



Splunk Workshop by Adrian Dolinay: Log Analysis

In this workshop we will focus on Splunk Enterprise. Splunk Enterprise is a platform to collect, analyze and visualize machine generated data. This data includes log files, application metrics, network data and industrial data to name a few.

In this workshop we will utilize Docker to run Splunk. Once Splunk is running we will upload synthetic logs to the platform and analyze the logs with Splunk commands.

Server Log

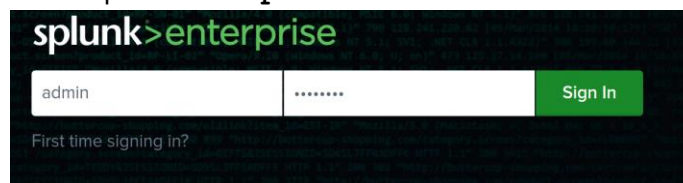
The `"log_data.csv"` is a synthetic log that contains data on server events. The log is saved as a comma separated values (csv) format. There are four fields in the log: **timestamp** (date/time the event occurred), **host** (server or device for the event), **event_type** (login, logout, error) and **event_description** (description of the event). We will upload this log into Splunk and analyze the data.

1. PULLING THE DOCKER IMAGE AND RUNNING SPLUNK

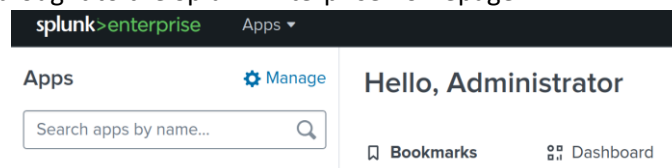
- To pull and run the Splunk image in a container use the command `"docker run -d -p 8000:8000 -p 8088:8088 -p 9997:9997 -e SPLUNK_START_ARGS="--accept-license" -e SPLUNK_PASSWORD="password" --name splunk splunk/splunk:latest"` in your terminal. Docker pulls down the Splunk image, it will take a few minutes.
- Run `"docker ps -a"`. You should see a Splunk container running. Under the `"status"` header, once the status displays as `"Up (healthy)"` you can access Splunk. If you see `"health: starting"`, then Splunk has not fully started yet.

2. ACCESSING SPLUNK

- Bring up a web browser and in the search bar run `"http://localhost:8000/"`.
- The Username is `"Admin"` and the password is `"password"`.

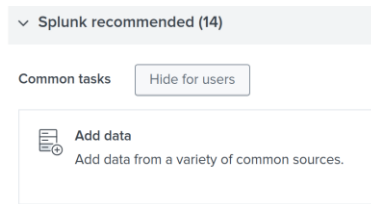


- Once logged in you will be brought to the Splunk Enterprise homepage.



3. ADDING DATA TO SPLUNK

- On the homepage select **"Add Data"**.

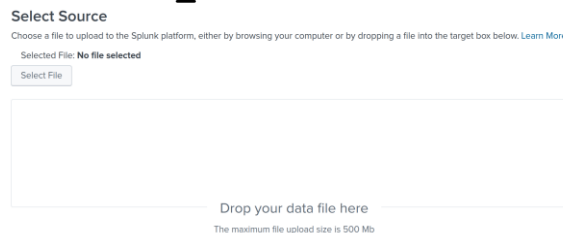


- On the next page select **"Upload files from my computer"**.

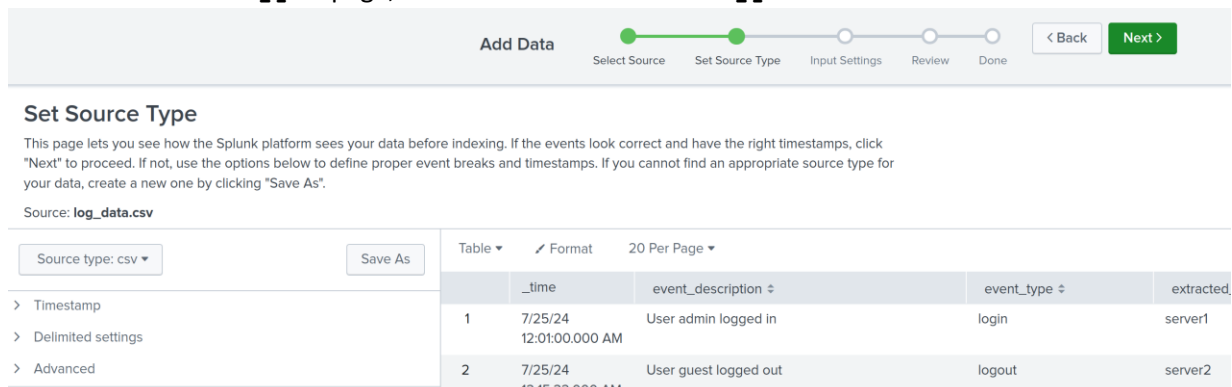


Upload
files from my computer

- Select **"Choose File"** and upload the **"log_data.csv"** file or drag and drop the file into the page.



- On the **"Set Source Type"** page, ensure that the **"Source type"** is set as a CSV. Select next.



- On the **"Input Settings"** in the **"Host field value"** type **"on_prem_servers"**. The **"Host field value"** denotes the machine where the uploaded log came from.

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

Host field value

on_prem_servers

- Still on the **"Input Settings"** page select **"Create a new index"**. Type **"servers"** for the Index Name. Use the default setting for the rest of the fields and select **"Save"**. Select **"Review"**.

New Index X

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

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

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

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 GB ▾
Maximum target size of entire index.

Save
Cancel

- Select **“Submit”** on the **“Review”** page.

Add Data

●
●
●
●
○

Select Source
Set Source Type
Input Settings
Review
Done

< Back
Submit >

Review

Input Type Uploaded File

File Name log_data.csv

Source Type csv

Host on_prem_servers

Index Default

4. BASIC SPLUNK SEARCHES

- Go back to the home page by selecting the **“Splunk Enterprise”** logo.
- Under **“Apps”** select **“Search and Reporting”**.
- Next to the magnifying glass icon, select the **“Last 24 hours”** and change it to **“All time”** under **“Other”**.
- Run the command **“index=servers”** in the search. This search retrieves all events that are stored in the **“servers”** index. It does not apply any additional filters or constraints, so it will return every event in the **“servers”** index.
- For the next search run **“index= servers | stats count by event_type”**. This search will bucket all of the **“events”** by their corresponding **“event types”**. Within our log we have 37 **“errors”**, 72 **“logins”** and 38 **“logouts”**. To look further into a specific event, select the event type and select **“View Events”**.

5. CREATING GRAPHS

- To visualize all of the events in the given log we can search for **“index=servers earliest="07/25/2024:00:00:00" latest="07/30/2024:23:59:59" | timechart span=1d count”**. To visualize the search select the **“Visualization”** tab. This will list out all the dates from July 25th to the 30th in single day intervals.

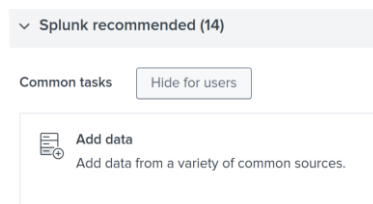
- We can breakout the bar charts by event type (login, logout, error) by searching the command `"index=servers earliest="07/25/2024:00:00:00" latest="07/30/2024:23:59:59" | timechart span=1d count by event_type"`.
- Under the same search, change the chart type by selecting **"Column Chart"** and select the **"Line Chart"**.
- To get the proportions of events by type we can run `"index=servers earliest="07/25/2024:00:00:00" latest="07/30/2024:23:59:59" | stats count by event_type | sort -count"`. Change the chart type by selecting **"Line Chart"** and select the **"Pie Chart"**.

Azure Cloud Log

The **"azure_log.json"** is a synthetic log that contains events for Microsoft Azure cloud activities. The log includes a range of information about operations and activities that occur within an Azure subscription, such as resource modifications, service health events, and other management operations. The log is formatted as a JavaScript Object Notation (JSON) file. We will upload this log into Splunk and analyze the data.

1. ADD AWS DATA TO SPLUNK

- On the homepage select **"Add Data"**.

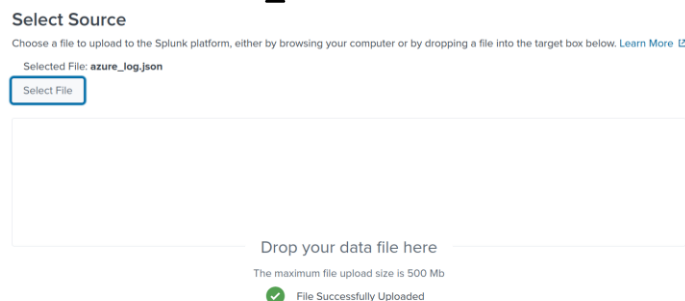


- On the next page select **"Upload files from my computer"**.



Upload
files from my computer

- Select **"Choose File"** and upload the **"azure_log.json"** file or drag and drop the file into the page.



- On the **"Set Source Type"** page, ensure that the **"Source type"** is set as **"_json"**. Select next.

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

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: `azure_log.json`

Source type: _json Save As

Timestamp

Advanced

	_time	callerIpAddress	category	correlationId	durationMs	identityType	identityUserId	operationName
1	8/1/24 8:00:00.000 AM	192.168.1.10	Administrative	abcdef12-3456-7890- abcd-ef1234567890	1300	User	user1@example.com	Microsoft.Compute/ virtualMachines/start/ action
2	8/1/24 8:30:00.000 AM	192.168.1.10	Administrative	abcdef12-3456-7890- abcd-ef1234567890	1500	User	user1@example.com	Microsoft.Compute/ virtualMachines/restart/ action

- On the "Input Settings" in the "Host field value" type "azure_instances". The "Host field value" denotes the machine where the uploaded log came from.

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

Host field value

- Still on the "Input Settings" page select "Create a new index". Type "azure" for the Index Name. Use the default setting for the rest of the fields and select "Save". Select "Review".

New Index X

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

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

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

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 GB ▾
Maximum target size of entire index.

Max Size of Hot/Warm/Cold Bucket GB ▾
Maximum target size of buckets. Enter 'auto_high_volume' for high-volume indexes.

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

App

Storage Optimization

Tsidx Retention Policy ☒ Enable Reduction ☐ Disable Reduction
Warning: Do not enable reduction without understanding the full implications. It is extremely difficult to rebuild reduced buckets. [Learn More](#)

Reduce tsidx files older Days ▾

Save
Cancel

- Select "Submit" on the "Review" page.

Add Data

Progress: Select Source, Set Source Type, Input Settings, **Review**, Done

Review

Input Type Uploaded File
 File Name azure_log.json
 Source Type _json
 Host azure_instances
 Index azure

< Back Submit >

2. SPLUNK SEARCHES FOR AZURE LOG

- Go back to the home page by selecting the **"Splunk Enterprise"** logo.
- Under **"Apps"** select **"Search and Reporting"**.
- Next to the magnifying glass icon, select the **"Last 24 hours"** and change it to **"All time"** under **"Other"**.
- Run the command **"index=azure"** in the search. This search retrieves all events that are stored in the **"azure"** index. It does not apply any additional filters or constraints, so it will return every event in the **"azure"** index.
- Run the command **"index="azure " earliest="08/01/2024:00:00:00" latest="08/05/2024:23:59:59"**. This will retrieve all of the logs for a certain date range.

3. SPLUNK ADVANCED SEARCHES FOR AZURE LOG

- Run **"index="azure" | stats count by operationName"**. This retrieves the count of each operation performed. To explain some events, the **"Microsoft.Compute/virtualMachines/restart/action"** is when a virtual machine restarts. The **"Microsoft.Network/networkSecurityGroups/securityRules/delete"** is when a security rule is deleted.
- Run **"index="azure" | stats count by callerIpAddress | sort -count"**. This command identifies which IP addresses are most active within the Azure environment by counting how many events are associated with each IP address.
- Run **"index="azure_logs" operationName="Microsoft.Storage/storageAccounts/delete"**. The command is used in identifying specific events where storage accounts were deleted within an Azure environment. It helps in monitoring and auditing Azure resources. Running **"index="azure_logs" operationName="Microsoft.Storage/storageAccounts/delete" | sort -time"** sorts the storage account deletions by time.

4. CREATING GRAPHS

- To visualize the activities by resource type, run **"index="azure" | rex field=resourceId "(?<resourceType>Microsoft\\.\\w+\\/\\w+)" | stats count by resourceType | sort -count"**. Select the **"Visualization"** tab and select **"Bar Chart"**. From the charts we can see that there have been a high number of events for Azure Virtual Machine and a single event for the Azure Network Interface resource.
- To visualize the number of operations by IP Address, run **"index="azure" | stats count by callerIpAddress | sort -count"** and make sure the **"Bar Chart"** graph is selected.