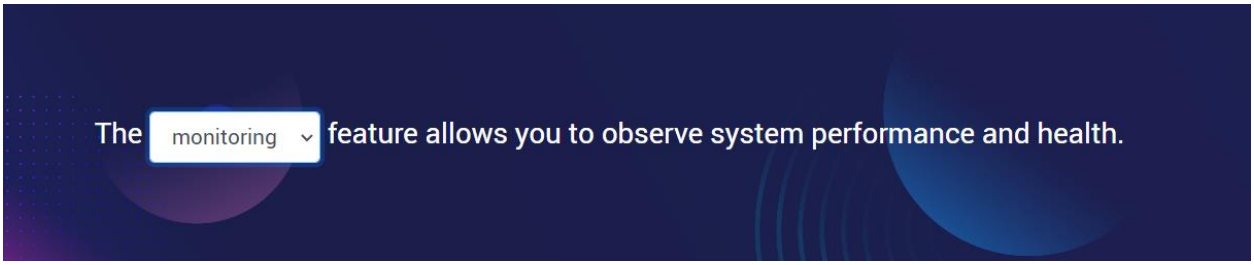
Assignment-3

1) What is Splunk?

Splunk is a versatile software tool designed for searching, monitoring, analyzing, and visualizing data through a web-based interface. It efficiently captures and organizes data into events based on source, time, and date, allowing for precise filtering and searches tailored to organizational needs. Splunk accommodates both structured and unstructured data, offering a powerful solution for various industries. Its applications span alerting, reporting, investigation, data visualization, operational intelligence, and even machine learning. From finance and marketing analysis to healthcare analytics, Splunk proves invaluable for understanding consumer habits, market trends, securing patient data, and real-time threat detection. The tool's core features include indexing, robust searching capabilities, customizable dashboards for visual representation, and proactive monitoring with automated alerting. Splunk's adaptability makes it a go-to solution for organizations seeking comprehensive data analytics.



Splunk is a useful tool that can help with detection of potential threats.

Search results can be narrowed down by specifying a time frame for the searched events.

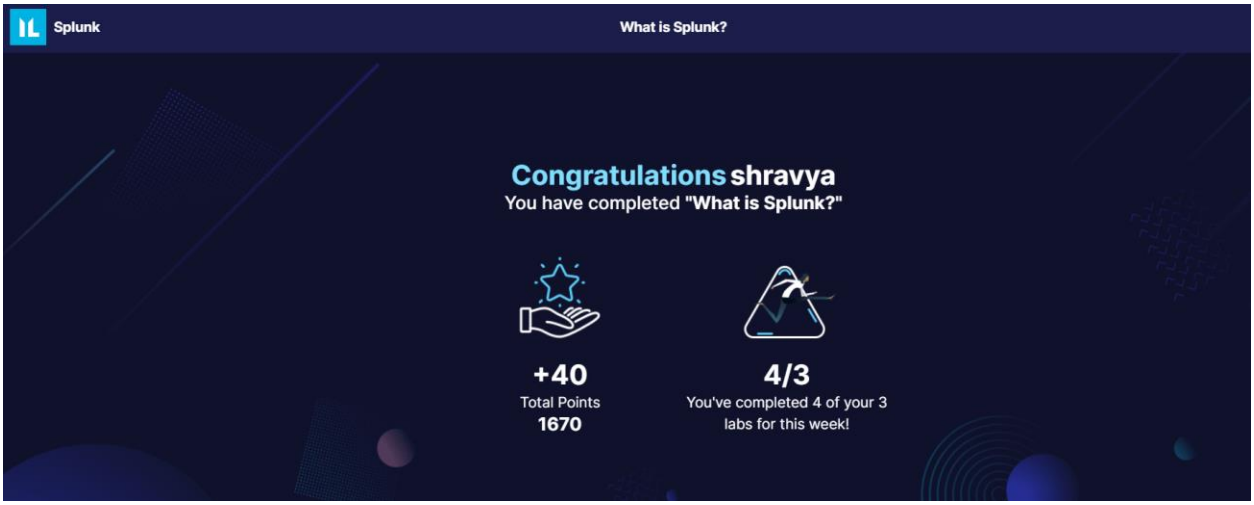
Splunk can process both struct data derived from devices.

Alerts can be configured to automatically respond to specific actions.

Splunk organises data into events based on the data source with the appropriate time and date.

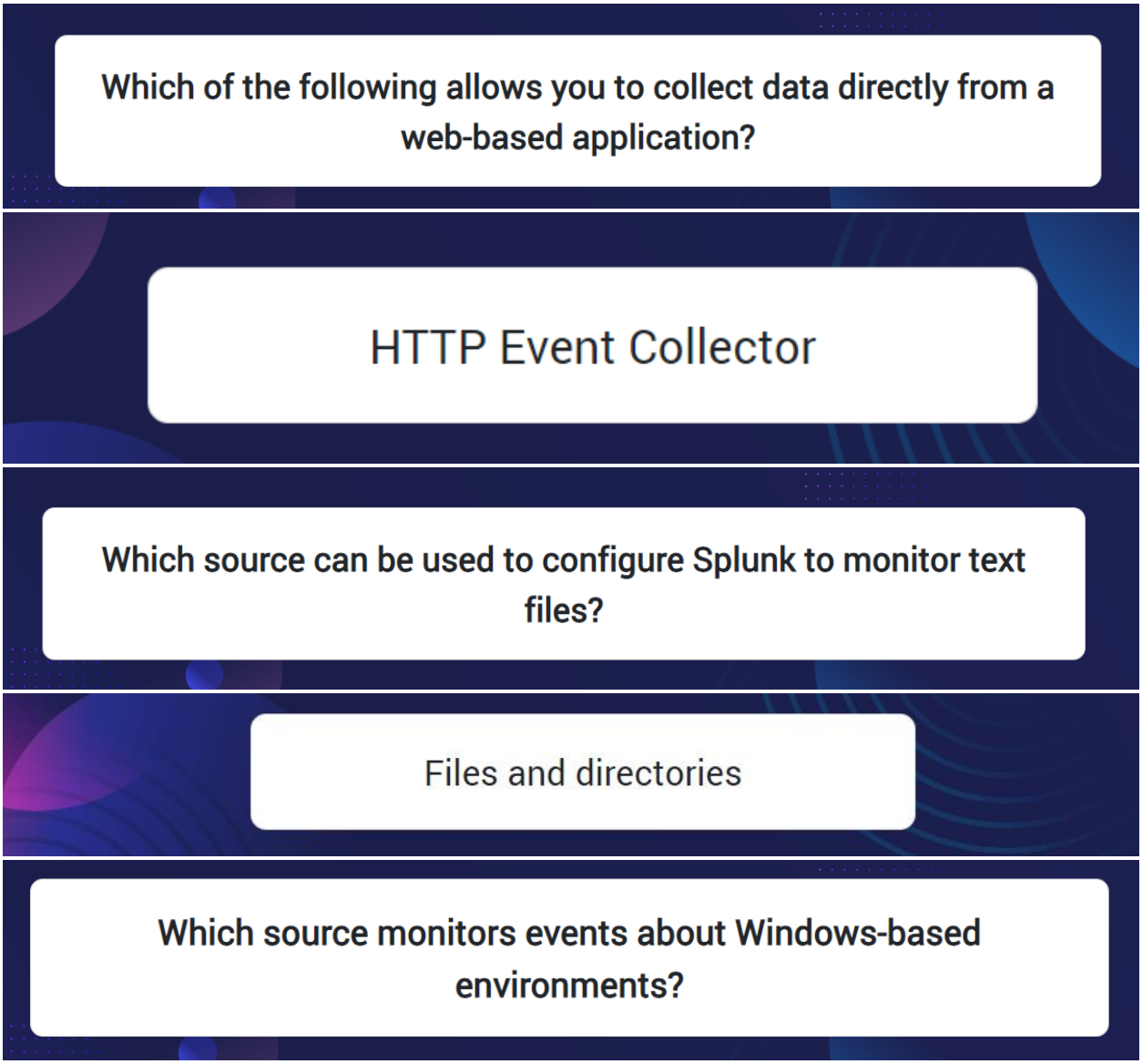
The dashboards and visualisatic features convey data in the form of graphs, charts, and tables.

Splunk offers the option to search and analyse data from a centralised repository.



2) Splunk Basics Ep-2: Data Sources

Splunk efficiently manages data from diverse organizational sources, including files, network traffic, scripts, Windows event logs, and HTTP applications. It indexes data based on source, automatically determining source types for formatting and indexing. Source types are crucial, defining how Splunk organizes data into events. For instance, a web server log generates fields like errors, logins, and authentication. Source and source type, though easily confused, differ in that the former is the data's file name, while the latter dictates data formatting. Understanding these distinctions enhances administration, filtering, and troubleshooting, facilitating efficient data search and threat identification for more streamlined organizational processes.



Windows event logs

Which user role can add data to Splunk?

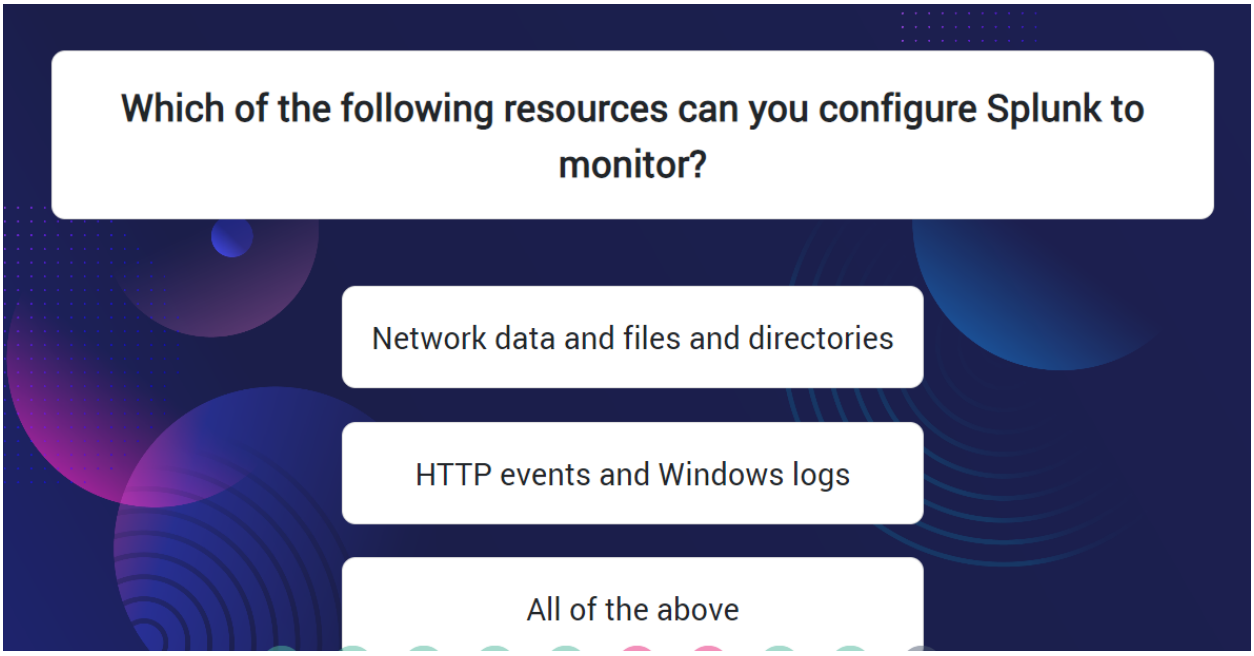
Administrator

If enough data is present, which of the following will Splunk automatically determine?

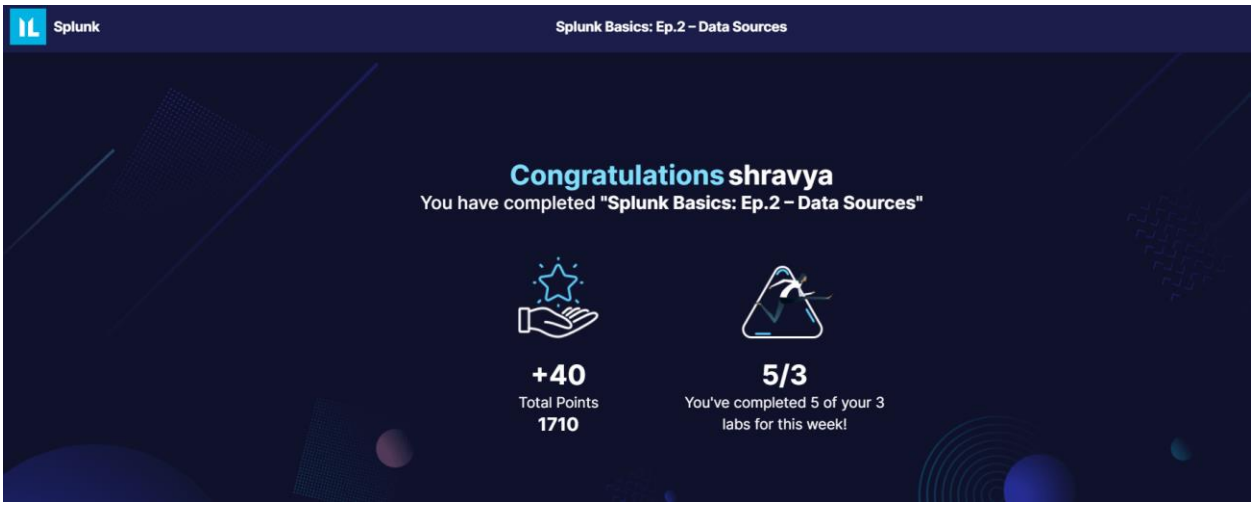
Source type

In Splunk terms, “source” indicates which of the following?

The file that the data has originated from



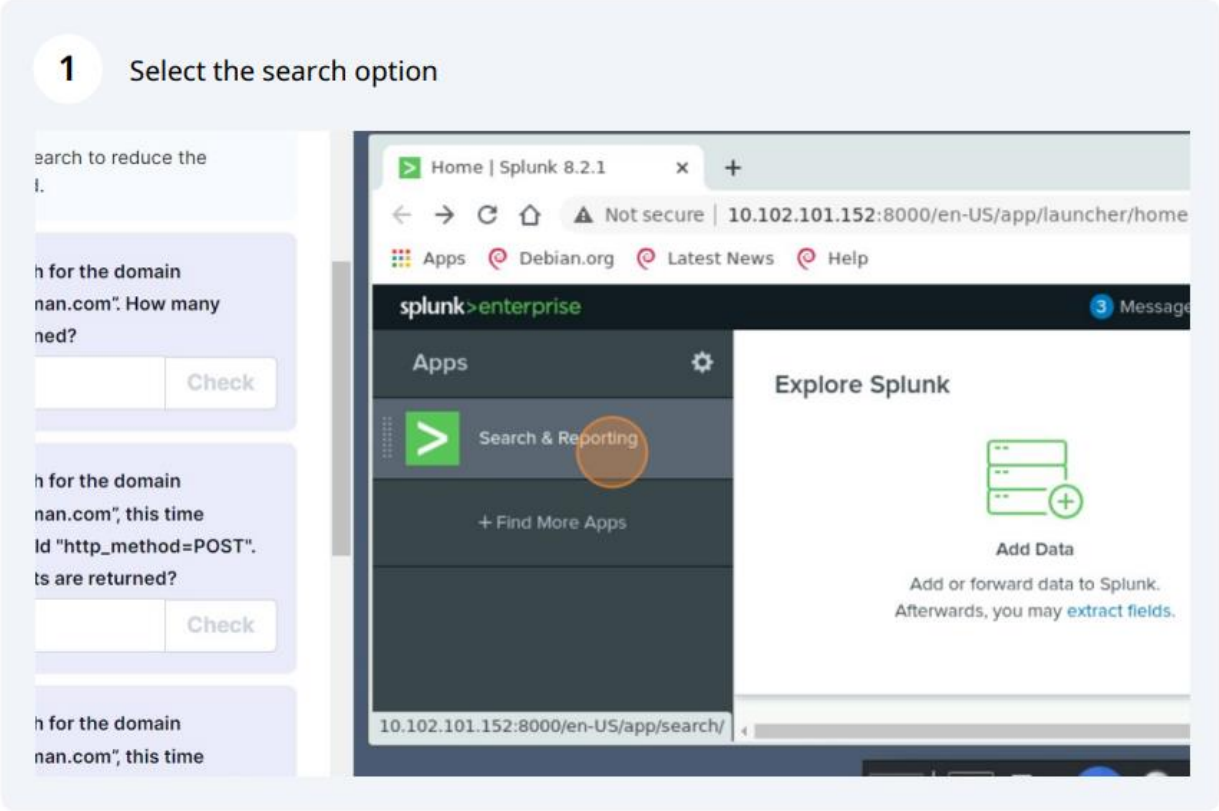
All of the above

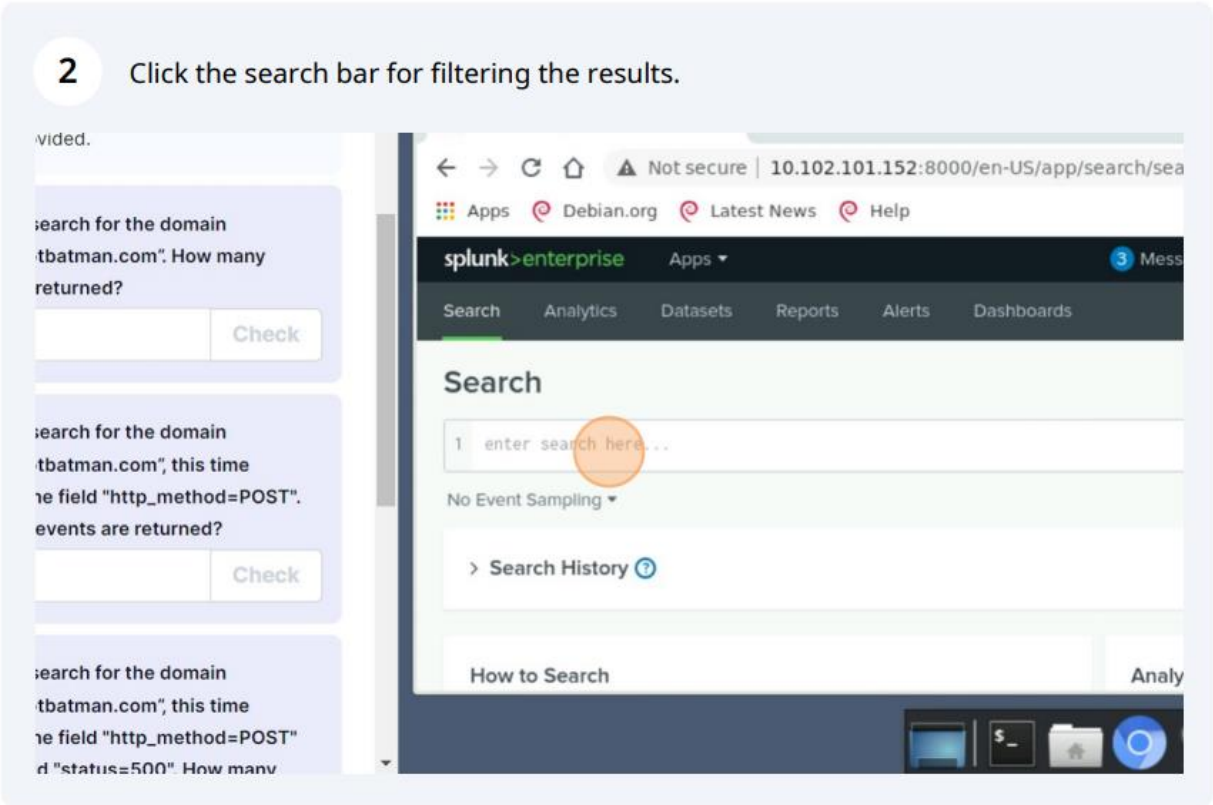


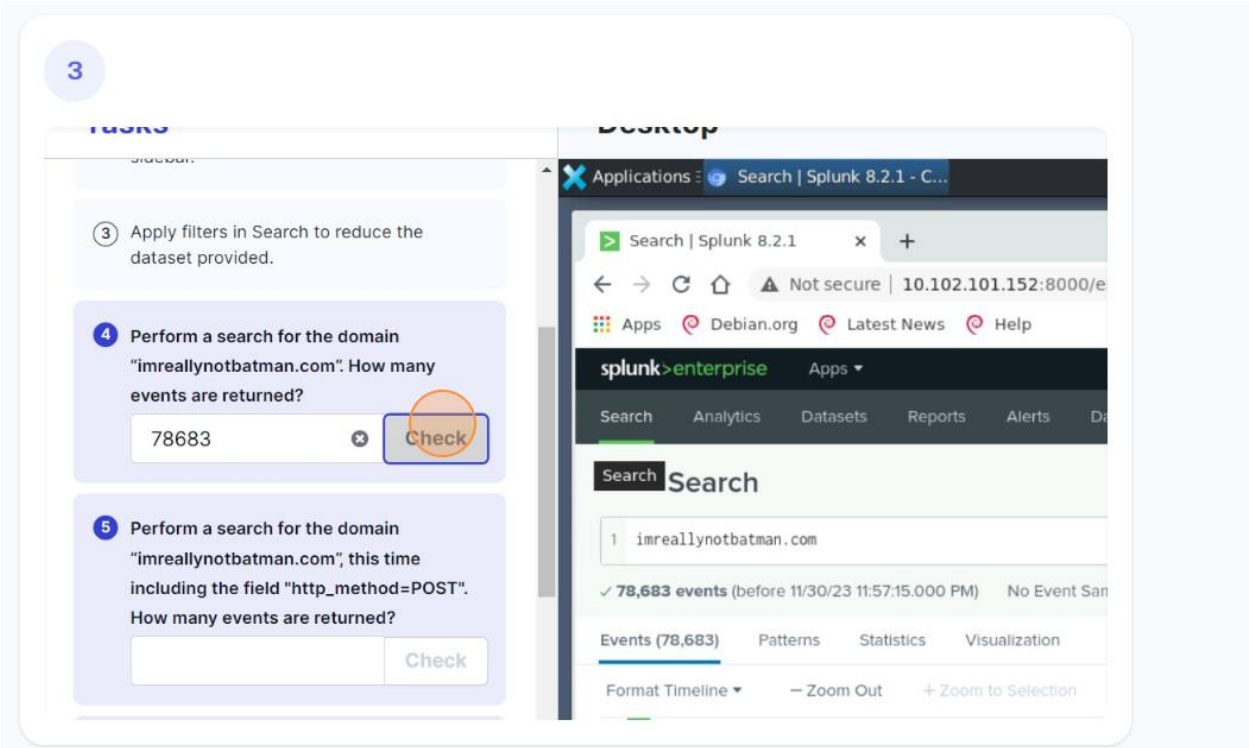
3) Splunk Basics Ep-3: Search

Splunk empowers users to conduct free-form searches and explore log files, raw events, and statistical data using its Search Processing Language (SPL). SPL transforms data

into charts and visualizations for efficient analysis. The search pipeline, composed of consecutive commands separated by pipes, enables refined searches by using output as input for subsequent commands. Fields, key/value pairs specific to ingested event data, aid in data filtering. Quotes and backslashes are utilized to handle special characters and spaces in search queries. Boolean expressions and wildcards enhance search flexibility. The Fields sidebar and comparison operators facilitate precise searches, while the Events Viewer displays raw event data for in-depth analysis.

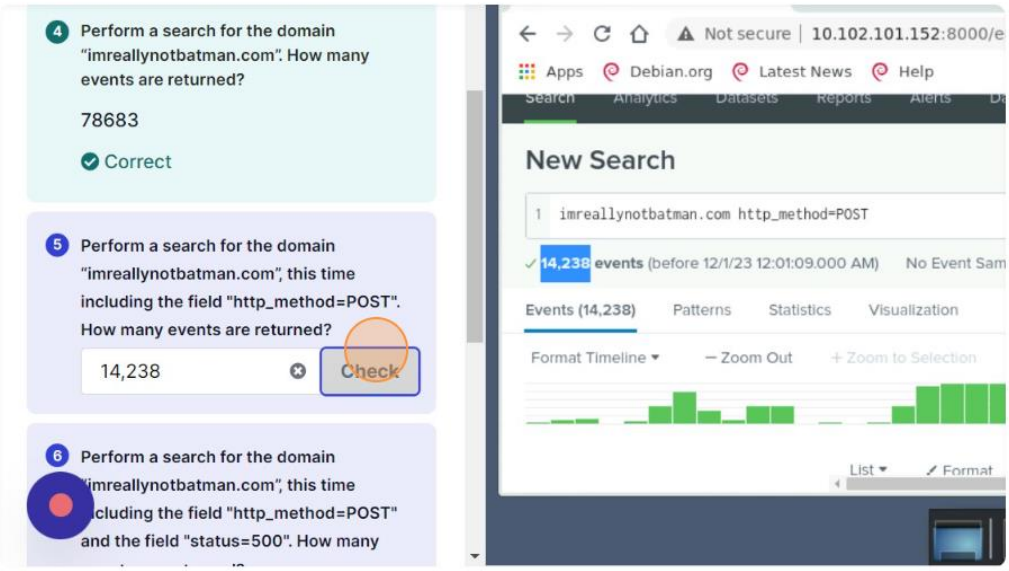




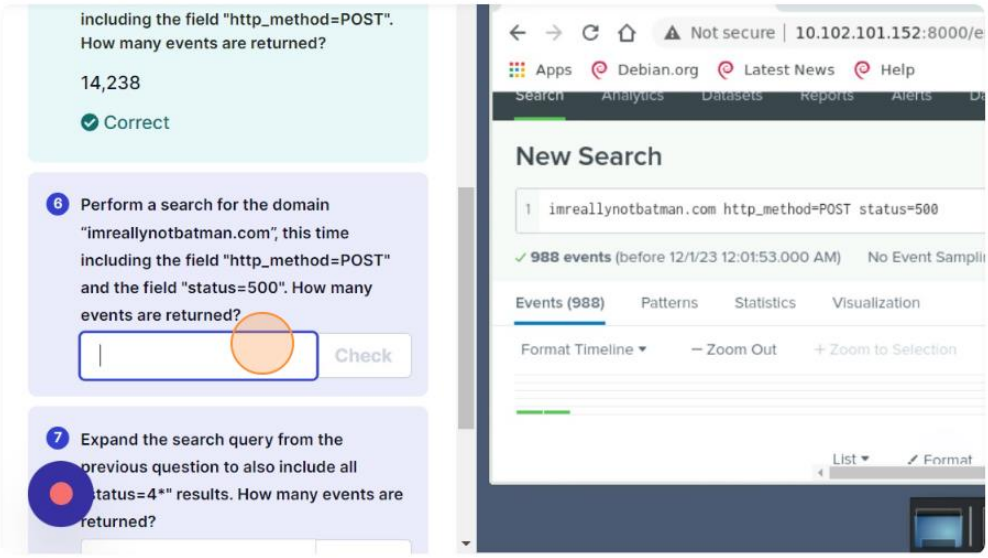


Performing a domain search using the search bar and domain name as mentioned above gives the required number of events.

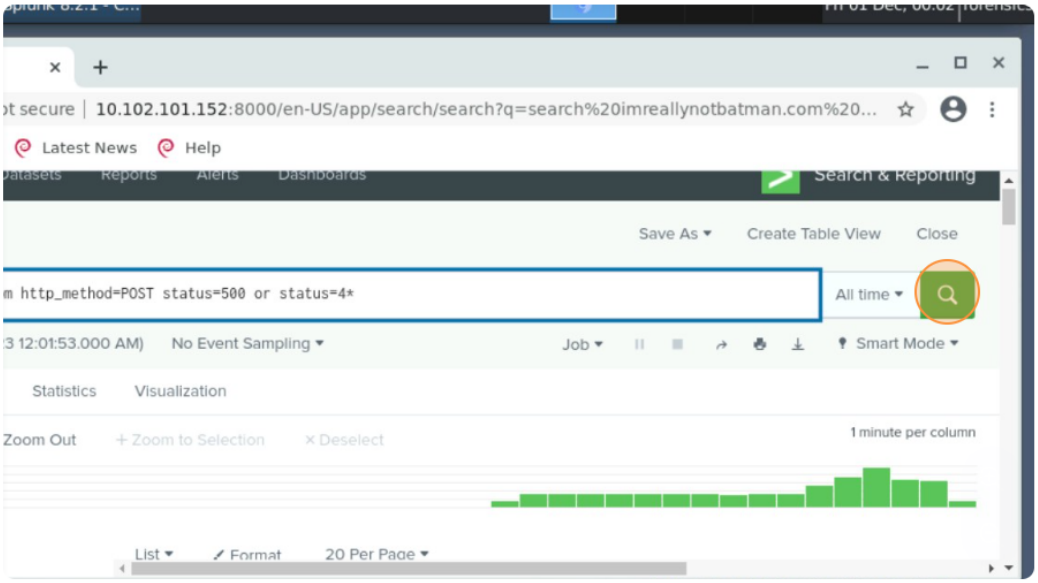
4 "Perform a search for the domain " imreallynotbatman.com ", this time including the field "http_method=POST". Returns 14,238



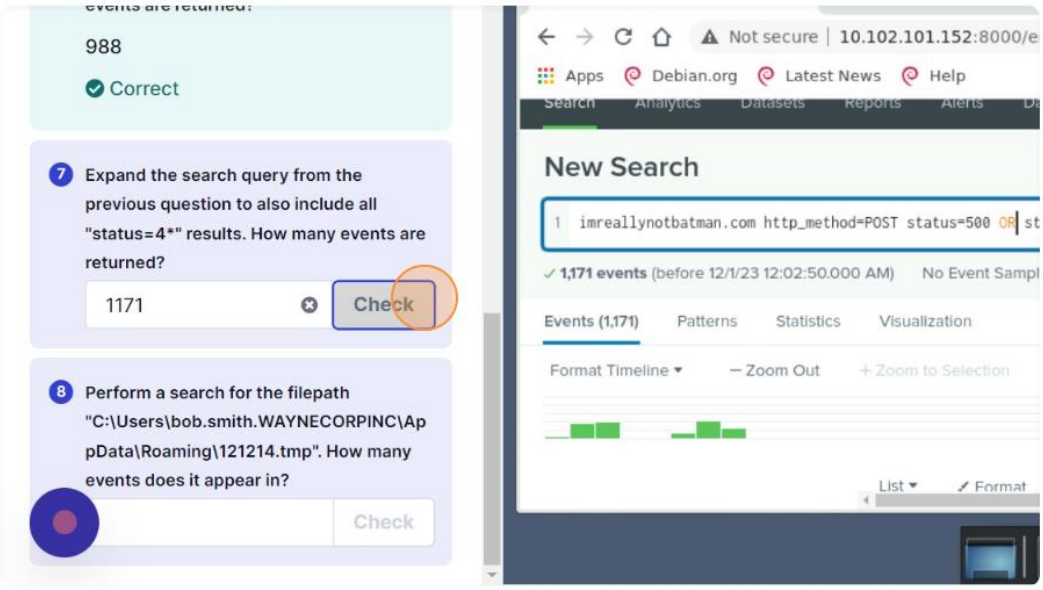
5 "Perform a search for the domain " imreallynotbatman.com ", this time including the field "http_method=POST" and the field "status=500". field.



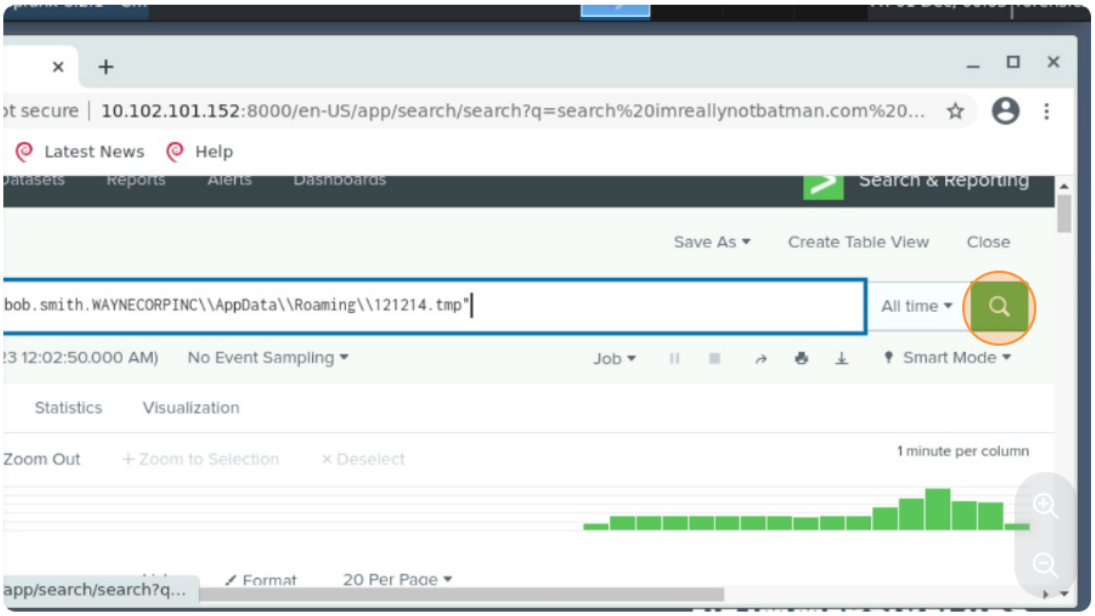
6 Expand the search query from the previous question to also include all "status=4*" results.



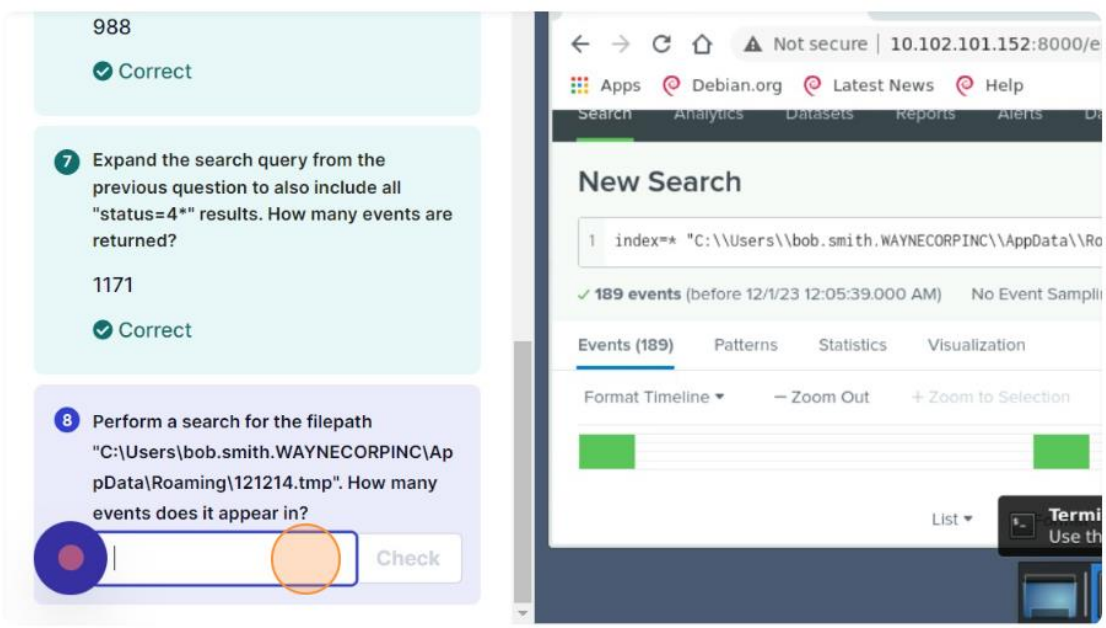
7

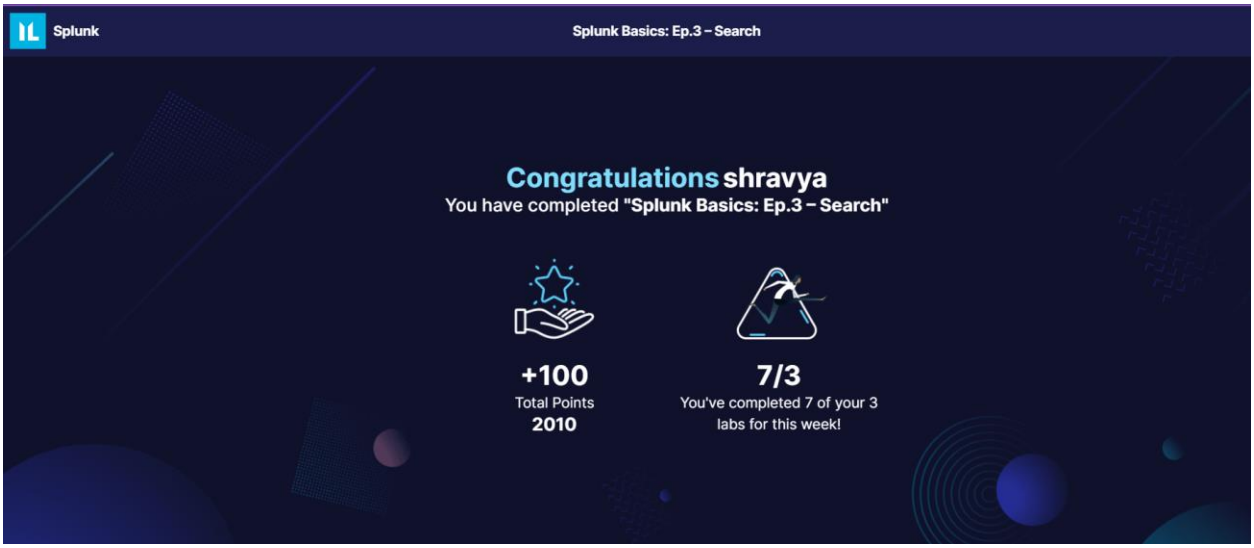


8 "Perform a search for the filepath "C:\Users\bob.smith.WAYNECORPINC\AppData\Roaming\121214.tmp".



9





The learning outcomes for basic Splunk search include:

1. Identifying Key Structure:
   - Understanding the search pipeline: Recognizing the consecutive command structure connected by pipes for refining searches.
   - Fields comprehension: Grasping the concept of fields as searchable key/value pairs in event data.
   - Quotes and escaping characters: Knowing the use of quotes and backslashes to handle special characters and spaces in search queries.

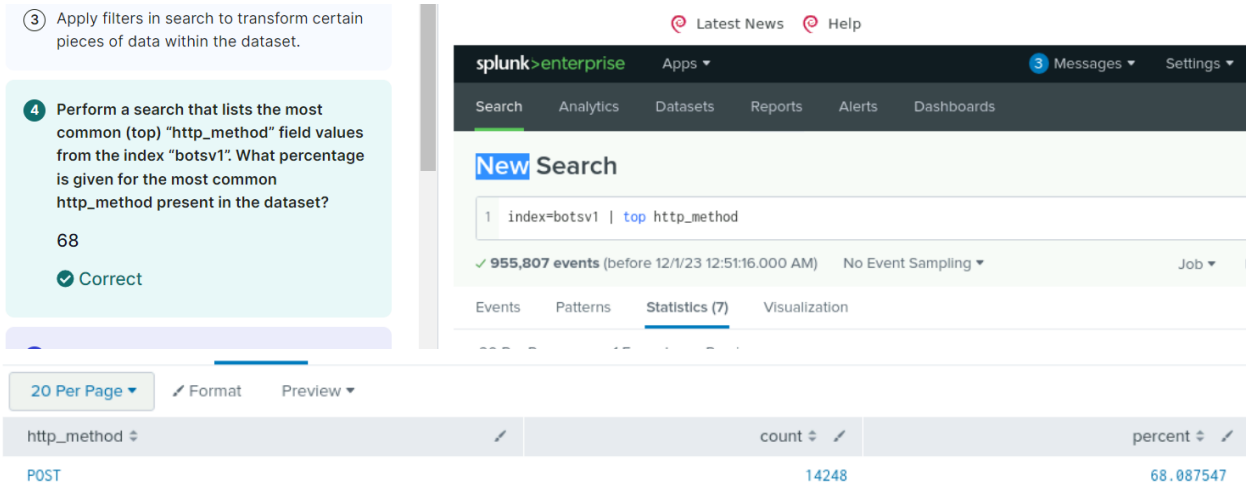
2. Demonstrating Basic Search Techniques:
   - Simple queries: Executing basic search queries to filter results using keywords or phrases.
   - Boolean expressions: Applying Boolean operators (AND, OR, NOT) for combining and refining search queries.
   - Wildcards: Utilizing asterisk (*) as a wildcard placeholder to match varying characters in a string.
   - Field-based searches: Efficiently using fields in searches, including syntax and comparison operators (e.g., =, !=, <, >).

3. Results Interpretation:
   - Interpreting search results: Understanding the Events Viewer and interpreting raw event data.
   - Adding to searches: Extending searches based on data-specific fields through the Add to search dialogue.

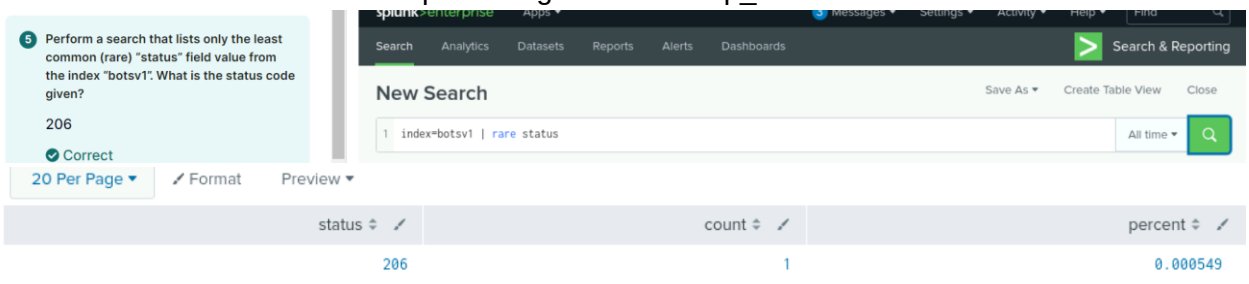
These outcomes collectively enable users to navigate and effectively utilize basic search functionalities in Splunk, facilitating data analysis and exploration.

4) Splunk Basics Ep-4: Advanced searching (SPL and Transforming)

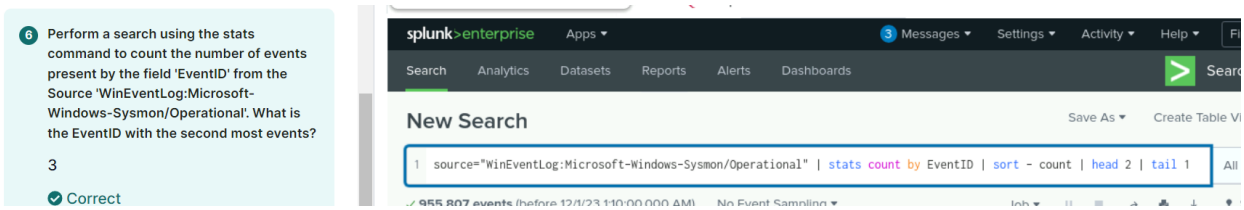
The Advanced Search lab in Splunk delves into the intricacies of the Search Processing Language (SPL), a powerful tool for interacting with event data. SPL encompasses search terms, commands, functions, arguments, and clauses. Search terms filter desired information, commands perform actions on search results, functions compute on fields, and arguments provide optional or required parameters. Clauses group or rename fields for result formatting. Transforming commands, including Chart, Timechart, Stats, Top, and Rare, organize data for statistical purposes and visualization. The lab explores subsearches, narrowing down events by using results as arguments for primary searches, facilitating complex data analysis and integration.



This search will return the most common values for the "http_method" field in the "botsv1" index. The result will include the percentage for each http_method value in the dataset



This search will return the least common values for the "status" field in the "botsv1" index. The result will include the status code associated with the rarest status value in the dataset.

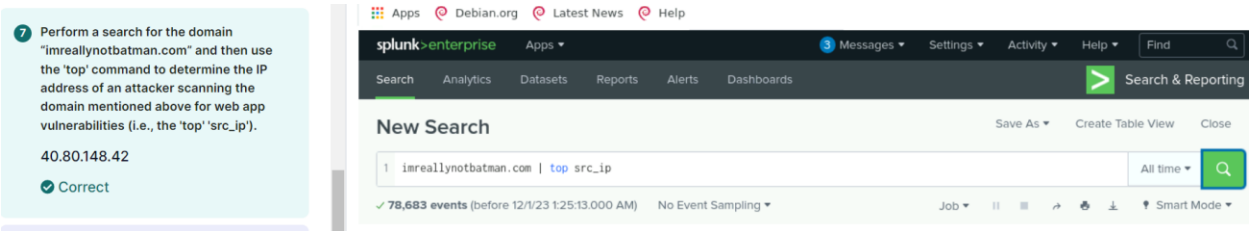


EventID	count
3	99320

This search query does the following:

1. It filters events from the specified source.
2. It uses the `stats` command to count the number of events for each unique EventID.
3. It sorts the results in ascending order based on the count.
4. It uses `head 2` to select the top two results.
5. It uses `tail 1` to choose the second result, which represents the EventID with the second most events.

The result will show the EventID with the second most events in the specified source.

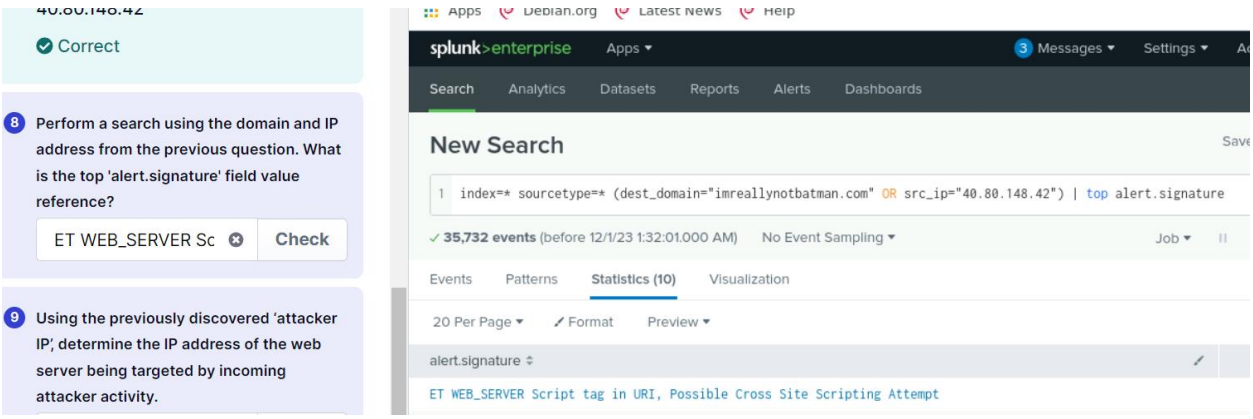


src_ip	count
40.80.148.42	34967

This search query does the following:

1. It filters events where the destination domain is "imreallynotbatman.com" (`dest_domain="imreallynotbatman.com"`).
2. It uses the `top` command to identify the top source IP addresses (`src_ip`) associated with the specified domain.

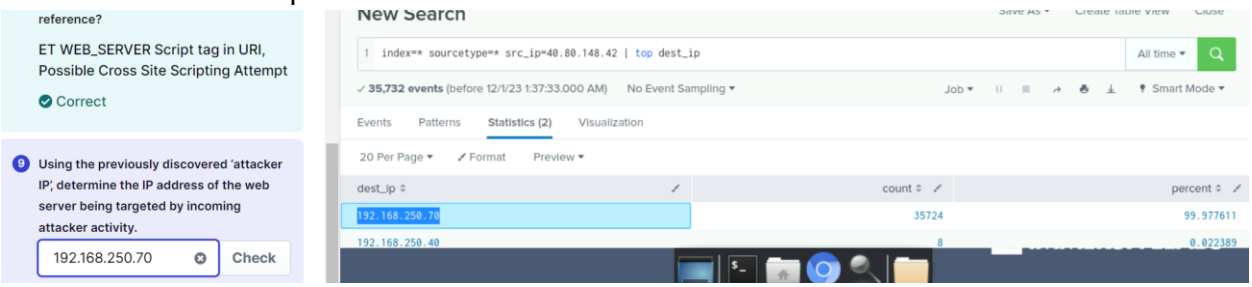
The result will show you the source IP address of the attacker that has been scanning the domain "imreallynotbatman.com" for web app vulnerabilities.



This search query does the following:

1. It searches across all indexes and sourcetypes (`index=* sourcetype=*`).
2. It filters events where the destination domain is "imreallynotbatman.com" or the source IP address is "40.80.148.42".
3. It uses the `top` command to identify the top values for the 'alert.signature' field.

The result will show you the top 'alert.signature' field value reference associated with events related to the specified domain and IP address.



This search query does the following:

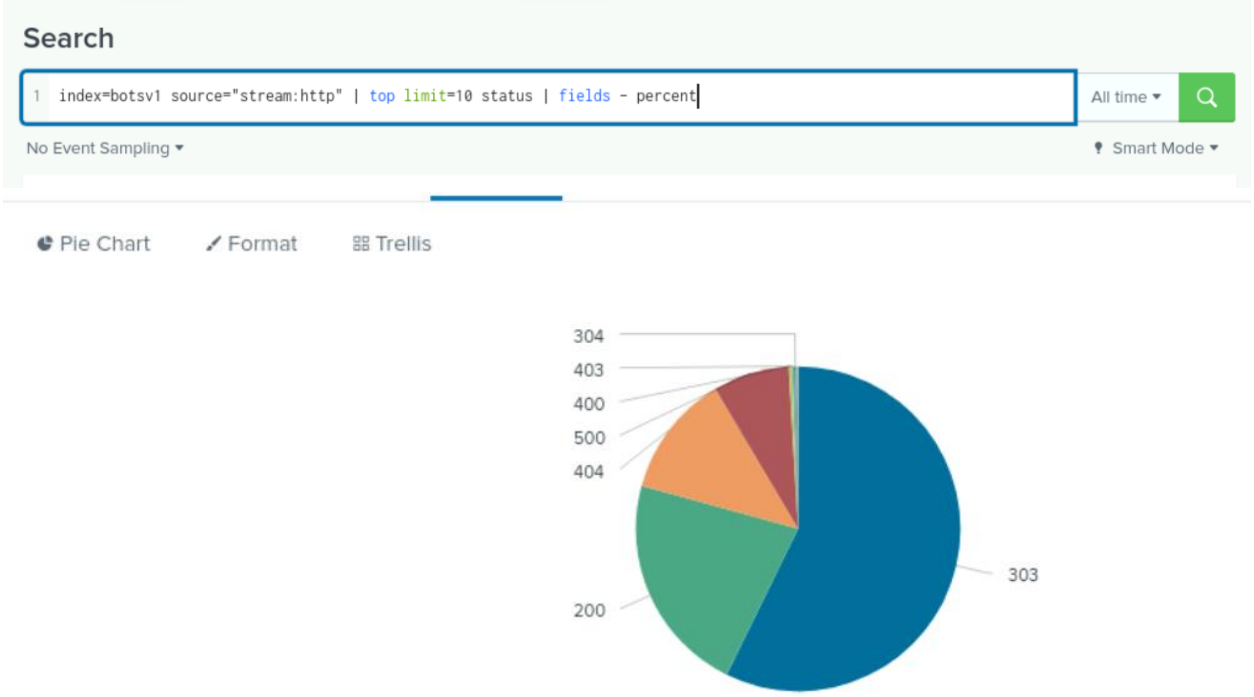
1. It searches across all indexes and sourcetypes (`index=* sourcetype=*`).
2. It filters events where the source IP address is the previously discovered attacker IP (`src_ip=40.80.148.42`).
3. It uses the `top` command to identify the top destination IP addresses (`dest_ip`) associated with the specified attacker IP.

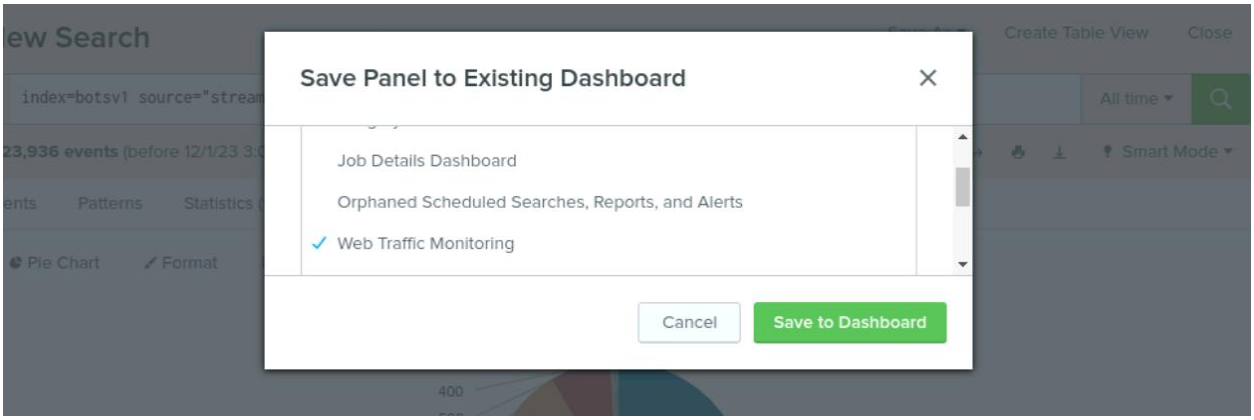
The result will show you the IP address of the web server that is being targeted by the incoming attacker activity.


Splunk
Splunk Basics: Ep.4 – Advanced Searching (SPL & Transforming)
Congratulations shravya
You have completed "Splunk Basics: Ep.4 – Advanced Searching (SPL & Transforming)"
+200
Total Points
2010
7/3
You've completed 7 of your 3 labs for this week!

5) Splunk Basics Ep-5: Dashboard and Visualizations

Data visualization in Splunk is a feature designed to present search results in a concise manner, aiding both technical and non-technical users in comprehending data. Users can generate visualizations on the Search page during event searches or dashboard creation. Splunk offers diverse visualization options, including charts, tables, maps, and custom visualizations. Dashboards, aggregating saved visualizations and panels, offer a summarized view of crucial information in a visual format. They streamline data monitoring, ideal for scenarios like data center oversight, SOC network monitoring, and intrusion detection. Dashboards enhance communication by providing effective summaries, fostering efficient collaboration among teams. Adding visualizations to dashboards involves selecting the desired visualization and saving it to an existing dashboard.





Follow the steps below:

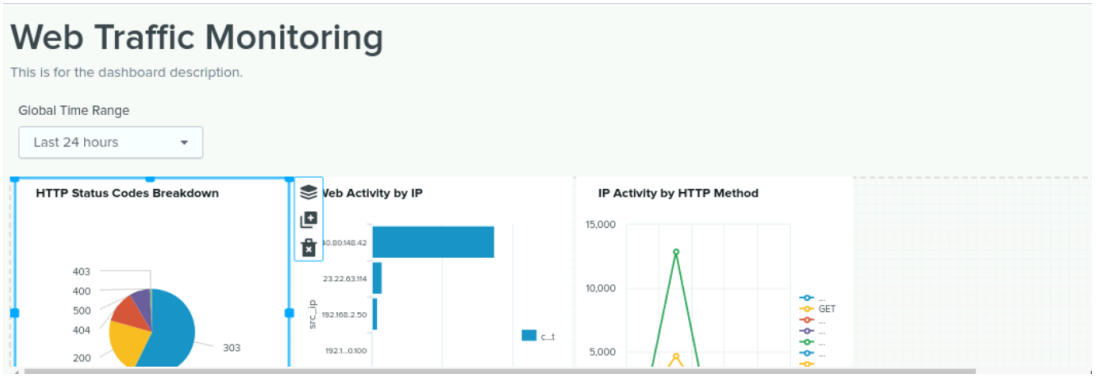
1. HTTP Status Codes Breakdown (Pie Chart):
   - Use the search: `index=botsv1 source="stream:http" | top limit=10 status | fields - percent`
   - Save the visualization as a "Pie Chart."
   - Name it "HTTP Status Codes Breakdown."
   - Add it to the existing dashboard "Web Traffic Monitoring."
   - Close the “View Dashboard” prompt.

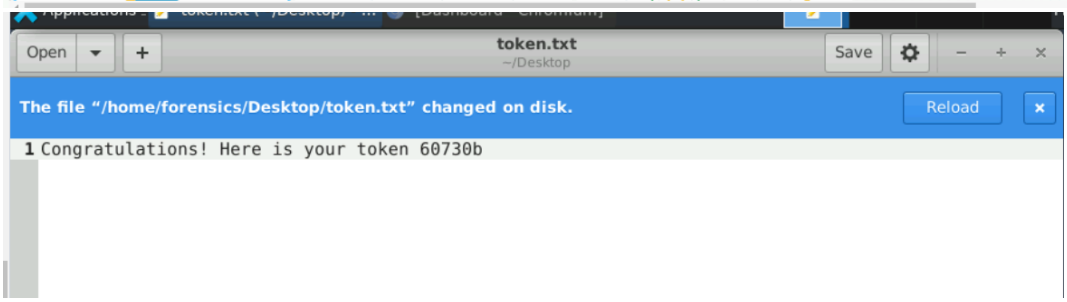
2. Web Activity by IP (Bar Chart):
   - Use the search: `index=botsv1 source="stream:http" | stats count by src_ip | sort - count`
   - Save the visualization as a "Bar Chart."
   - Name it "Web Activity by IP."
   - Add it to the existing dashboard "Web Traffic Monitoring."

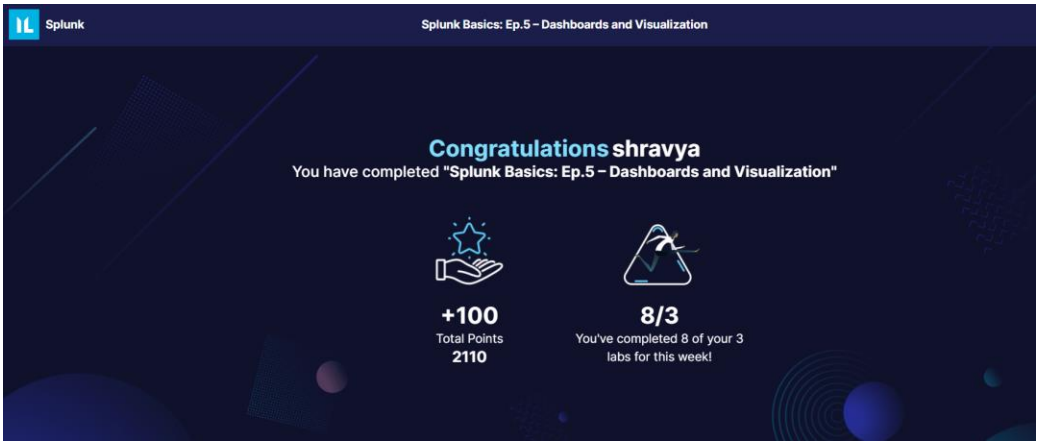
- Close the “View Dashboard” prompt.

3. IP Activity by HTTP Method (Line Chart):
   - Use the search: `index=botsv1 source="stream:http" | chart count over src_ip by http_method`
   - Save the visualization as a "Line Chart."
   - Name it "IP Activity by HTTP Method."
   - Add it to the existing dashboard "Web Traffic Monitoring."
   - Select “View dashboard” when prompted.

After adding all three visualizations, retrieve the token from the `token.txt` file on your desktop.

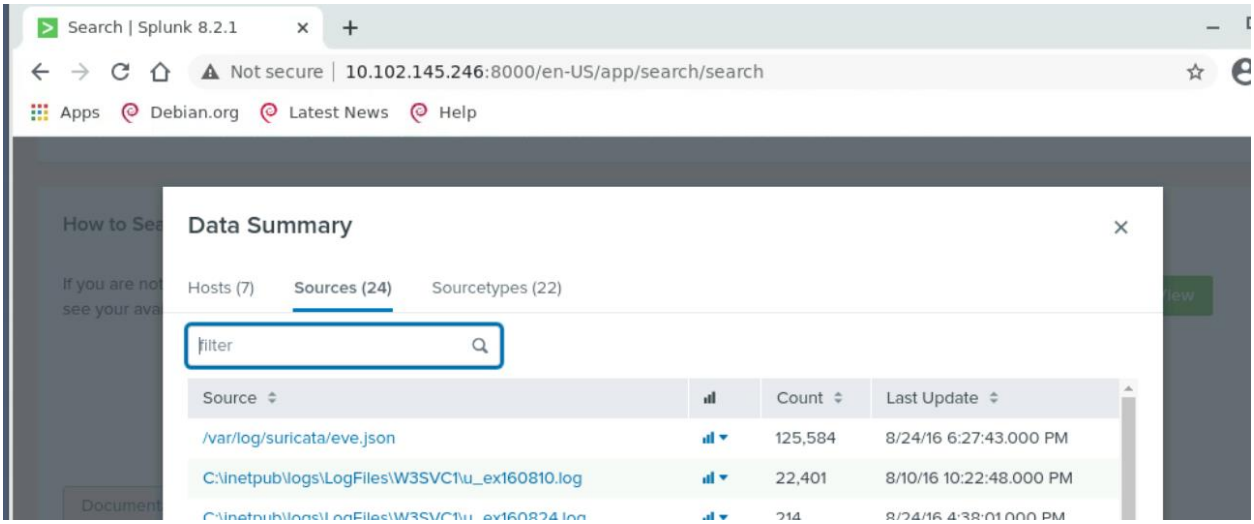






6) Demonstrate Your Skills:Splunk Basics

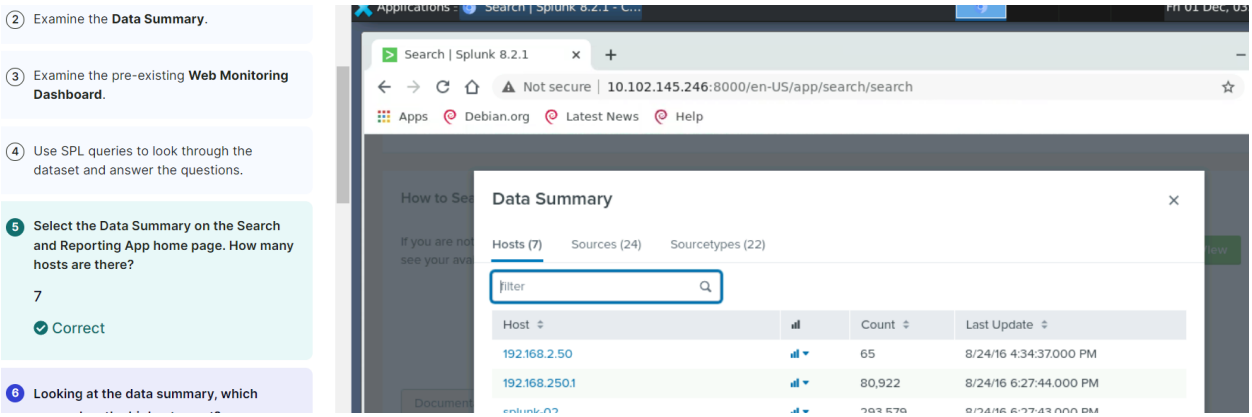
The lab utilizes logs from diverse sources, including Windows Sysmon for process insights, Suricata for intrusion detection, Windows Event Logs for system notifications, Windows Registry for configuration changes, Stream for live data monitoring, and fgt for Fortinet FortiGate firewall data. Splunk's 'Boss of the SOC' CTF competition provides these logs, empowering you to enhance your skills. To explore and enumerate these logs, the Splunk search command '| metadata type=sourcetypes index=botsv1' proves valuable. This scenario emphasizes the crucial role of a SOC analyst in navigating and analyzing diverse logs to ensure effective threat detection and response.



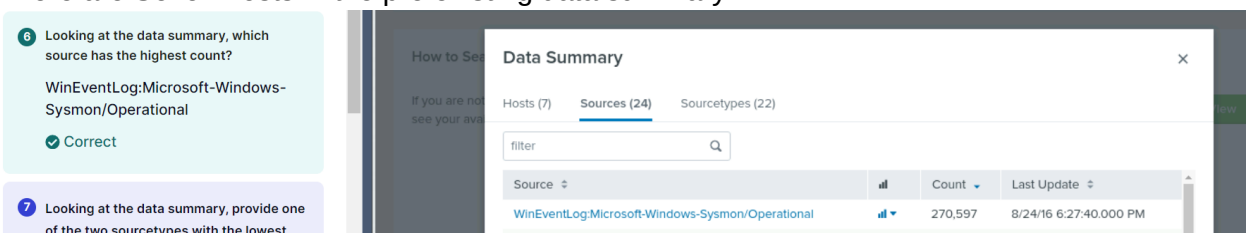
Monitoring and searching of data from various log sources, including Windows Sysmon, Suricata, Windows Event Logs, Windows Registry, Stream, and Fortinet FortiGate firewall data. To enumerate these logs in Splunk, you can use the following command in the Splunk search:

| metadata type=sourcetypes index=botsv1

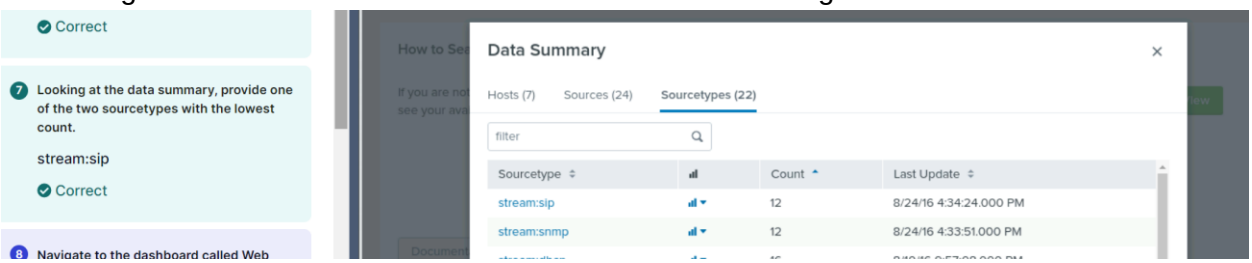
This command utilizes the Splunk metadata command to list the sourcetypes present in the "botsv1" index. Sourcetypes represent different types of data sources, helping you understand the variety of logs available for analysis. This initial enumeration is crucial for familiarizing with the data landscape and preparing for further exploration and investigation.



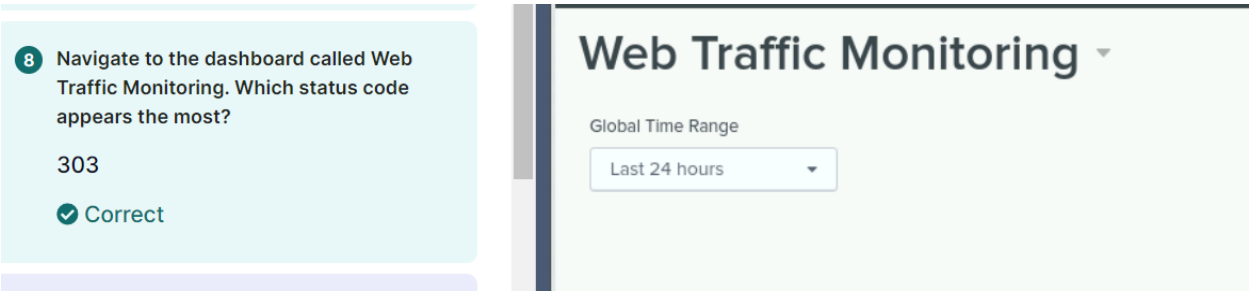
There are Seven hosts in the pre existing data summary.



On sorting the count we can see that THe above Source has highest count.

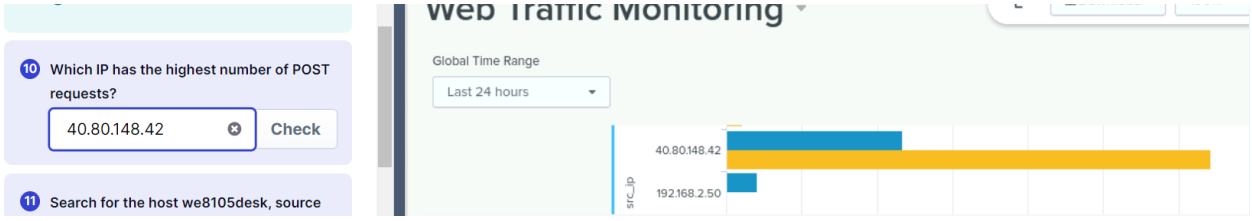


On sorting the source types we see that above sip and snmp has least count.

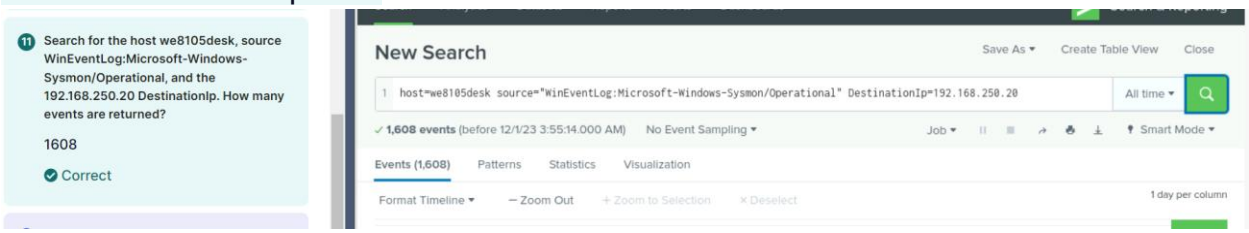


From the pie chart above in the web monitoring we can conclude that 303 status code appears most.

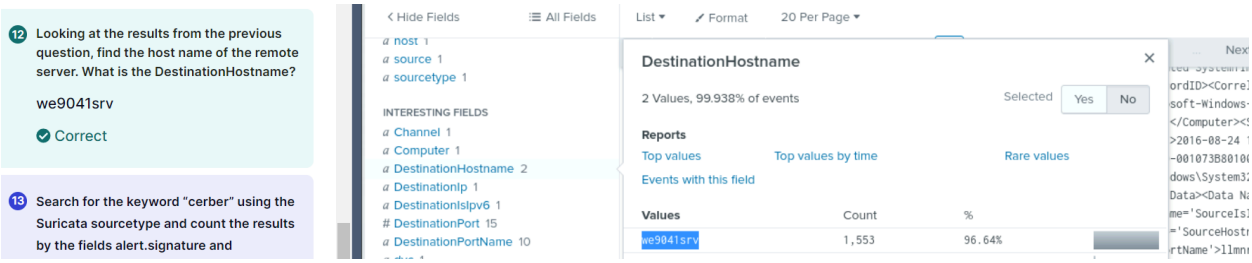
From the bar graph it can be observed that 192.168.250.70 was served the lowest number of requests.



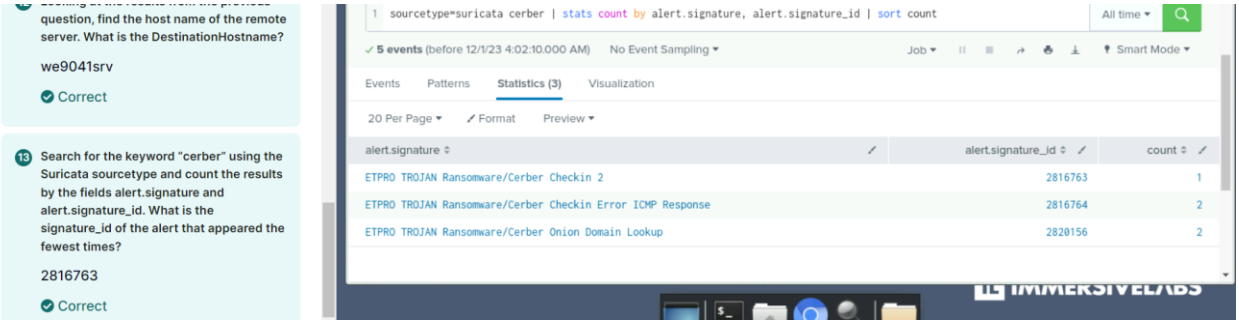
From the bar graph it can be observed that 40.80.148.42 was served the highest number of POST requests.



This search filters events in the "botsv1" index where the host is "we8105desk," the source is "WinEventLog:Microsoft-Windows-Sysmon/Operational," and the destination IP is "192.168.250.20." The number of events returned will be displayed in the search results.



Under the interesting fields in the results, we find the details of host along with Destination host name.



This search filters events in the "botsv1" index with the Suricata sourcetype containing the keyword "cerber." It then uses the `stats` command to count the occurrences based on the `alert.signature` and `alert.signature_id` fields. The results are sorted by the count in ascending order, allowing you to identify the alert with the fewest occurrences. The signature_id of the alert that appeared the fewest times will be visible in the search results.

Conclusion

Gained hands-on experience in log analysis and threat detection, utilizing diverse log sources such as Windows Sysmon, Suricata, Windows Event Logs, Windows Registry, Stream, and Fortinet FortiGate firewall data. The assignment begins with log enumeration, emphasizing the crucial skill of understanding log sources and their metadata.

Tasked with conducting intricate searches, honing their query-building skills to filter events based on specific criteria. This includes searches focused on HTTP traffic analysis, Sysmon operational logs, and Suricata alerts containing the keyword "cerber." By employing advanced Splunk commands like `metadata`, `stats`, and `table`, to distill meaningful insights from vast datasets.

Furthermore, the assignment extends to the creation of visualizations and dashboards, emphasizing the significance of presenting data in a comprehensible format. This not only explores the functionalities of various visualization types such as Pie Charts, Bar Charts, and Line Charts but also practices incorporating them into dashboards for effective data monitoring.

Ultimately, the assignment cultivates a highly technical skill set in cybersecurity, empowering to navigate complex log landscapes, conduct precise threat searches, and leverage visualization tools for efficient data communication and analysis in a Security Operations Center (SOC) setting.