



HTTP Log Analysis using Splunk Project

By Gaurav Ghandat

Hands-on cybersecurity lab focused on analyzing HTTP traffic using **Splunk** and **Zeek-style JSON logs** to detect errors, anomalies, and suspicious web activity.

Objective

In this lab, you will:

- Ingest and analyze HTTP logs using **Splunk**
 - Detect **client-side (4xx)** and **server-side (5xx)** HTTP errors
 - Identify **suspicious User-Agents and URIs**
 - Detect **large file transfers** that may indicate data exfiltration
 - Gain practical experience with **SPL (Search Processing Language)**
-

Tech Stack & Tools

- **SIEM Tool:** Splunk Enterprise
 - **Log Source:** Zeek-style HTTP logs (JSON format)
 - **Index:** http_lab (custom index)
 - **Sourcetype:** json or zeek:http
-

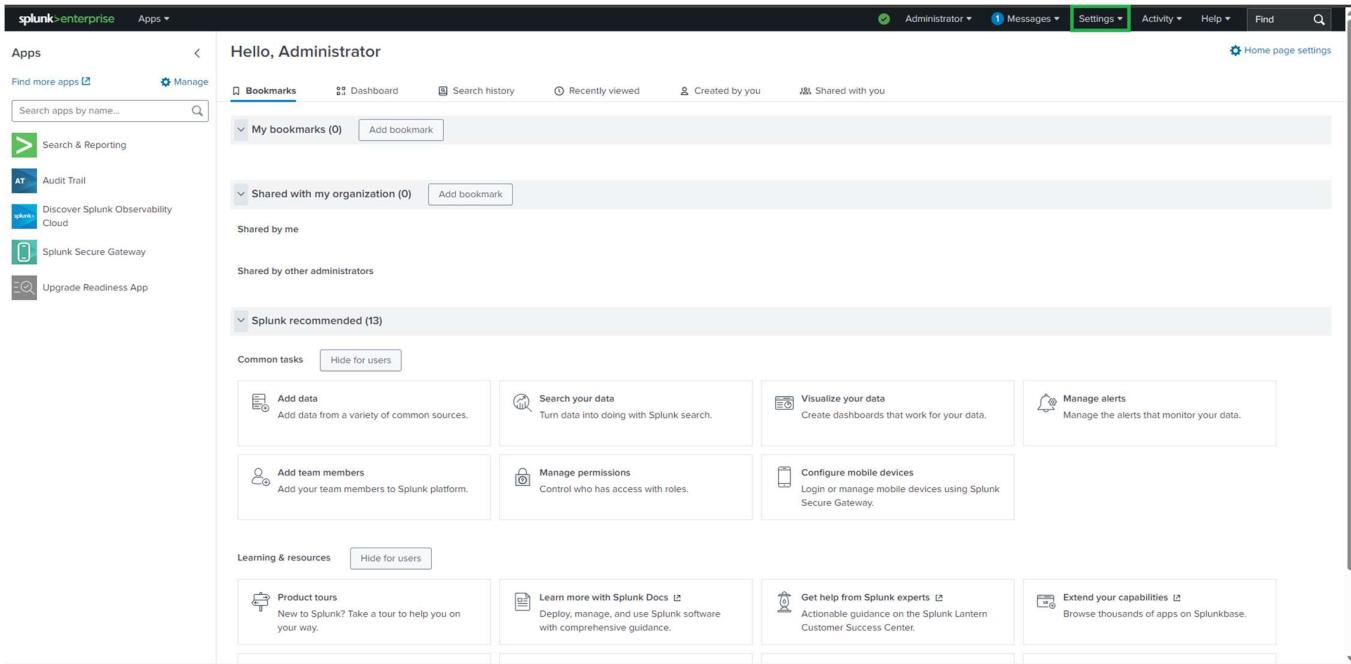
Lab Setup

Prerequisites

- Splunk installed and accessible via Splunk Web
- HTTP logs in JSON format (http_logs.json)

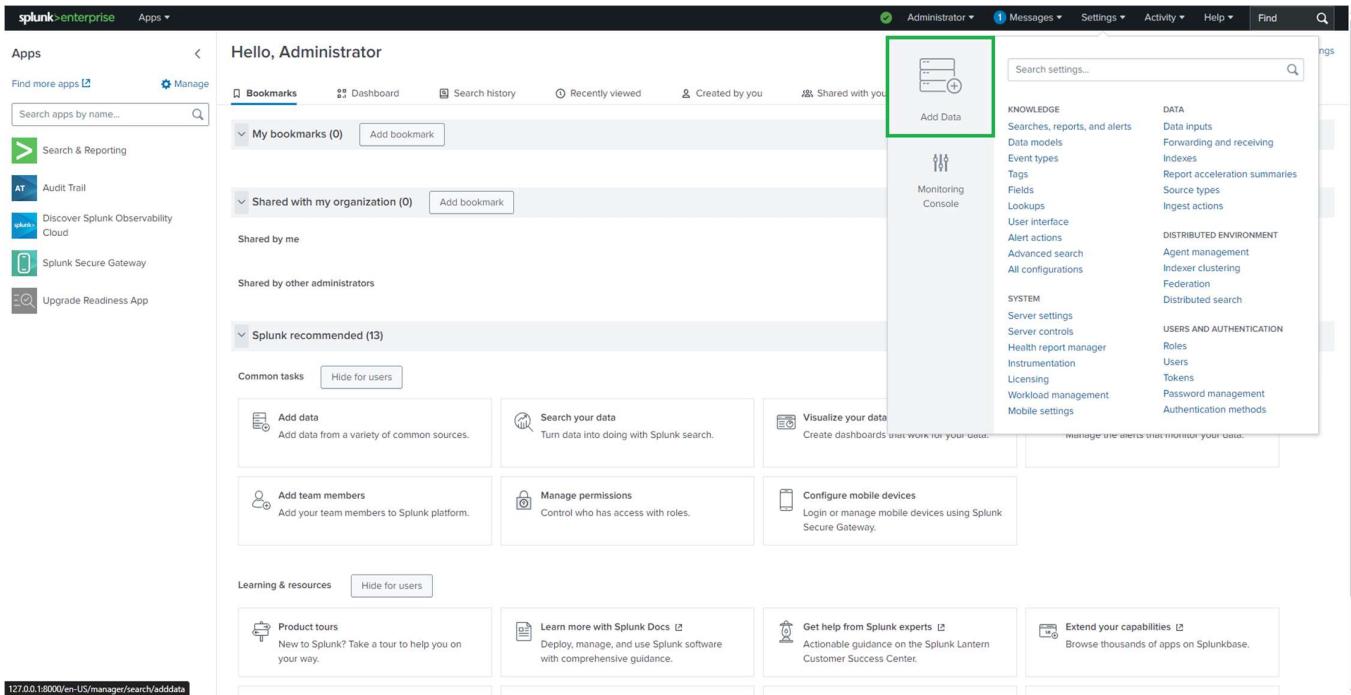
Data Ingestion Steps

1. Open Splunk Web



The screenshot shows the Splunk Web interface. At the top, there's a navigation bar with links for 'Administrator', 'Messages', 'Settings' (which is highlighted with a green box), 'Activity', 'Help', and 'Find'. Below the navigation is a search bar labeled 'Home page settings'. The main content area is titled 'Hello, Administrator'. It features several sections: 'Bookmarks' (My bookmarks, Shared with my organization), 'Splunk recommended' (Common tasks like Add data, Search your data, Visualize your data, Manage alerts, Add team members, Manage permissions, Configure mobile devices), and 'Learning & resources' (Product tours, Learn more with Splunk Docs, Get help from Splunk experts, Extend your capabilities). On the left, there's a sidebar with 'Apps' and a search bar.

2. Navigate to Settings → Add Data



This screenshot is similar to the first one, showing the 'Hello, Administrator' dashboard. However, the sidebar on the right is expanded to show the 'Monitoring Console' section. Within this section, the 'Add Data' link is highlighted with a green box. The sidebar also lists other monitoring-related options like 'Searches, reports, and alerts', 'Data inputs', 'Forwarding and receiving', etc. The rest of the interface is identical to the first screenshot, including the 'Common tasks' and 'Learning & resources' sections.

3. Select Upload

splunk>enterprise Apps *

Administrator Messages Settings Activity Help Find

What data do you want to send to the Splunk platform?

Follow guides for onboarding popular data sources

Cloud computing Networking Operating System Security

Get your cloud computing data in to the Splunk platform. Get your networking data in to the Splunk platform. Get your operating system data in to the Splunk platform. Get your security data in to the Splunk platform.

10 data sources 2 data sources 1 data source 3 data sources

4 data sources in total

Or get data in with the following methods

Upload Monitor Forward

files from my computer files and ports on this Splunk platform instance data from a Splunk forwarder

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

Tutorial for adding data ↗

127.0.0.1:8000/en-US/manage/search/adddatamethods/selectforwarders

splunk>enterprise Apps *

Administrator Messages Settings Activity Help Find

Add Data

Select Source Set Source Type Input Settings Review Done

Next >

Select Source

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 ↗

Selected File: No file selected

Select File

Drop your data file here

The maximum file upload size is 500 Mb

FAQ

> What kinds of files can the Splunk platform index?
> What is a source?
> How do I get remote data onto my Splunk platform instance?

4. Upload http_logs.json

The screenshot shows the Splunk Add Data interface for uploading a file named "http_logs". The interface consists of several panels:

- Left Panel:** A file browser window titled "Open" showing the file "http_logs" in the "Downloads" folder. The "Open" button is highlighted with a green box.
- Middle Panel:** A large input field labeled "Drop your data file here" with a note: "The maximum file upload size is 500 Mb".
- Top Bar:** Progress steps: "Source" (green dot), "Set Source Type", "Input Settings", "Review", and "Done". Buttons: "< Back", "Next >" (highlighted with a green box).
- FAQ Panel:** Questions about file types, sources, and remote data upload.
- Bottom Status:** URL "127.0.0.1:8000/en-US/manager/search/adddata?method=selectsource&input_mode=0&f" and a message: "File Successfully Uploaded" with a checkmark icon.

5. Configure:

- o **Source type:** json or create zeek:http
- o **Index:** http_lab (recommended)

The screenshot shows the Splunk Add Data wizard with the title 'Add Data' at the top. The progress bar is at the second step, 'Set Source Type'. Below the bar, there are four buttons: 'Select Source' (green), 'Set Source Type' (green), 'Input Settings' (white), 'Review' (white), and 'Done' (white). To the right of the bar are 'Back' and 'Next >' buttons. The main area is titled 'Set Source Type' with a sub-instruction: 'This page lets you see how the Splunk platform sees your data before indexing. If the events look correct and have the right timestamps, click "Next" to proceed. If not, use the options below to define proper event breaks and timestamps. If you cannot find an appropriate source type for your data, create a new one by clicking "Save As".' A table titled 'Source: http_logs.json' displays 12 log entries. The columns include _time, event_type, id.orig_h, id.resp_h, method, resp_body_len, status_code, ts, uid, uri, and user_agent. The first entry is highlighted.

Set Source Type

This page lets you see how the Splunk platform sees your data before indexing. If the events look correct and have the right timestamps, click "Next" to proceed. If not, use the options below to define proper event breaks and timestamps. If you cannot find an appropriate source type for your data, create a new one by clicking "Save As".

Source type: _json											Save As	View Event Summary									
	_time	event_type	id.orig_h	id.resp_h	method	resp_body_len	status_code	ts	uid	uri	user_agent	< Prev	1	2	3	4	5	6	7	8	... Next >
> Timestamp	4/25/25 4:16:18.860 PM	Large Transfer	10.0.0.49	10.0.1.6	GET	1958305	200	2025-04-25T10:46:18.860765Z	HT1031308	/index.html	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7)										
> Advanced	4/25/25 4:16:18.860 PM	Unexpected Method	10.0.0.46	10.0.1.1	OPTIONS	2464	200	2025-04-25T10:46:18.860837Z	HT8956786	/index.html	Mozilla/5.0 (Windows NT 10_0; Win64; x64)										
1	4/25/25 4:16:18.860 PM	Standard	10.0.0.24	10.0.1.3	POST	6163	200	2025-04-25T10:46:18.860855Z	HT9025844	/index.html	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7)										
2	4/25/25 4:16:18.860 PM	Standard	10.0.0.30	10.0.1.5	GET	11580	200	2025-04-25T10:46:18.860869Z	HT8250492	/index.html	Mozilla/5.0 (X11; Linux x86_64)										
3	4/25/25 4:16:18.860 PM	Large Transfer	10.0.0.13	10.0.1.5	POST	1223825	200	2025-04-25T10:46:18.860888Z	HT8686201	/index.html	Mozilla/5.0 (Windows NT 10_0; Win64; x64)										
4	4/25/25 4:16:18.860 PM	Large Transfer	10.0.0.46	10.0.1.7	POST	1550707	200	2025-04-25T10:46:18.860905Z	HT7409193	/index.html	Mozilla/5.0 (X11; Linux x86_64)										
5	4/25/25 4:16:18.860 PM	Standard	10.0.0.43	10.0.1.11	POST	7756	200	2025-04-25T10:46:18.860915Z	HT6831662	/index.html	Mozilla/5.0 (X11; Linux x86_64)										
6	4/25/25 4:16:18.860 PM	Large Transfer	10.0.0.23	10.0.1.2	GET	1191977	200	2025-04-25T10:46:18.860930Z	HT7221204	/index.html	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7)										
7	4/25/25 4:16:18.860 PM	Client Error	10.0.0.36	10.0.1.5	GET	12998	404	2025-04-25T10:46:18.860942Z	HT1157835	/index.html	Mozilla/5.0 (X11; Linux x86_64)										
8	4/25/25 4:16:18.860 PM	Server Error	10.0.0.10	10.0.1.2	POST	1380	503	2025-04-25T10:46:18.860951Z	HT3509295	/index.html	Mozilla/5.0 (X11; Linux x86_64)										
9	4/25/25 4:16:18.860 PM	Standard	10.0.0.44	10.0.1.11	POST	14481	200	2025-04-25T10:46:18.860960Z	HT5761373	/index.html	Mozilla/5.0 (Windows NT 10_0; Win64; x64)										
10	4/25/25 4:16:18.860 PM	Standard	10.0.0.48	10.0.1.4	POST	4427	200	2025-04-25T10:46:18.860968Z	HT5435674	/index.html	Mozilla/5.0 (Windows NT 10_0; Win64; x64)										

The screenshot shows the Splunk Add Data wizard with the title 'Add Data' at the top. The progress bar is at the third step, 'Input Settings'. Below the bar, there are five buttons: 'Select Source' (green), 'Set Source Type' (green), 'Input Settings' (green), 'Review' (green), and 'Done' (white). To the right of the bar are 'Back' and 'Review >' buttons. The main area is titled 'Input Settings' with a sub-instruction: 'Optionally set additional input parameters for this data input as follows:'. It contains sections for 'Host', 'Index', and 'FAQ'.

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: LAPTOP-EHJ3QFJI

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: Default [Create a new index](#)

FAQ

- > How do indexes work?
- > How do I know when to create or use multiple indexes?

splunk enterprise Apps *

Administrator * Messages * Settings * Activity * Help * Find

New Index

General Settings

Index Name

Set index name (e.g., INDEX_NAME). Search using index=INDEX_NAME.

Index Data Type Events Metrics

The type of data to store (event-based or metrics).

Home Path optional Hot/warm db path. Leave blank for default (\$\$SPLUNK_DB/INDEX_NAME/db).

Cold Path optional Cold db path. Leave blank for default (\$\$SPLUNK_DB/INDEX_NAME/colddb).

Thawed Path optional Thawed/resurrected db path. Leave blank for default (\$\$SPLUNK_DB/INDEX_NAME/thawedb).

Data Integrity Check Enable Disable

Enable this if you want Splunk to compute hashes on every slice of your data for the purpose of data integrity.

Max Size of Entire Index 500 GB ▾

Maximum target size of entire index.

Max Size of Hot/Warm/Cold Bucket auto GB ▾

Maximum target size of buckets. Enter 'auto_high_volume' for high-volume indexes.

Frozen Path optional Frozen bucket archive path. Set this if you want Splunk to automatically archive frozen buckets.

App Search & Reporting ▾

Storage Optimization

Save Cancel

Input Sets

Host

Index

FAQ

> How do I ...
> How do I ...
> How do I ...

splunk enterprise Apps *

Administrator * Messages * Settings * Activity * Help * Find

New Index

General Settings

Index Name

Set index name (e.g., INDEX_NAME). Search using index=INDEX_NAME.

Index Data Type Events Metrics

The type of data to store (event-based or metrics).

Home Path optional Hot/warm db path. Leave blank for default (\$\$SPLUNK_DB/INDEX_NAME/db).

Cold Path optional Cold db path. Leave blank for default (\$\$SPLUNK_DB/INDEX_NAME/colddb).

Thawed Path optional Thawed/resurrected db path. Leave blank for default (\$\$SPLUNK_DB/INDEX_NAME/thawedb).

Data Integrity Check Enable Disable

Enable this if you want Splunk to compute hashes on every slice of your data for the purpose of data integrity.

Max Size of Entire Index 500 GB ▾

Maximum target size of entire index.

Max Size of Hot/Warm/Cold Bucket auto GB ▾

Maximum target size of buckets. Enter 'auto_high_volume' for high-volume indexes.

Frozen Path optional Frozen bucket archive path. Set this if you want Splunk to automatically archive frozen buckets.

App Search & Reporting ▾

Storage Optimization

Save Cancel

Input Sets

Host

Index

FAQ

> How do I ...
> How do I ...
> How do I ...

splunk>enterprise Apps ▾

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

Add Data Select Source Set Source Type Input Settings Review Done < Back Review >

Input Settings

Optional set additional input parameters for this data input as follows:

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: LAPTOP-EHJ3QFJI

Index

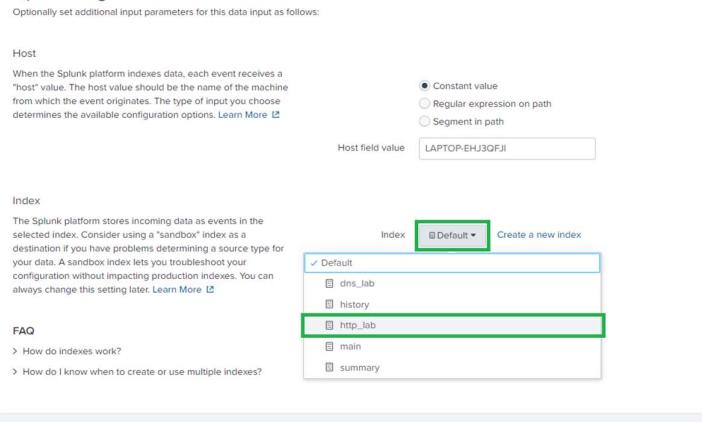
The Splunk platform stores incoming data as events in the selected index. Consider using a "sandbox" index as a destination if you have problems determining a source type for your data. A sandbox index lets you troubleshoot your configuration without impacting production indexes. You can always change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

Index: Default ▾ Create a new index

Default
dns_lab
history
http_lab
main
summary

FAQ

> How do indexes work?
> How do I know when to create or use multiple indexes?



127.0.0.1:8000/en-USmanager/search/adddatamethods/inputsettings#

splunk>enterprise Apps ▾

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

Add Data Select Source Set Source Type Input Settings Review Done < Back Review >

Input Settings

Optional set additional input parameters for this data input as follows:

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: LAPTOP-EHJ3QFJI

Index

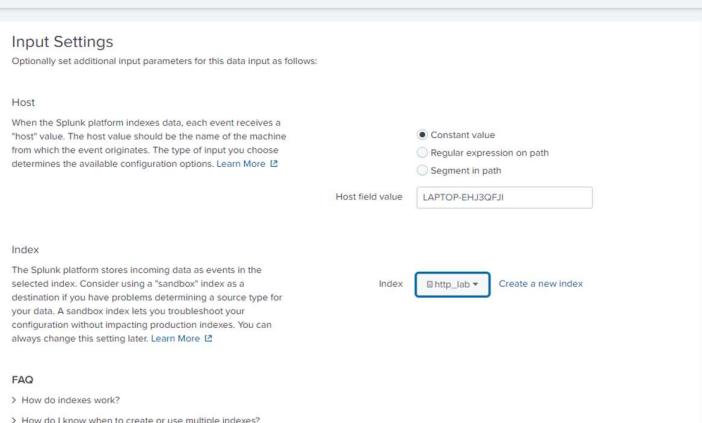
The Splunk platform stores incoming data as events in the selected index. Consider using a "sandbox" index as a destination if you have problems determining a source type for your data. A sandbox index lets you troubleshoot your configuration without impacting production indexes. You can always change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

Index: http_lab ▾ Create a new index

http_lab

FAQ

> How do indexes work?
> How do I know when to create or use multiple indexes?



6. Finish upload and verify indexing

The screenshot shows the Splunk Add Data interface. At the top, there is a navigation bar with links for Administrator, Messages, Settings, Activity, Help, and Find. Below the navigation bar, a progress bar indicates the current step: Select Source (green dot), Set Source Type (green dot), Input Settings (green dot), Review (green dot), and Done (white circle). A green rectangular box highlights the "Review" section. Inside the "Review" section, the following data is displayed:

Input Type	Uploaded File
File Name	http_logs.json
Source Type	json
Host	LAPTOP-EHJ3QFJI
Index	http_log

At the bottom right of the "Review" section is a "Submit" button.

The screenshot shows the Splunk Add Data interface at the final "Done" step. The progress bar at the top shows all five steps (Select Source, Set Source Type, Input Settings, Review, Done) as completed, indicated by green dots. A green checkmark icon is positioned next to the "Done" step. A green rectangular box highlights the "File has been uploaded successfully." message. The message text is: "✓ File has been uploaded successfully. Configure your inputs by going to Settings > Data Inputs". Below this message are several buttons with corresponding descriptions:

- Start Searching**: Search your data now or see examples and tutorials.
- Extract Fields**: Create search-time field extractions. Learn more about fields.
- Add More Data**: Add more data inputs now or see examples and tutorials.
- Download Apps**: Apps help you do more with your data. Learn more.
- Build Dashboards**: Visualize your searches. Learn more.

Lab Tasks & SPL Queries

◆ Task 1: Top 10 Endpoints Generating Web Traffic

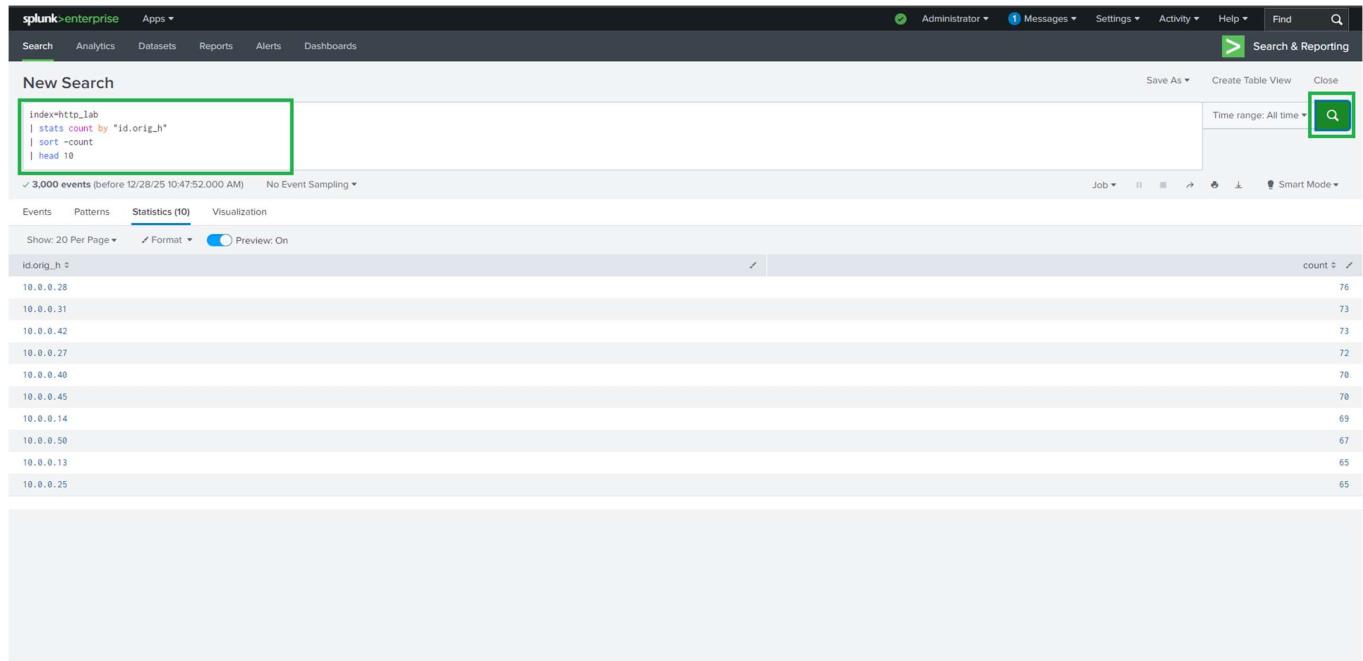
```
index=http_lab
```

```
| stats count by "id.orig_h"
```

```
| sort -count
```

```
| head 10
```

 *Identifies the most active source IPs generating HTTP requests.*



The screenshot shows the Splunk Enterprise search interface. The search bar at the top contains the following SPL query:

```
index=http_lab  
| stats count by "id.orig_h"  
| sort -count  
| head 10
```

The results table below the search bar displays 10 rows of data, each representing a source IP address and its count. The columns are labeled "id.orig_h" and "count". The data is sorted by count in descending order. The first row shows 10.0.0.28 with a count of 76.

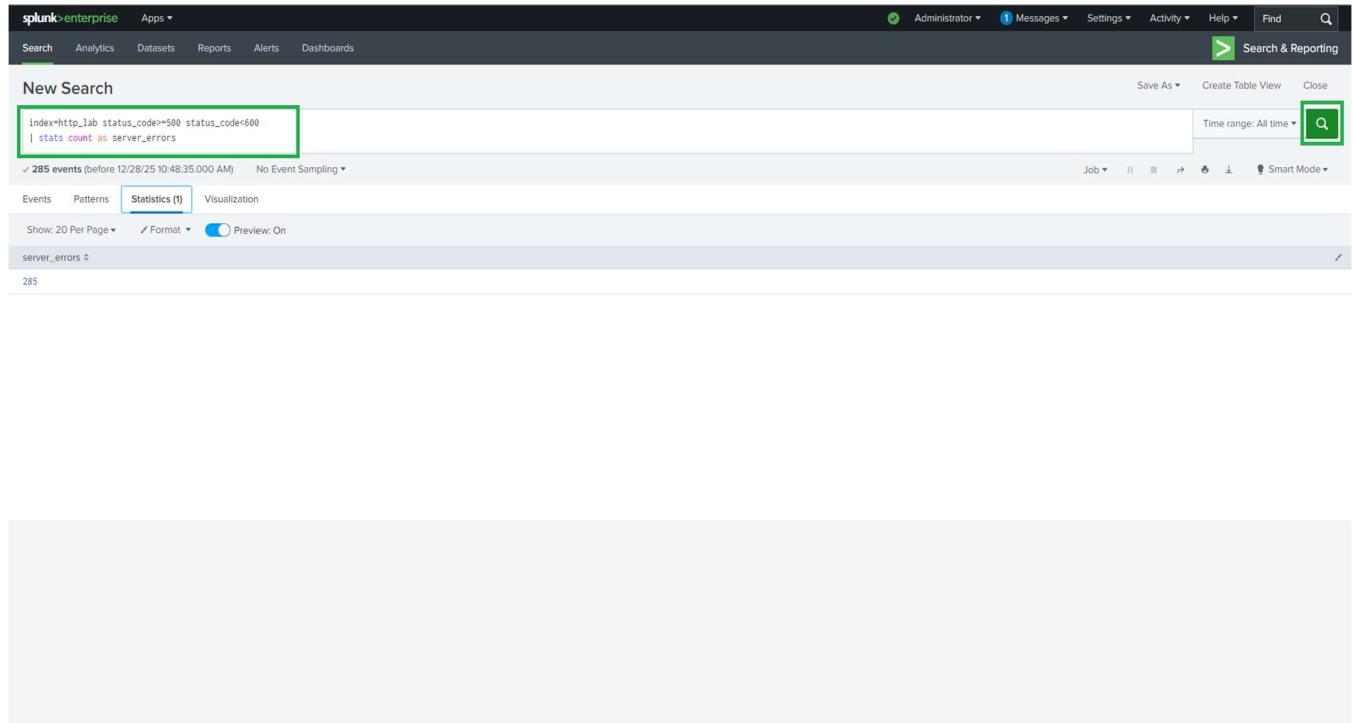
id.orig_h	count
10.0.0.28	76
10.0.0.31	73
10.0.0.42	73
10.0.0.27	72
10.0.0.40	70
10.0.0.45	70
10.0.0.14	69
10.0.0.50	67
10.0.0.13	65
10.0.0.25	65

◆ Task 2: Count Server Errors (HTTP 5xx)

```
index=http_lab status_code>=500 status_code<600
```

```
| stats count as server_errors
```

 Helps detect backend failures or application crashes.



The screenshot shows the Splunk Enterprise search interface. The search bar contains the command: `index=http_lab status_code>=500 status_code<600 | stats count as server_errors`. A green box highlights this search bar. Below the search bar, it says "285 events (before 12/28/25 10:48:35.000 AM) No Event Sampling". The "Statistics (1)" tab is selected. The results table has one row with the value "server_errors 285". The interface includes a header with user information (Administrator), a "Search & Reporting" button, and various navigation and search controls.

◆ Task 3: Detect Suspicious / Scripted User-Agents

index=http_lab

user_agent IN ("sqlmap/1.5.1", "curl/7.68.0", "python-requests/2.25.1", "botnet-checker/1.0")

| stats count by user_agent

📌 *Identifies automated tools commonly used in attacks or scanning.*

The screenshot shows the Splunk Enterprise search interface. The search bar contains the query: `index=http_lab user_agent IN ("sqlmap/1.5.1", "curl/7.68.0", "python-requests/2.25.1", "botnet-checker/1.0") | stats count by user_agent`. The results table has one row:

user_agent	count
sqlmap/1.5.1	79
curl/7.68.0	69
python-requests/2.25.1	78
botnet-checker/1.0	70

◆ Task 4: Find Large File Transfers (> 500 KB)

```
index=http_lab resp_body_len>500000
```

```
| table ts "id.orig_h" "id.resp_h" uri resp_body_len  
| sort -resp_body_len
```

📌 Useful for spotting potential data leaks or suspicious downloads.

The screenshot shows a log analysis interface with a search bar at the top containing the query: `index=http_lab resp_body_len>500000 | table ts "id.orig_h" "id.resp_h" uri resp_body_len | sort -resp_body_len`. Below the search bar, it displays 323 events found between December 28, 2025, and January 1, 2026. The results are presented in a table with columns: ts, id.orig_h, id.resp_h, uri, and resp_body_len. The table is sorted by resp_body_len in descending order. The first few rows of data are as follows:

ts	id.orig_h	id.resp_h	uri	resp_body_len
2025-04-25T10:46:18.879523Z	10.0.0.23	10.0.1.4	/index.html	1377613
2025-04-25T10:46:18.879312Z	10.0.0.49	10.0.1.7	/index.html	1376613
2025-04-25T10:46:18.888207Z	10.0.0.50	10.0.1.3	/index.html	1374922
2025-04-25T10:46:18.863095Z	10.0.0.30	10.0.1.5	/index.html	1368142
2025-04-25T10:46:18.861078Z	10.0.0.48	10.0.1.12	/index.html	1360325
2025-04-25T10:46:18.860765Z	10.0.0.49	10.0.1.6	/index.html	1358305
2025-04-25T10:46:18.862785Z	10.0.0.57	10.0.1.10	/index.html	1351520
2025-04-25T10:46:18.882673Z	10.0.0.36	10.0.1.7	/index.html	1345121
2025-04-25T10:46:18.877884Z	10.0.0.15	10.0.1.10	/index.html	1344872
2025-04-25T10:46:18.889246Z	10.0.0.35	10.0.1.2	/index.html	1341685
2025-04-25T10:46:18.874811Z	10.0.0.43	10.0.1.11	/index.html	1329783
2025-04-25T10:46:18.882569Z	10.0.0.28	10.0.1.5	/index.html	1324376
2025-04-25T10:46:18.887168Z	10.0.0.32	10.0.1.4	/index.html	1316823
2025-04-25T10:46:18.883728Z	10.0.0.51	10.0.1.9	/index.html	1309326
2025-04-25T10:46:18.880276Z	10.0.0.40	10.0.1.8	/index.html	1302154
2025-04-25T10:46:18.893626Z	10.0.0.36	10.0.1.5	/index.html	1301515
2025-04-25T10:46:18.869238Z	10.0.0.57	10.0.1.8	/index.html	1293936
2025-04-25T10:46:18.883666Z	10.0.0.35	10.0.1.1	/index.html	1289812
2025-04-25T10:46:18.864666Z	10.0.0.21	10.0.1.12	/index.html	1283715
2025-04-25T10:46:18.865466Z	10.0.0.36	10.0.1.7	/index.html	12879333

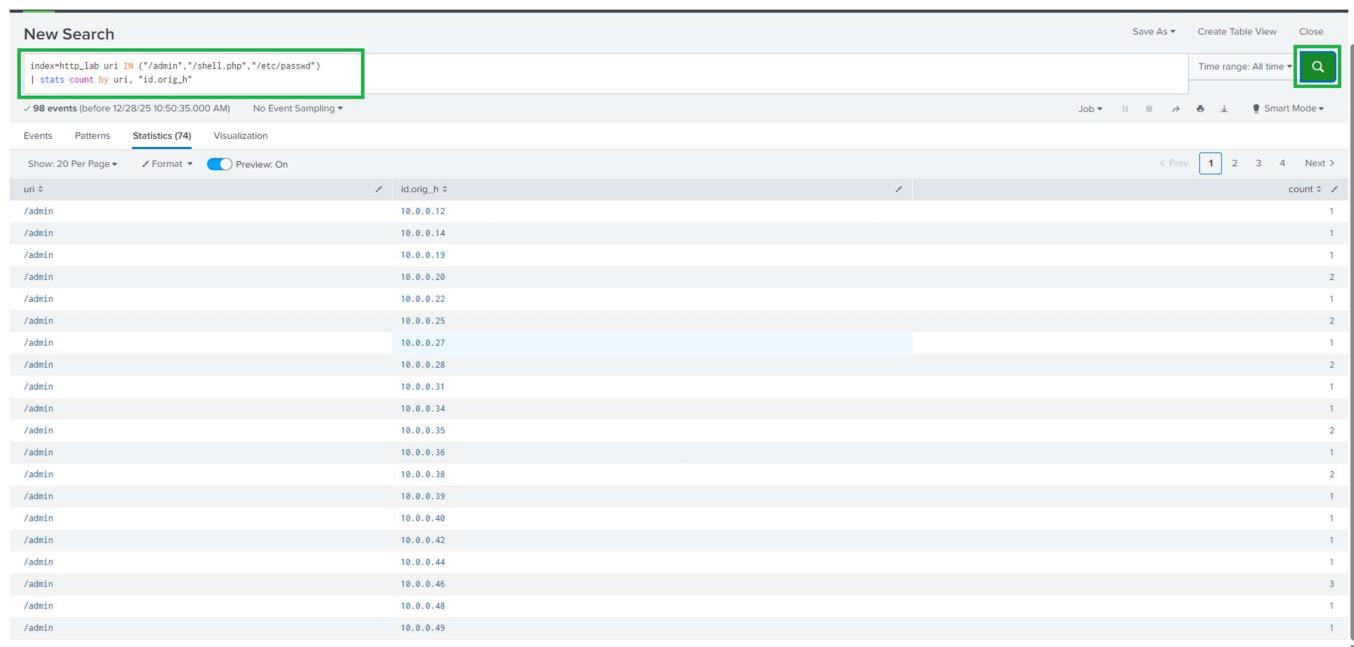
◆ Task 5: Detect Suspicious URI Access Attempts

index=http_lab

uri IN ("/admin","/shell.php","/etc/passwd")

| stats count by uri, "id.orig_h"

📌 *Highlights attempts to access admin panels, shells, or sensitive files.*



💡 Security Insights Gained

Through this lab, you can:

- Detect **malicious web reconnaissance**
- Identify **brute-force or scanning tools**
- Monitor **abnormal HTTP behavior**
- Recognize **potential data exfiltration**
- Build detection logic for **SOC use cases**

Possible Enhancements

-  Create **Splunk alerts** for:
 - Repeated 5xx errors
 - Large file transfers
 - Suspicious User-Agents
 -  Build dashboards for:
 - HTTP status trends
 - Top URLs and IPs
 -  Integrate with IDS / threat intelligence feeds
-

Learning Outcome

By completing this project, you have:

- Gained hands-on experience with **Splunk SPL**
 - Learned **real-world HTTP log analysis**
 - Practiced **threat detection techniques**
 - Strengthened **SIEM & SOC analyst skills**
-

Conclusion

This lab demonstrates how **SIEM tools like Splunk** can be effectively used to analyze web traffic, detect attacks, and improve organizational security posture using log data.

 A must-have project for **Cybersecurity, SOC Analyst**, and **Blue Team** portfolios.
