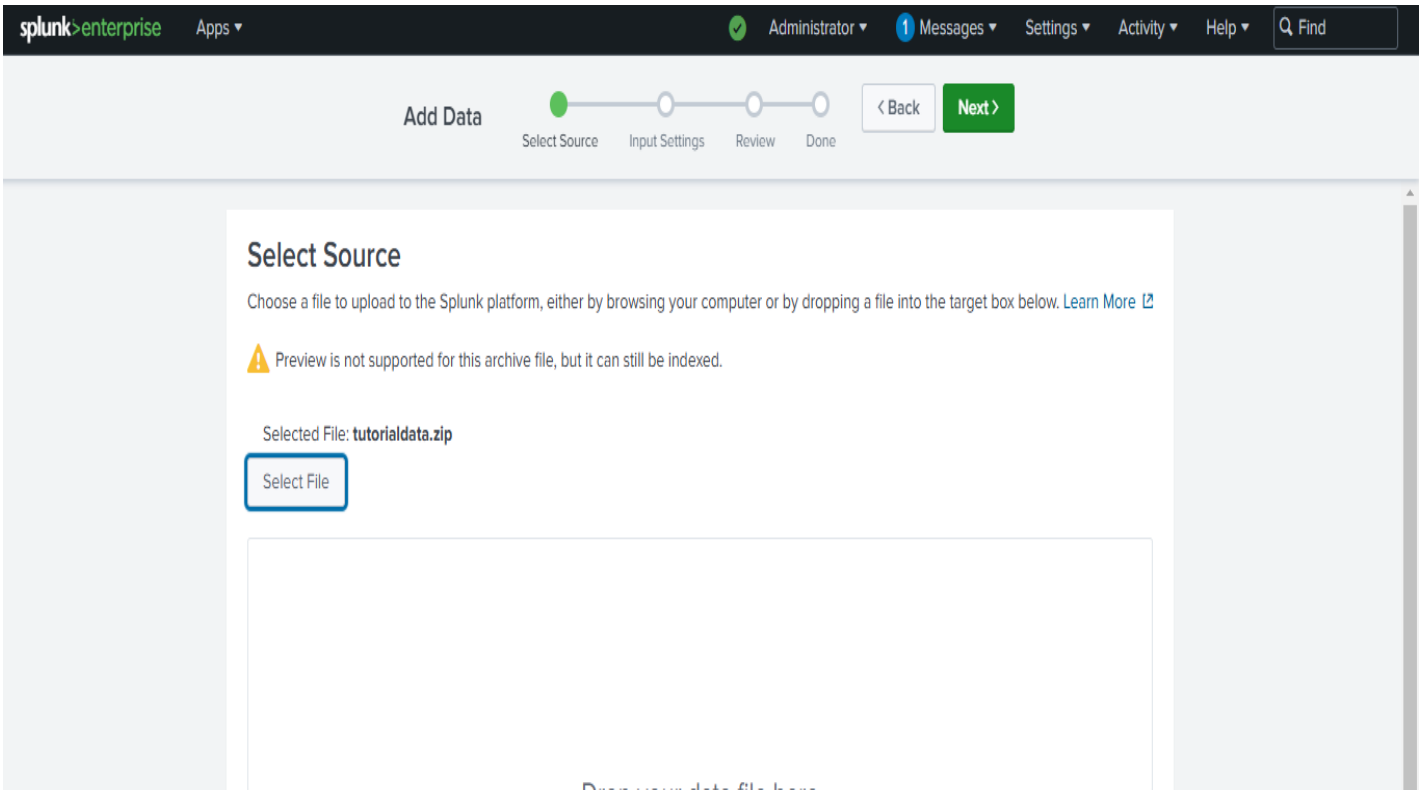
Monitor a Source: Specify a directory, file path, or network source to monitor.

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.



Input Settings

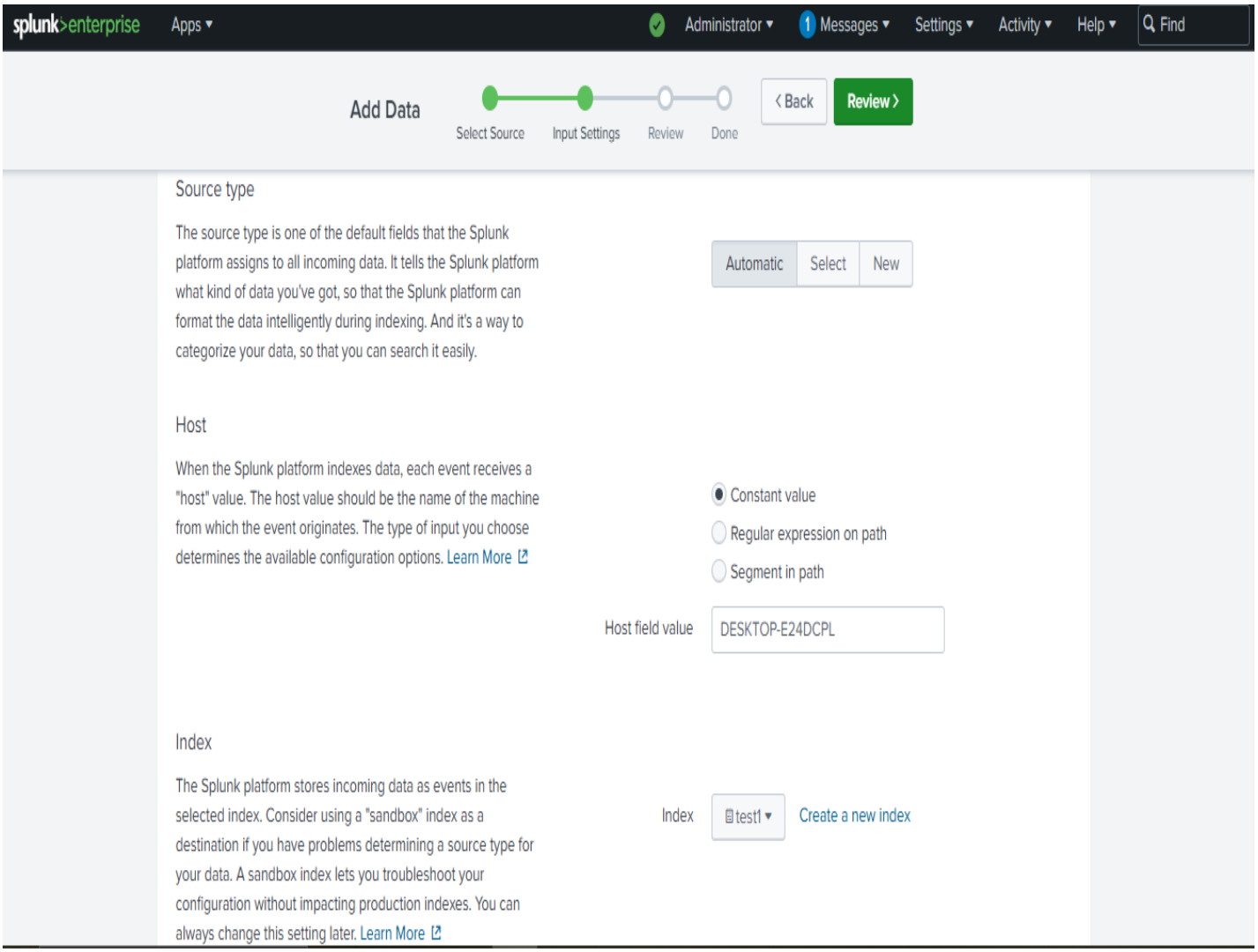
In this step, you define how Splunk processes the ingested data.

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.



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).

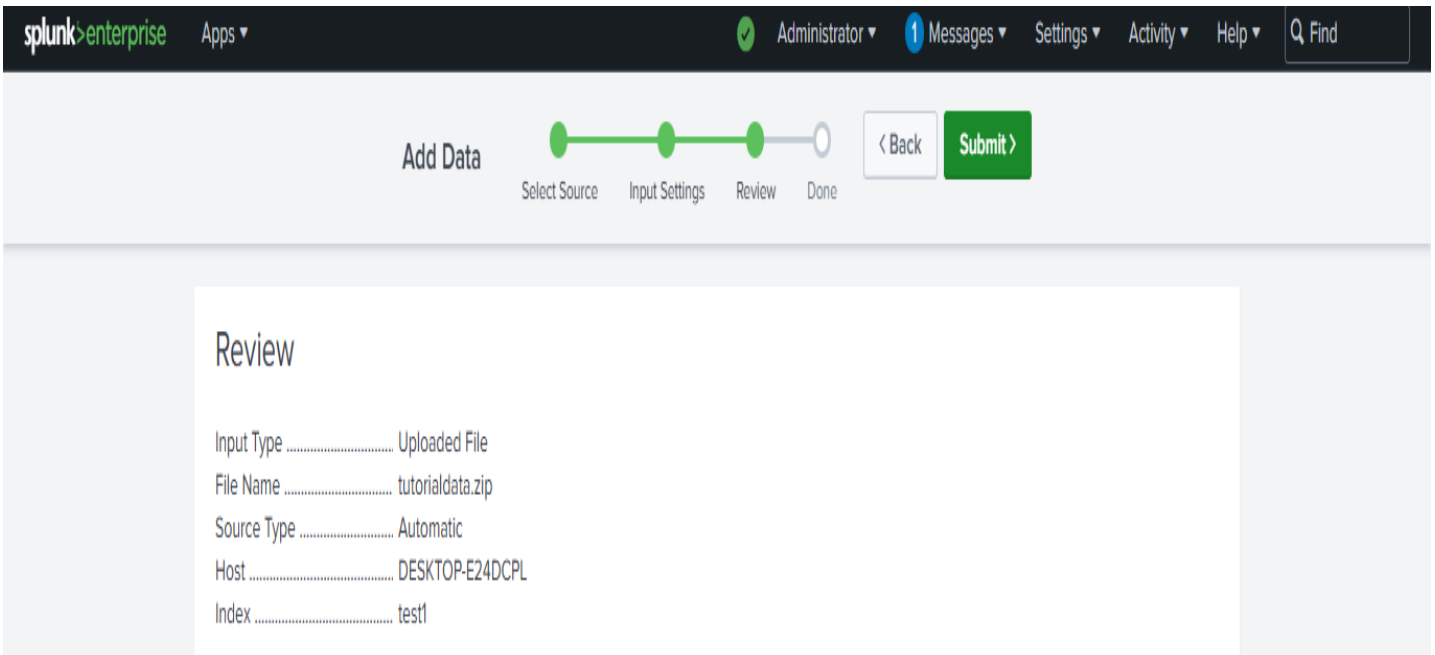
Set Index

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.



Review

The final step before ingestion.

Verify Settings:

Confirm your configurations:

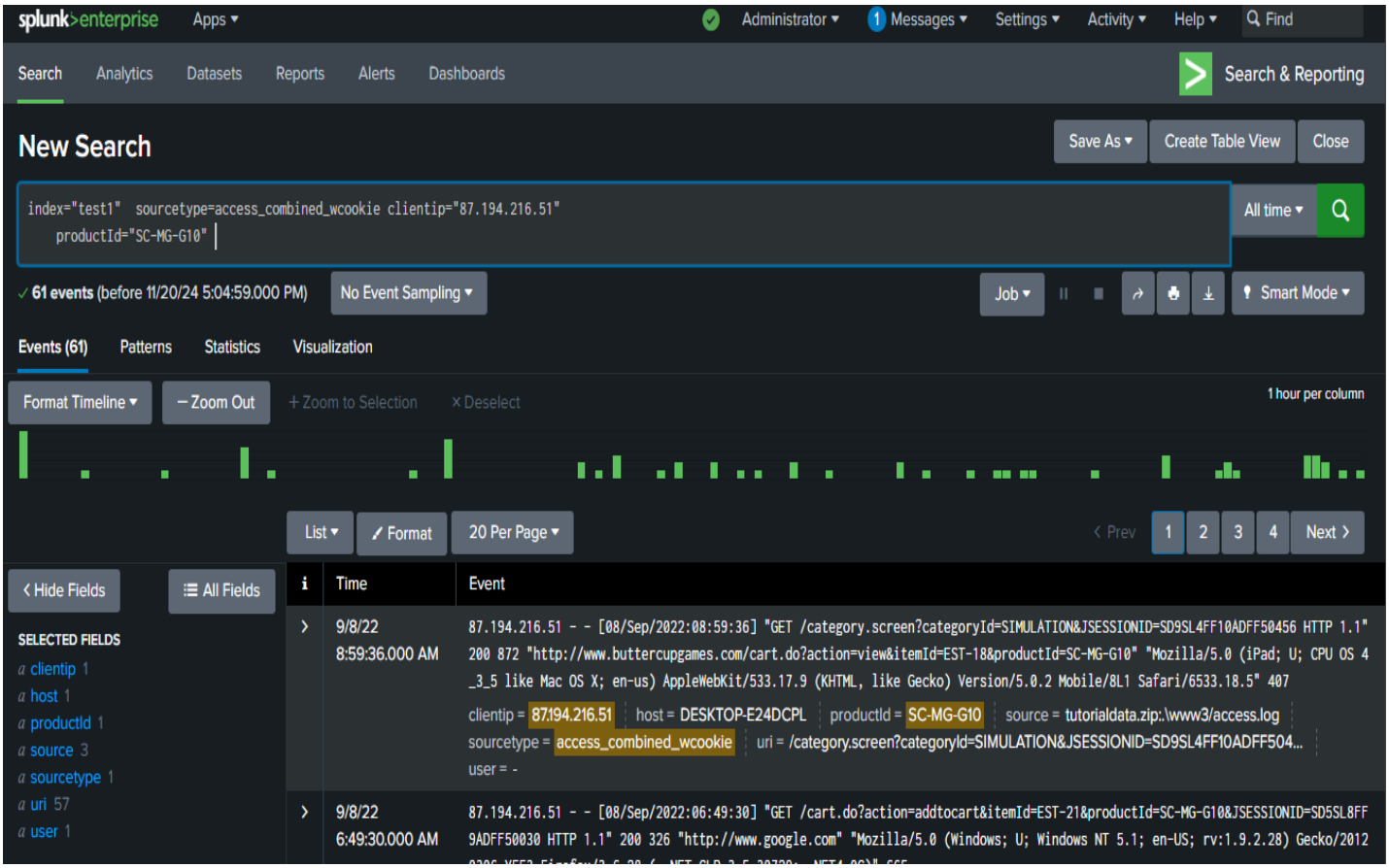
Selected data source.

Source type.

Index and host settings.

Click Finish or Submit to start data ingestion.

Splunk Search and Reporting



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

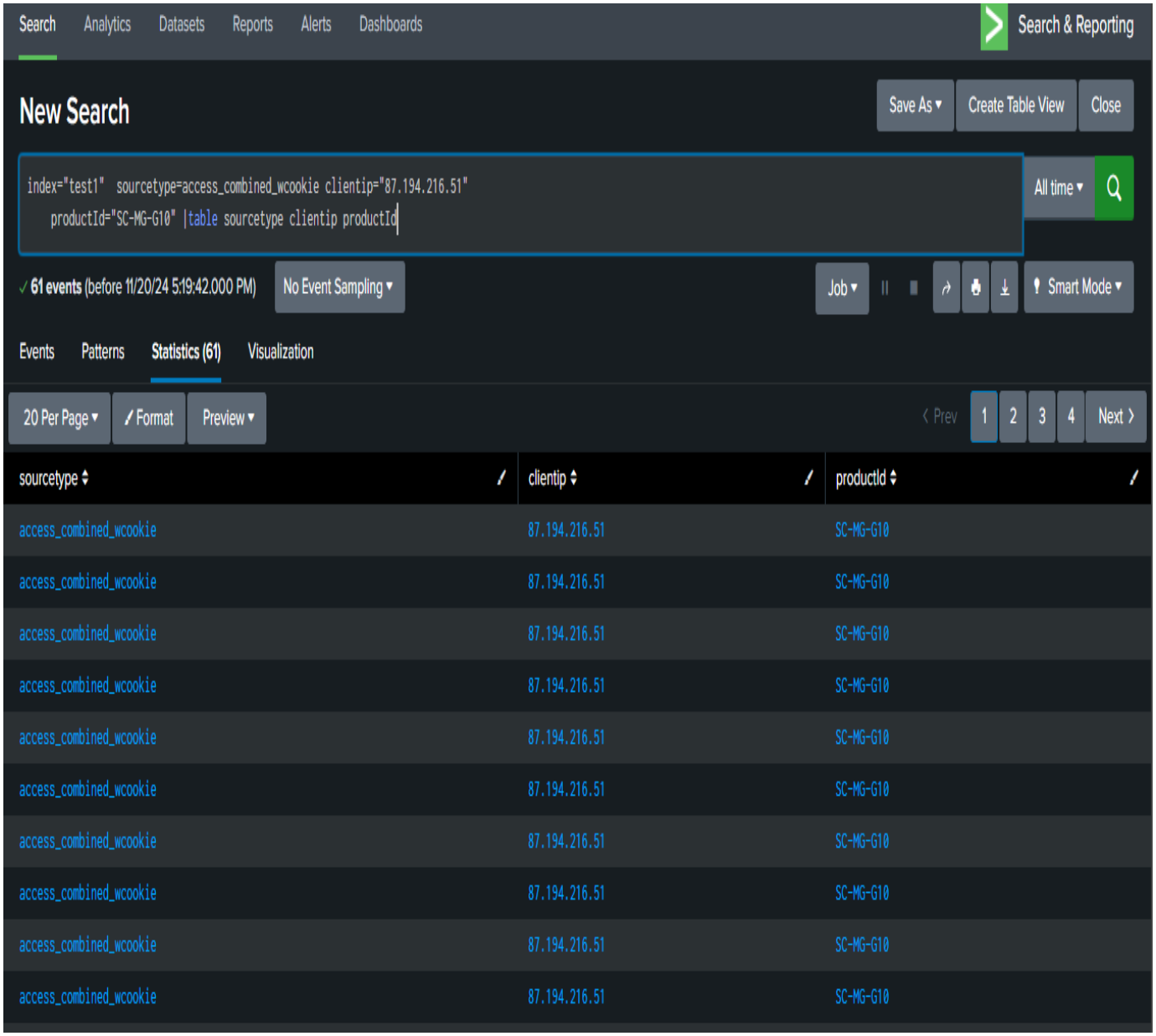
index="test1" sourcetype="access_combined" | stats count by clientip

- 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.

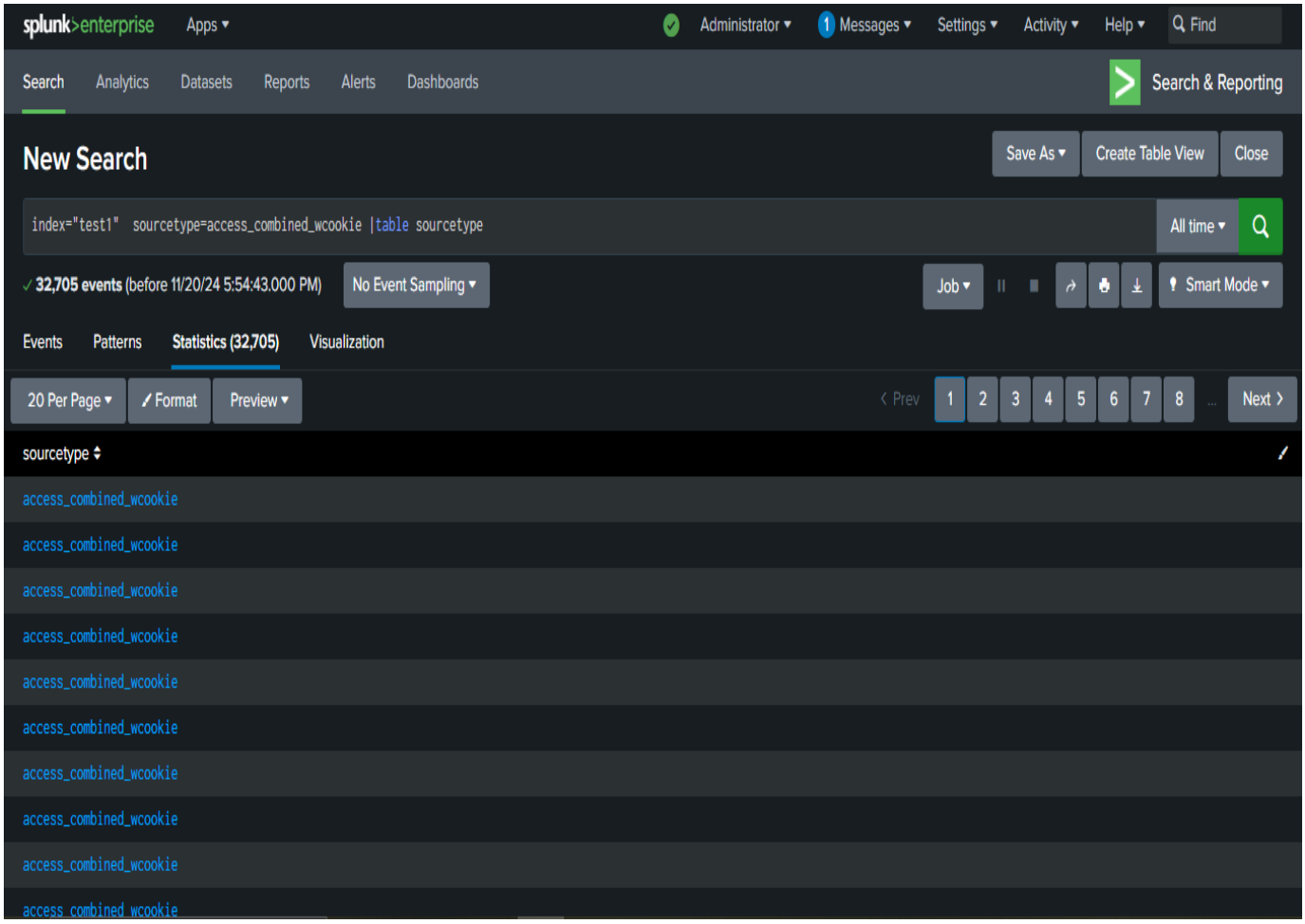
Basic Query Commands in Splunk

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.



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.

Search Analytics Datasets Reports Alerts Dashboards

Search & Reporting

New Search

Save As ▾ Create Table View Close

```
index="test1" sourcetype=access_combined_wcookie |table sourcetype|dedup sourcetype
```

All time ▾

✓ **32,705 events** (before 11/20/24 5:53:02.000 PM) No Event Sampling ▾ Job ▾ Smart Mode ▾

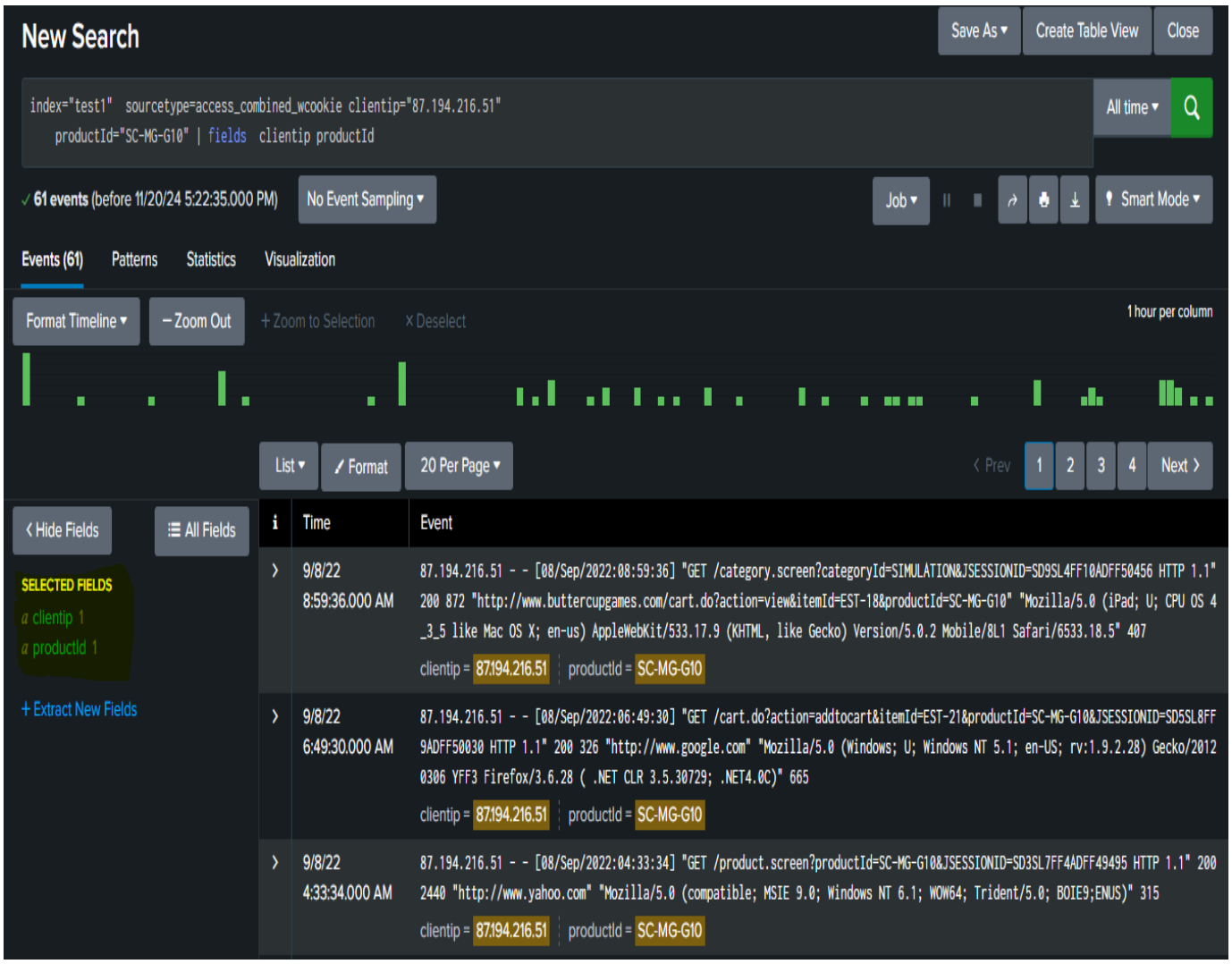
Events Patterns **Statistics (1)** Visualization

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.

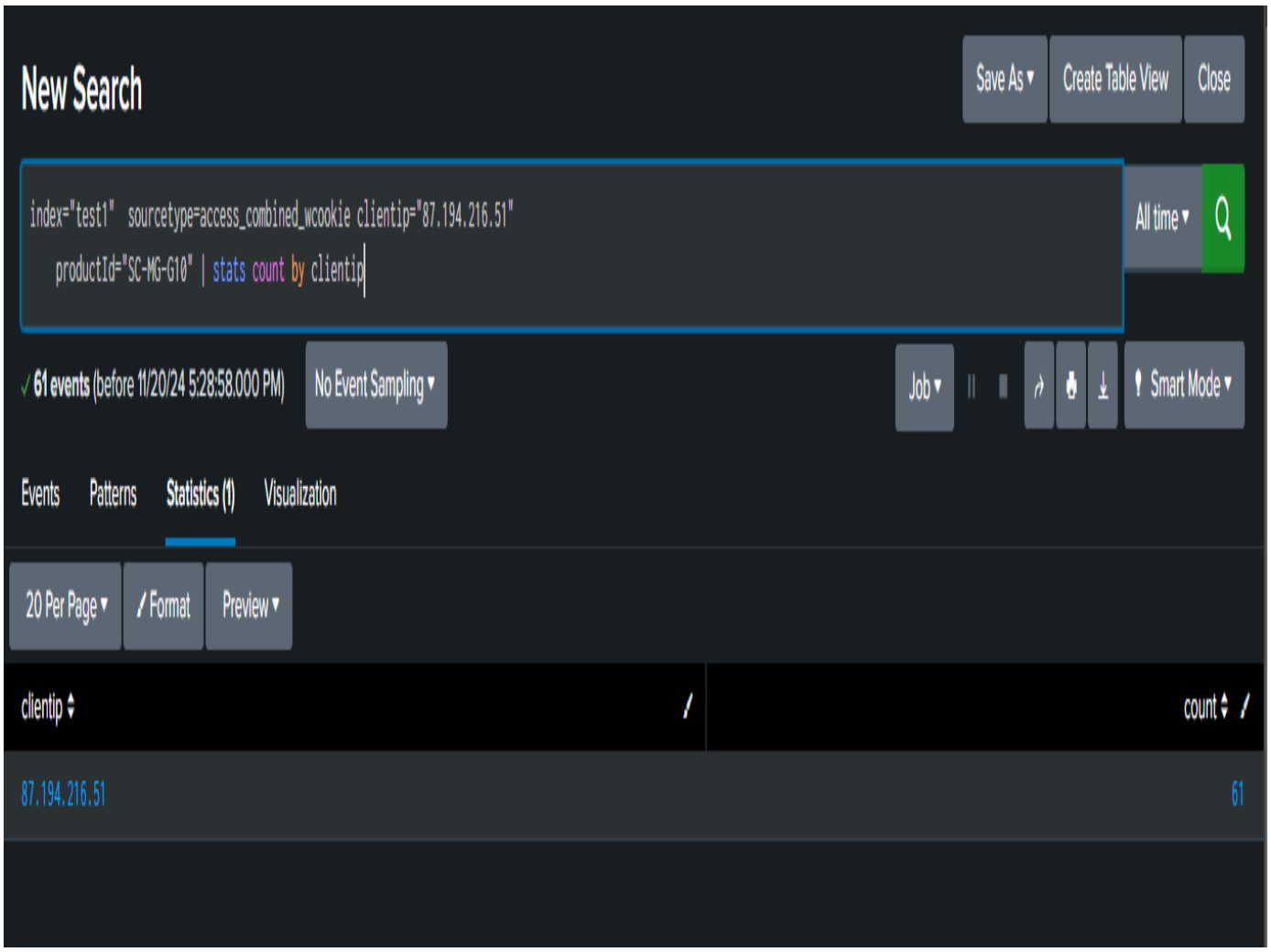
You can use fields to list the fields you want to keep (or exclude), improving performance and clarity in search results.



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.

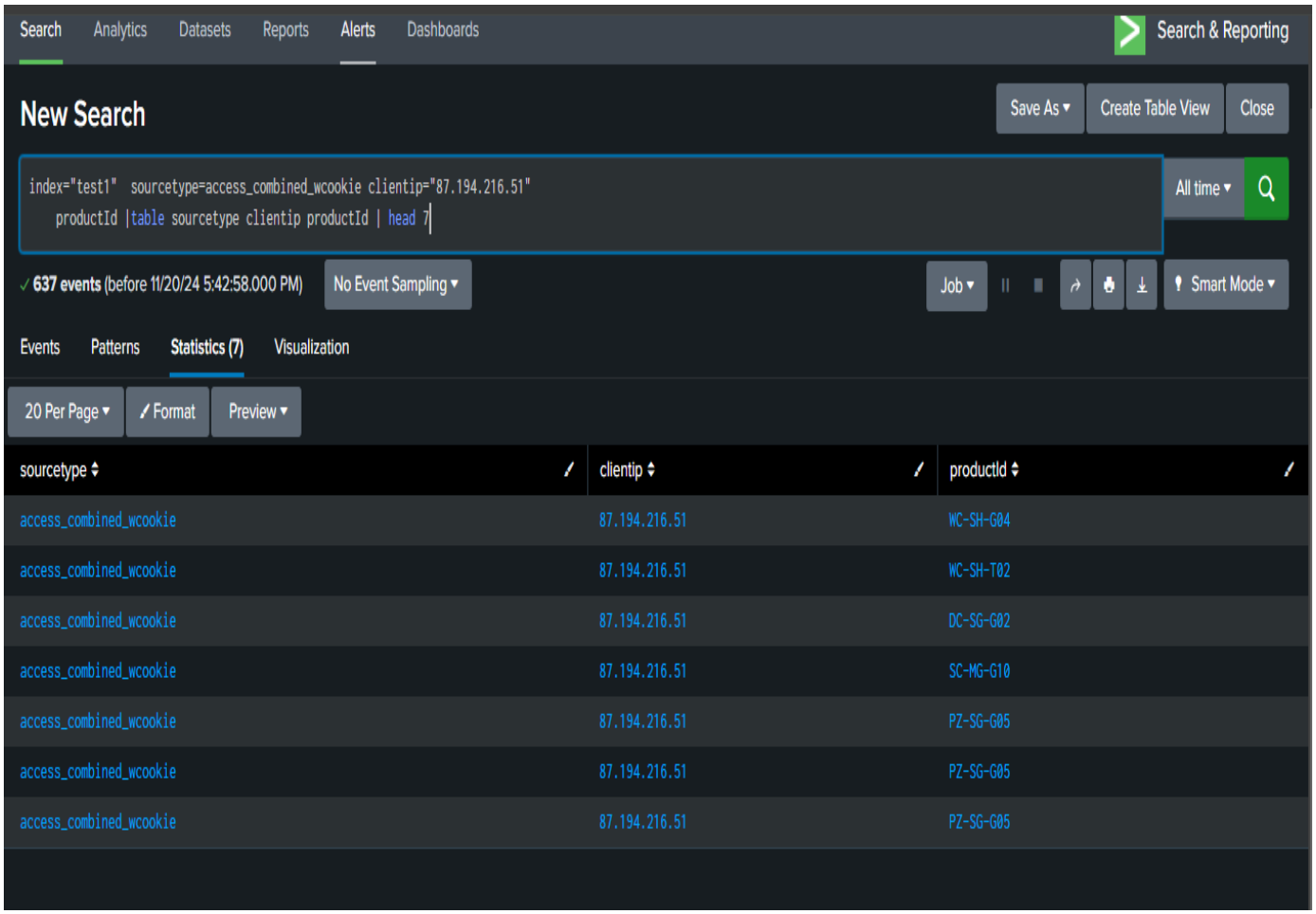
4 **Stats-** The stats command in Splunk is used to calculate statistics over your data, such as counts, sums, averages, etc.

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



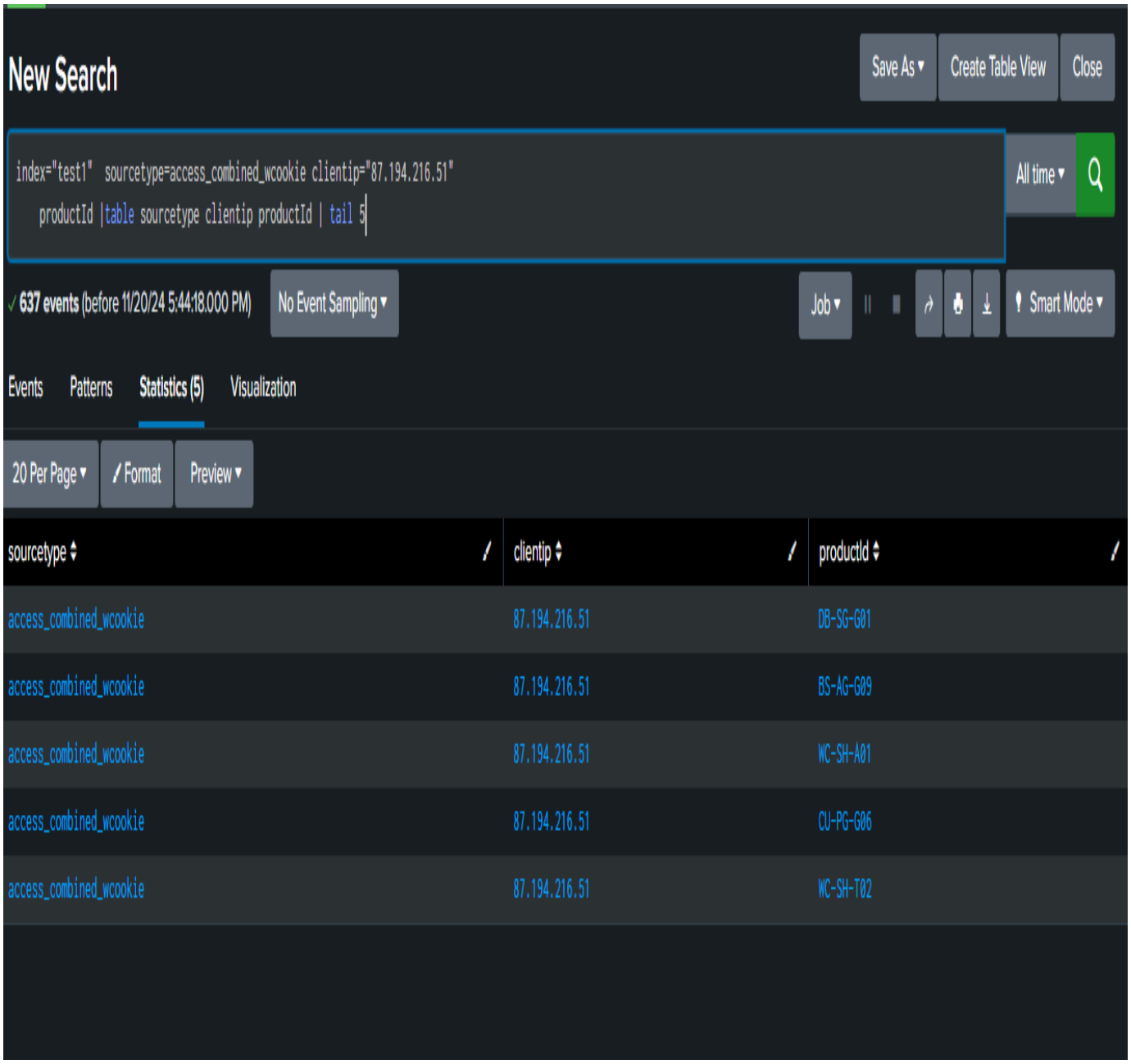
5.head- Retrieves the first N results from your search. This command is useful when you want to quickly examine the top records in your data.

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



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.

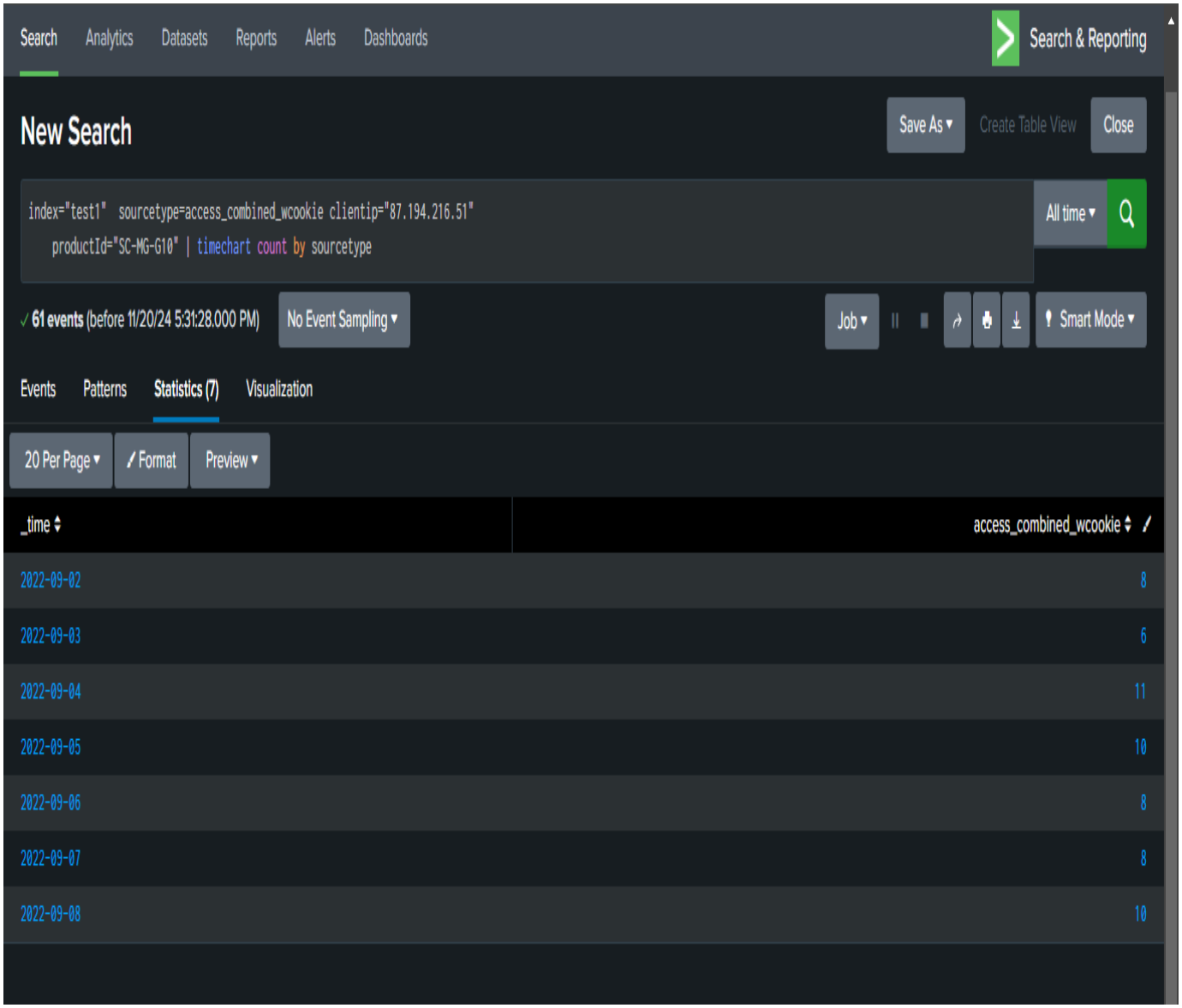
By default, tail returns the last 10 events, but you can adjust the number of results.

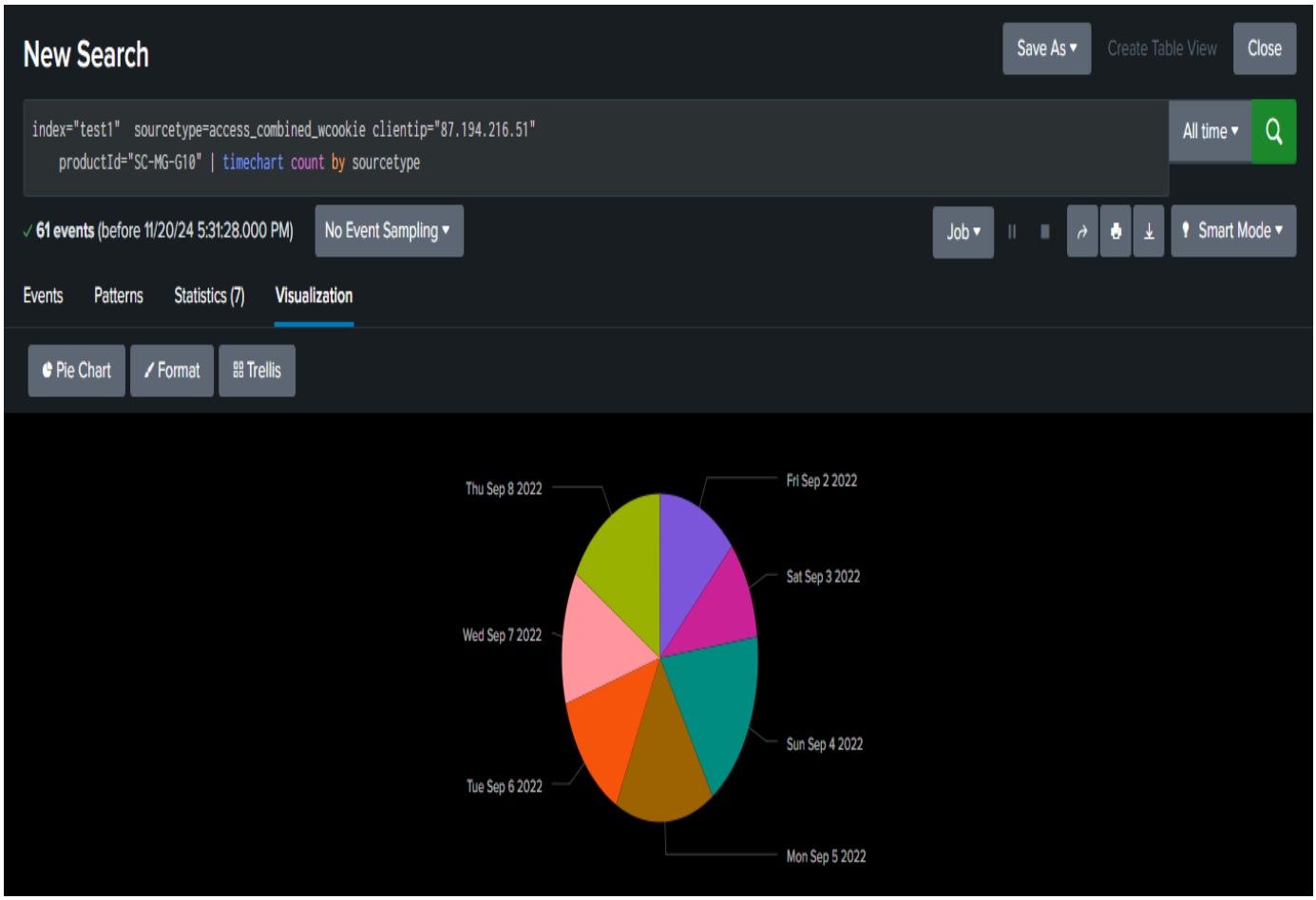


7 **timechart**- command in Splunk is used to create time-based visualizations from your event data.

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.