Introduction to SIEMs using Splunk

Cybersecurity

SIEMs Unit, Day 1

Class Objectives

By the end of class today, students will be able to:

- □ Recognize the role SIEMs play in protecting an organization's security.
- □ Explain how logs are filtered, normalized, and correlated for events.
- □ Demonstrate how to use basic features of the Splunk User Interface.
- □ Explain basic database terms and query functions.
- □ Use the Splunk Processing Language (SPL) for simple queries.

Introduction to SIEMS

What are SIEMs?

The Scenario:

An organization's IT infrastructure has a variety of systems and applications on their networks, including host systems, product applications, network devices, firewalls, etc.

The Problem:

The Solution:

What are SIEMs?

The Scenario:

An organization's IT infrastructure has a variety of systems and applications on their networks, including host systems, product applications, network devices, firewalls, etc.

The Problem:

It’s challenging for organizations to have full visibility of their network, making **suspicious behavior more difficult to detect**.

The Solution:

What are SIEMs?

The Scenario:

An organization's IT infrastructure has a variety of systems and applications on their networks, including host systems, product applications, network devices, firewalls, etc.

The Problem:

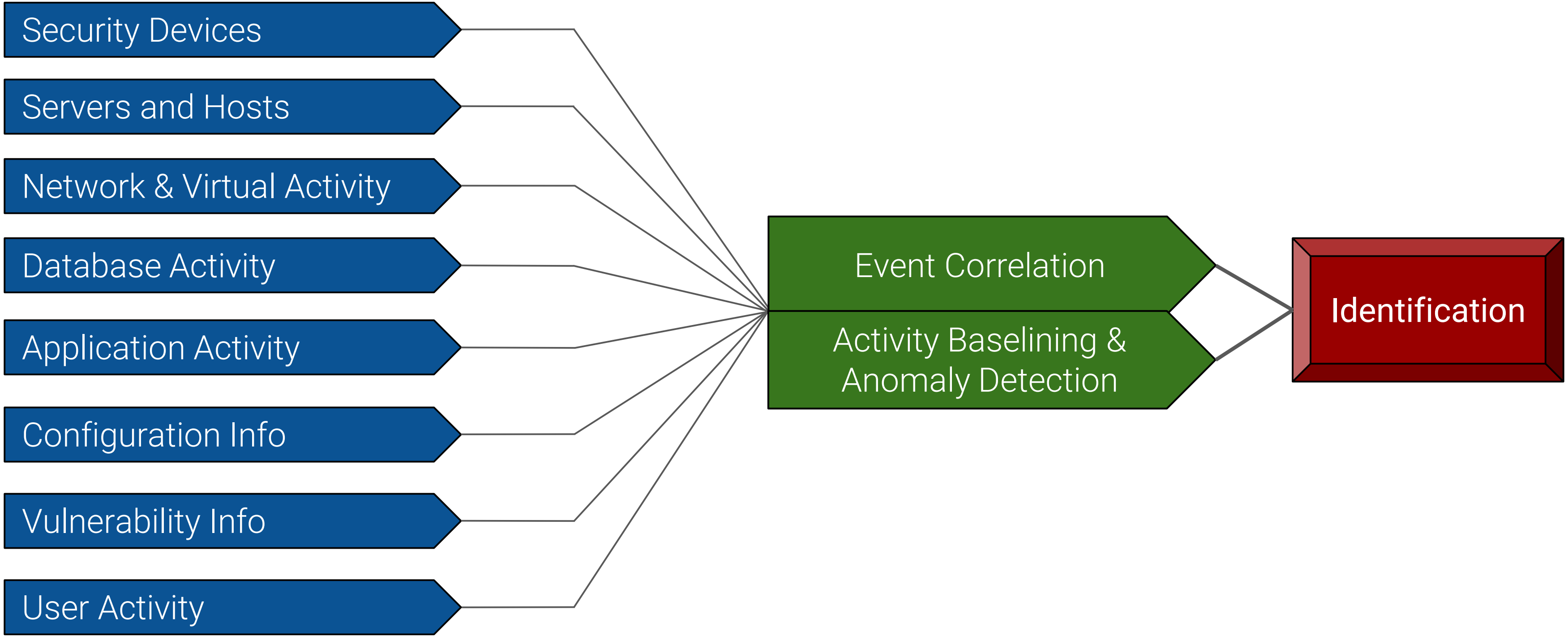
It's challenging for organizations to have full visibility of their network, making **suspicious behavior more difficult to detect**.

The Solution:

Organizations use **SIEMs (Security Information and Event Management)** to monitor odd behavior and irregular traffic on their network, allowing infosec teams to detect suspicious activity.

SIEMS

A SIEM connects and unifies the information from various sources, allowing data to be analyzed and cross referenced.



SIEMS

A SIEM connects and unifies the information from various sources, allowing data to be analyzed and cross referenced.

- ✅ SIEMs provide **real-time monitoring** of machine data to correlate events, baselines, notifications, alerts, reports and visualization.

- ❌ SIEMs *do not* support **security analytics**, which create behaviors and profiles using machine learning and applies statistical analysis to detect anomalies that could indicate potential threats.

The Benefits of SIEMS

Using SIEMs provides the following benefits:

- Increased efficiency
- Prevention of potential security threats
- Reduced impact of security breaches
- Reduced costs
- Better reporting, log analysis, and retention
- IT Compliance

SIEMs Use-Cases

Security Operations Center (SOC) and **Security Analysts** (SA) Teams use SIEMs to analyze data in order to detect malicious cyber attacks across devices, systems, applications and network infrastructures. In this context, SIEMs are used for:

Threat Hunting:

- Developing and testing a hypothesis by exploring logs and searching data for security patterns similar to current incidents.

Detecting Data Exfiltration:

- Detecting emails sent to persons other than the intended recipient.
- Identifying and reporting on excessive `print` and `send` alerts to trouble ticket systems.

SIEMs Use-Cases

Security Operations Center (SOC) and **Security Analysts** (SA) Teams use SIEMs to analyze data in order to detect malicious cyber attacks across devices, systems, applications and network infrastructures. In this context, SIEMs are used for:

Internet of Things (IoT) Security

- Identifying unusual traffic that results in Denial of Service (DoS) attacks.
- Identifying suspicious behavior on devices.

Privileged Access Abuse

- Monitoring suspicious access to sensitive data.
- Reporting on users that are accessing data outside their regular profile.
- Reporting on activity of terminated user accounts.

SIEMS is not one single application…

Like Linux distributions, SIEMS are a class of tools that come in a variety of products and vendors depending on a user’s specific need.

Student Activity: SIEM Vendor Research

In this activity, you will research other SIEM vendor options.

💡 Although we'll focus on Splunk today, it is important to familiarize yourself with a variety of SIEM products and vendors.

Suggested Time:
15 minutes

Your Turn! SIEM Vendor Research

Instructions:

1. Break out into groups of 2-3 students.

1. Using Google, research 3-5 SIEM vendors (besides Splunk).

1. List the following information:

   - Company Name

   - Product Name

   - Capabilities (Advantages/Disadvantages)

SIEM Vendor Research Review

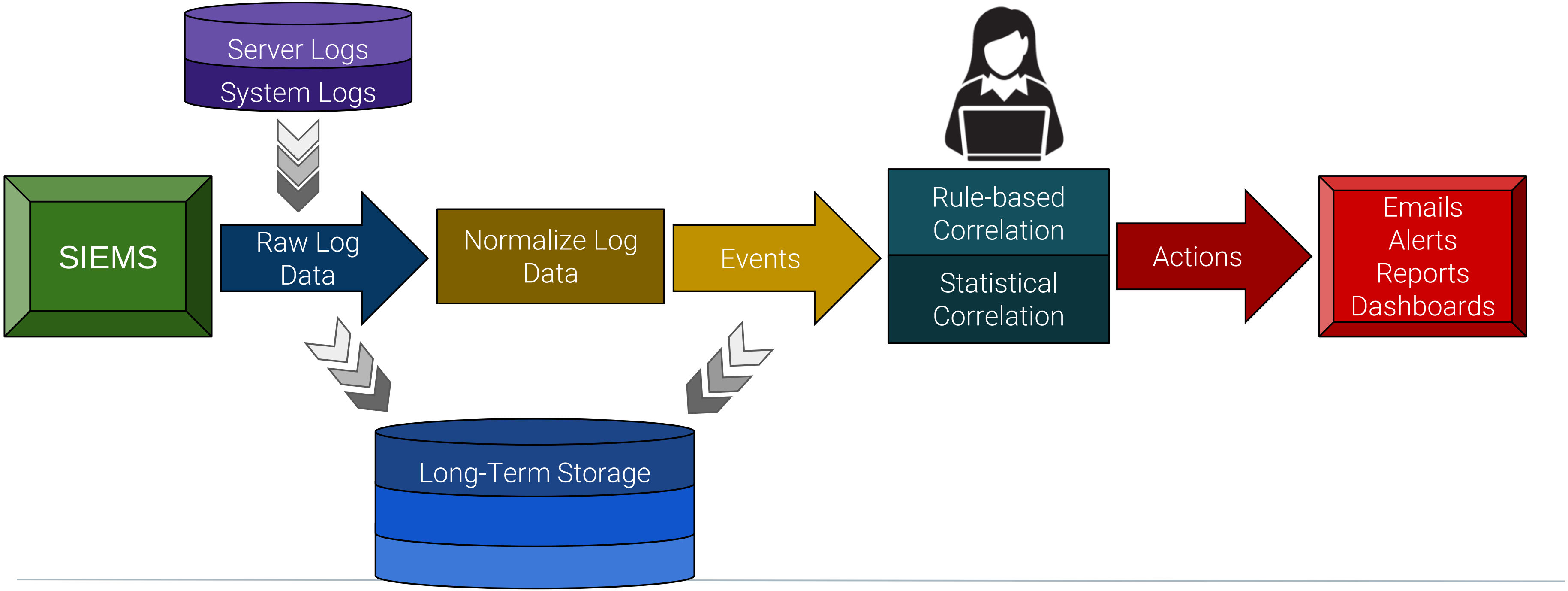
Top SIEM Vendors

●●●● Best ●●● Very Good ●● Good ● Fair

	Threats Blocked	Sources Ingested	Performance	Value	Implementation	Management	Support	Scalability
splunk > ES	●●●●	●●●	●●●	●●●	●●	●●●	●●	●●●
LogRhythm	●●●	●●●●	●●●	●●	●●●	●●●	●●●	●●
ALIEN VAULT	●●●	●●●	●●●	●●●●	●●●	●●	●●	●●●
MICRO FOCUS ArcSight	●●	●●●	●●●	●●	●●●	●●●●	●●	●●●
MICRO FOCUS Sentinel	●●	●●	●●	●●●	●●●	●●●	●●	●●●
McAfee ESM	●●●	●●●	●●●	●●●	●●	●●	●●●	●●●
Trustwave SIEM	●●●	●●●	●●●	●●●	●●	●●●	●●	●●●●
IBM QRadar	●●●	●●●	●●●●	●●●	●●	●●●	●●●	●●●
RSA NetWitness	●●	●●	●●●	●●	●●	●●	●●●	●●●
solarwind LEM	●●	●●●	●●	●●	●●●●	●●	●●●	●●

Logs, Events, Analysis and Response

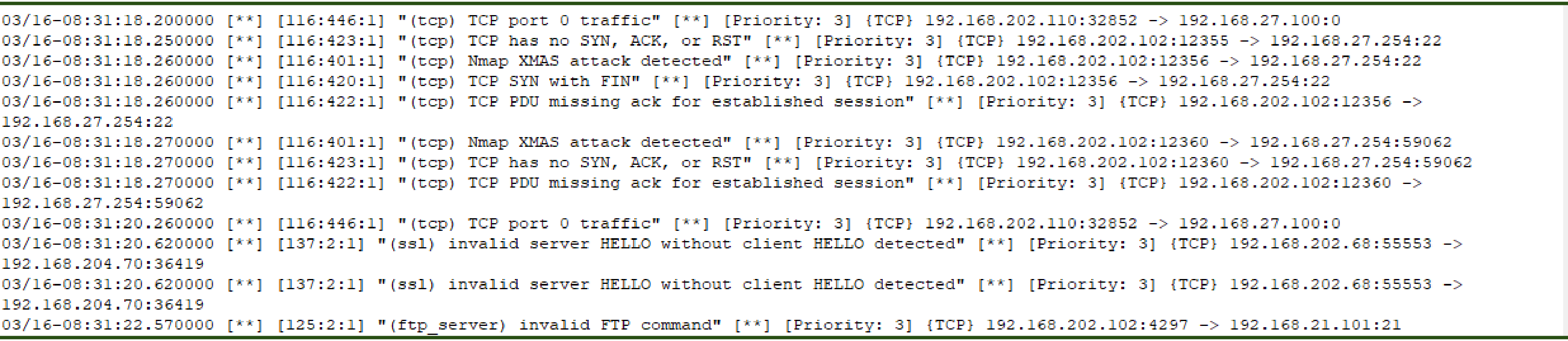
Now, we'll focus on how log data is processed and analyzed as events.



How do SIEMs Parse, Identify, and Assist in Event Analysis

Collecting Raw Log Data

- Raw log data is collected in real-time from various sources.
- Below is an example of TCP, SSL, and FTP events from a network log.



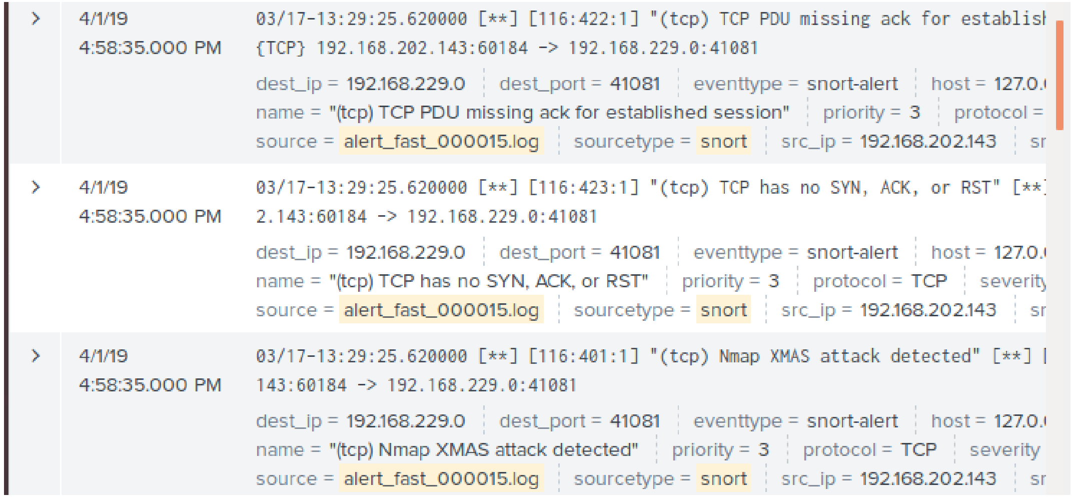
```
03/16-08:31:18.200000 [**] [116:446:1] "(tcp) TCP port 0 traffic" [**] [Priority: 3] {TCP} 192.168.202.110:32852 -> 192.168.27.100:0
03/16-08:31:18.250000 [**] [116:423:1] "(tcp) TCP has no SYN, ACK, or RST" [**] [Priority: 3] {TCP} 192.168.202.102:12355 -> 192.168.27.254:22
03/16-08:31:18.260000 [**] [116:401:1] "(tcp) Nmap XMAS attack detected" [**] [Priority: 3] {TCP} 192.168.202.102:12356 -> 192.168.27.254:22
03/16-08:31:18.260000 [**] [116:420:1] "(tcp) TCP SYN with FIN" [**] [Priority: 3] {TCP} 192.168.202.102:12356 -> 192.168.27.254:22
03/16-08:31:18.260000 [**] [116:422:1] "(tcp) TCP PDU missing ack for established session" [**] [Priority: 3] {TCP} 192.168.202.102:12356 ->
192.168.27.254:22
03/16-08:31:18.270000 [**] [116:401:1] "(tcp) Nmap XMAS attack detected" [**] [Priority: 3] {TCP} 192.168.202.102:12360 -> 192.168.27.254:59062
03/16-08:31:18.270000 [**] [116:423:1] "(tcp) TCP has no SYN, ACK, or RST" [**] [Priority: 3] {TCP} 192.168.202.102:12360 -> 192.168.27.254:59062
03/16-08:31:18.270000 [**] [116:422:1] "(tcp) TCP PDU missing ack for established session" [**] [Priority: 3] {TCP} 192.168.202.102:12360 ->
192.168.27.254:59062
03/16-08:31:20.260000 [**] [116:446:1] "(tcp) TCP port 0 traffic" [**] [Priority: 3] {TCP} 192.168.202.110:32852 -> 192.168.27.100:0
03/16-08:31:20.620000 [**] [137:2:1] "(ssl) invalid server HELLO without client HELLO detected" [**] [Priority: 3] {TCP} 192.168.202.68:55553 ->
192.168.204.70:36419
03/16-08:31:20.620000 [**] [137:2:1] "(ssl) invalid server HELLO without client HELLO detected" [**] [Priority: 3] {TCP} 192.168.202.68:55553 ->
192.168.204.70:36419
03/16-08:31:22.570000 [**] [125:2:1] "(ftp_server) invalid FTP command" [**] [Priority: 3] {TCP} 192.168.202.102:4297 -> 192.168.21.101:21
```

How do SIEMs Parse, Identify, and Assist in Event Analysis

Normalize Raw Log Data

- The raw log data is mapped to various elements, such as source and destination IP in order to produce a *common format* or *metadata* for event types.
- The data is **indexed.** Extended term storage is provided for both raw and event data.

- **Right:** For example, in this event, the *destination ip* and *destination port* have been parsed, identified and labeled.



```
>   4/1/19            03/17-13:29:25.620000 [**] [116:422:1] "(tcp) TCP PDU missing ack for establis
    4:58:35.000 PM    {TCP} 192.168.202.143:60184 -> 192.168.229.0:41081
                      dest_ip = 192.168.229.0 | dest_port = 41081 | eventtype = snort-alert | host = 127.0.
                      name = "(tcp) TCP PDU missing ack for established session" | priority = 3 | protocol =
                      source = alert_fast_000015.log | sourcetype = snort | src_ip = 192.168.202.143 | sr

>   4/1/19            03/17-13:29:25.620000 [**] [116:423:1] "(tcp) TCP has no SYN, ACK, or RST" [**
    4:58:35.000 PM    2.143:60184 -> 192.168.229.0:41081
                      dest_ip = 192.168.229.0 | dest_port = 41081 | eventtype = snort-alert | host = 127.0.
                      name = "(tcp) TCP has no SYN, ACK, or RST" | priority = 3 | protocol = TCP | severity
                      source = alert_fast_000015.log | sourcetype = snort | src_ip = 192.168.202.143 | sr

>   4/1/19            03/17-13:29:25.620000 [**] [116:401:1] "(tcp) Nmap XMAS attack detected" [**] |
    4:58:35.000 PM    143:60184 -> 192.168.229.0:41081
                      dest_ip = 192.168.229.0 | dest_port = 41081 | eventtype = snort-alert | host = 127.0.
                      name = "(tcp) Nmap XMAS attack detected" | priority = 3 | protocol = TCP | severity
                      source = alert_fast_000015.log | sourcetype = snort | src_ip = 192.168.202.143 | sr
```

How do SIEMs Parse, Identify, and Assist in Event Analysis

Correlate Event and State Data

- Correlation is the driver for analysis and the actions.
- SIEMs capture *state data* which differentiates SIEMs from log management systems.
  - State data is information regarding the *full state* of a system: configurations, current applications, active users, processes, registry settings and vulnerabilities.
  - Understanding the full state of the system is the foundation for all security-related decisions.
- Event correlation is the process of finding relationships between seemingly unrelated events in data.
- We can use **rules** and **statistical analysis** to correlate data.

Rules-based Event Correlation

A SIEM correlation rule indicates which sequence of events could be indicative of anomalies.

```
A connection event

AND

 A failed login event

AND

An application launched in some place in the system

Action: Create an Alert

This may be a system compromise or insider abuse of system privileges.
```

Statistical Event Analysis

Statistical analysis can be used to flag more latent data and events that resemble other events.

Statistical methods use tools like:

- **Frequencies**
- **Baselines**
- **Thresholds**

Statistical data can also be visualized using charts, dashboards, metrics, and other methods.

Let’s take a look at some statistical techniques...

Statistical Techniques

Frequencies, in general, are an event’s number of occurrences in a specific period of time.

- In SIEMs, frequencies are useful because we can count source and destination IP addresses in an incoming log across all log sources.

- For example, a spike in the number of occurrences of a destination IP address can be an early warning sign that someone is targeting an attack on this system.

Statistical Techniques

Baselines provide a measure of what is normal in a SIEMs data set.

Baselines are calculated over an elongated period of time.

- Once the baseline is set, we can monitor unusual data that falls outside of the baseline. This data can serve as an early warning for targeted attacks to the system.

Baseline usage includes:

- User logons and logoffs, both successful and failed.
- Network traffic bytes, inbound and outbound.
- Network traffic to particular ports, services and protocols.
- Administrative account usage and access.
- Processes running on a server.

Statistical Techniques

Thresholds are used for determining when something exceeds a baseline value.

For example: time and frequency thresholds include:

- A server that normally receives an average of 20 failed logins *per hour* and is now receiving 50.
- The number of hits on port 443 over the *last week*.
- User logins to a server X *times per day*.
- Use of `su` command *per hour of day*.

Activity: Working with Logs and Events

In this activity, you will review a log event in order to see data and understand the normalization process.

Instructions sent via Slack.

Suggested Time:
20 Minutes

Times Up! Let’s Review.

Working with Logs and Events

Working with Logs and Events Review

Part 1:

1. List two benefits of using SIEMS:

2. How are logs used in a SIEMS?

3. What is a baseline and how is it used in a SIEMs

Working with Logs and Events Review

Part 1:

1. List two benefits of using SIEMS: Answers may include:

Increased efficiency; preventing potential security threats; reducing the impact of security breaches; reducing costs; better reporting log analysis, and retention; and IT compliance

2. How are logs used in a SIEMS?

3. What is a baseline and how is it used in a SIEMs

Working with Logs and Events Review

Part 1:

1. List two benefits of using SIEMS: Answers may include:

Increased efficiency; preventing potential security threats; reducing the impact of security breaches; reducing costs; better reporting log analysis, and retention; and IT compliance

2. How are logs used in a SIEMS?

SIEMs collect logs from different sources in real-time and security teams use the data to detect and respond to security incidents.

Log data is processed into events, which are then used in analysis.

3. What is a baseline and how is it used in a SIEMs

Working with Logs and Events Review

Part 1:

1. List two benefits of using SIEMS: Answers may include:

Increased efficiency; preventing potential security threats; reducing the impact of security breaches; reducing costs; better reporting log analysis, and retention; and IT compliance

2. How are logs used in a SIEMS?

SIEMs collect logs from different sources in real-time and security teams use the data to detect and respond to security incidents.

Log data is processed into events, which are then used in analysis.

3. What is a baseline and how is it used in a SIEMs

A baseline provides a measure calculated over an extended period of time in order to determine what is normal in a SIEMS data set.

Working with Logs and Events Review

Part 2: Use the sample log entry to find the following field names and values:

```
192.188.106.240 - - [06/Feb/2019:23:59:45]
"GET http://www.buttercupgames.com/category.screen?
categoryId=TEE&JSESSIONID=SD2SL4FF9ADFF4959 HTTP 1.1" 200 2958
```

Field names and values:

Working with Logs and Events Review

Part 2: Use the sample log entry to find the following field names and values:

```
192.188.106.240 - - [06/Feb/2019:23:59:45]
"GET http://www.buttercupgames.com/category.screen?
categoryId=TEE&JSESSIONID=SD2SL4FF9ADFF4959 HTTP 1.1" 200 2958
```

Field names and values:

client_ip = 192.188.106.240

timestamp = 06/Feb/2019:23:59:45

http_method = GET

http_url= GET http://www.buttercupgames.com/category.screen?categoryId=TEE&JSESSIONID=SD2SL4FF9ADFF4959 HTTP 1.1

status = 200

bytes = 2958

Working with Logs and Events Review

Part 3: Write a pseudocode statement(s) that checks if the `GET` request sent to the server was unsuccessful for the source IP address and the date in our sample log entry.

Working with Logs and Events Review

Part 3: Write a pseudocode statement(s) that checks if the GET request sent to the server was unsuccessful for the source IP address and the date in our sample log entry.

Potential Solution:

```
client_ip is equal to `192.188.106.240`

AND

timestamp is equal to `06/Feb/2019:23:59:45`

AND

status does not equal `200`
```

Working with Logs and Events Review

Extra Challenge: Send an alert if the count of unsuccessful GET responses are greater than 50 in one minute.

```
Potential Solution:
```

Working with Logs and Events Review

Extra Challenge: Send an alert if the count of unsuccessful GET responses are greater than 50 in one minute.

Potential Solution:

```
Store the count of unsuccessful responses in count
if count is greater than 50
AND
time is one minute
send an alert to the trouble ticket system
```

Break

Introduction to Splunk

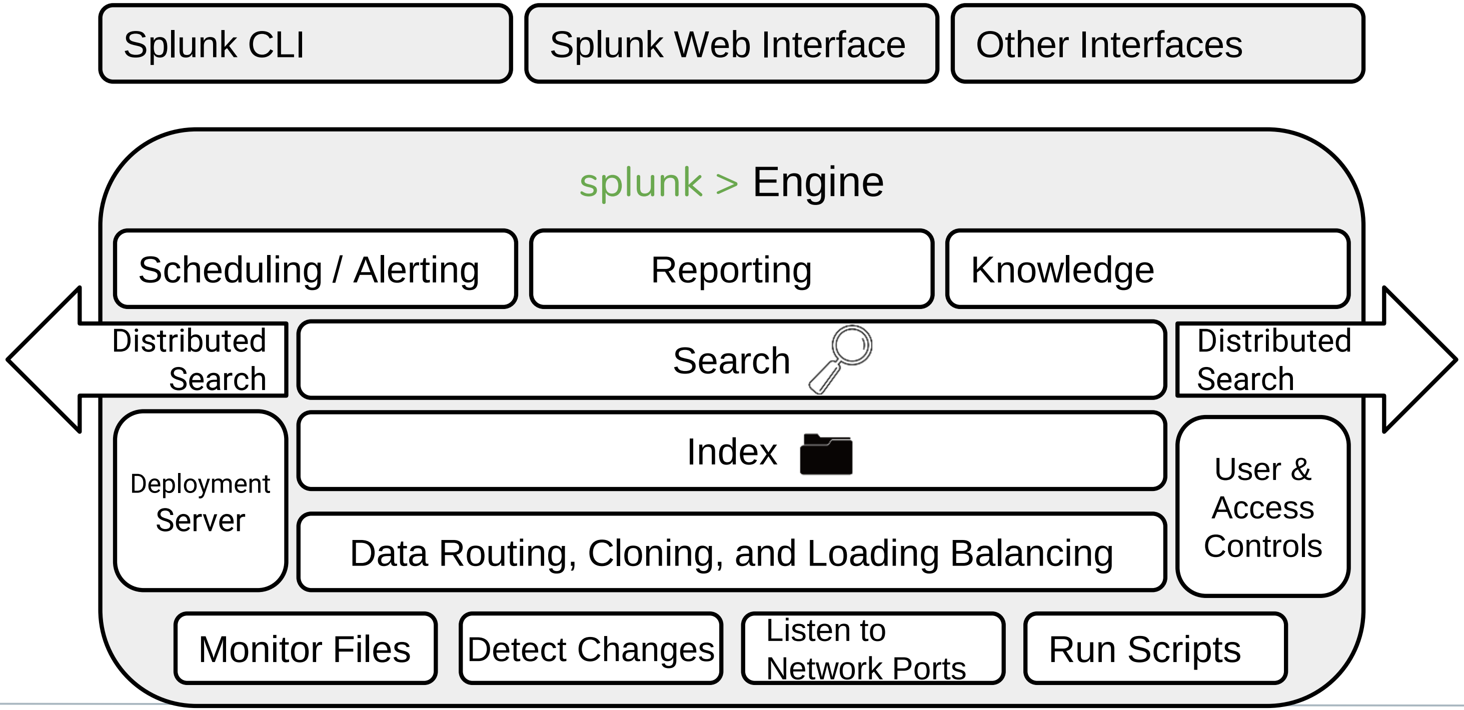
What’s Splunk?

Splunk is a SIEM software platform used to **search**, **analyze** and **visualize** machine generated data gathered from websites, applications, sensors, and other devices that comprise a business’s IT infrastructure.

Splunk offers **one single interface** to view logs from many different sources in **real time.**

Splunk’s Interface

Splunk offers **one single interface** to view logs from many different sources in **real time.**



Splunk CLI

Splunk Web Interface

Other Interfaces

splunk > Engine

Scheduling / Alerting

Reporting

Knowledge

Distributed Search

Search

Distributed Search

Deployment Server

Index

User & Access Controls

Data Routing, Cloning, and Loading Balancing

Monitor Files

Detect Changes

Listen to Network Ports

Run Scripts

Splunk

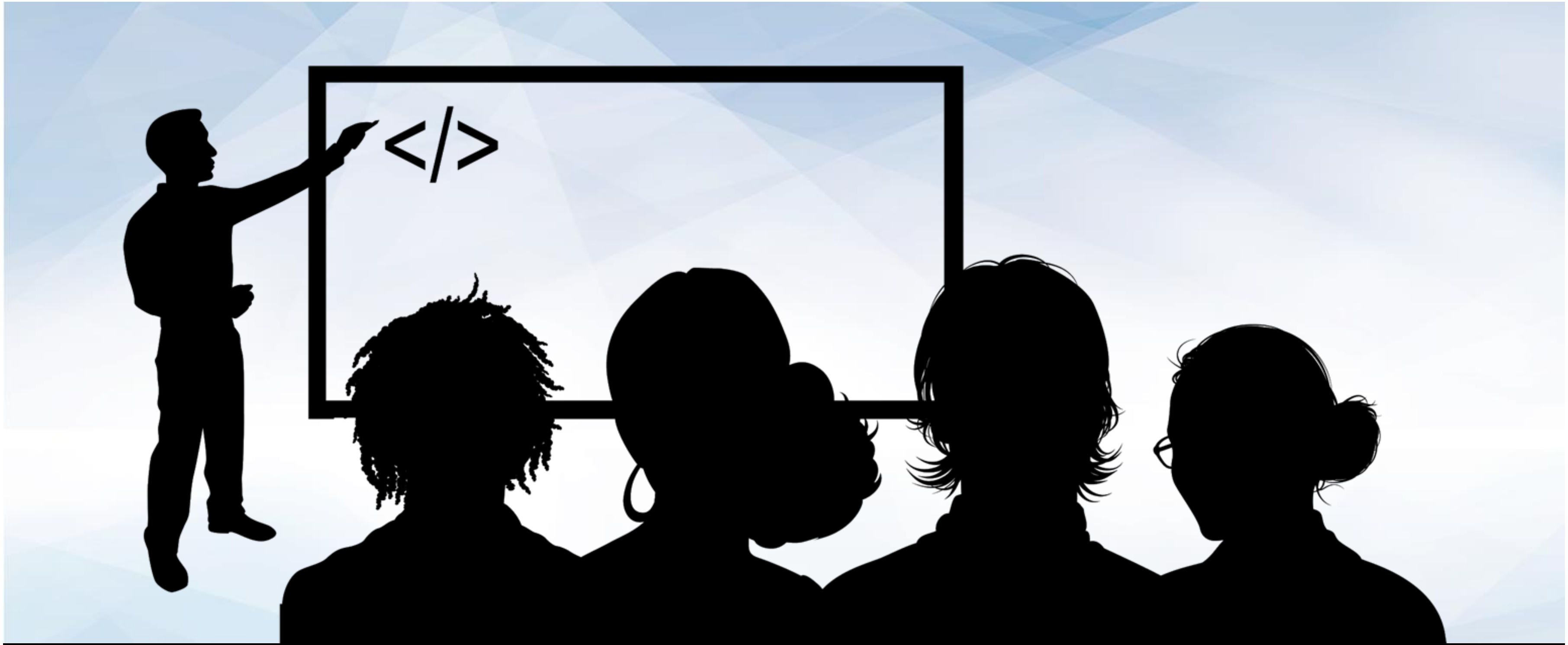
Splunk's Key Features and Functionality:

- Analyze system performance
- Troubleshoot failures
- Monitor business metrics
- Search and investigate a particular outcome
- Create dashboards to visualize and analyze results
- Store and retrieve data for later use

Splunk

Some benefits include the abilities to...

- Input data in any format, like `.csv, json`, etc.
- Splunk can be configured to give alert / event notifications at the onset of a machine state.
- Splunk can accurately predict the resources needed for scaling up the infrastructure.
- Create knowledge objects for Operational Intelligence.



Instructor Demonstration

Activating and Logging into Splunk

Activating Splunk

Your VM already contains the Splunk installation. However, in order to activate it, you will need to complete the following steps:

- Open your terminal window:
- Type: `start_splunk`
- Type: `Y` (to accept the terms of service)
- Type: `student` for the username
- Type: `cybersecurity` for the password

Log into Splunk

- Open a web browser
- Navigate to:``http://127.0.0.1:8000`` once Splunk has finished setting up.
- Enter your username and password.

You're now using Splunk!

Student Activity: Activate Splunk

Splunk should already be installed on your machines.
In this activity, you will activate and log into Splunk.

Instructions sent via Slack.

Suggested Time:
10 minutes

Your Turn! Activating Splunk

Instructions

Your VM already contains the Splunk installation. However, in order to activate it, you will need to complete the following steps:

- Open your terminal window:
- Type: `start_splunk`
- Type: `Y` (to accept the terms of service)
- Type: `student` for the username
- Type: `cybersecurity` for the password

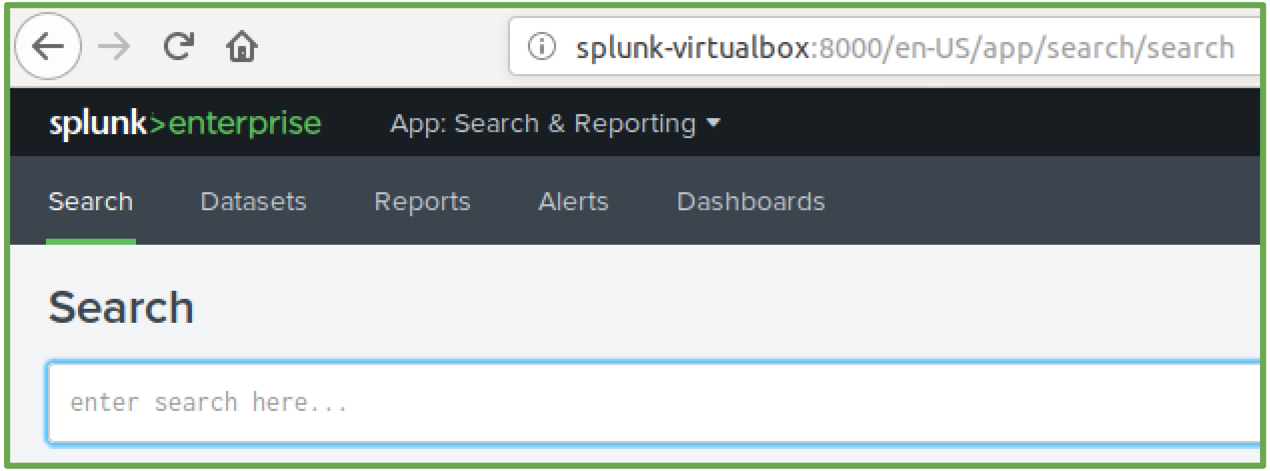
Log into Splunk

- Navigate to:``http://127.0.0.1:8000`` once Splunk has finished setting up.
- Enter your username and password.

Splunk User Interface

Introduction to the Splunk User Interface

The **Search & Reporting Application** is the primary interface for running searches, saving reports, and creating alerts and dashboards.



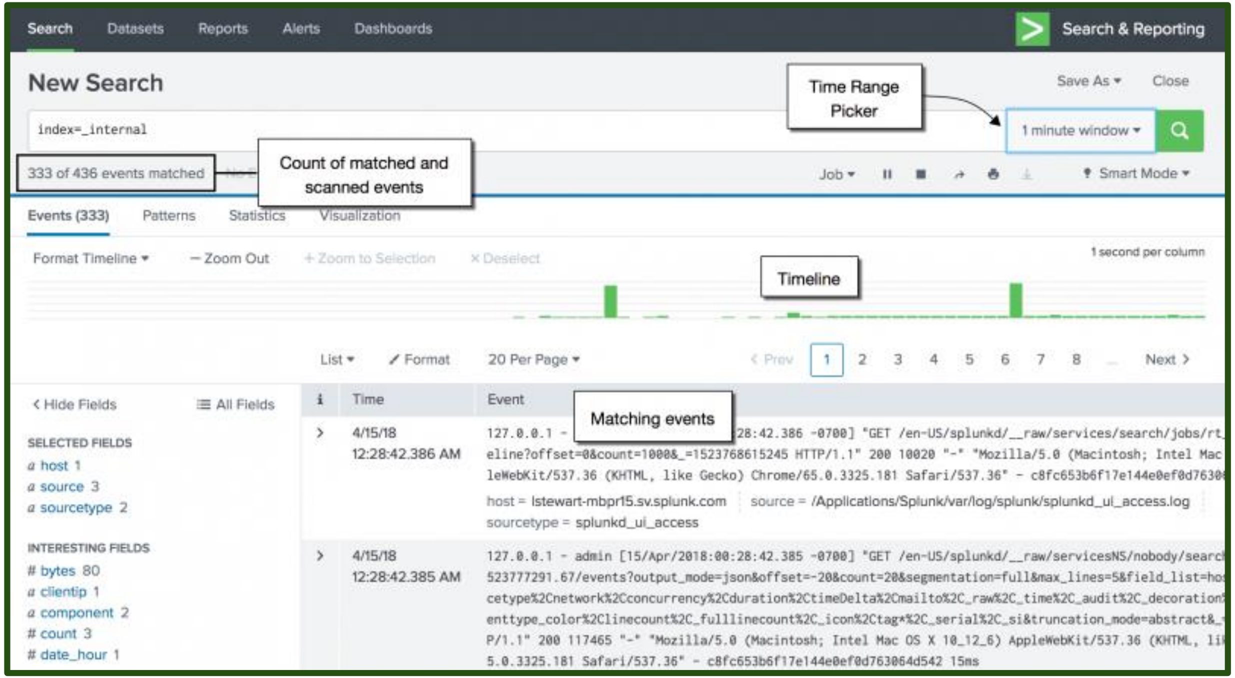
splunk-virtualbox:8000/en-US/app/search/search

Search

The **Search** option provides the ability to search events from indexed data . (For example: Windows event logs, web server logs, live application logs, network feeds.)

Security teams run a series of time-based searches to investigate and identify abnormal activity.

The timeline is used to drill into specific **time periods** (minutes, hours, days) in order to correlate events that occur around the same time.



Time Range Picker

Count of matched and scanned events

Timeline

Matching events

```
index=_internal
```

333 of 436 events matched

SELECTED FIELDS
a host 1
a source 3
a sourcetype 2

INTERESTING FIELDS
# bytes 80
a clientip 1
a component 2
# count 3
# date_hour 1

```
> 4/15/18          127.0.0.1 -                       28:42.386 -0700] "GET /en-US/splunkd/__raw/services/search/jobs/rt
  12:28:42.386 AM  eline?offset=0&count=1000&_=1523768615245 HTTP/1.1" 200 10020 "-" "Mozilla/5.0 (Macintosh; Intel Mac
                   leWebKit/537.36 (KHTML, like Gecko) Chrome/65.0.3325.181 Safari/537.36" - c8fc653b6f17e144e0ef0d7630
                   host = lstewart-mbpr15.sv.splunk.com   source = /Applications/Splunk/var/log/splunk/splunkd_ui_access.log
                   sourcetype = splunkd_ui_access

> 4/15/18          127.0.0.1 - admin [15/Apr/2018:00:28:42.385 -0700] "GET /en-US/splunkd/__raw/servicesNS/nobody/searc
  12:28:42.385 AM  523777291.67/events?output_mode=json&offset=-20&count=20&segmentation=full&max_lines=5&field_list=ho
                   cetype%2Cnetwork%2Cconcurrency%2Cduration%2CtimeDelta%2Cmailto%2C_raw%2C_time%2C_audit%2C_decoration
                   enttype_color%2Clinecount%2C_fulllinecount%2C_icon%2Ctag*%2C_serial%2C_si&truncation_mode=abstract&_
                   P/1.1" 200 117465 "-" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_12_6) AppleWebKit/537.36 (KHTML, li
                   5.0.3325.181 Safari/537.36" - c8fc653b6f17e144e0ef0d763064d542 15ms
```

Datasets

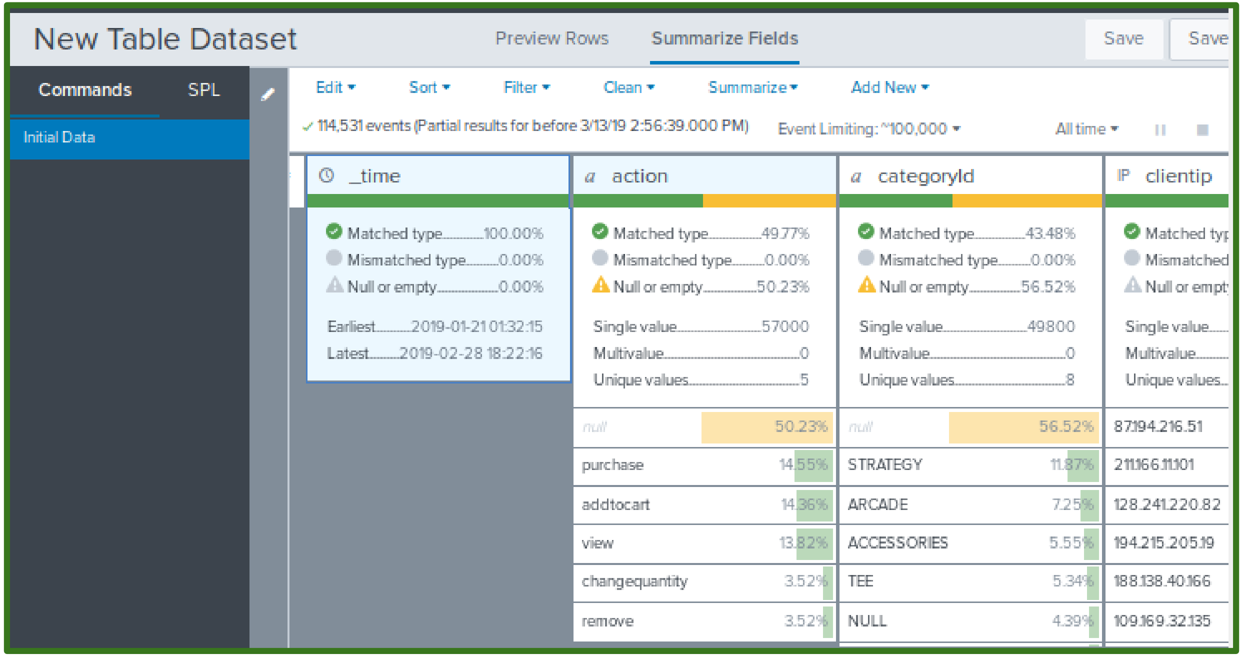
A **dataset** is a collection of data that is defined and maintained for a specific business purpose.

Datasets are used when:

1. You do NOT know the Search Processing Language (SPL).

2. You want to avoid spending time designing complicated searches.

Datasets are created using a **Table Editor** in which events fields are selected for a search.



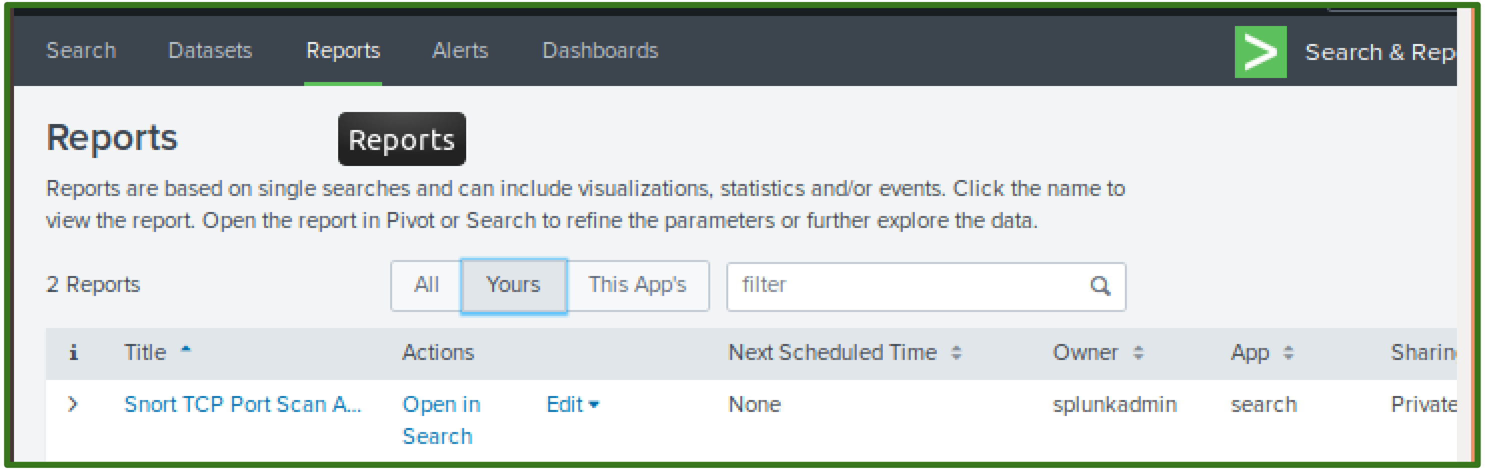
✓ 114,531 events (Partial results for before 3/13/19 2:56:39.000 PM)

◷ _time	a action	a categoryId	IP clientip
Matched type 100.00%	Matched type 49.77%	Matched type 43.48%	Matched typ
Mismatched type 0.00%	Mismatched type 0.00%	Mismatched type 0.00%	Mismatched
Null or empty 0.00%	Null or empty 50.23%	Null or empty 56.52%	Null or empt
Earliest 2019-01-21 01:32:15	Single value 57000	Single value 49800	Single value
Latest 2019-02-28 18:22:16	Multivalue 0	Multivalue 0	Multivalue
	Unique values 5	Unique values 8	Unique values
	null 50.23%	null 56.52%	87.194.216.51
	purchase 14.55%	STRATEGY 11.87%	211.166.11.101
	addtocart 14.36%	ARCADE 7.25%	128.241.220.82
	view 13.82%	ACCESSORIES 5.55%	194.215.205.19
	changequantity 3.52%	TEE 5.34%	188.138.40.166
	remove 3.52%	NULL 4.39%	109.169.32.135

Reports

Reports are created when you save a search. They can also be created manually. They can be scheduled to run at intervals and used in the dashboard.

PDF documents can also be generated from reports.



Reports

Reports are based on single searches and can include visualizations, statistics and/or events. Click the name to view the report. Open the report in Pivot or Search to refine the parameters or further explore the data.

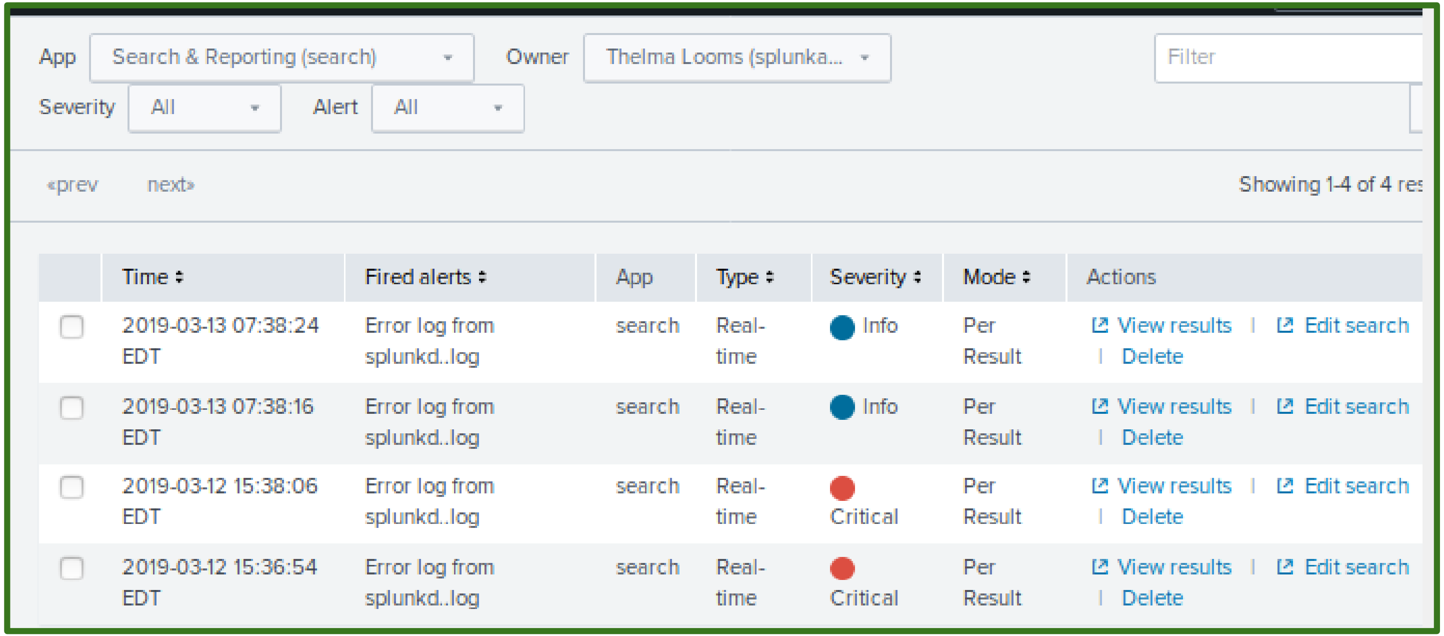
2 Reports

i	Title	Actions		Next Scheduled Time	Owner	App	Sharin
›	Snort TCP Port Scan A...	Open in Search	Edit ▾	None	splunkadmin	search	Private

Alerts

Alerts are used to monitor for and respond to specific events.

- They can look for events in *real time* or on at *scheduled intervals*.
- Alerts can be assigned a priority, such as Informational or Critical.



App: Search & Reporting (search) Owner: Thelma Looms (splunka... Filter

Severity: All Alert: All

«prev next» Showing 1-4 of 4 res

	Time ⇕	Fired alerts ⇕	App	Type ⇕	Severity ⇕	Mode ⇕	Actions
□	2019-03-13 07:38:24 EDT	Error log from splunkd..log	search	Real-time	● Info	Per Result	View results | Edit search | Delete
□	2019-03-13 07:38:16 EDT	Error log from splunkd..log	search	Real-time	● Info	Per Result	View results | Edit search | Delete
□	2019-03-12 15:38:06 EDT	Error log from splunkd..log	search	Real-time	● Critical	Per Result	View results | Edit search | Delete
□	2019-03-12 15:36:54 EDT	Error log from splunkd..log	search	Real-time	● Critical	Per Result	View results | Edit search | Delete

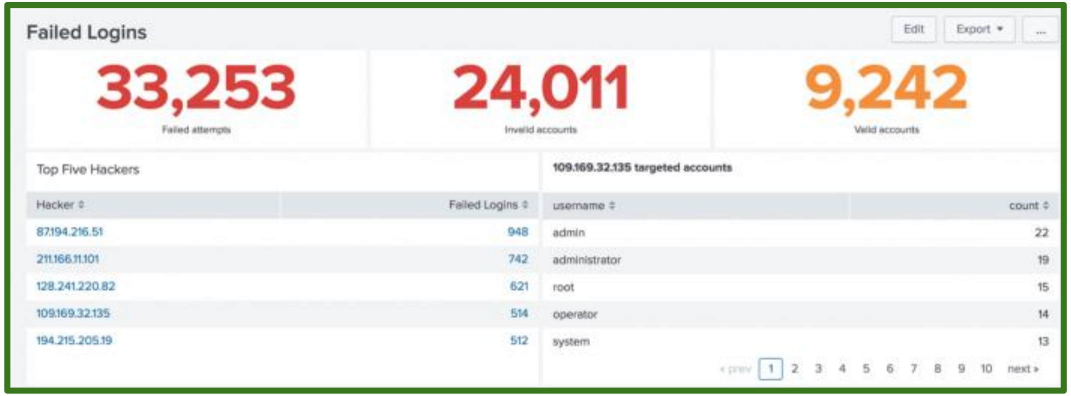
Dashboards

Dashboards integrate **real-time searches, charts, reports, gauges, maps**, and **tables** in panels to display the most relevant information for different teams and use cases.

Scheduled searches are used to create real-time dashboards and visualizations.

Some benefits for security analysts and SOC teams include:

- Dashboards accelerate **time-to-time** and **time-to-action**.
- Dashboards use the same event data for security analysts and operations teams to visualize events.



Failed Logins

33,253 Failed attempts

24,011 Invalid accounts

9,242 Valid accounts

Top Five Hackers

Hacker ⇕	Failed Logins ⇕
87.194.216.51	948
211.166.11.101	742
128.241.220.82	621
109.169.32.135	514
194.215.205.19	512

109.169.32.135 targeted accounts

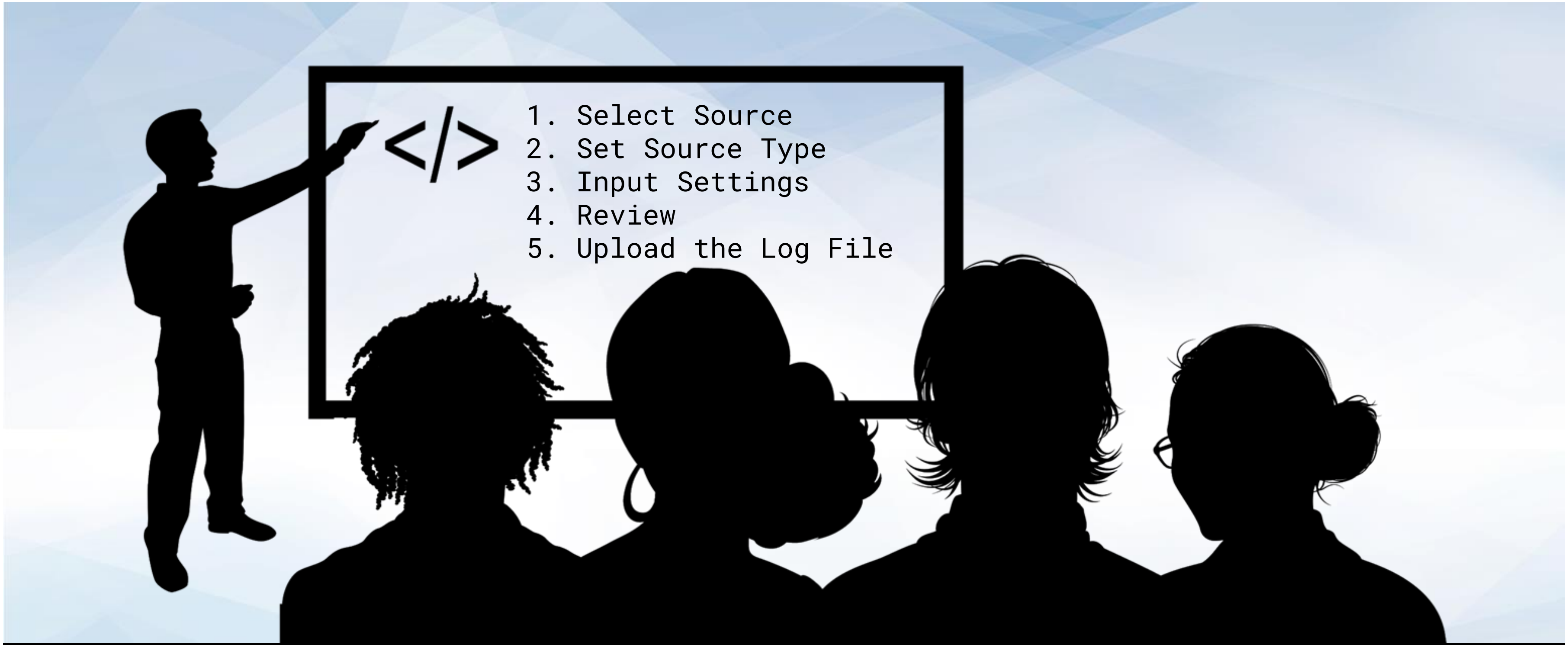
username ⇕	count ⇕
admin	22
administrator	19
root	15
operator	14
system	13

Getting Data into Splunk

There are different methods of getting data into Splunk:

- **Uploading** files (e.g. cvs, json, log, zip) from a system. The maximum size is 500 mb.
- **Monitoring** files and network ports on the host that runs the Splunk instance.
- **Forwarding** data to another Splunk instance or to a third-party system.

Next, we’ll launch into a demo of the workflow for **uploading data** into Splunk.



1. Select Source
2. Set Source Type
3. Input Settings
4. Review
5. Upload the Log File

Instructor Demonstration

Uploading a Web Log into Splunk

Demo Summary

Uploading a Web Log into Splunk:

- We grabbed an existing log file that was created from a different machine.
- We uploaded that log file into Splunk.
- We can now filter and search for different events based on our customized criteria.
- When it comes to searching, quotations after an equals (=) sign are not needed, however adding the quotations makes it easier for others to read your work.

Activity: Uploading a Web Server Log into Splunk

You will be uploading a web server log file into Splunk. This is an important exercise, as you will be uploading various files into Splunk in the coming classes.

Instructions sent via Slack.

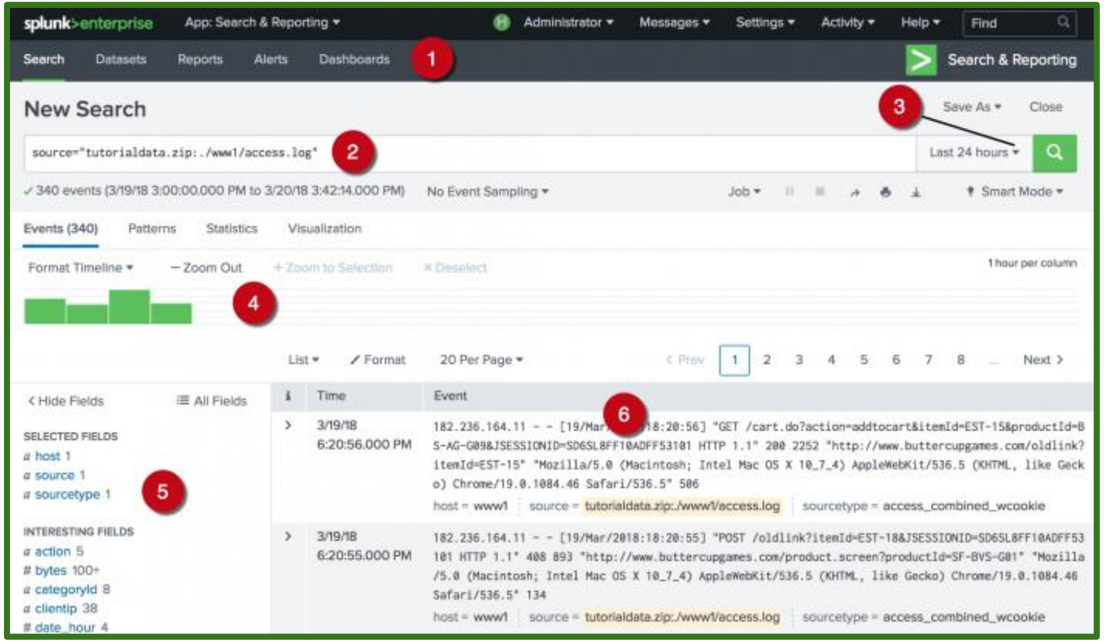
Suggested Time:
10 minutes

Times Up! Let’s Review.

Uploading a Web Server Log

Web Server Log Review

Let’s review the parts of the splunk web server.



```
source="tutorialdata.zip:./www1/access.log"
```

✓ 340 events (3/19/18 3:00:00.000 PM to 3/20/18 3:42:14.000 PM)

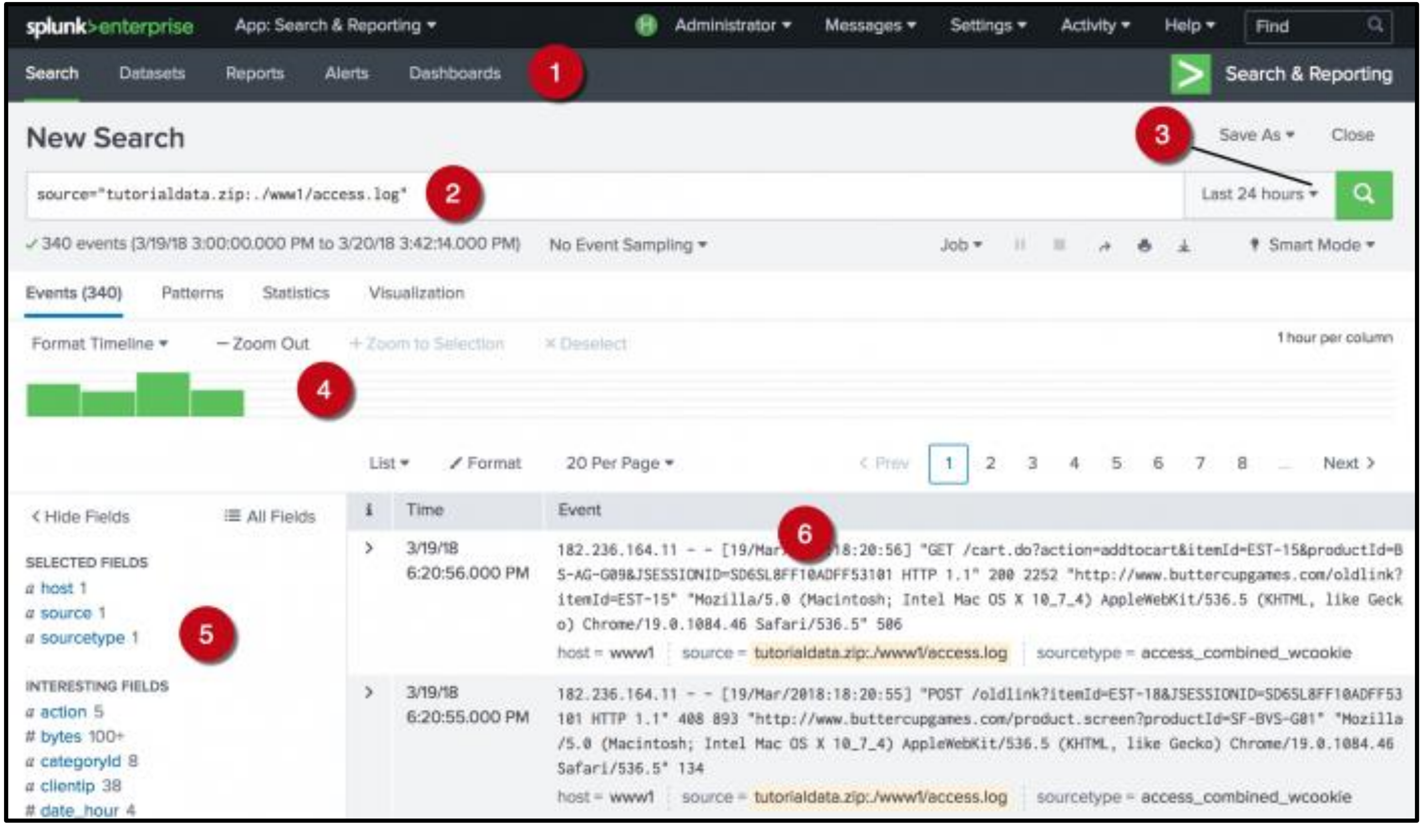
SELECTED FIELDS
- a host 1
- a source 1
- a sourcetype 1

INTERESTING FIELDS
- a action 5
- # bytes 100+
- a categoryId 8
- a clientip 38
- # date_hour 4

```
> 3/19/18
  6:20:56.000 PM
182.236.164.11 - - [19/Mar 8:20:56] "GET /cart.do?action=addtocart&itemId=EST-15&productId=B
S-AG-G09&JSESSIONID=SD6SL8FF10ADFF53101 HTTP 1.1" 200 2252 "http://www.buttercupgames.com/oldlink?
itemId=EST-15" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/536.5 (KHTML, like Geck
o) Chrome/19.0.1084.46 Safari/536.5" 506
host = www1 | source = tutorialdata.zip:./www1/access.log | sourcetype = access_combined_wcookie

> 3/19/18
  6:20:55.000 PM
182.236.164.11 - - [19/Mar/2018:18:20:55] "POST /oldlink?itemId=EST-18&JSESSIONID=SD6SL8FF10ADFF53
101 HTTP 1.1" 408 893 "http://www.buttercupgames.com/product.screen?productId=SF-BVS-G01" "Mozilla
/5.0 (Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46
Safari/536.5" 134
host = www1 | source = tutorialdata.zip:./www1/access.log | sourcetype = access_combined_wcookie
```

Web Server Log Review



```
source="tutorialdata.zip:./www1/access.log"
```

✓ 340 events (3/19/18 3:00:00.000 PM to 3/20/18 3:42:14.000 PM)

SELECTED FIELDS
- a host 1
- a source 1
- a sourcetype 1

INTERESTING FIELDS
- a action 5
- # bytes 100+
- a categoryId 8
- a clientip 38
- # date_hour 4

i	Time	Event
›	3/19/18 6:20:56.000 PM	182.236.164.11 - - [19/Mar[illegible]8:20:56] "GET /cart.do?action=addtocart&itemId=EST-15&productId=BS-AG-G09&JSESSIONID=SD6SL8FF10ADFF53101 HTTP 1.1" 200 2252 "http://www.buttercupgames.com/oldlink?itemId=EST-15" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 506 host = www1 | source = tutorialdata.zip:./www1/access.log | sourcetype = access_combined_wcookie
›	3/19/18 6:20:55.000 PM	182.236.164.11 - - [19/Mar/2018:18:20:55] "POST /oldlink?itemId=EST-18&JSESSIONID=SD6SL8FF10ADFF53101 HTTP 1.1" 408 893 "http://www.buttercupgames.com/product.screen?productId=SF-BVS-G01" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 134 host = www1 | source = tutorialdata.zip:./www1/access.log | sourcetype = access_combined_wcookie

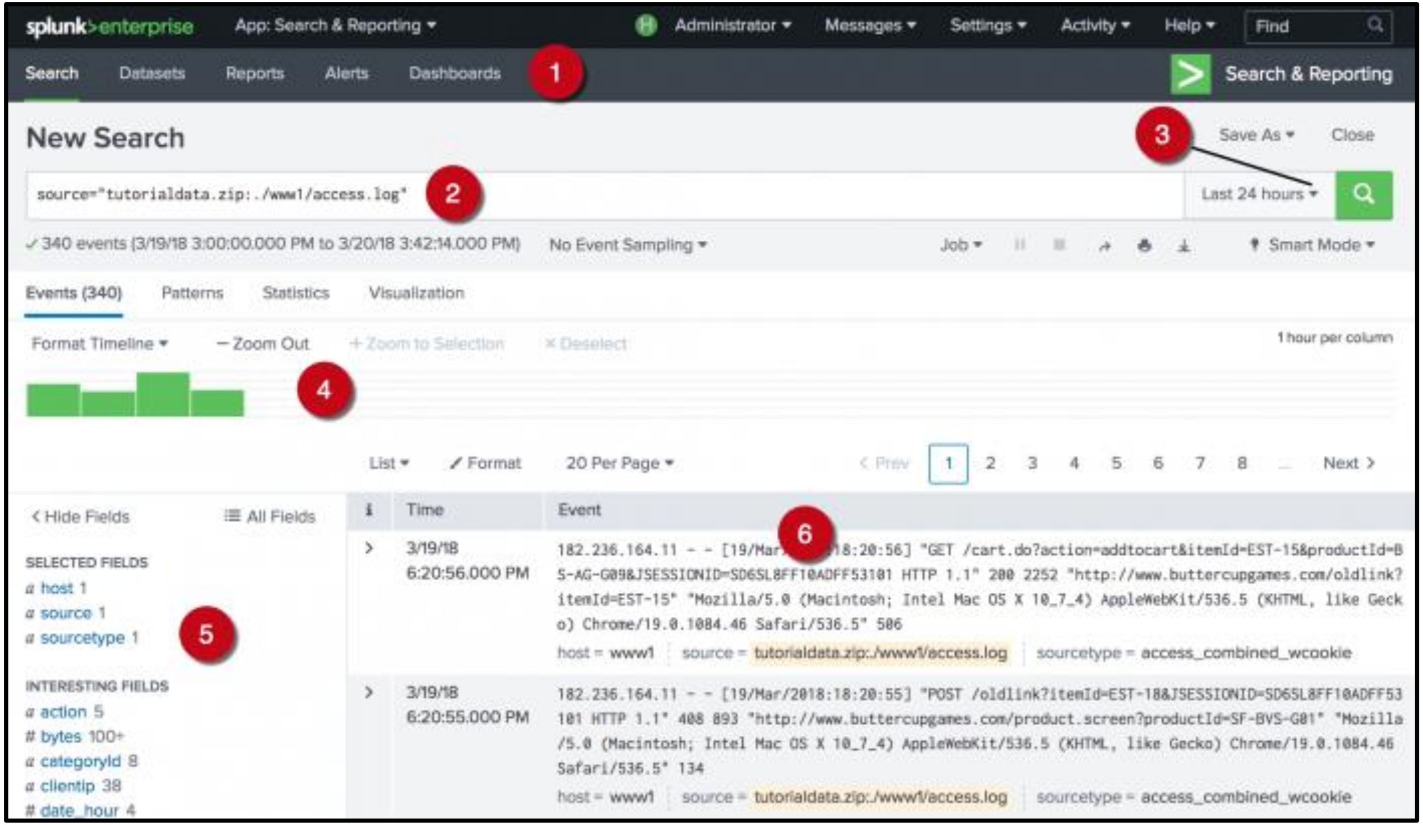
1. Application bar contains main functions.

2. Search bar specifies your search criteria.

3. **Time range picker** specifies the time period for the search.

4. The peaks and valleys in the **timeline** can indicate spikes in activity or server downtime.

Web Server Log Review



```
source="tutorialdata.zip:./www1/access.log"
```

✓ 340 events (3/19/18 3:00:00.000 PM to 3/20/18 3:42:14.000 PM)

SELECTED FIELDS
- a host 1
- a source 1
- a sourcetype 1

INTERESTING FIELDS
- a action 5
- # bytes 100+
- a categoryId 8
- a clientip 38
- # date_hour 4

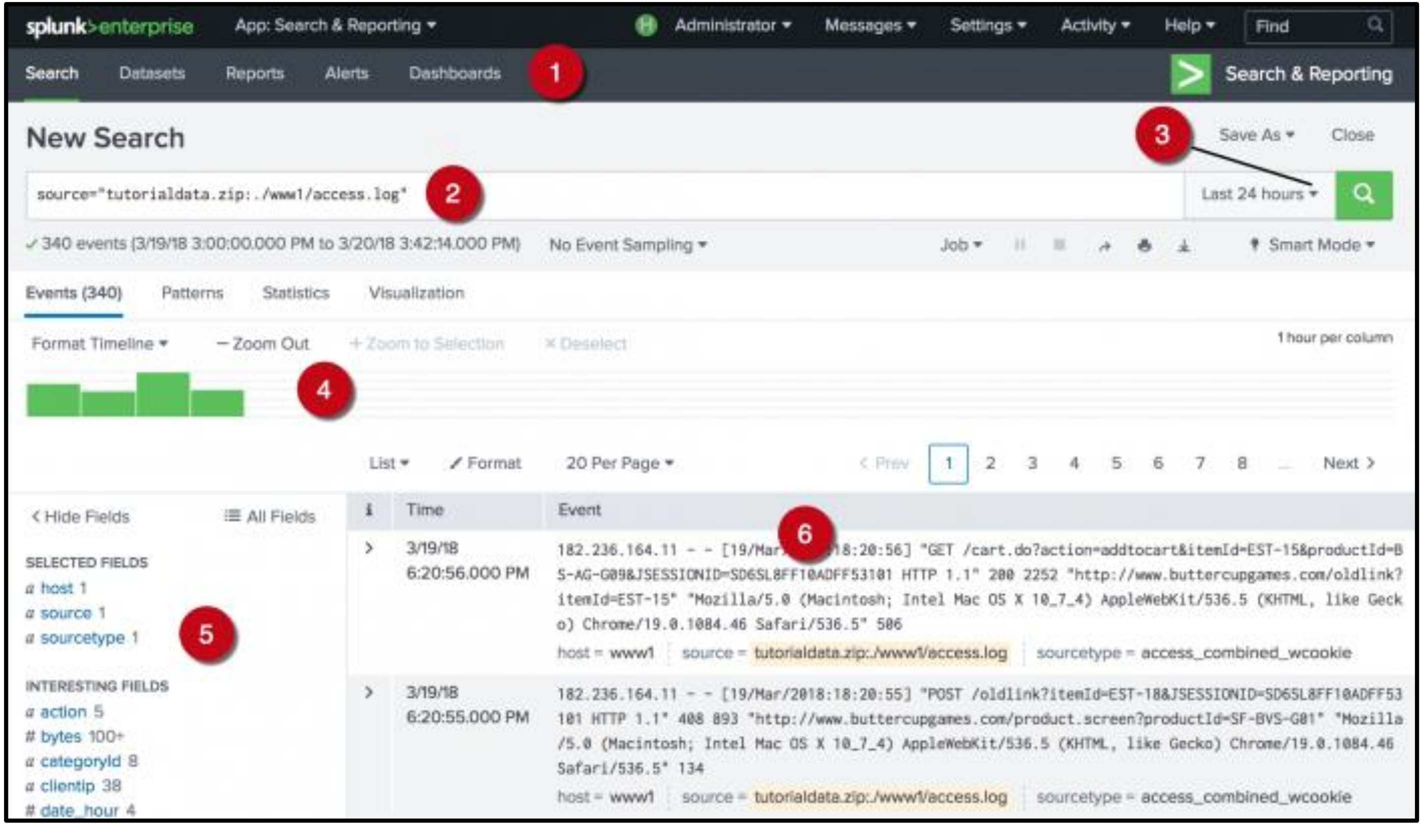
i	Time	Event
>	3/19/18 6:20:56.000 PM	182.236.164.11 - - [19/Mar 18:20:56] "GET /cart.do?action=addtocart&itemId=EST-15&productId=BS-AG-G09&JSESSIONID=SD6SL8FF10ADFF53101 HTTP 1.1" 200 2252 "http://www.buttercupgames.com/oldlink?itemId=EST-15" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 506 host = www1 | source = tutorialdata.zip:./www1/access.log | sourcetype = access_combined_wcookie
>	3/19/18 6:20:55.000 PM	182.236.164.11 - - [19/Mar/2018:18:20:55] "POST /oldlink?itemId=EST-18&JSESSIONID=SD6SL8FF10ADFF53101 HTTP 1.1" 408 893 "http://www.buttercupgames.com/product.screen?productId=SF-BVS-G01" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 134 host = www1 | source = tutorialdata.zip:./www1/access.log | sourcetype = access_combined_wcookie

5. **Select Fields and Interesting Fields**: This sidebar displays a list of the fields discovered in the event.

When you first run a search, the **Selected Fields** list contains the default fields host, source, and source type.

Interesting Fields are fields that appear in at least 20% of the events.

Web Server Log Review



```
source="tutorialdata.zip:./www1/access.log"
```

✓ 340 events (3/19/18 3:00:00.000 PM to 3/20/18 3:42:14.000 PM)

SELECTED FIELDS
- a host 1
- a source 1
- a sourcetype 1

INTERESTING FIELDS
- a action 5
- # bytes 100+
- a categoryId 8
- a clientip 38
- # date_hour 4

Time	Event
3/19/18 6:20:56.000 PM	182.236.164.11 - - [19/Mar 8:20:56] "GET /cart.do?action=addtocart&itemId=EST-15&productId=BS-AG-G09&JSESSIONID=SD6SL8FF10ADFF53101 HTTP 1.1" 200 2252 "http://www.buttercupgames.com/oldlink?itemId=EST-15" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 506 host = www1 source = tutorialdata.zip:./www1/access.log sourcetype = access_combined_wcookie
3/19/18 6:20:55.000 PM	182.236.164.11 - - [19/Mar/2018:18:20:55] "POST /oldlink?itemId=EST-18&JSESSIONID=SD6SL8FF10ADFF53101 HTTP 1.1" 408 893 "http://www.buttercupgames.com/product.screen?productId=SF-BVS-G01" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 134 host = www1 source = tutorialdata.zip:./www1/access.log sourcetype = access_combined_wcookie

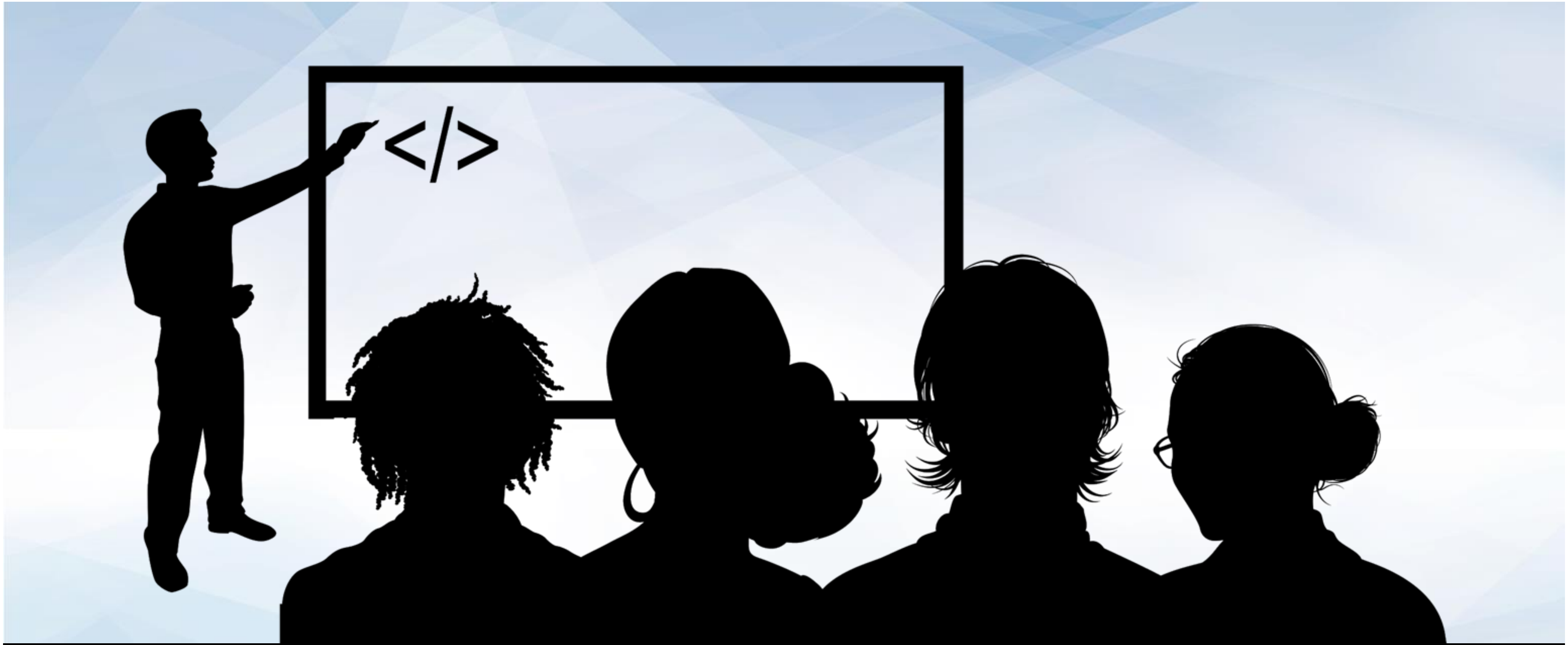
6. **Events viewer** displays the events that match your search.

By default, the most recent event is listed first and the matching search terms are highlighted.

Search Processing Language

So far we've looked at the basic functions of the Search and Reporting App and how to upload a weblog into Splunk and generate events.

Now, it’s time to search events using the **Search Processing Language (SPL).**



Instructor Demonstration

Using SPL to Retrieve Events

The Anatomy of a Search

Search Processing Language (SPL) is a language based on the Unix Pipeline and the Standard Query Language (SQL).

When working with SPL:

- We are seeking to match search terms against segments of the data in order to return events from indexes.

- Search terms use **`keywords`**, **`field name`** and **`value pairs`**, **`boolean expressions`**, **`logical / relational operators`** and **`wildcards`** that specify which events we want to retrieve from the indexes.

Keywords

Keywords are fields that have been automatically extracted from the log file that can be used for searches.

- These keywords are chosen based on the frequency they appear in the log file.

Keywords are also known as **selected** and **interesting fields** in Splunk and appear on the left hand side of the search page.

- On the ***right side*** of the field, we can see the number of instances each field shows up in the log file. This is useful when conducting searches.

Field / Value Pairs

Field / Value Pairs match keywords with the specific information that you are searching for.

Syntax looks like:

```
<field> = <value>

clientip = 87.194.216.51
```

Examples include:

- `user=AJ`
  - This field / value pair would find the user named AJ.
- `domain=facebook.com`
  - This pair would find the website facebook.

Logical Operators and Boolean Expressions

Three boolean expressions that we covered in the past are applicable in Splunk

`NOT`: "I want all instances of the error 403 **NOT** forbidden access" :
I only want to see error 403 that are not also errors of forbidden access.

`OR`: "I want all instances of the errors 403 **OR** forbidden access" :
I want to see singular instances of one or the other, but not both.

`AND`: “I want all instances of the errors 403 **AND** forbidden access":
I want both errors.

Relational Operators

Splunk also uses relational operators

Not Equals Operator:

```
!= 200
```

“I want all events where the status does not equal to 200”

Greater Than Operator:

```
> 4
```

“I want all events where the line count is greater than 4”

Wildcards and Boolean Expression

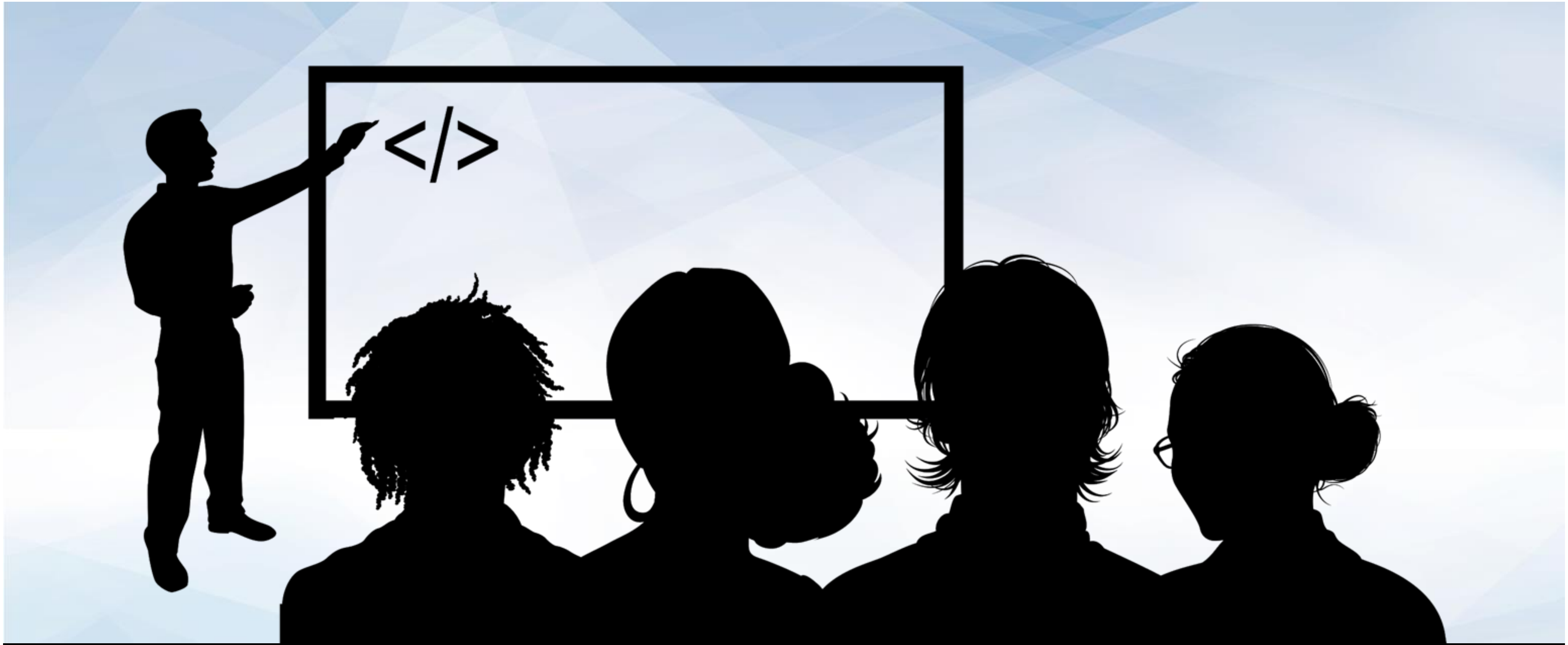
Wildcards use the symbol “*” to match any character in that string”:

“you*” would match with you’re, young, your, youth

“*ing” would match with matching, fighting, talking

“*each*” would match with reaching, breach, preach

Wildcards are used in the same way that they are used when operating a Linux machine.



Instructor Demonstration

Boolean Expression Demo

Splunk Searching

Key Takeaways:

- Machine data that comes into Splunk is **unruly** and **unorganized**.
- **Keywords**, **wildcards**, and **filed / value pairs** allow us to filter the data.
- **Boolean expressions** take our filtration techniques a step further, to more precisely receive what we need.

Student Activity: Introduction to SPL

In this activity, you will practice writing and implementing a simple SPL statement that uses field/value pairs, boolean expressions, and fields.

Instructions sent via Slack.

Suggested Time:
15 Minutes

Times Up! Let’s Review.

Introduction to SPL

Introduction to SPL Review Part 1

1. What is an event?

2. What is a source type?

3. What is the most important search parameter to specify?

4. Identify what this search will do: `src="10.9.165.*" OR dst="10.9.165.8"`

Introduction to SPL Review Part 1

1. What is an event?

Events are a single record of activity or instance data that has been indexed by Splunk.

For example, a single log entry in a log file might be an event.

2. What is a source type?

3. What is the most important search parameter to specify?

4. Identify what this search will do: `src="10.9.165.*" OR dst="10.9.165.8"`

Introduction to SPL Review Part 1

1. What is an event?

Events are a single record of activity or instance data that has been indexed by Splunk.

For example, a single log entry in a log file might be an event.

2. What is a source type?

Splunk assigns a source type to determine how to format the event data during the indexing process.

3. What is the most important search parameter to specify?

4. Identify what this search will do: `src="10.9.165.*" OR dst="10.9.165.8"`

Introduction to SPL Review Part 1

1. What is an event?

Events are a single record of activity or instance data that has been indexed by Splunk.

For example, a single log entry in a log file might be an event.

2. What is a source type?

Splunk assigns a source type to determine how to format the event data during the indexing process.

3. What is the most important search parameter to specify?

Time

4. Identify what this search will do: `src="10.9.165.*" OR dst="10.9.165.8"`

Introduction to SPL Review Part 1

1. What is an event?

Events are a single record of activity or instance data that has been indexed by Splunk.

For example, a single log entry in a log file might be an event.

2. What is a source type?

Splunk assigns a source type to determine how to format the event data during the indexing process.

3. What is the most important search parameter to specify?

Time

4. Identify what this search will do: `src="10.9.165.*" OR dst="10.9.165.8"`

Returns events where the source IP addresses that start with `10.9.165` or destination IP address is `10.9.165.8`.

Introduction to SPL Review: Part 2

For the search, provide the following:

The field / value pairs:

- `source=“access_30DAY.log”`
- `sourcetype=“access_combined_wcookie”`
- `status=“505”`

The boolean operator:

The boolean expression

Introduction to SPL Review: Part 2

For the search, provide the following:

The field / value pairs:

The boolean operator:

The boolean expression

Introduction to SPL Review: Part 2

For the search, provide the following:

The field / value pairs:

- `source=“access_30DAY.log”`
- `sourcetype=“access_combined_wcookie”`
- `status=“505”`

The boolean operator:

The boolean expression

Introduction to SPL Review: Part 2

For the search, provide the following:

The field / value pairs:

- `source=“access_30DAY.log”`
- `sourcetype=“access_combined_wcookie”`
- `status=“505”`

The boolean operator: AND

The boolean expression

Introduction to SPL Review: Part 2

For the search, provide the following:

The field / value pairs:

- `source=“access_30DAY.log”`
- `sourcetype=“access_combined_wcookie”`
- `status=“505”`

The boolean operator: **AND**

The boolean expression

```
source=“access_30DAY.log” sourcetype= “access_combined_wcookie” AND
status=“505”
```

Introduction to SPL Review Part 3

1. Write the search term that will first display events for ALL server response errors.

Introduction to SPL Review Part 3

1. Write the search term that will first display events for ALL server response errors.

This search requires a wildcard to specify the *5xx family of the HTTP server responses.

```
source=“access_30DAY.log” sourcetype=“access_combined_wcookie” AND
status=”5**”
```

2. Next, narrow your search down to HTTP versions not supported events.

Introduction to SPL Review Part 3

1. Write the search term that will first display events for ALL server response errors.

This search requires a wildcard to specify the *5xx family of the HTTP server responses.

```
source=“access_30DAY.log” sourcetype=“access_combined_wcookie”AND
status=”5**”
```

2. Next, narrow your search down to HTTP versions not supported events.

```
source=“access_30DAY.log” sourcetype=“access_combined_wcookie” AND
status=“505”
```

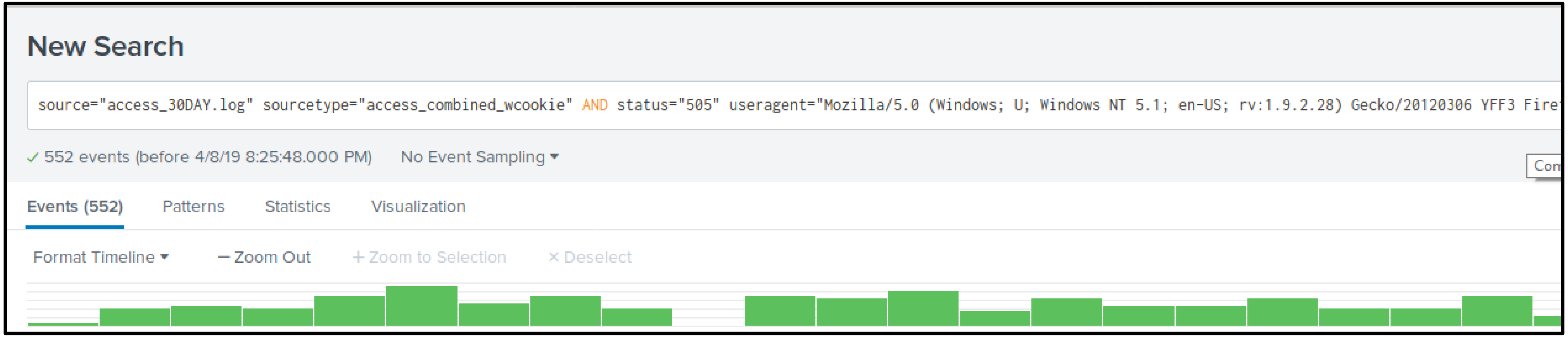
Introduction to SPL Review Part 3

3. Now, isolate the highest client software that is originating the response:

Introduction to SPL Review Part 3

3. Now, isolate the highest client software that is originating the response:

Use `Interesting Fields` and select the highest value from the `useragent` field.



New Search

```
source="access_30DAY.log" sourcetype="access_combined_wcookie" AND status="505" useragent="Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.2.28) Gecko/20120306 YFF3 Fire
```

✓ 552 events (before 4/8/19 8:25:48.000 PM)

Introduction to SPL Review Part 3

4. Record the number of events that are returned.

5. Record the peak times when the events occur.

Introduction to SPL Review Part 3

4. Record the number of events that are returned.

552

5. Record the peak times when the events occur.

Peak times will vary.

Class Objectives

By the end of class today, students will be able to:

- 🟩 Recognize the role SIEMs play in protecting an organization's security.
- 🟩 Explain how logs are filtered, normalized, and correlated for events.
- 🟩 Use basic features of the Splunk User Interface.
- 🟩 Explain basic database terms and query functions.
- 🟩 Use the Splunk Processing Language (SPL) for simple queries.