



Data Onboarding Where Do I Begin?

Luke Netto

Staff Professional Services Consultant | Splunk

Forward-Looking Statements



During the course of this presentation, we may make forward-looking statements regarding future events or plans of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results may differ materially. The forward-looking statements made in the this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, it may not contain current or accurate information. We do not assume any obligation to update any forward-looking statements made herein.

In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only, and shall not be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionalities described or to include any such feature or functionality in a future release.

Splunk, Splunk>, Turn Data Into Doing, The Engine for Machine Data, Splunk Cloud, Splunk Light and SPL are trademarks and registered trademarks of Splunk Inc. in the United States and other countries. All other brand names, product names, or trademarks belong to their respective owners. © 2019 Splunk Inc. All rights reserved.

Who Are You?

You have Splunk installed, either in your datacenter or on your laptop

You have data you want to onboard into Splunk

Your data comes from syslog, wineventlog, custom modular inputs, and/or flat files such as .csv | .log | .json | .txt

Who Am I?

5+ years of Splunk experience

10+ years of systems engineering

7+ years of data analytics

systems engineering + data analytics = **splunk**>

Agenda

Why is this important?

Basic Splunk overview

What do I ask before I onboard data?

How Splunk makes it easier & faster

Creating your own sourcetype

Onboarding data (inputs.conf/props.conf)

The importance of normalizing (props.conf/eventtypes.conf/tags.conf)

Why is This Important?

Many Reasons!

1. Your organization wants to become data-driven
2. Data needs to be collected, accessible, and queryable
3. Decisions without **quality** data is simply guessing

In Data Science, **80%** of time spent **prepare data**, **20%** of time spent complain about need for prepare data.”

@BigDataBorat

<https://twitter.com/bigdataborat/status/306596352991830016>

Data Life Cycle

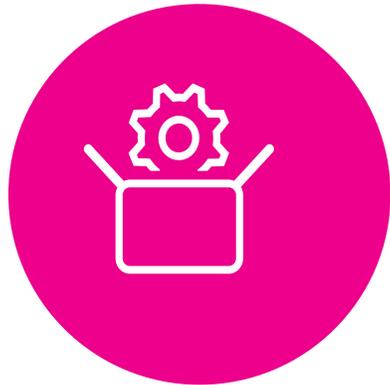
What your executives want



Making Machine Data Accessible, Usable and Valuable to Everyone.

Reducing That 80%

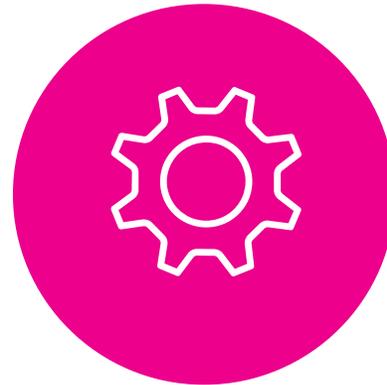
How Splunk is helping you!



Splunkbase
Apps &
Add-ons



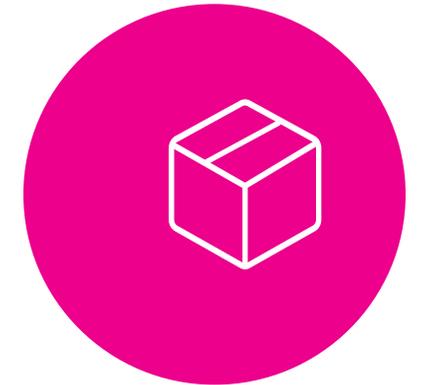
Guided Data
Onboarding
(GDO)



Splunk
Connect



Splunk Data
Stream
Processor (DSP)



Splunk Add-on
Builder

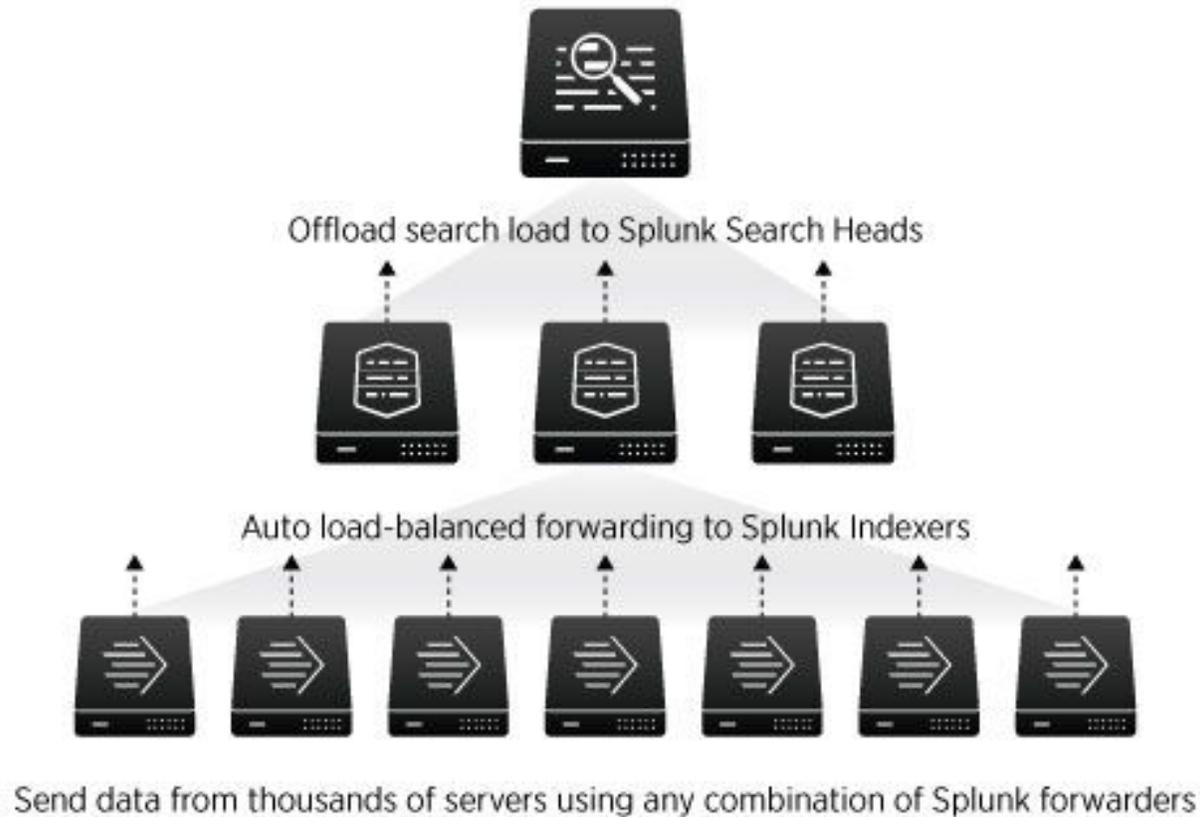


Splunk's Data Collection

Basic Architecture Refresher

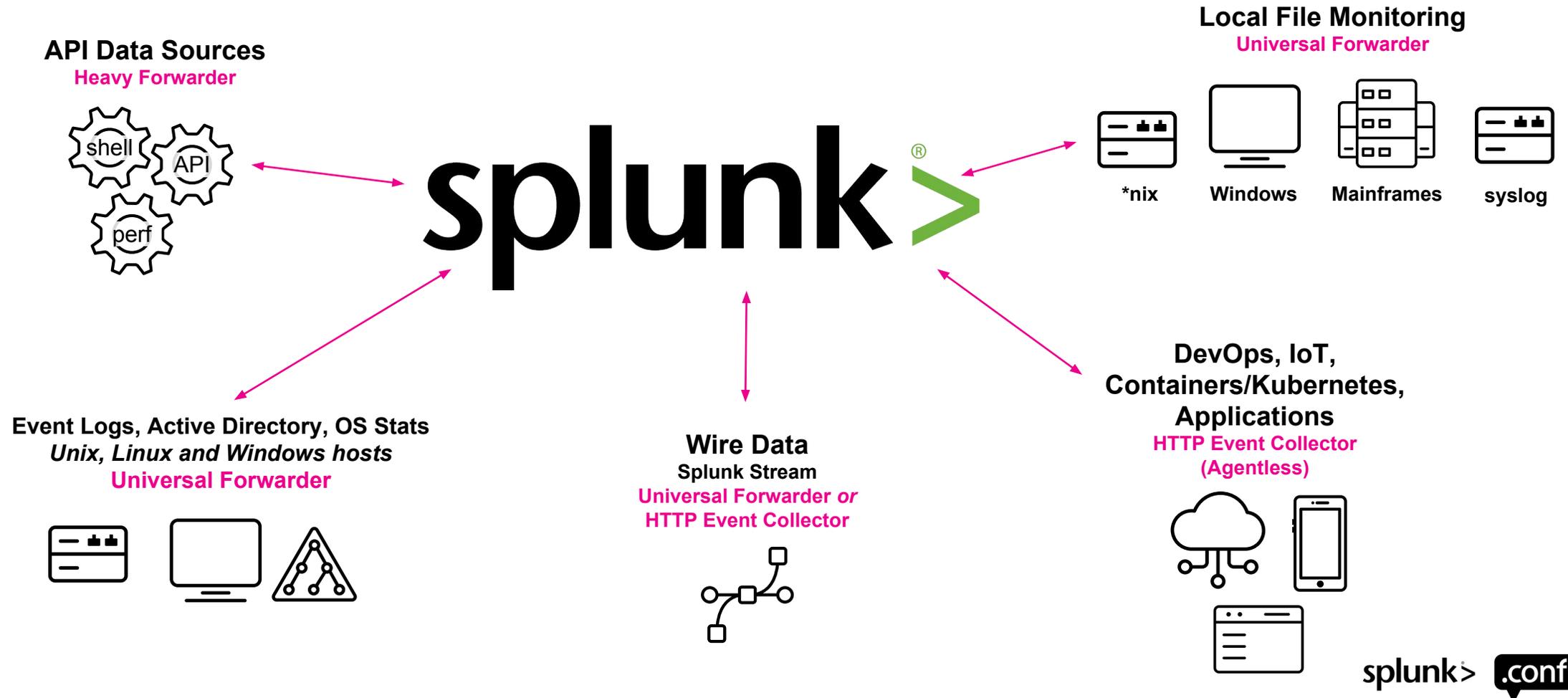
Basic Architecture

How Splunk works at a high level



What Can Splunk Ingest?

Data from anything!





Let's Begin

Index Time

Metadata

Your Data's Data

- Index
- Host
- Source
- Sourcetype
- Date/Timestamp
- Event Boundary/Line Breaker

Index

The repository for data

When the Splunk platform indexes raw data, it transforms the data into searchable events

Indexes reside in flat files on the indexer

Used to organize your data, role-based access control (RBAC), and retention

- winlog, winlogsec, winperf, nixlog, nixlogsec, nixperf
- firewall, proxy, dhcp, dns, dblog, web, app

Host

A default field that contains the hostname or IP address of the device that generated the event

Use the host field in searches to narrow the search results to events that originate from a specific device

Allows you to locate the originating device

If you are forwarding data from many hosts through a single host, you may need to create a new "indexed field" called `orig_host` or something similar

Source

A default field that identifies the source of an event (where the event originated)

For data monitored from files and directories, the source consists of the full pathname of the file or directory

- /var/log/messages
- /var/log/messages.1
- /var/log/secure

For network-based sources, the source field consists of the protocol and port

- UDP:514
- TCP:1514
- HTTP:<token name>

Timestamp

Splunk uses timestamps to

- correlate events by time
- create the timeline histogram in Splunk Web
- set time ranges for searches

Usually automatic or defined by sourcetype (recommended)

Event Boundary/Line Breaker

Allows Splunk to break the incoming stream of bytes into separate events

Supports single-line and multi-line

Splunk can usually do this automatically or it can be defined by sourcetype (recommended)

Sourcetype

Defining your data's data

A default field that identifies the data structure of an event including datetime extraction and linebreaking

Determines how Splunk extracts & calculates fields during search time

Use the sourcetype field in searches to find all data of a certain type (as opposed to all data from a certain source)

Important – syslog, csv, json, xml are not sourcetypes!

Examples: cisco:asa, cisco:ios, pan:firewall, wineventlog, linux_secure

Source vs. Sourcetype

Events with the same sourcetype can come from different sources

- /var/log/messages
- /var/log/messages.1
- udp:514

sourcetype=linux_messages_syslog may retrieve events from both of those sources

What Happens with Bad Sourcetypes

Same regex, same sourcetype, bad results 😞

Squid
Proxy

Blue Coat
ProxySG

REGULAR EXPRESSION 2 matches, 145 steps (~2ms)

:/ GET\s.+?\s"*(?P<user>.+)?\s /gm

TEST STRING SWITCH TO UNIT TESTS ▶

```
1503884030.208725 49 172.20.14.112 TCP_REFRESH_HIT/200 10635 GET
http://www.cambridgestatebank.com/images/bankfplow.jpg "brooks@demo.com"
DIRECT/www.cambridgestatebank.com image/jpeg DEFAULT_CASE-DefaultGroup-
Demo_Clients-NONE-NONE-NONE-DefaultRouting
<IW_fnnc,5.0,0,-,-,-,-,0,-,-,-,-,-,IW_fnnc,-> - "SonyEricssonW810i/R4EA
Browser/NetFront/3.3 Profile/MIDP-2.0 Configuration/CLDC-1.1 UP.Link/6.3.0.0.0"
http://www.cambridgestatebank.com/ "Finance"
```

```
2017-08-27 20:31:52 10.180.6.76 64507 465 TCP_MISS 200 200 64445 326 - -
OBSERVED GET boostifun.cellmania.com HTTP/1.1 291
http://boostifun.cellmania.com/ifuncb/ContentDownload?
orderid=56660425&itemid=222067&fromProd=true - ringtone/mp3;charset=ISO-8859-1
"Java/1.4.2_08" -
```

With Correct Sourcetypes

Same regex, different sourcetypes, good results 😊

REGULAR EXPRESSION Cisco Squid 1 match, 95 steps (~3ms)

:/ GET\s.+?\s"*(?P<user>.+?)\s /gm

TEST STRING SWITCH TO UNIT TESTS ▶

```
1503884030.208725 49 172.20.14.112 TCP_REFRESH_HIT/200 10635 GET
http://www.cambridgestatebank.com/images/bankfplow.jpg
"brooks@demo.com" DIRECT/www.cambridgestatebank.com image/jpeg
DEFAULT_CASE-DefaultGroup-Demo_Clients-NONE-NC
<IW_fnnc,5.0,0,-,-,-,-,0,-,-,-,-,-,-,-,IW_fnnc
"SonyEricssonW810i/R4EA Browser/NetFront/3.3 P
Configuration/CLDC-1.1 UP.Link/6.3.0.0.0"
http://www.cambridgestatebank.com/ "Finance"
```

REGULAR EXPRESSION Bluecoat 1 match, 48 steps (~1ms)

:/ GET\s.+?\s"*(?P<protocol_version>.+?)\s /gm

TEST STRING SWITCH TO UNIT TESTS ▶

```
2017-08-27 20:31:52 10.180.6.76 64507 465 TCP_MISS 200 200 64445 326 -
- OBSERVED GET boostifun.cellmania.com HTTP/1.1 291
http://boostifun.cellmania.com/ifuncb/ContentDownload?
orderid=56660425&itemid=222067&fromProd=true -
ringtone/mp3;charset=ISO-8859-1 "Java/1.4.2_08" -
```

Configuring the Metadata

- props.conf

```
[your:sourcetype]
```

```
TIME_PREFIX =
```

```
MAX_TIMESTAMP_LOOKAHEAD =
```

```
TIME_FORMAT =
```

```
SHOULD_LINEMERGE = false
```

```
LINE_BREAKER =
```

```
EVENT_BREAKER_ENABLE = true
```

```
EVENT_BREAKER =
```

```
TRUNCATE = 10000
```

- inputs.conf

```
[monitor:///your/logs]
```

```
host = <optional/depends>
```

```
index = <your index>
```

```
sourcetype = your:sourcetype
```

```
source = <optional/depends>
```

*host & source are usually
automagically set*

Time to Explore!

Everything else happens at search!

Start exploring and understanding your data

Apply different lenses to your data for use cases not yet discovered

Maintaining the original log ensures integrity



Data Discovery

Using what we now know, what should we ask before data onboarding?

Find the Data

What is producing the data?

- Appliance
- Application

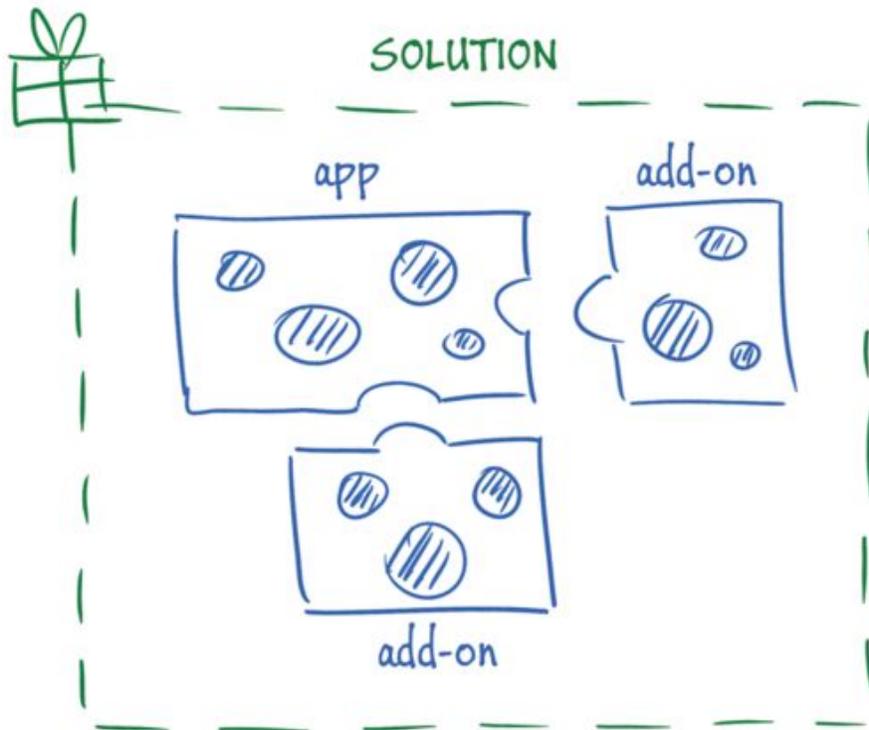
Where is the data?

- Flat file
- Network/Syslog feed
- REST API
- Database
- Wineventlog

Can you get a sample?

Is there a logging document or manual? – these sometimes do exist!

Apps & Add-Ons (TA's)



Your **first** choice when onboarding new data

Includes relevant config files (props/transforms) and ancillary scripts & binaries

Where Do You Get Apps? Splunkbase!

The screenshot displays the Splunkbase app marketplace interface. It is divided into two main sections: 'Browse by Category' and 'Browse by Technology'.

Browse by Category

- DevOps**: 41 Apps
- IT Operations**: 634 Apps
- Security, Fraud & Compliance**: 572 Apps
- Business Analytics**: 94 Apps
- IoT & Industrial Data**: 75 Apps
- Utilities**: 565 Apps

Browse by Technology

Logos for Cisco, Dell EMC, Amazon Web Services, and Palo Alto Networks are visible. A link 'See all Cisco apps >' is present.

App listings under 'Browse by Technology':

- Cisco Networks App for Splunk Enterprise**: 1661 Installs
- Splunk Add-on for Cisco UCS**: 223 Installs
- TA-meraki**: 183 Installs
- Cisco ACI Add-on for Splunk Enterprise**: 88 Installs
- Cisco AnyConnect Network Visibility**: 85 Installs
- Cisco ACI App for Splunk Enterprise**: 83 Installs

Splunk Built Apps

A link 'See all apps >' is present.

App listings under 'Splunk Built Apps':

- Splunk UBA RHEL 7.2 Software for Bare**
- Splunk UBA RHEL 6.7 Software for Bare**
- Splunk App for PCI Compliance - Splunk**
- Splunk UBA OVA Software**
- Splunk UBA Software Update**
- Splunk IT Service Intelligence**

What If There is No App/Add-On?





Let's Do It

Getting Data In!

Data Discovery

Understand the data source

Your networking organization wants to bring their firewall logs into Splunk

Ask yourself...

- Are the logs coming from an appliance or application?
- How will I access the logs?
- Who needs access to the logs in Splunk?
- How long do I need to retain the logs?
- Can I get a log sample?

Data Discovery

Our answers...

Your network team uses Cisco ASA and Palo Alto firewalls

Both firewalls are appliances, outputting their logs using syslog

The networking team also has a syslog server

Your networking team is the only team that needs access until your security team becomes aware of the data in Splunk (same data, difference lenses, multiple use-cases)

Your organization would like 12 months of logs for any network related equipment for compliance purposes

A log sample has been provided

Questions For Us

What index do I use?

Do you have any existing indexes for the network team?

- Yes, but all of their other data is in there and security will probably want to see the firewall logs

Do their existing indexes have the required retention?

- No, we keep their other logs for only 90 days

Sounds like we need a new index...maybe “firewall”?

Questions For Us

How am I going to receive the data?

Leverage the networking team's existing syslog server?

Use my own syslog server?

Use Splunk Connect for Syslog?

Any answer is acceptable!

Questions For Us

What sourcetype do I use?

Are there any existing apps on Splunkbase? YES!



Source type	Description
<code>cisco:asa</code>	The system logs of Cisco ASA record user authentication, user session, VPN, and intrusion messages.
<code>cisco:fws</code>	The system logs of Cisco FWSM record user authentication, user session, and firewall messages.
<code>cisco:pix</code>	The system logs of Cisco PIX record user authentication, user session, and intrusion messages.



Log source	Source Type
Only Firewall logs	pan:firewall
Only Traps Management Service logs	pan:traps
Only Traps 4.x logs	pan:traps4
This input receives Firewall and Traps logs	pan:log

Read the docs, we did most of the 80% already for you!

What Are Our Inputs?

Making some assumptions here...

```
[monitor:///var/log/cisco_asa/<host>]
```

```
index = firewall
```

```
sourcetype = cisco:asa
```

```
disabled = false
```

```
host_segment = 4
```

```
[monitor:///var/log/pan_firewall/<host>]
```

```
index = firewall
```

```
sourcetype = pan:firewall
```

```
disabled = false
```

```
host_segment = 4
```



Value Extraction

Exploring Our Data!

If We Used the Incorrect Sourcetype

We would have a lot of work ahead of us 😞

New Search Save As ▾ Close

sourcetype="syslog" All time ▾ 🔍

✓ 9,961 events (before 9/9/19 10:11:15.000 AM) No Event Sampling ▾ Job ▾ ⏸ ⏹ ↶ 🖨 ⏴ Smart Mode ▾

Events (9,961) Patterns Statistics Visualization

Format Timeline ▾ — Zoom Out + Zoom to Selection × Deselect 1 minute per column

List ▾ ✎ Format 20 Per Page ▾ < Prev 1 2 3 4 5 6 7 8 ... Next >

< Hide Fields ☰ All Fields

SELECTED FIELDS
 a host 14
 a source 2
 a sourcetype 1

INTERESTING FIELDS
 # date_hour 2
 # date_mday 1
 # date_minute 49
 a date_month 1
 # date_second 60
 a date_wday 1
 # date_year 1
 a date_zone 1
 a index 1
 # linecount 1
 a punct 12
 a splunk_server 1
 # timeendpos 1

i	Time	Event	
✓	9/9/19 9:22:14.000 AM	Sep 09 09:22:14 1,2019/09/09 09:22:14,001606001116, TRAFFIC, end, 1, 2019/09/09 09:22:14, 192.168.0.2, 205.171.2.25, 0.0.0.0, 0.0.0.0, rule1, tng\crusher, , dns, vsys1, trust, untrust, ethernet1/2, ethernet1/1, forwardAll, 2019/09/09 09:22:14, 38090, 1, 61910, 53, 0, 0, 0x200000, udp, allow, 326, 75, 251, 2, 2019/09/09 09:22:14, 0, any, 0, 0, 0x0, 192.168.0.0-192.168.255.255, United States, 0, 1, 1	
Event Actions ▾			
Type	Field	Value	Actions
Selected	✓ host ▾	lnmbp.local	▾
	✓ source ▾	sample1.txt	▾
	✓ sourcetype ▾	syslog	▾
Time ⚙	_time ▾	2019-09-09T09:22:14.000+01:00	
Default	index ▾	main	▾
	linecount ▾	1	▾
	punct ▾	⏏	▾
	splunk_server ▾	lnmbp.local	▾

> 9/9/19 Sep 09 09:22:14 1,2019/09/09 09:22:14,001606001116, THREAT, ur1, 1, 2019/09/09 09:22:14, 192.168.0.6, 65.55.12.249, 0.0.0.0, 0.0.0.0.

Using the Add-On: Palo Alto

<https://splunkbase.splunk.com/app/2757/>

New Search

source="pan:threat"

5,736 events (before 9/9/19 10:13:37.000 AM) No Event Sampling

Events (5,736) Patterns Statistics Visualization

Format Timeline Zoom Out Zoom to Selection Deselect 1 minute per column

List Format 20 Per Page

Selected Fields: host, source, sourcetype

Interesting Fields: action, action_flags, app, app:able_to_transfer_file, app:category, app:default_ports, app:evasive, app:excessive_bandwidth, app:has_known_vulnerability, app:is_saas, app:is_sanctioned_saas, app:pervasive_use, app:prone_to_misuse, app:risk, app:subcategory, app:technology, app:tunnels_other_application, app:used_by_malware, application, category, client_ip

9/9/19 9:22:14.192 AM Sep 09 09:22:14 1,2019/09/09 09:22:14,001606001116,THREAT,url,1,2019/09/09 09:22:14,192.168.0.6,65.55.12.249,0.0.0.0,0.0.0.0,rule1,tng\picard,web-browsing,vsys1,trust,untrust,ethernet1/2,ethernet1/1,forwardAll,2019/09/09 09:22:14,63454,1,2102,80,0,0,0x208000,tcp>alert,"www.microsoft.com/en-us/default.aspx", (9999),business-and-economy,informational,client-to-server,0x0,192.168.0.0-192.168.255.255,United States,0,text/html

Event Actions

Type	Field	Value	Actions
Selected	host	inmbp.local	
	source	sample1.txt	
	sourcetype	pan:threat	
Event	action	allowed	
	action_flags	0x0	
	app	web-browsing	
	app:able_to_transfer_file	yes	
	app:category	general-internet	
	app:default_ports	tcp/80	
	app:evasive	no	
	app:excessive_bandwidth	no	
	app:has_known_vulnerability	yes	
	app:is_saas	no	
	app:is_sanctioned_saas	no	
	app:pervasive_use	yes	

Using the App: Palo Alto

<https://splunkbase.splunk.com/app/491/>

Palo Alto Networks Activity Threats Operations Search

All Incidents Edit Export ...

Product: ANY Vsys: Client IP: Server IP: Log Subtype:

Severity: ANY Action: All AutoFocus Tag: ANY Submit Hide Filters

Search produced no results.

0 Correlated Incidents Per Hour

0 Network Incidents Per Hour

0 Endpoint Incidents Per Hour

0 Aperture SaaS Incidents Per Hour

52 Malicious WildFire Submissions Per Hour

Threat Subtypes Over Time

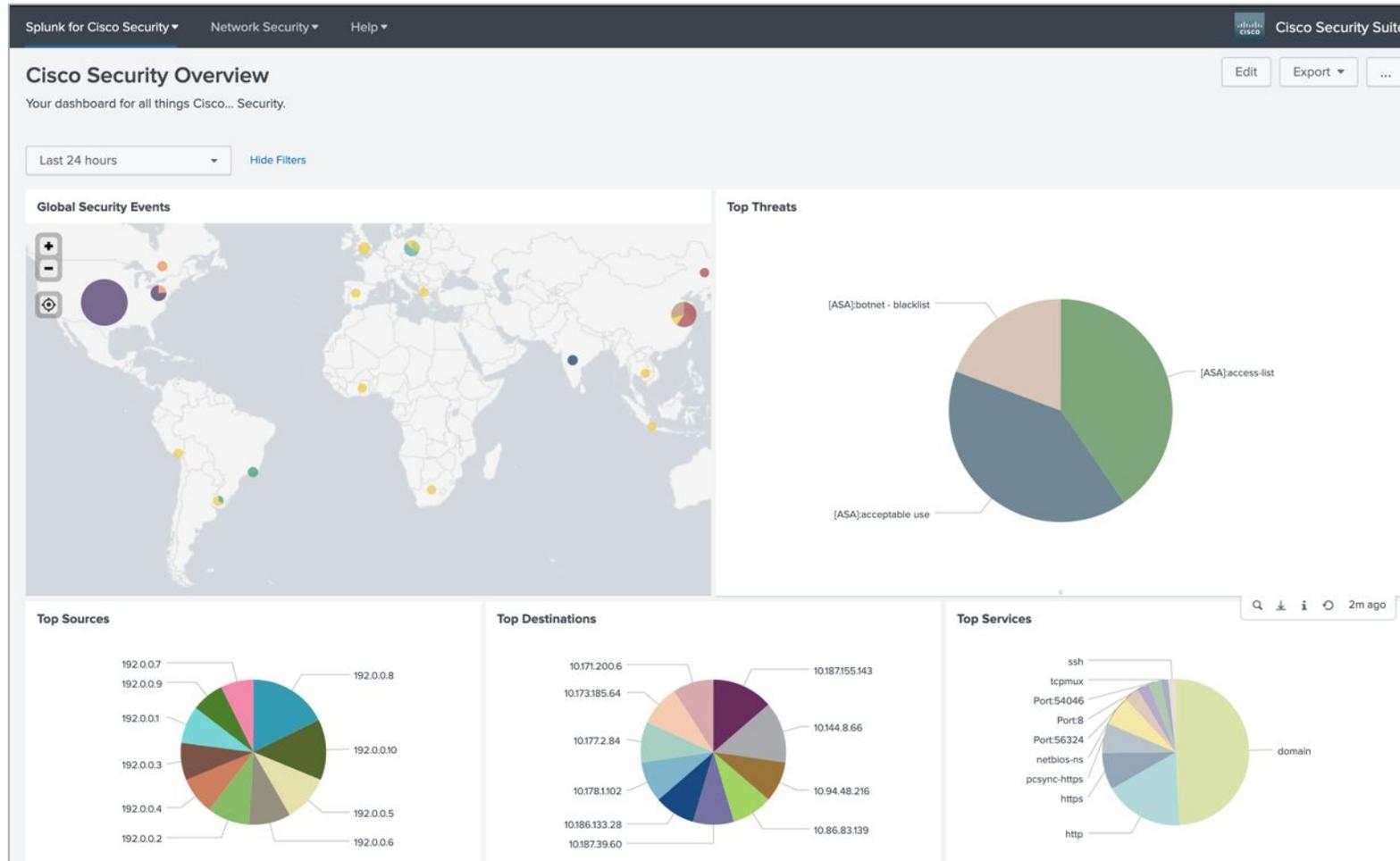
Threat Severity Over Time

Incident Investigation Feed

_time	log_subtype	threat_name	severity	action	app	client_ip	server_ip	autofocus_tags	user	file_name
2019-09-09 09:21:17	wildfire	254796918	medium	allowed	web-browsing	192.168.0.3	64.39.66.153		tng\crusher	zKo.exe
2019-09-09 09:20:44	wildfire	254796028	medium	allowed	web-browsing	192.168.0.2	50.23.163.176		tng\jordy	filescoutsetup.exe
2019-09-09 09:19:59	wildfire	254801358	medium	allowed	web-browsing	192.168.0.3	93.115.95.168		tng\crusher	movie1080p.mkv.exe

Using the App: Cisco Security Suite

<https://splunkbase.splunk.com/app/525/>





What If There is No Add-on?

That's OK!

Data Discovery

Understand the data source

Your networking organization wants another set of logs into Splunk

Ask yourself...

- Are the logs coming from an appliance or application?
- How will I access the logs?
- Who needs access to the logs in Splunk?
- How long do I need to retain the logs?
- Can I get a log sample?

Data Discovery

Our answers...

You discovered these are logs from a new application called “Acme Photon”

The logs are written to a directory on the application server

There is no add-on available on Splunkbase (yet)

Sample logs and the complete file path have been provided

```
/opt/acme/photon/audit.log
```

```
/opt/acme/photon/db.log
```

```
/opt/acme/photon/scan.log
```

```
/opt/acme/photon/web.log
```

What Are the Sourcetypes?

<https://docs.splunk.com/Documentation/AddOns/released/Overview/Sourcetypes>

Naming format: vendor:product:technology:format

/opt/acme/photon/audit.log

- acme:photon:audit

/opt/acme/photon/db.log

- acme:photon:db

/opt/acme/photon/scan.log

- acme:photon:scan

/opt/acme/photon/web.log

- acme:photon:web

Sample Logs

```
2019-09-24 13:04:04 device=something1 user login (paynealbert)
result="<code>1</code>" from 192.168.1.10.136:51750 to 192.168.10.1:443
uri="https://acme.com/"
```

```
2019-09-24 13:04:06 device=something2 user login (steven81)
result="<code>2</code>" from 192.168.1.14:10607 to 192.168.10.2:443
uri="https://acme.com/"
```

Configuring the Metadata

- props.conf

```
[acme:photon:audit]
TIME_PREFIX = ^
MAX_TIMESTAMP_LOOKAHEAD = 19
TIME_FORMAT = %Y-%m-%d %H:%M:%S
```

```
SHOULD_LINEMERGE = false
LINE_BREAKER = ([\r\n]+)
EVENT_BREAKER_ENABLE = true
EVENT_BREAKER = ([\r\n]+)
```

```
TRUNCATE = 10000
```

- inputs.conf

```
[monitor:///opt/acme/photon/audit.log]
index = myindex
sourcetype = acme:photon:audit
```

*host & source are usually
automagically set*

Where are the Fields?

auto vs custom

2019-09-24 13:04:04 device=something1 user login (paynealbert) result="<code>1</code>" from 192.168.1.10.136:51750 to 192.168.10.1:443 uri="https://acme.com/"

- `_time = 2019-09-24 13:04:04` `TIME_PREFIX, TIME_FORMAT, etc.`
- `device = something1` auto extracted
- `user = paynealbert` ???
- `result = <code>1</code>` auto extracted
- `result_code = 1` ???
- `src_ip = 192.168.1.10.136` ???
- `src_port = 51750` ???
- `dest_ip = 192.168.10.1` ???
- `dest_port = 443` ???

Where are the Fields?

Custom extractions

```
[acme:photon:audit]
```

```
# continued...
```

```
EXTRACT-username = user\slogin\s\((?P<username>[^\)]+)\)
```

```
EXTRACT-result_code = result="<code>(P<result_code>\d+)\<
```

```
EXTRACT-src_ip_port = from\s(?P<src_ip>.+?):(?P<src_port>\d+)\s
```

```
EXTRACT-dest_ip_port = to\s(?P<dest_ip>.+?):(?P<dest_port>\d+)\s
```

Common Information Model (CIM)

<http://docs.splunk.com/Documentation/CIM/latest/User/Overview>

A way of normalizing your data for maximum efficiency at search time

Splunk Certified TA's typically include necessary normalizations

Allows end-users to search using common fields such as "user" across many sourcetypes

Extract your fields, normalize, then tag your data

CIM Data Models

- Alerts
- Application State
- Authentication
- Certificates
- Databases
- Data Loss Prevention
- Email
- Interprocess Messaging
- Intrusion Detection
- Inventory
- Java Virtual Machines
- Malware
- Network Resolution (DNS)
- Network Sessions
- Network Traffic
- Performance
- Ticket Management
- Updates
- Vulnerabilities
- Web

What Do You Do?

Making your data usable

Already have a proper sourcetype

Extract your fields

Create field aliases

- username AS user
- src_ip as src, dest_ip as dest

Create calculations

- `action=if(action="OK","success","failure")`

Create tags

Use the Data Models as a guide

Fields for Authentication event datasets

The following table lists the extracted and calculated fields for the event datasets in the model. The table does not include any inherited fields. For more information, see [How to use these reference tables](#).

Dataset name	Field name	Data type	Description	Expected values
Authentication	<code>action</code>	string	The action performed on the resource.	<code>success</code> , <code>failure</code>
Authentication	<code>app</code>	string	The application involved in the event (such as <code>ssh</code> , <code>splunk</code> , <code>win:local</code>).	
Authentication	<code>dest</code>	string	The target involved in the authentication. You can alias this from more specific fields, such as <code>dest_host</code> , <code>dest_ip</code> , or <code>dest_nt_host</code> .	

Tags used with Authentication event datasets

The following tags act as constraints to identify your events as being relevant to this data model. For more information, see [How to use these reference tables](#).

Dataset name	Tag name
Authentication	authentication
<code> __ Default_Authentication</code>	default
<code> __ Insecure_Authentication</code>	cleartext OR insecure
<code> __ Privileged_Authentication</code>	privileged

Applying CIM

Normalizing your fields

```
[acme:photon:audit]
```

```
# continued...
```

```
FIELDALIAS-user = username AS user
```

```
FIELDALIAS-src = src_ip AS src
```

```
FIELDALIAS-dest = dest_ip AS dest
```

```
EVAL-action = case(result_code==1,"success",result_code==2,"failure",1=1,null())
```

Tagging Your Data

- eventtypes.conf

```
[acme_photon_audit_authentication]
```

```
search = sourcetype=acme:photon:audit "user login"
```

- tags.conf

```
[eventtype=acme_photon_audit_authentication]
```

```
authentication = enabled
```

Become Famous

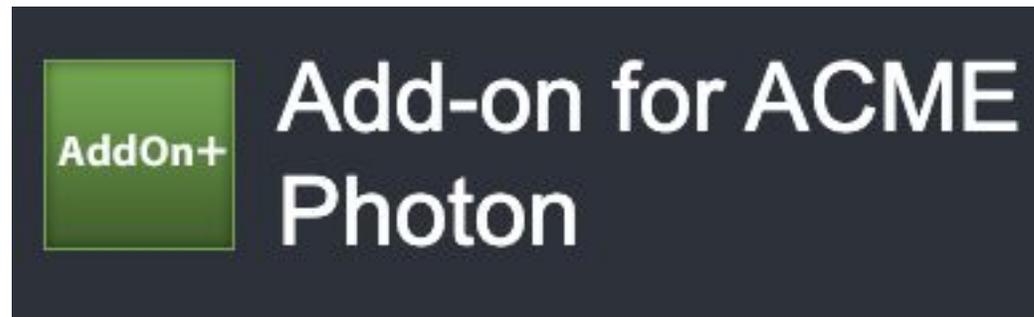
Submit your TA to Splunkbase 😊

Submit content to Splunkbase with the web UI

The user account that you use to submit your content becomes the account that owns the app. Log in to that user account to respond to questions and manage Splunkbase content.

1. Navigate to <https://splunkbase.splunk.com/develop/> and click **Submit Your App**.
2. Review and agree to the Splunk Apps Developer Distribution License.

You are now on the Hosting page.



BUILT BY
[Your Name Here](#)



Recap

Don't Forget!

Recap

Always set a sourcetype!

Don't use syslog as a sourcetype!

Don't use csv, json, xml as a sourcetype!

Use Splunkbase as a starting point!

Make your own sourcetype if you must, but use the naming format:
vendor:product:technology:format

Make the data usable – CIM!

Do You Have Bad Sourcetypes?

inputs.conf without a sourcetype defined

- | tstats count where index=* (sourcetype=*-* OR sourcetype=*too_small) by sourcetype

Don't make the puppy sad ☹...

...so you can sleep good at night 😊



What's Next?

Splunk Enterprise Data Administration

- <https://www.splunk.com/view/SP-CAAAPSE>

Go see these sessions....

- SEC2280: Break Down Silos By Ingesting Multi-Purpose Data Sources into Splunk
- BA2766: Machine Data Alchemy: Transforming Digital Exhaust into Campus Gold
- FN2103: Exploratory Data Analysis on Aviation Safety Data
- FN1921: Saving the Nation's Food Supply with Data-Driven Analytics
- DEV1140: Next Generation Data Ingestion and Preparation with Splunk

Resources

<https://docs.splunk.com/Documentation/Splunk/latest/Data/Whysourcetypesmatter>

<https://docs.splunk.com/Documentation/AddOns/released/Overview/Sourcetypes>

<https://www.splunk.com/blog/2012/08/10/sourcetypes-whats-in-name.html>

<https://www.splunk.com/blog/2010/02/11/sourcetypes-gone-wild.html>



splunk>

Thank

You



Go to the .conf19 mobile app to

RATE THIS SESSION

