Splunk4Ninjas - Data Onboarding

splunk>

Forward-Looking Statements

This presentation may contain forward-looking statements regarding future events, plans or the expected financial performance of our company, including our expectations regarding our products, technology, strategy, customers, markets, acquisitions and investments. These statements reflect management's current expectations, estimates and assumptions based on the information currently available to us. These forward-looking statements are not guarantees of future performance and involve significant risks, uncertainties and other factors that may cause our actual results, performance or achievements to be materially different from results, performance or achievements expressed or implied by the forward-looking statements contained in this presentation.

For additional information about factors that could cause actual results to differ materially from those described in the forward-looking statements made in this presentation, please refer to our periodic reports and other filings with the SEC, including the risk factors identified in our most recent quarterly reports on Form 10-Q and annual reports on Form 10-K, copies of which may be obtained by visiting the Splunk Investor Relations website at www.investors.splunk.com or the SEC's website at www.sec.gov. The forward-looking statements made in this presentation are made as of the time and date of this presentation. If reviewed after the initial presentation, even if made available by us, on our website or otherwise, it may not contain current or accurate information. We disclaim any obligation to update or revise any forward-looking statement based on new information, future events or otherwise, except as required by applicable law.

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. We undertake no obligation either to develop the features or functionalities described, in beta or in preview (used interchangeably), or to include any such feature or functionality in a future release.

splunk>

Presenter Name

Job title

splunk>

Meet Your Presenters!

Presenter Name

Job title

Presenter Name

Job title

splunk>

Agenda

Data Sources
What and Where?

Getting Data In (GDI)
How?

Tuning the Data
Configuration Options

Lab Exercises
Getting Hands-On

What Happens Next?
What and Why?

splunk>

REGISTRATION 5 MINS

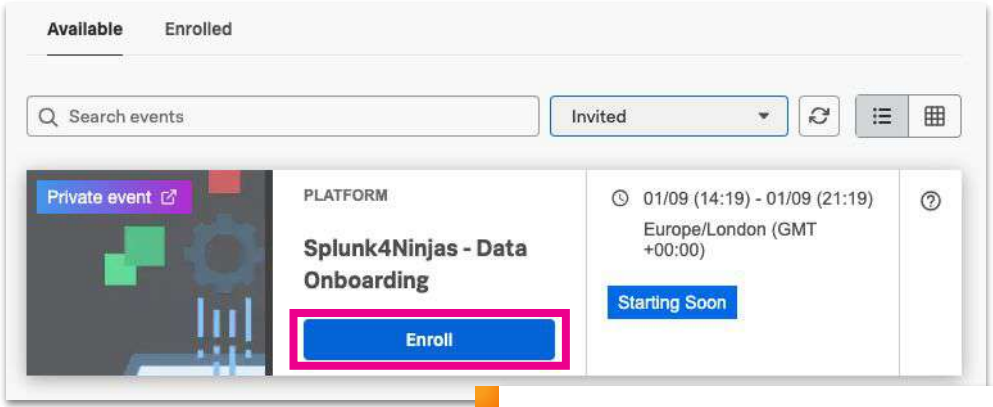
Enroll in Today's Workshop

Tasks

1. Get a splunk.com account if you don't have one yet: https://splk.it/SignUp
2. Enroll in the Splunk Show workshop event: https://show.splunk.com/event/<eventID>
3. Download the hands-on lab guide: https://splk.it/S4N-DO-Lab-Guide

   Contains step-by-step instructions for all of today's exercises!
4. Download a copy of today's slide deck: https://splk.it/S4N-DataOnboarding

Goal



Enroll in today's event

splunk>

How Data Moves Through Splunk

INDEXING

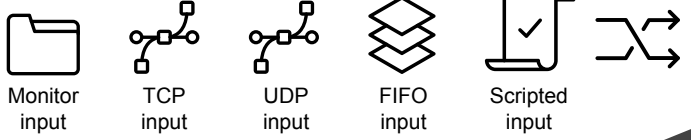
- Parsed events written to index on disk
- Writes both compressed raw data & the corresponding index files

PARSING

- Event processing occurs here
- Examine, analyse & transform data
- Identify, parse & set timestamps
- Apply regex transform rules

INPUT

- Raw data stream from source
- Contents ignored, no events



DATA RETENTION

- Buckets: **HOT**, **WARM**, **COLD**, **FROZEN** and **THAWED**
- Cloud: Smart Store, Active Archive

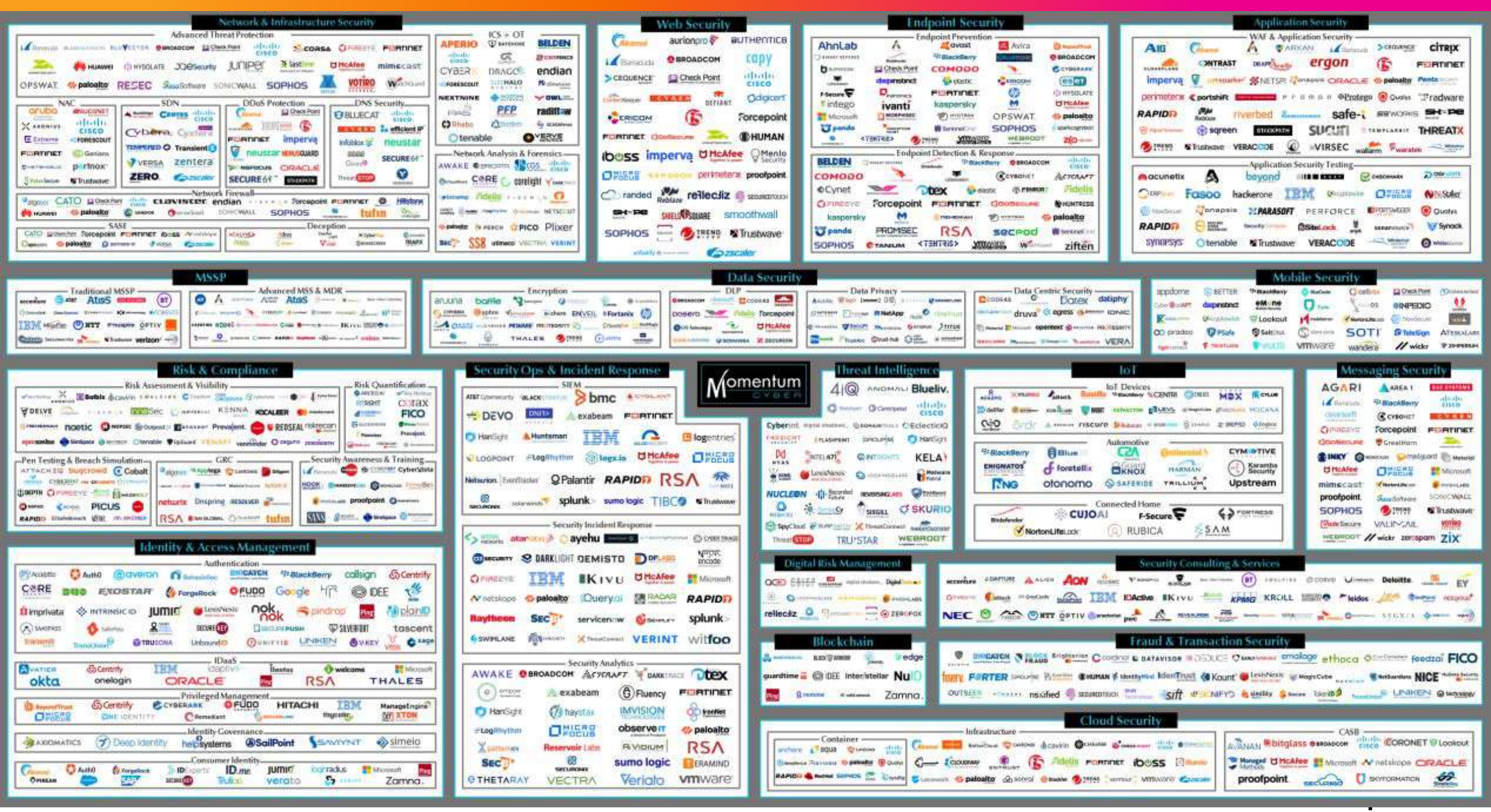
SEARCH

- Manages how the user accesses, views and searches the indexed data
- Stores knowledge objects: reports, event types, dashboards, alerts and field extractions

Data Sources

and the importance of data quality

splunk>



Source: Momentum Cyber

What Can Splunk Ingest?

Local File Monitoring

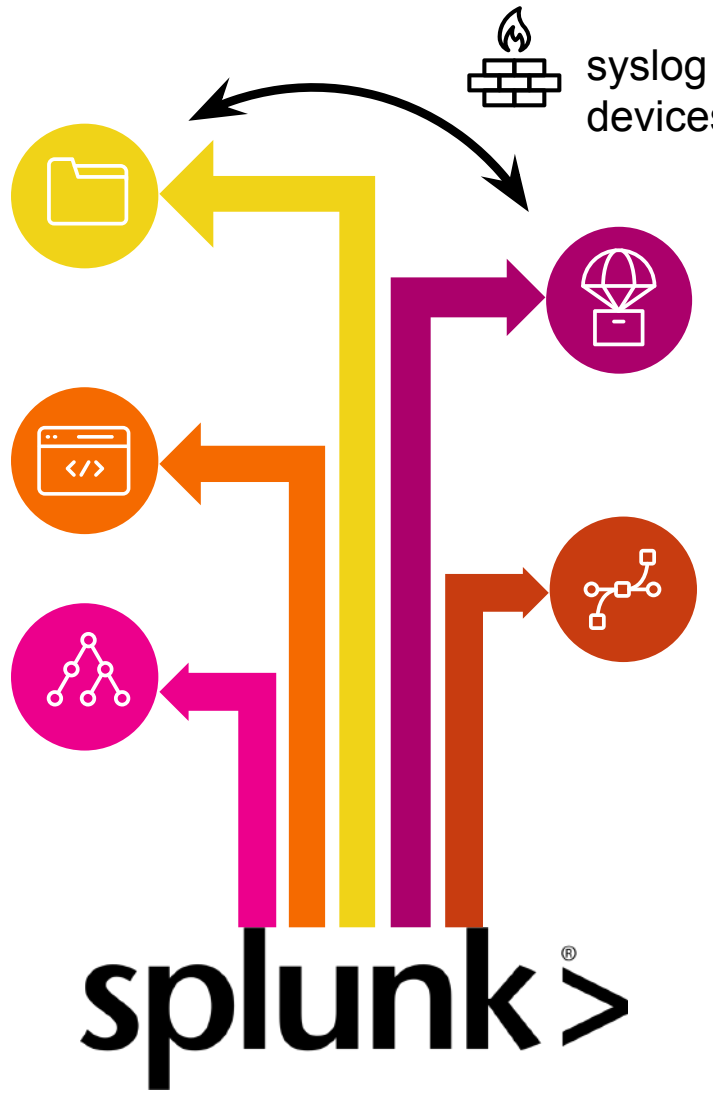
- Universal Forwarder

Aggregated / API Sources

- Heavy Forwarder

Event Logs, Active Directory, OS Stats, Unix/Linux/ Windows Hosts

- Universal Forwarder



DevOps, IoT, Containers

- HTTP Event Collector (HEC)
- Agentless

Wire Data

- via Splunk Stream
- HTTP Event Collector (HEC) or Universal Forwarder

splunk>

Data Quality

Data collection is the foundation of any Splunk platform.
Decisions without quality is simply guessing.

Poor Data Quality
Increases difficulty with searching data sources and often search performance

Performance
Improved indexing times and reduced indexing latency

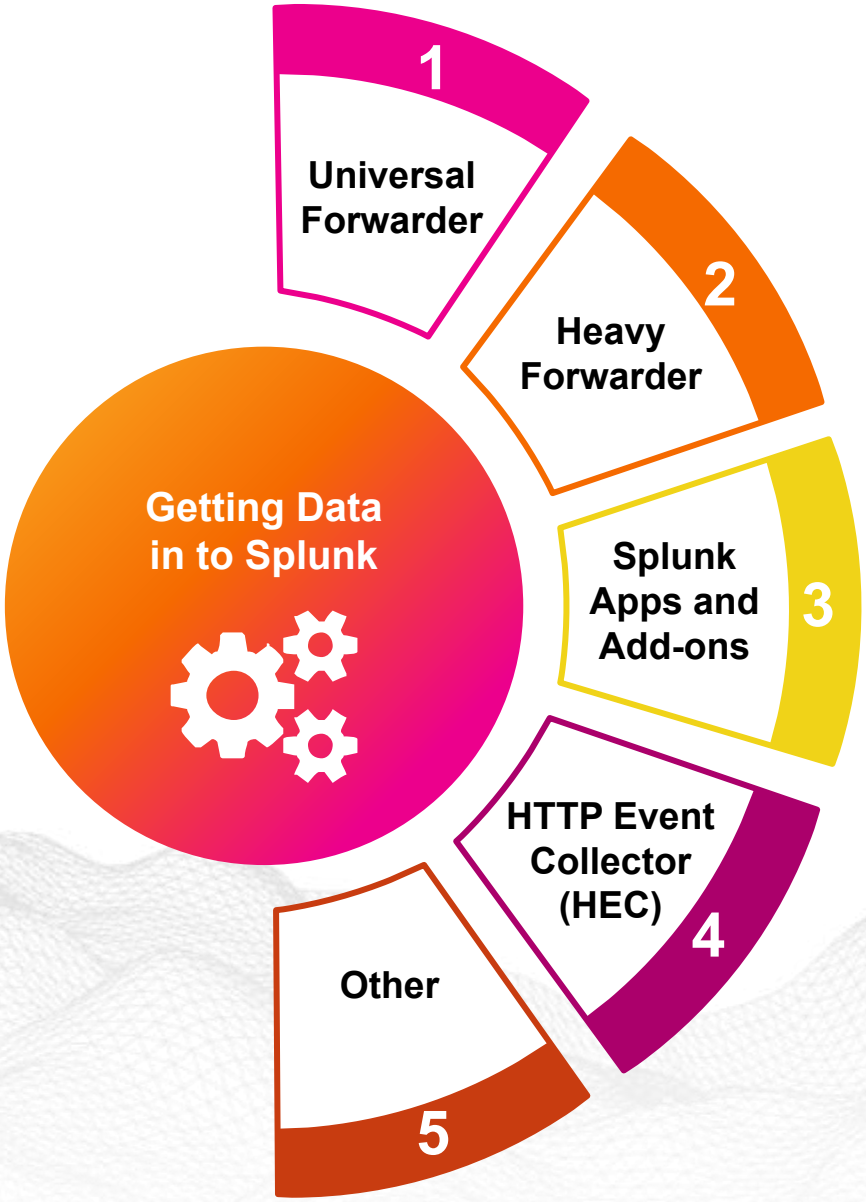
Source Types Matter
Auto assigned source types may add inefficiencies and potential gaps to matching manipulation rules

Timestamps
Improperly timestamped events can lead to early deletion

splunk>

Getting Data In (GDI)

splunk>



Get Started with GDI

Universal Forwarder
- Dedicated executable, smallest footprint
- Replaces Light Forwarder (deprecated)
- Best method for most situations

Heavy Forwarder
- Splunk Enterprise instance minus some features
- Able to index locally while forwarding
- Slower than Universal Forwarder

Splunk Apps and Add-ons (Splunkbase)
- Apps and add-ons with preconfigured inputs, views and knowledge objects
- Examples: DB Connect, Splunk Stream, Add-on for Microsoft Windows

HTTP Event Collector (HEC)
- For Splunk Cloud Platform
- Get data directly from a source with the HTTP or HTTPS protocols

Other
- Direct upload via Splunk Web, CLI, API, scripts, FIFO queues, collectd, Data Manager for Splunk Cloud

splunk>

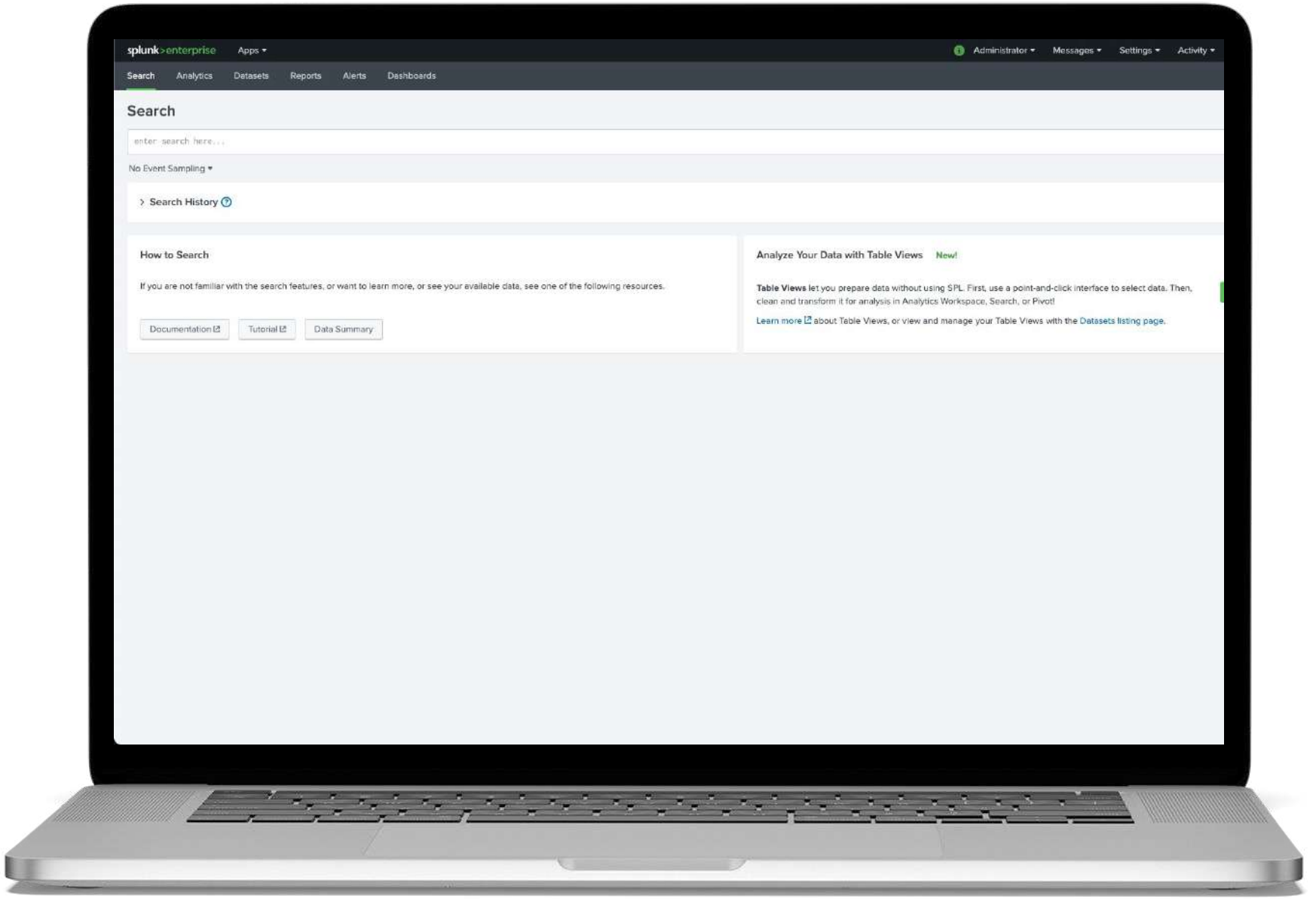
Universal vs Heavy Forwarder

Features and capabilities	Universal forwarder	Heavy forwarder
Type of Splunk Enterprise instance	Dedicated executable	Full Splunk Enterprise, with some features disabled
Footprint (memory, CPU load)	Smallest	Medium-to-large (depending on enabled features)
Bundles Python?	No	Yes
Handles data inputs?	All types (but scripted inputs might require Python installation)	All types
Forwards to Splunk Enterprise?	Yes	Yes
Forwards to 3rd party systems?	Yes	Yes
Serves as intermediate forwarder?	Yes	Yes
Indexer acknowledgment (guaranteed delivery)?	Optional	Optional (version 4.2 and later)
Load balancing?	Yes	Yes

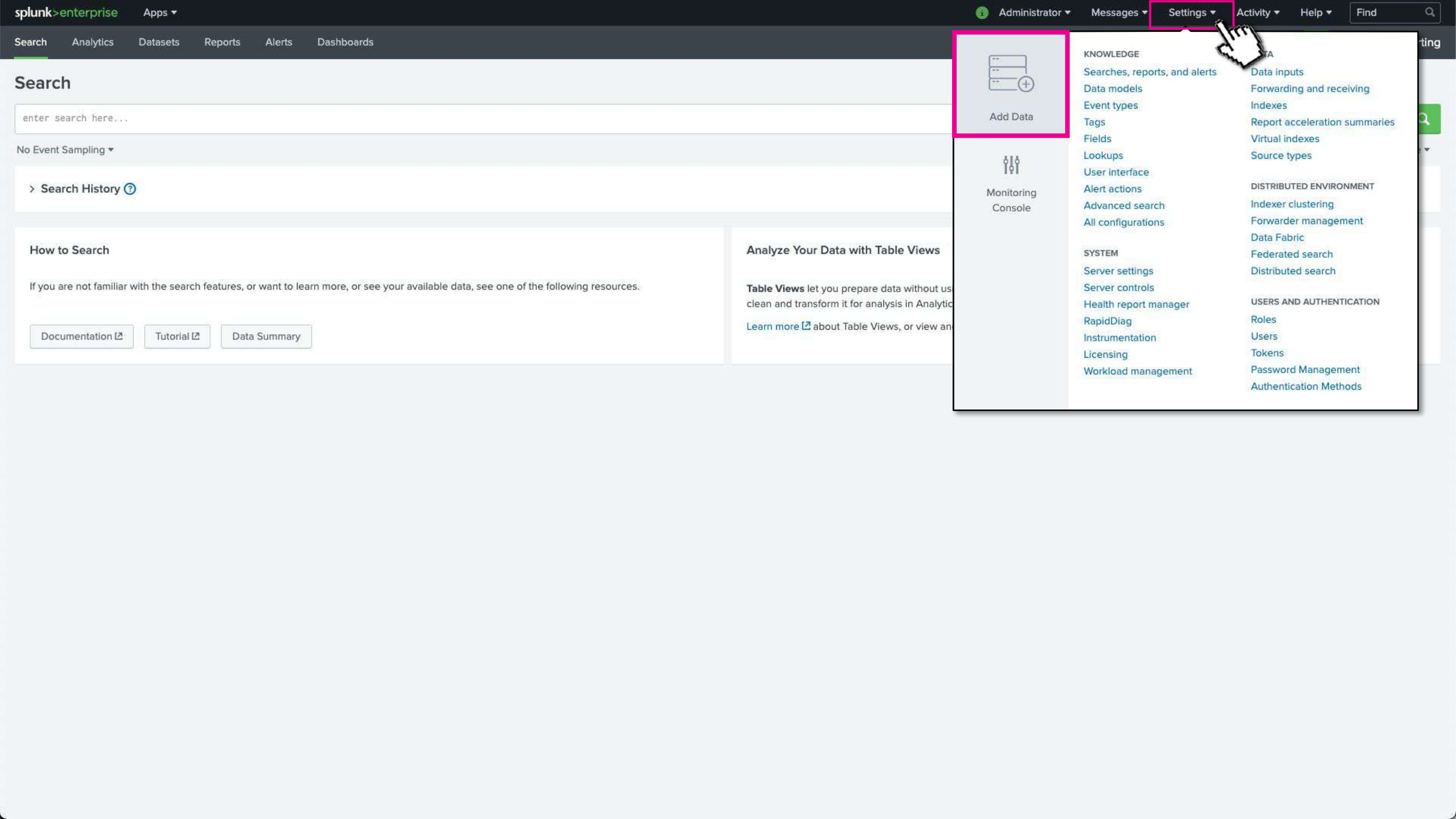
A full list of features and capabilities can be found at:
https://docs.splunk.com/Documentation/Splunk/latest/Forwarding/Typesofforwarders

Getting Files into Splunk

- Direct Upload
- File Monitor
- Custom scripts
- syslog and plenty more!



splunk>


splunk>enterprise
Apps
Administrator
Messages
Settings
Activity
Help
Find
Search
Analytics
Datasets
Reports
Alerts
Dashboards
Search
enter search here...
No Event Sampling
Search History
How to Search
If you are not familiar with the search features, or want to learn more, or see your available data, see one of the following resources.
Documentation
Tutorial
Data Summary
Analyze Your Data with Table Views
Table Views let you prepare data without us
clean and transform it for analysis in Analytic
Learn more about Table Views, or view an
Add Data
Monitoring Console
KNOWLEDGE
Searches, reports, and alerts
Data models
Event types
Tags
Fields
Lookups
User interface
Alert actions
Advanced search
All configurations
SYSTEM
Server settings
Server controls
Health report manager
RapidDiag
Instrumentation
Licensing
Workload management
DATA
Data inputs
Forwarding and receiving
Indexes
Report acceleration summaries
Virtual indexes
Source types
DISTRIBUTED ENVIRONMENT
Indexer clustering
Forwarder management
Data Fabric
Federated search
Distributed search
USERS AND AUTHENTICATION
Roles
Users
Tokens
Password Management
Authentication Methods

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
Get your cloud computing data in to the Splunk platform.
10 data sources

Networking
Get your networking data in to the Splunk platform.
2 data sources

Operating System
Get your operating system data in to the Splunk platform.
1 data source

Security
Get your security data in to the Splunk platform.
3 data sources

4 data sources in total

Or get data in with the following metho

Upload
files from my computer
Local log files
Local structured files (e.g. CSV)
Tutorial for adding data

files and port
Files -
Modular

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

Add Data

Select Source — Set Source Type — Input Settings — Review — Done

< Back Next >

- **Files & Directories**
  Upload a file, index a local file, or monitor an entire directory.
- **HTTP Event Collector**
  Configure tokens that clients can use to send data over HTTP or HTTPS.
- **TCP / UDP**
  Configure the Splunk platform to listen on a network port.
- **Scripts**
  Get data from any API, service, or database with a script.
- **DBX Live Query Server**
  Run database statement in a live way
- **Systemd Journald Input for Splunk**
  This is the input that gets data from journald (systemd's logging component) into Splunk.
- **SA-Eventgen**
  This modular input generates data for Splunk.
- **Splunk Secure Gateway**
  Initializes the Splunk Secure Gateway application to talk to mobile clients over websockets
- **DB Connect Task Server**
  Task server running scheduled jobs (inputs, outputs)
- **Splunk Secure Gateway Mobile Alerts TTL**
  Cleans up storage of old mobile alerts
- **Splunk Secure Gateway Deleting Expired Tokens**
  Delete expired or invalid tokens created by Secure Gateway from Splunk
- **Splunk Secure Gateway Role Based Notification Manager**
  Used for sending mobile alerts to users by role
- **Splunk Secure Gateway Enable**
  Determine if Splunk Secure Gateway core modular inputs should be enabled
- **Splunk Secure Gateway Metrics Collector**
  Collects metrics for Splunk Secure Gateway

Configure this instance to monitor files and directories for data. To monitor all objects in a directory, select the directory. The Splunk platform monitors and assigns a single source type to all objects within the directory. This might cause problems if there are different object types or data sources in the directory. To assign multiple source types to objects in the same directory, configure individual data inputs for those objects. Learn More

File or Directory ? [ ] Browse

On Windows: c:\apache\apache.error.log or \\hostname\apache\apache.error.log. On Unix: /var/log or /mnt/www01/var/log.

Continuously Monitor | Index Once

Whitelist ?

Blacklist ?

FAQ

What kinds of files can the Splunk platform index?

Many kinds. The Splunk platform recognizes many different file formats, and you can configure it to recognize many more.

I can't access the file that I want to index. Why?

Make sure that the file is available on your system by checking mount points or mapped drives. Also, make sure the user account that the Splunk platform runs as has proper permissions to access the file.

How do I get remote data onto my Splunk platform instance?

If the data is on a machine on the same network, you can map or mount a drive to access the data. The most popular option is to forward the data by installing a universal forwarder on the machine that contains the data.

Can I monitor changes to files in addition to their content?

Yes. Best Practices suggest using native OS file auditing tools, like Audit Policy for Windows and auditd for UNIX, and then indexing the output of those tools into the Splunk platform.

What is a source type?

A source type is a field that defines how the Splunk platform handles a piece of incoming data. The source type defines specifications for line break behavior, timestamp location, and character set.

Input Settings

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

Source type

The source type is one of the default fields that the Splunk platform assigns to all incoming data. It tells the Splunk platform what kind of data you've got, so that the Splunk platform can format the data intelligently during indexing. And it's a way to categorize your data, so that you can search it easily.

Automatic | Select | New

App context

Application contexts are folders within a Splunk platform instance that contain configurations for a specific use case or domain of data. App contexts improve manageability of input and source type definitions. The Splunk platform loads all app contexts based on precedence rules. Learn More

App Context: Search & Reporting (search) ▾

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: so1

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?

Add Data

Select Source — Input Settings — Review — Done

< Back Next >

Files & Directories
Upload a file, index a local file, or monitor an entire directory.

HTTP Event Collector
Configure tokens that clients can use to send data over HTTP or HTTPS.

TCP / UDP
Configure the Splunk platform to listen on a network port.

Scripts
Get data from any API, service, or database with a script.

DBX Live Query Server
Run database statement in a live way

Systemd Journald Input for Splunk
This is the input that gets data from journald (systemd's logging component) into Splunk.

SA-Eventgen
This modular input generates data for Splunk.

Splunk Secure Gateway
Initializes th… clients over…

DB Conne…
Task server…

Splunk Se…
Cleans up s…

Splunk Se…
Delete expi… Splunk

Splunk Secure Gateway Role Based Notification Manager
Used for sending mobile alerts to users by role

Splunk Secure Gateway Enable
Determine if Splunk Secure Gateway core modular inputs should be enabled

Splunk Secure Gateway Metrics Collector
Collects metrics for Splunk Secure Gateway

Configure this instance to listen on any TCP or UDP port to capture data sent over the network (such as syslog). Learn More

TCP | UDP

Port ?
Example: 514

Source name override ?
optional
host:port

Only accept connection from ?
optional
example: 10.1.2.3, !badhost.splunk.com, *.splunk.com

FAQ

How should I configure the Splunk platform for syslog traffic?

The syslog service runs on UDP port 514 by default. If possible, send this traffic over TCP for better transmission reliability.

install a third-party syslog service on your Windows hosts, you can collect the data on the Splunk platform with syslog monitoring.

What is a source type?

A source type is a field that defines how the Splunk platform handles a piece of incoming data. The source type defines specifications for line break behavior, timestamp location, and character set.

FAQ

How should I configure the Splunk platform for syslog traffic?

The syslog service runs on UDP port 514 by default. If possible, send this traffic over TCP for better transmission reliability.

Add Data

Select Source — Input Settings — Review — Done

< Back Next >

Files & Directories
Upload a file, index a local file, or monitor an entire directory.

HTTP Event Collector
Configure tokens that clients can use to send data over HTTP or HTTPS.

TCP / UDP
Configure the Splunk platform to listen on a network port.

Scripts
Get data from any API, service, or database with a script.

DBX Live Query Server
Run database statement in a live way

Systemd Journald Input for Splunk
This is the input that gets data from journald (systemd's logging component) into Splunk.

SA-Eventgen
This modular input generates data for Splunk.

Splunk Secure Gateway
Initializes the Splunk Secure Gateway application to talk to mobile clients over

DB Conne
Task server

Splunk Se
Cleans up s

Splunk Se
Delete expi
Splunk

Splunk Se
Used for se

Splunk Se
Determine if Splunk Secure Gateway core modular inputs should be enabled

Splunk Secure Gateway Metrics Collector
Collects metrics for Splunk Secure Gateway

Configure this instance to execute a script or command and to capture its output as event data. Scripted inputs are useful when the data that you want to index is not available in a file to monitor.
Learn More

Script Path: -- Select script path --

Command ?

Interval Input ?: In Seconds

Interval ?: 60.0

Source name override ?: optional

FAQ

What kind of scripts can I run?

It depends on the operating system that this machine runs. If it runs a *nix OS, you can create and run shell scripts or binaries that send text output to the stdout or stderr output channels. If it runs Windows, you can deploy batch files or PowerShell scripts. You can

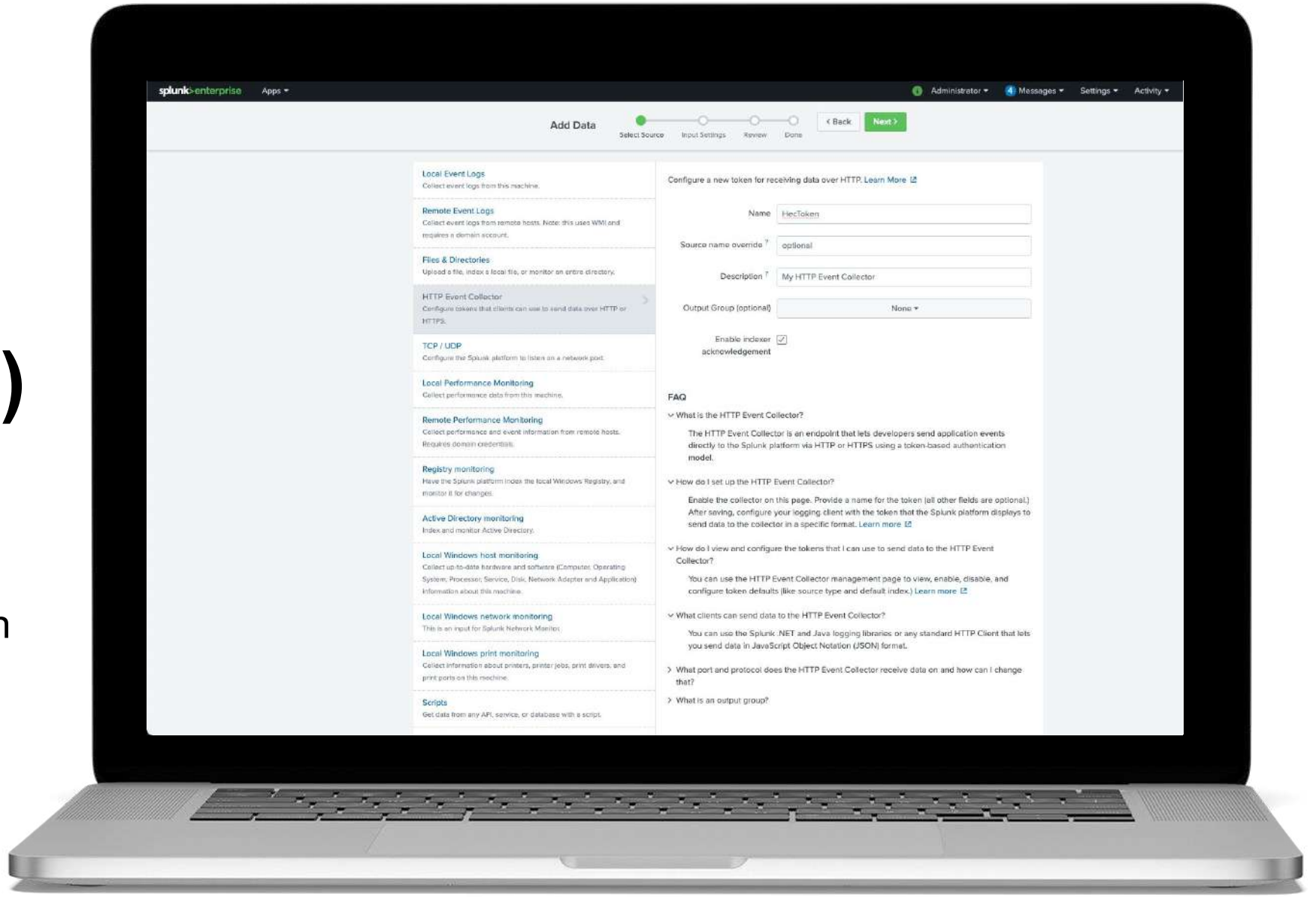
FAQ

What kind of scripts can I run?

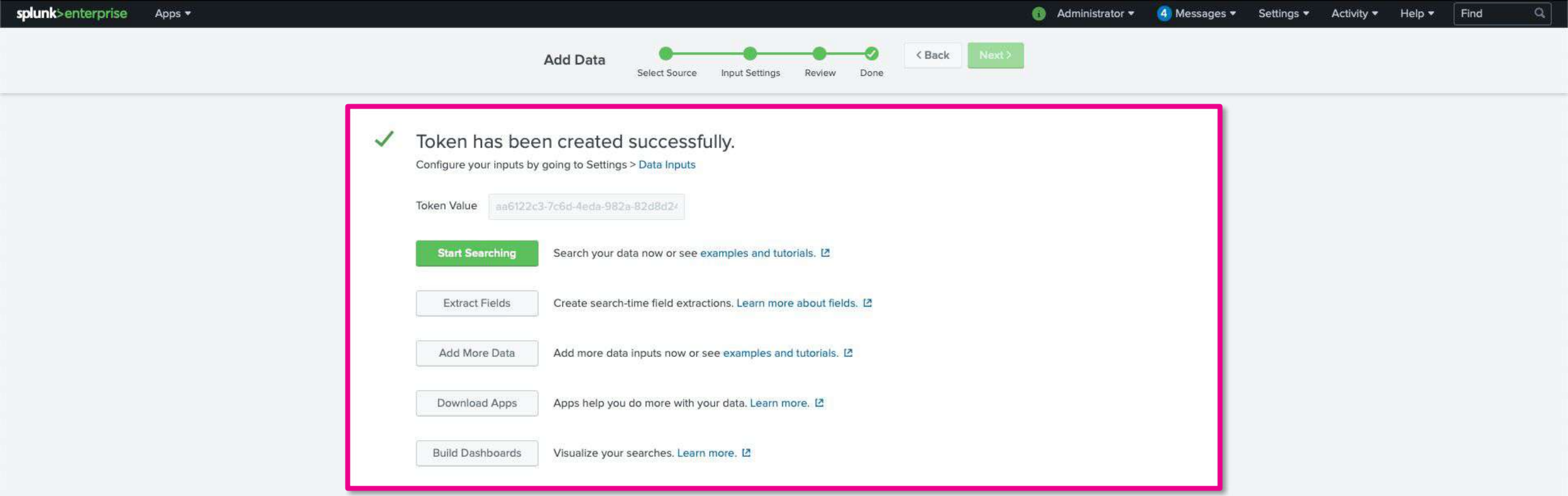
It depends on the operating system that this machine runs. If it runs a *nix OS, you can create and run shell scripts or binaries that send text output to the stdout or stderr output channels. If it runs Windows, you can deploy batch files or PowerShell scripts. You can create and use scripts to get data from APIs. You can also use a wrapper to execute a script that the Splunk platform would not otherwise support. Learn More

Get Data with HTTP Event Collector (HEC)

- Send data and app events
- Supports http and https
- Token-based authentication
- No need for Forwarder



splunk>




```
curl https://hec.example.com:8088/services/collector/event -H "Authorization: Splunk B5A
79AAD-D822-46CC-80D1-819F80D7BFB0" -d '{"event": "hello world"}'
```

Source Types

Formatting your data

splunk>

‘Magic 8’ Best Practice Settings

For tuning data ingestion

EVENT_BREAKER:
regular expression for event breaks*

EVENT_BREAKER_ENABLE:
`TRUE*`

TRUNCATE:
999999 (always a high number)

TIME_FORMAT:
`strptime` format of the timestamp

SHOULD_LINEMERGE:
`FALSE` (always)

LINE_BREAKER:
regular expression for event breaks

TIME_PREFIX:
regex of the text that leads up to the timestamp

MAX_TIMESTAMP_LOOKAHEAD:
how many characters for the timestamp

splunk>

Measurable Impact

Wall-Clock Seconds Ingesting 10M Events

01	**Defaults** Out of the box settings	190 seconds
02	**MAX_TIMESTAMP_LOOKAHEAD**	179 seconds
03	**MTL + LINEMERGE = FALSE**	105 seconds
04	**MTL + LM + TIME_PREFIX**	105 seconds
05	**MTL + LM + TIME_FORMAT**	51 seconds
06	**MTL + LM + TF + TIME_PREFIX**	51 seconds
07	**MTL + LM + TF + RF + ANNOTATE_PUNCT = FALSE**	44 seconds

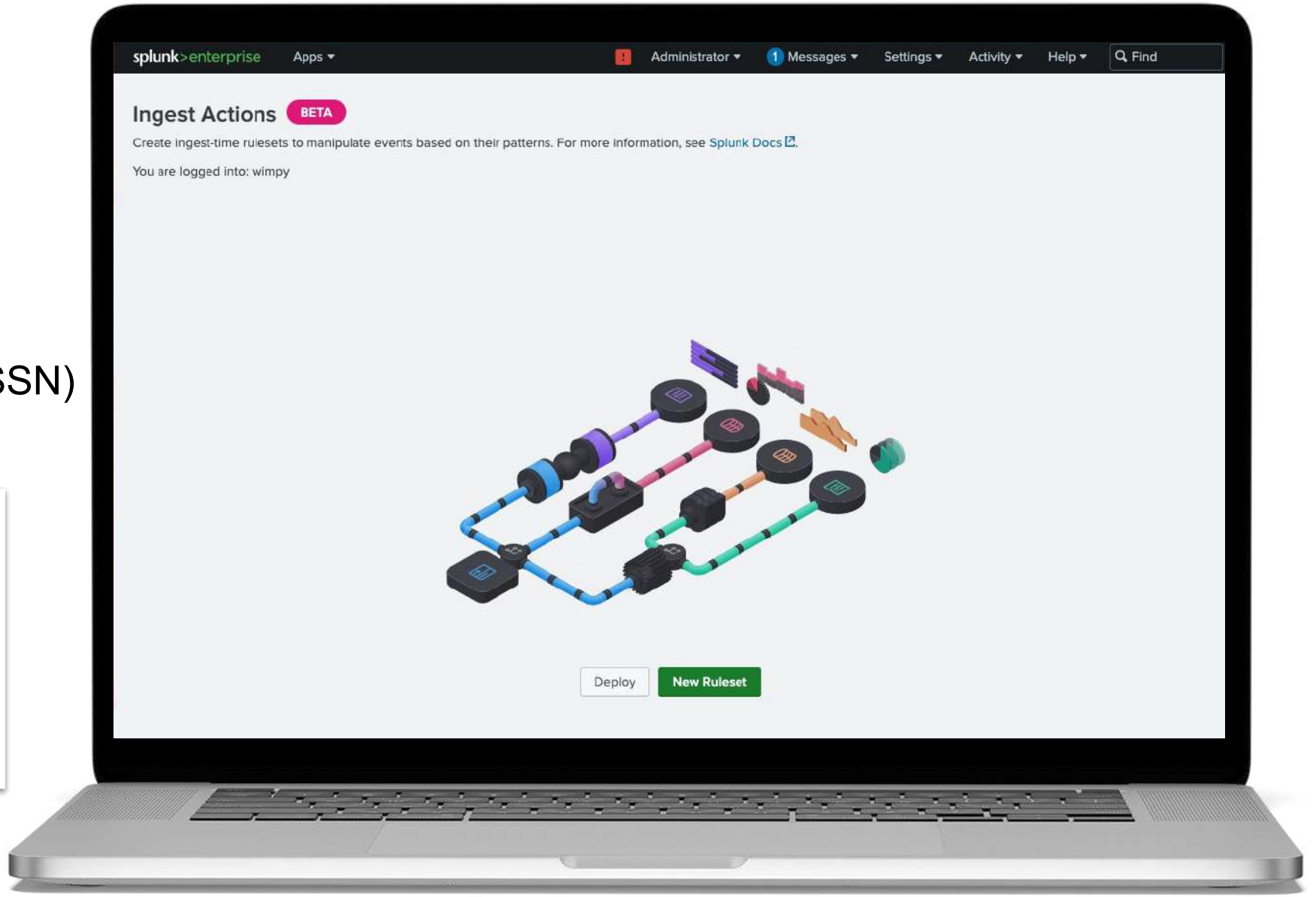
single line is an event, LM had little work to do

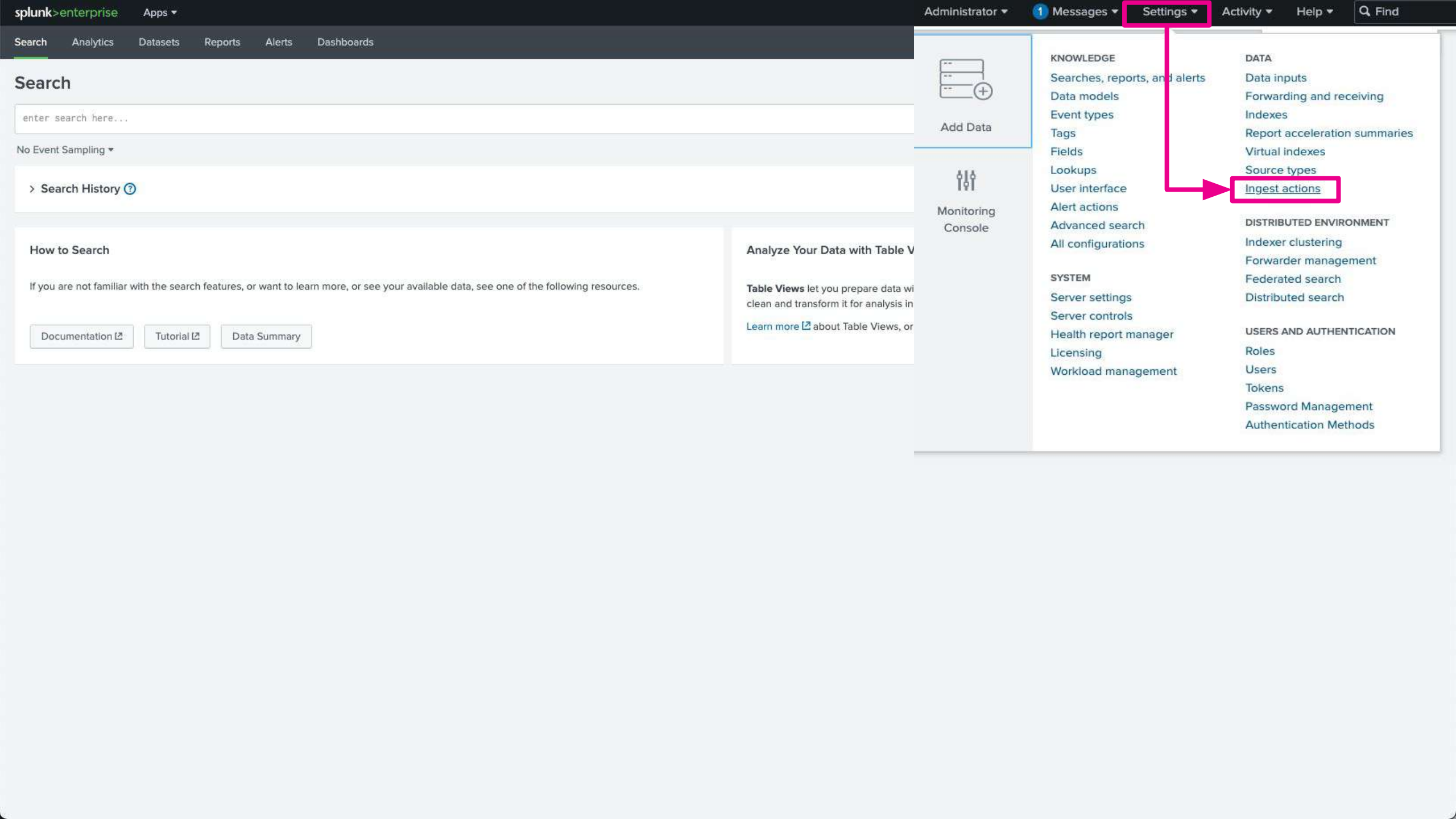
splunk>

Ingest Actions

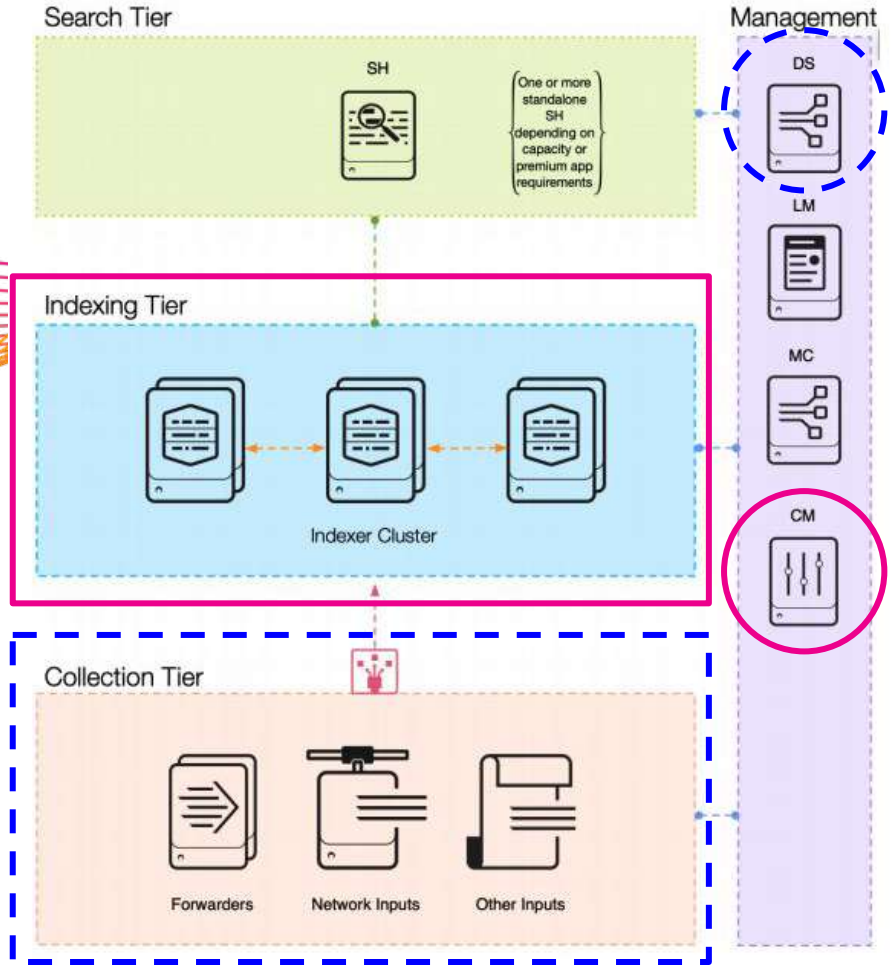
- Route Data to S3 Bucket
- Filter Data
- Mask Sensitive Data (CC,SSN)

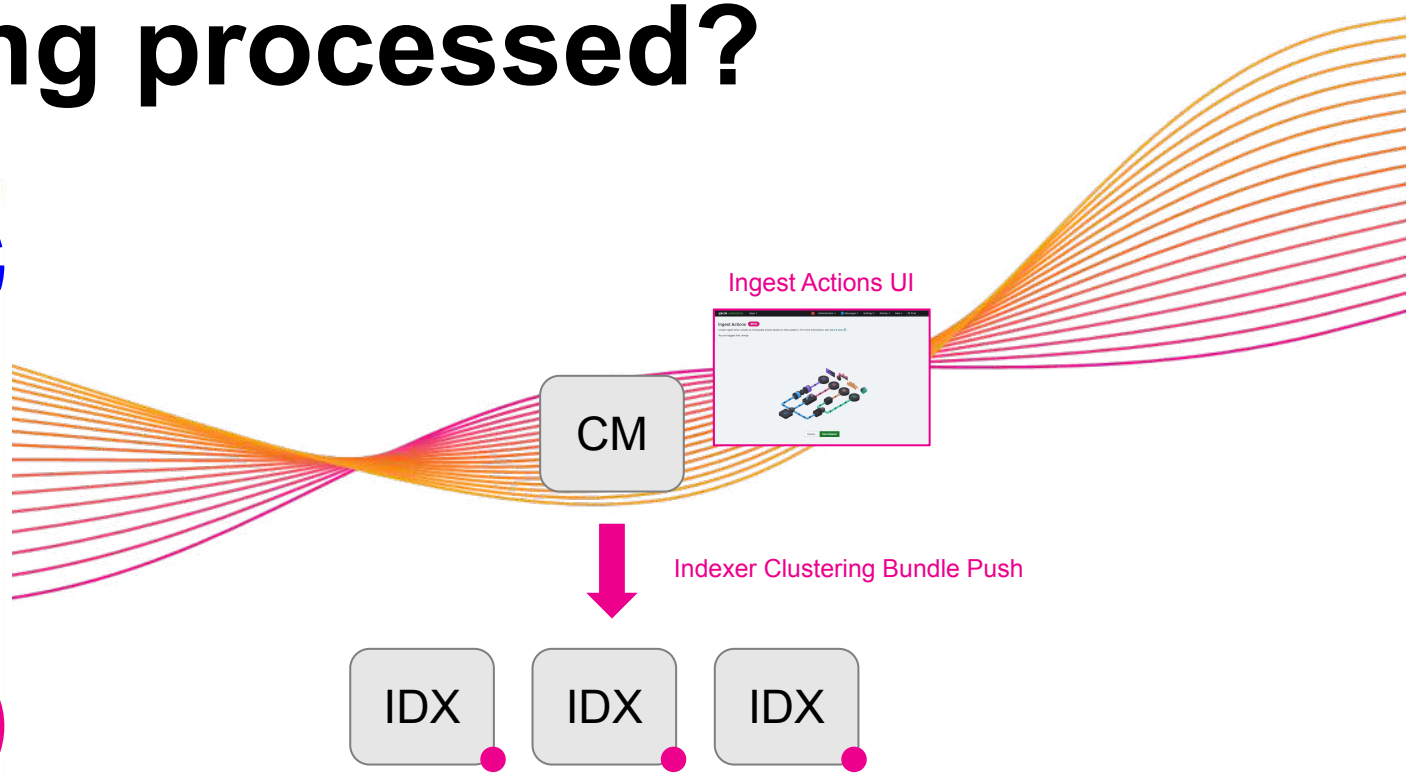
Goal: not get between data ingest and enrichment applied through Apps and Add-ons, but to offer additional routing and filtering options before the events are indexed.




splunk>enterprise
Apps
Administrator
1 Messages
Settings
Activity
Help
Find
Search
Analytics
Datasets
Reports
Alerts
Dashboards
Search
enter search here...
No Event Sampling
Search History
How to Search
If you are not familiar with the search features, or want to learn more, or see your available data, see one of the following resources.
Documentation
Tutorial
Data Summary
Analyze Your Data with Table V
Table Views let you prepare data wi
clean and transform it for analysis in
Learn more about Table Views, or
Add Data
Monitoring Console
KNOWLEDGE
Searches, reports, and alerts
Data models
Event types
Tags
Fields
Lookups
User interface
Alert actions
Advanced search
All configurations
SYSTEM
Server settings
Server controls
Health report manager
Licensing
Workload management
DATA
Data inputs
Forwarding and receiving
Indexes
Report acceleration summaries
Virtual indexes
Source types
Ingest actions
DISTRIBUTED ENVIRONMENT
Indexer clustering
Forwarder management
Federated search
Distributed search
USERS AND AUTHENTICATION
Roles
Users
Tokens
Password Management
Authentication Methods

Where are rules being processed?





`$SPLUNK_HOME/etc/master-apps/_cluster/splunk_ingest_actions`

Or on standalone:

`$SPLUNK_HOME/etc/apps/splunk_ingest_actions/local/`

splunk>

Interaction with TRANSFORMS

The `RULESET` setting is similar in behavior to the `TRANSFORMS` setting in props.conf

Considerations when using `RULESET`:

- `TRANSFORMS` is applied first if a source type matches both a `RULESET` and a `TRANSFORMS` stanza configuration on the cluster peers.
- A source type should *only* have one `RULESET` configuration. Additional rulesets is unsupported.
- Only use the Ingest Actions page, or the REST API `/services/data/ingest/rulesets` endpoint to create a `RULESET` configuration.
- A `RULESET` is applied to data streams from a UF and to cooked data from the HF. This represents a change in behavior from using the `TRANSFORMS` setting.

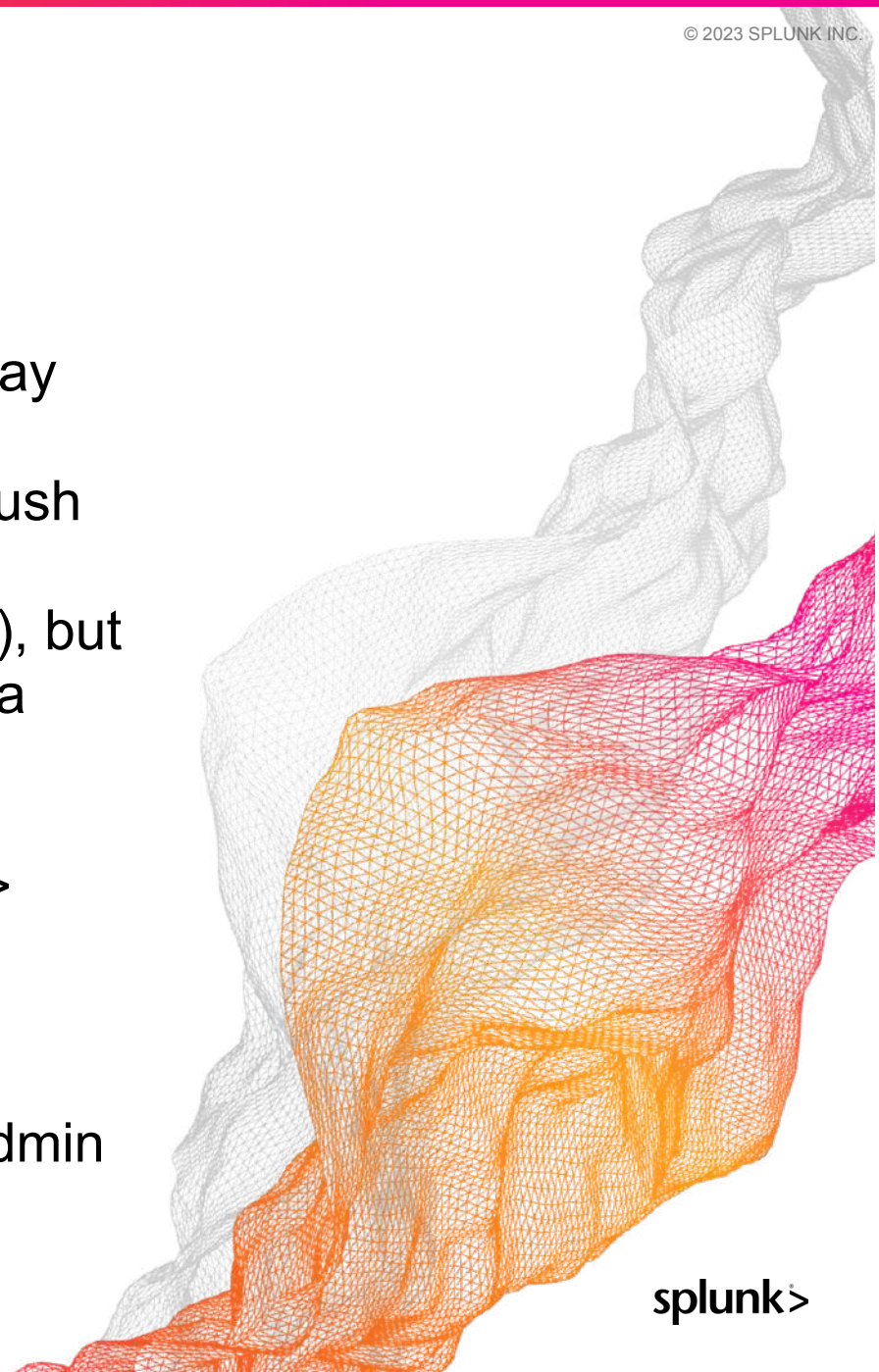
Order of operations for Ingest Actions on a HF:

1. Existing HWF TRANSFORMS
2. Ingest Actions HWF RULESET TRANSFORMS
3. *INDEXER TRANSFORMS (skipped if already touched by HF TRANSFORMS)*
4. Ingest Actions INDEXER RULESET TRANSFORMS (accepts cooked data)

splunk>

Deployment Gotchas

- Saved rulesets will be deployed on the next bundle push, important if you save now but decide not to deploy right away
- Deleted ruleset is only applied to the peers with a bundle push
- Rulesets are hot-reloadable (will not trigger Rolling-Restart), but if there are *other* configuration settings staged that require a RR, a RR will initiate
- Check the Configuration Bundle page for issues (Settings > Distributed Environment > Indexer Clustering > Edit > Configuration Bundle Actions)
- IA requires the *list_ingest_ruleset* and *edit_ingest_ruleset* admin capabilities to run



splunk>

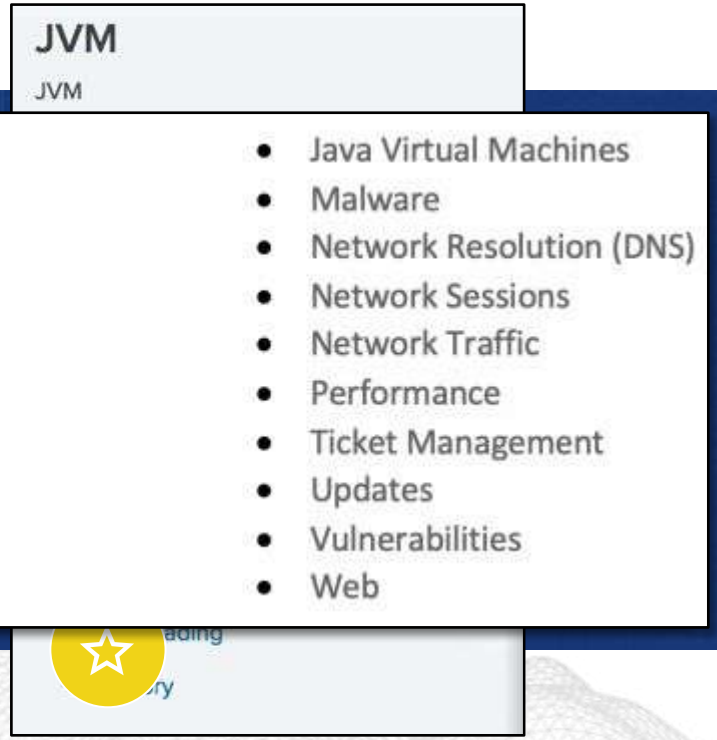
Data Models

Standardize your data

splunk>

What are Data Models?

A data model is a search-time mapping of data into a hierarchical structure



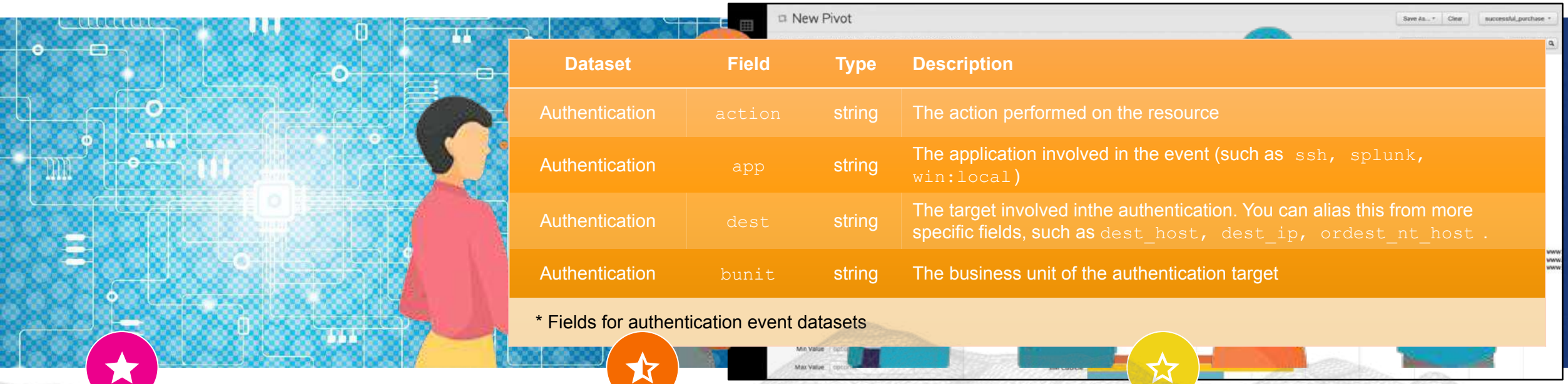
A collection of objects

That have constraints and attributes

Where child objects inherit those constraints and attributes

splunk>

Why Use Data Models?



Dataset	Field	Type	Description
Authentication	`action`	string	The action performed on the resource
Authentication	`app`	string	The application involved in the event (such as `ssh`, `splunk`, `win:local`)
Authentication	`dest`	string	The target involved inthe authentication. You can alias this from more specific fields, such as `dest_host`, `dest_ip`, `ordest_nt_host` .
Authentication	`bunit`	string	The business unit of the authentication target

* Fields for authentication event datasets

Fields

Common fields allow for unique insights across many sets of data

Pivot

Allow general users to build reports with an Excel-like interface

Common Language

CIM allows for normalized fields that can be tailored for many use cases

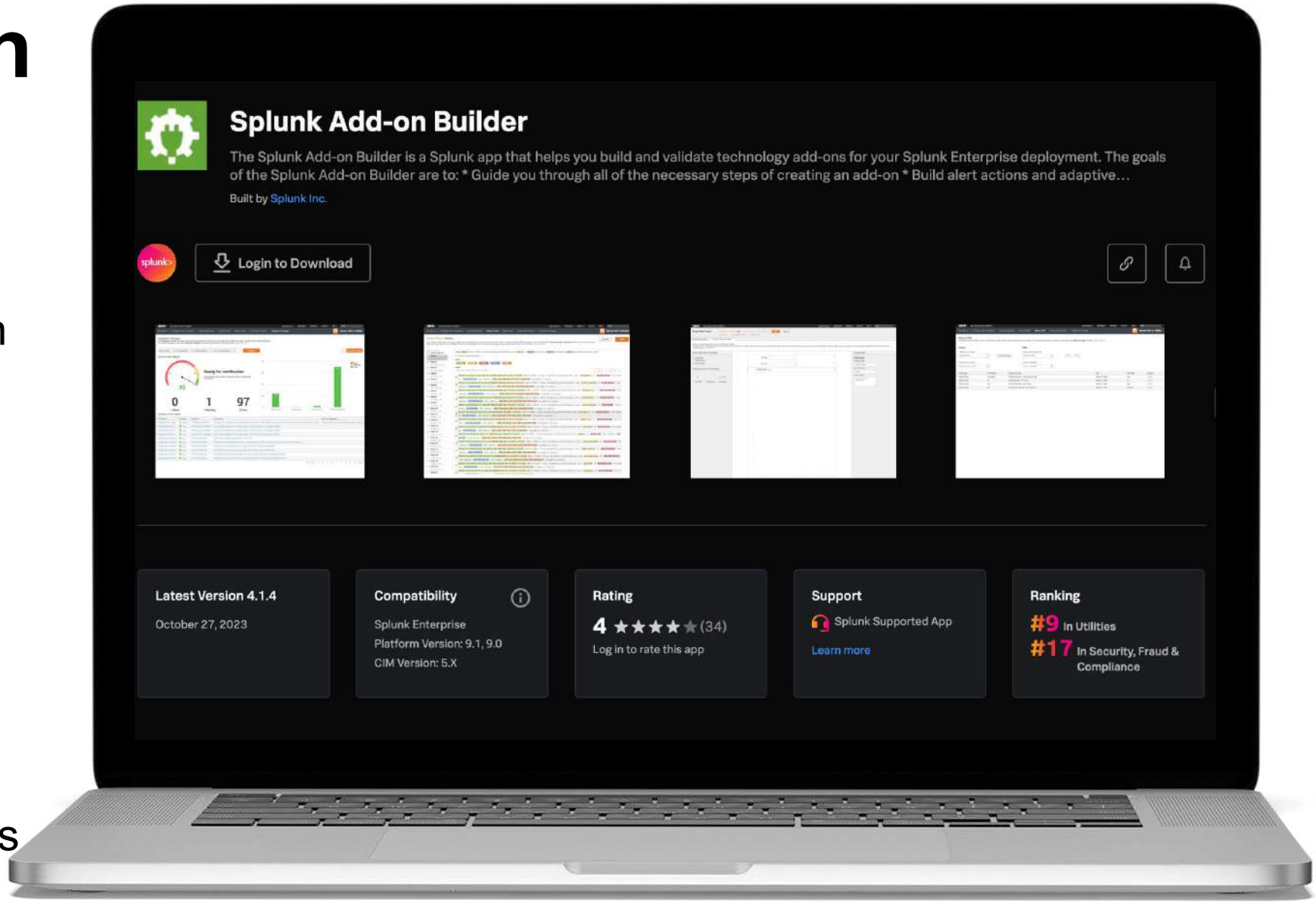
splunk>

Splunk Add-on Builder

splunk>

Splunk Add-on Builder

- Onboard additional data
- Build and validate add-on
- Custom configurations
  - field extractions
  - transforms
- Knowledge mapping
  - event types
  - tags
- Custom views and reports



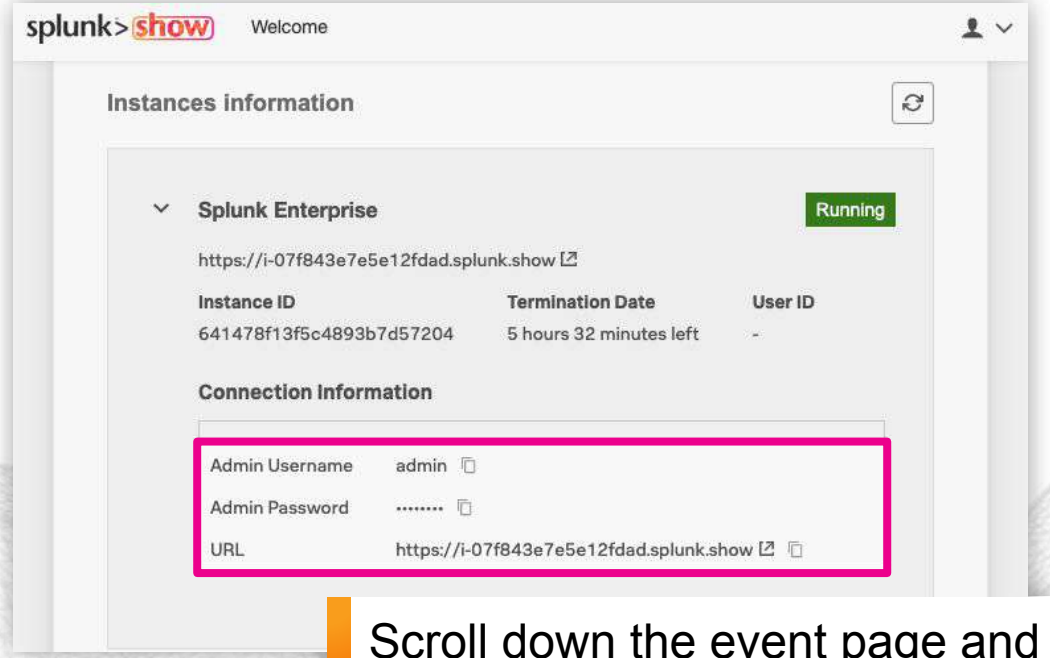
splunk>

Lab Access

splunk>

Login to Splunk

Locate your instance URL and credentials in the Splunk Show event https://show.splunk.com



Scroll down the event page and expand the **Splunk Enterprise** section to view your login details

Log in to your Splunk instance

Login using the credentials from Splunk Show

splunk>

Login to Splunk

Locate your instance URL and credentials in the list provided

<link to Gsheet with list of instances>

adminUsername	adminPassword
admin	changeme
admin	changeme
admin	changeme
admin	changeme
admin	changeme

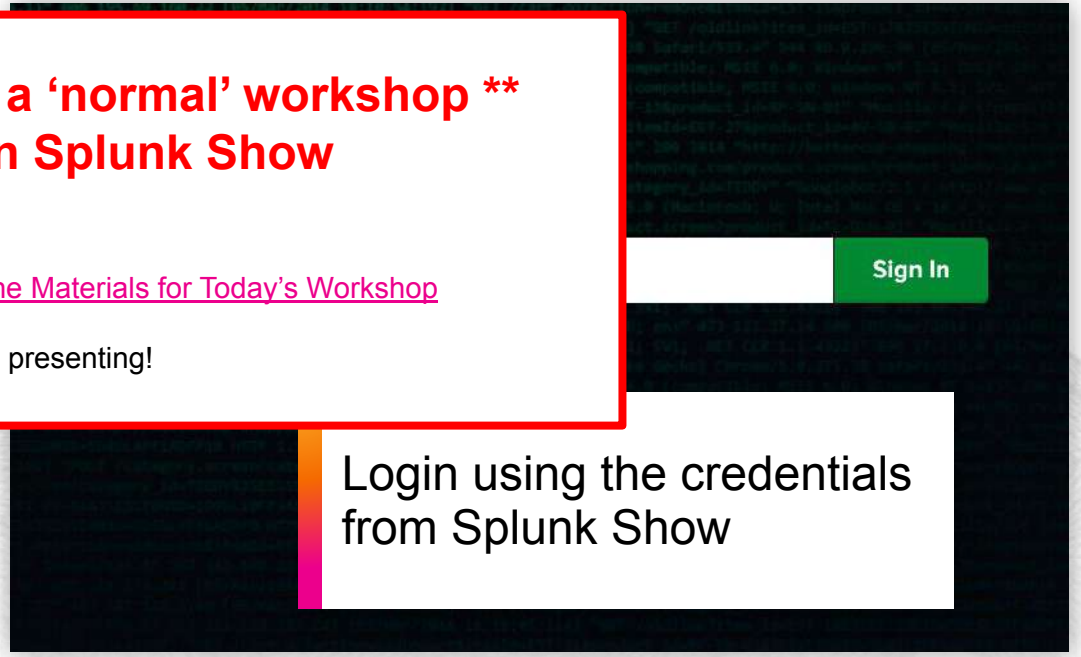
**** Optional slide for running a 'normal' workshop ****
(i.e. not an 'event') in Splunk Show

Presenter instructions:

1. **Follow the instructions on this slide:** Obtain the Materials for Today's Workshop
2. **Skip the previous slide** and **unskip this slide!**
3. **Delete or move this text box off screen** before presenting!

Locate your instance URL in the list provided - make sure you use the one with your name next to it!

Log in to your Splunk instance



Login using the credentials from Splunk Show

splunk>

Lab Exercises

splunk>

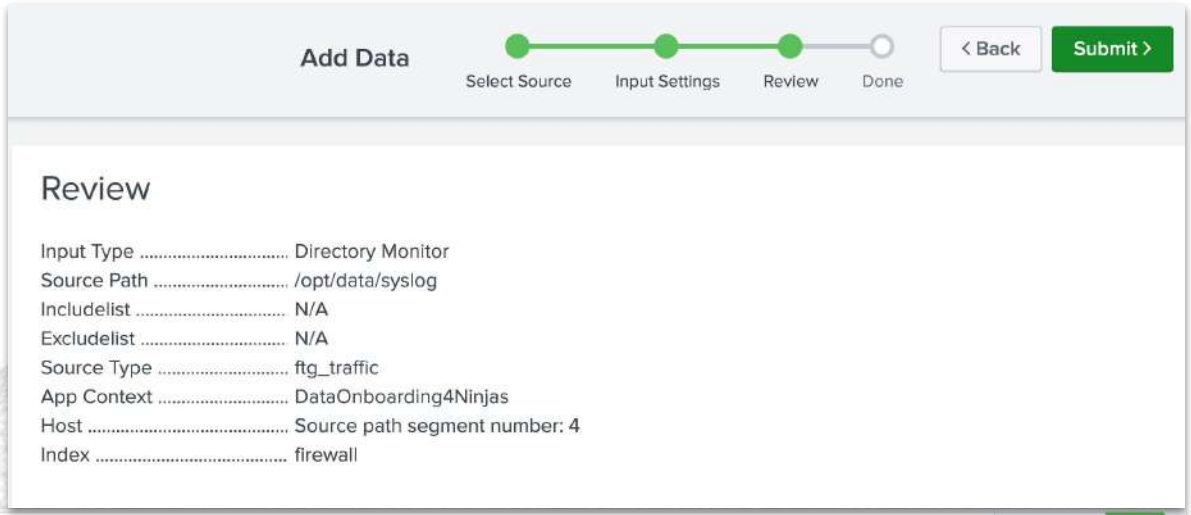
Lab 1: File Monitor

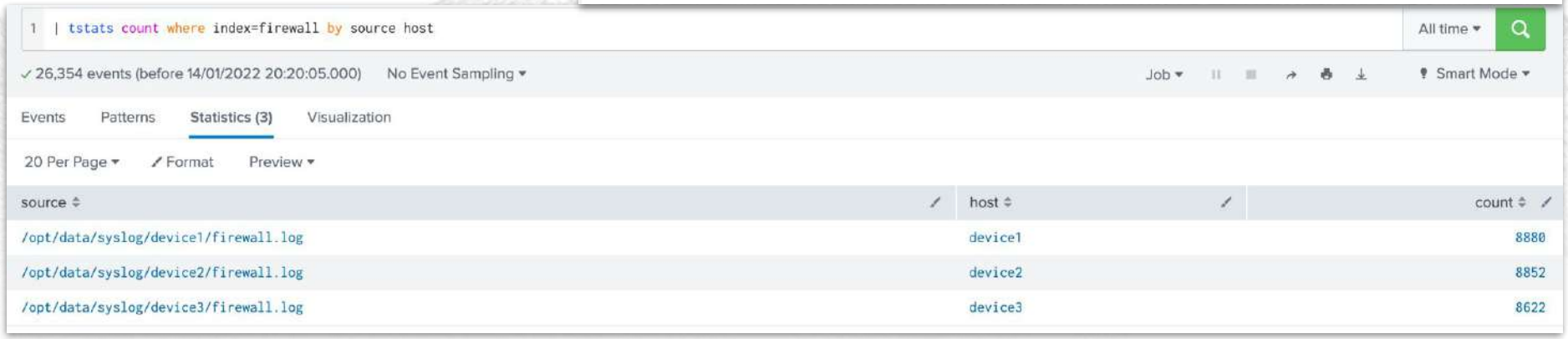
Collect some local files within different folder structure by using inputs.conf

Activity:

1. Monitor the syslog directory on the local machine
   - Each device has a dedicated folder
2. Adjust input settings
   - host to match the folder name
3. Confirm Input

Goal:





splunk>

Break for Lab 1

splunk>

Lab 1: Option 1 - File Monitor: GUIDED

Collect some local files within different folder structure by using inputs.conf

GUI:

Navigate to Settings > Add Data > Files and Directories

Select your firewall logs under /opt/data/syslog/

***KEY:* Choose continuous monitor for this lab!**

 # Input Settings

New Source Type: ftg_ttraffic

App Context: DataOnboarding4Ninjas

Host Value: Set ‘Segment in Path’ to 4

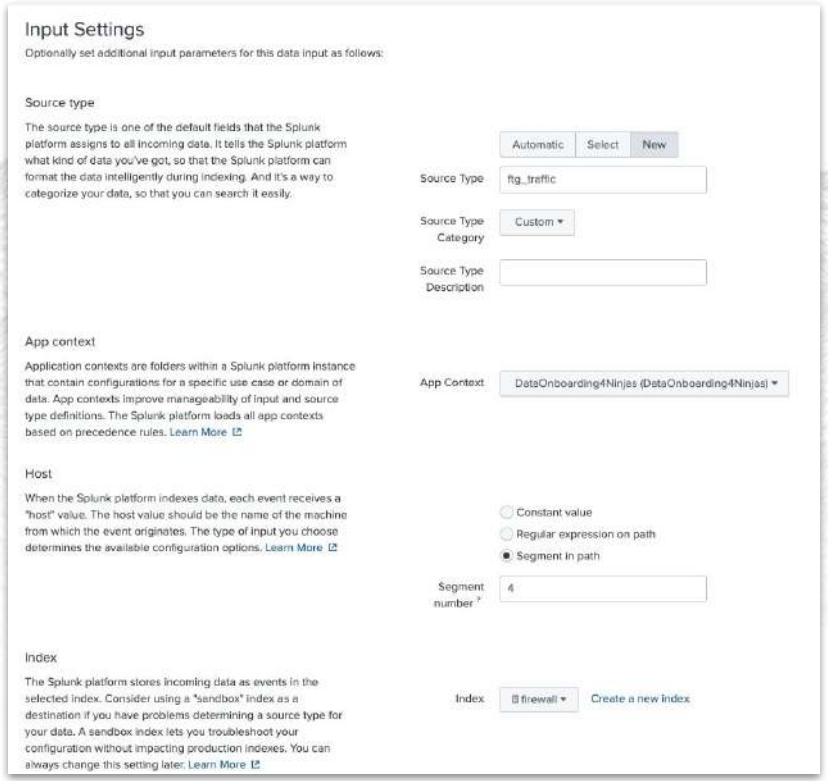
Index: firewall

 # Check result

```
| tstats count where index=firewall by host source
```

***KEY:* Host value now has the device name**

Goal:



Lab 1: Option 2 - File Monitor: GUIDED

Collect some local files within different folder structure by using inputs.conf

CLI:

Navigate to /opt/splunk/etc/system/local/inputs.conf

```
[monitor:///opt/data/syslog/*/firewall.log]
sourcetype = fgt_traffic
index = firewall
host_segment = 4
```

 # Reload Monitor

```
/opt/splunk/bin/splunk reload monitor

/opt/splunk/bin/splunk btool inputs list
--debug | grep firewall
```

 # Check result

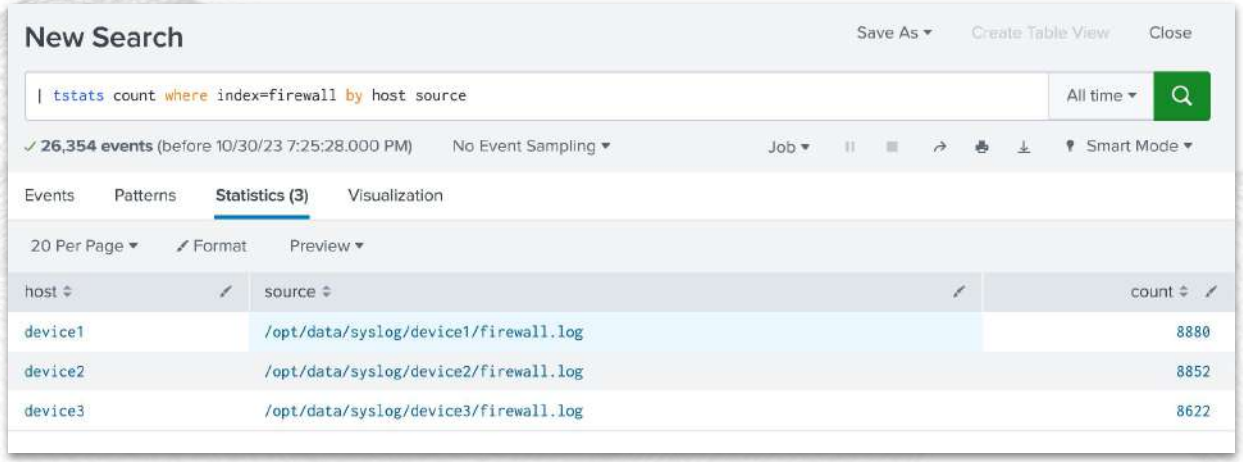
```
| tstats count where index=firewall by host source
```

KEY: Host value now has the device name

Goal:

```
splunk@show-demo-i-05d1683d07b1bccca:~$ sudo su
root@show-demo-i-05d1683d07b1bccca:/home/splunk# cd /opt/splunk/etc/apps/DataOnboarding4Ninjas/
root@show-demo-i-05d1683d07b1bccca:/opt/splunk/etc/apps/DataOnboarding4Ninjas# mkdir local
root@show-demo-i-05d1683d07b1bccca:/opt/splunk/etc/apps/DataOnboarding4Ninjas# cd local
root@show-demo-i-05d1683d07b1bccca:/opt/splunk/etc/apps/DataOnboarding4Ninjas/local# vi inputs.conf
```

```
[monitor:///opt/data/syslog/*/firewall.log]
sourcetype = fgt_traffic
index = firewall
host_segment = 4
~
```



splunk>

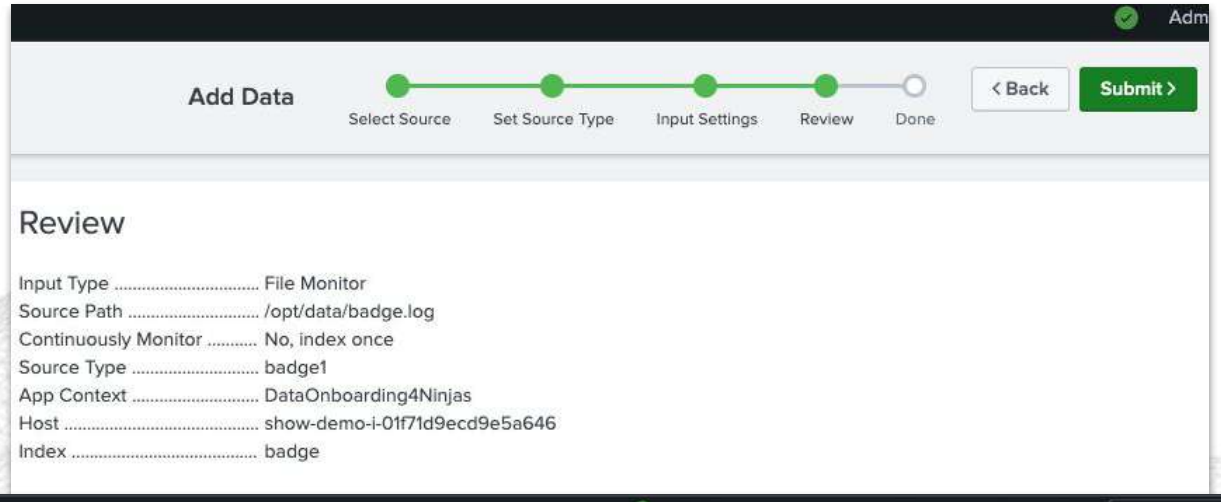
Lab 2: Data Parsing

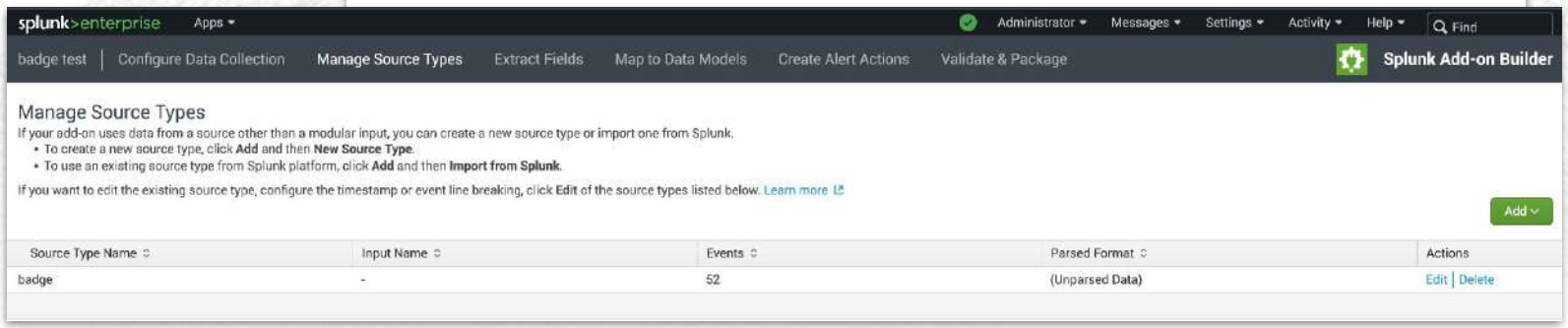
Collect and parse badge data through a sourcetype and create an add-on

Activity:

1. Collect badge data being stored on a local structure
2. Parse out the badge data via a new sourcetype
3. Select our input settings
4. Create a *badge* add-on
5. Import the new sourcetype into the add-on

Goal:





splunk>

Break for Lab 2

splunk>

Lab 2: Option 1- Data Parsing: Guided Steps

Collect and parse badge data through a new sourcetype

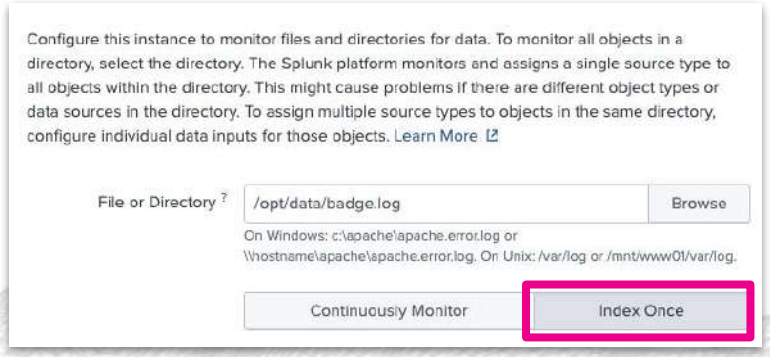
GUI:

Navigate to Settings > Add Data > Files and Directories

Select your badge logs under /opt/data/

***KEY:* Choose index once for this lab!**

Goal:



splunk>

Lab 2: Data Parsing: Guided Steps

Collect and parse badge data through a new sourcetype

GUI:

 # Save Source Type as badge

SHOULD_LINEMERGE = false

LINE_BREAKER = (##)

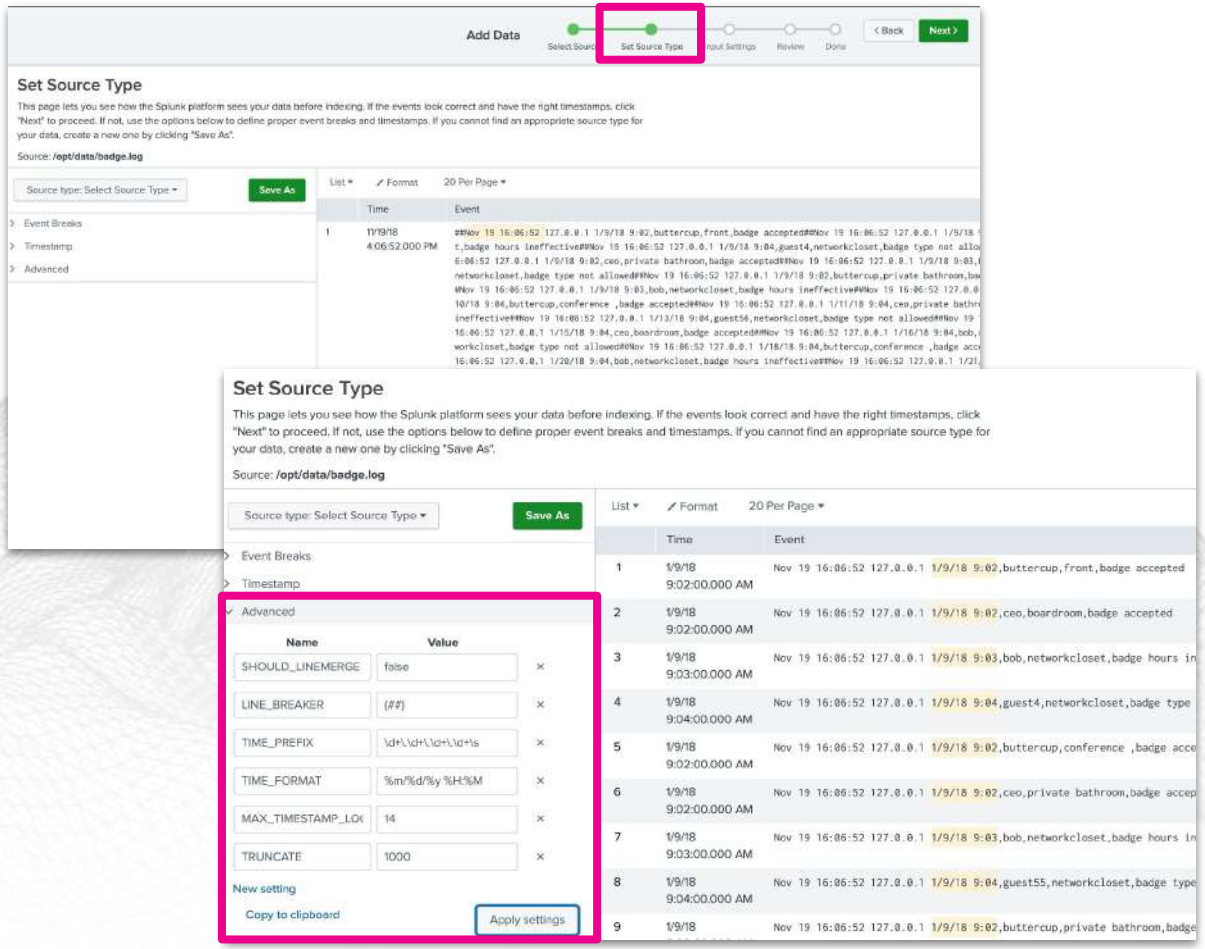
TIME_PREFIX = \d+\.\d+\.\d+\.\d+\s

TIME_FORMAT = %m/%d/%y %H:%M

MAX_TIMESTAMP_LOOKAHEAD = 14

TRUNCATE = 1000

Goal:



splunk>

Lab 2: Data Parsing: Guided Steps

Adjust input settings and check results

GUI:

 # Input Settings

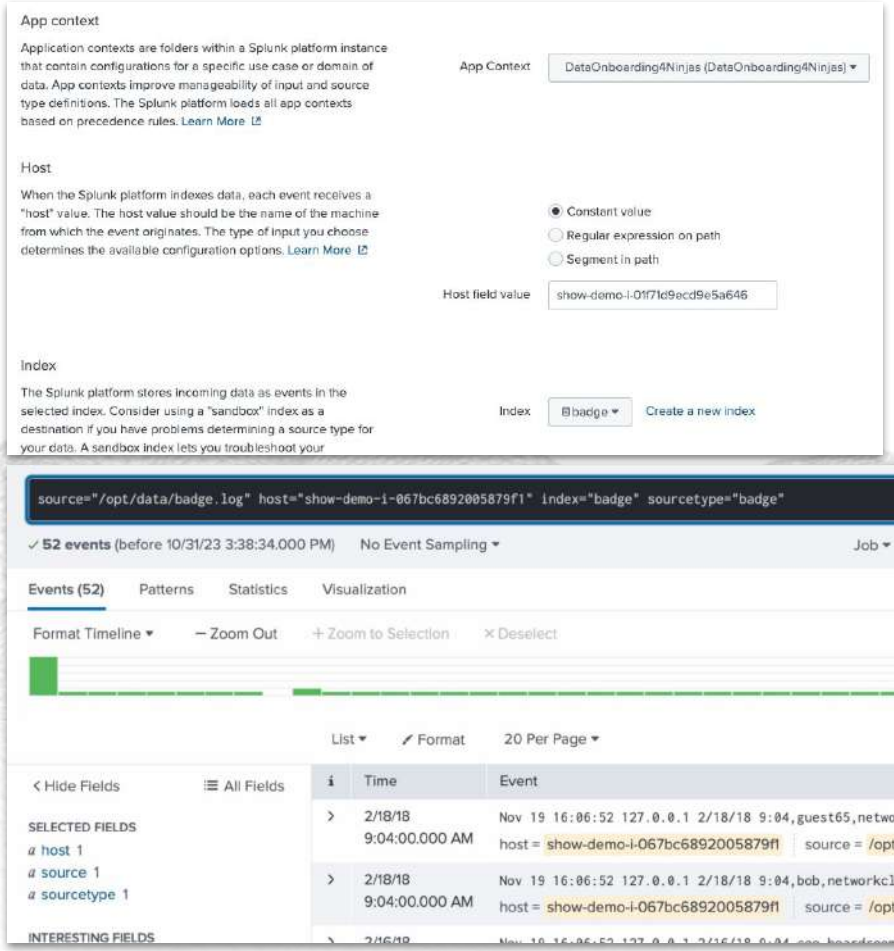
App Context: DataOnboarding4Ninjas

Index: badge

 # Check Results

Start searching after submitting input

Goal:



Lab 2: Data Parsing: Guided Steps

Create an add-on and import the badge source type

GUI:

 # Add-on Builder

Navigate to Apps > Splunk Add-on Builder

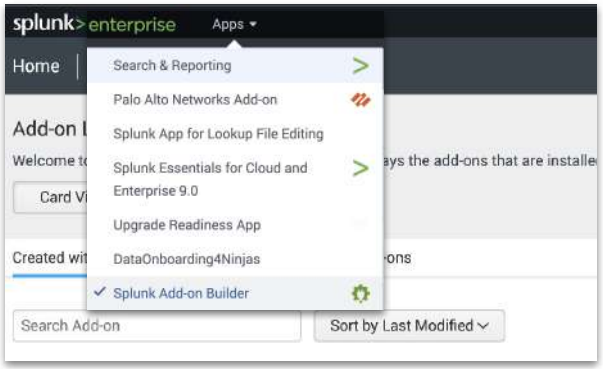
Name a new add-on *badge*

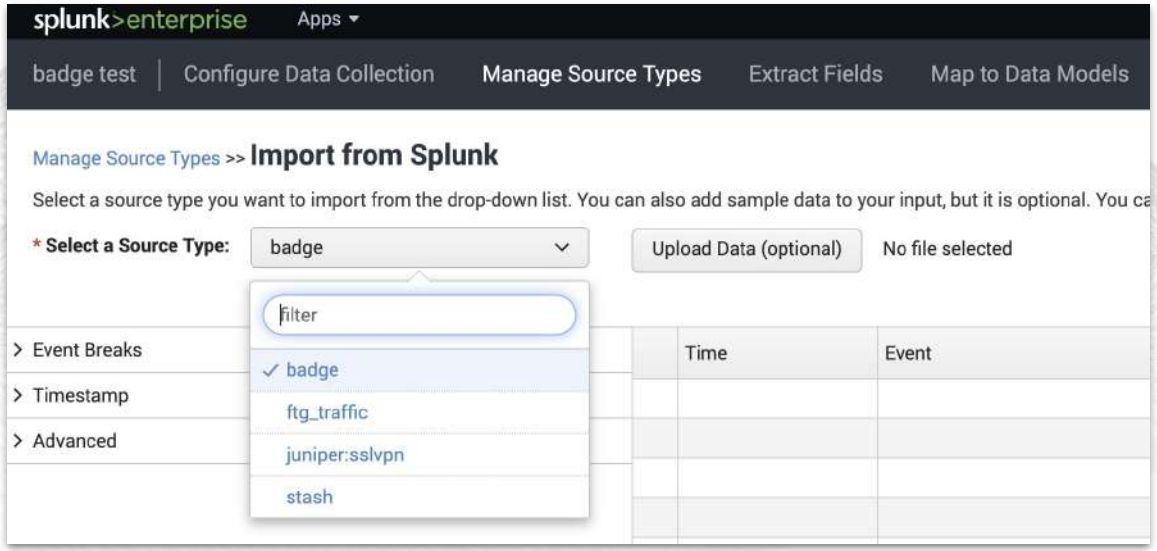
 # Import Source Source type

Navigate to *Manage Source Types*

import *badge* source type from splunk

Goal:





Lab 3: Field Extraction and CIM Compliance

Extract the new fields and make your data CIM compliant

Activities:

1. Extract new fields: employee, door, result
2. Apply badge data to the CIM “Authentication” data model
3. Display the badge data on the “Authentication Dashboard” in the “DataOnboarding4Ninjas” app

Goal:

splunk>

Break for Lab 3

splunk>

Lab 3: Field Extraction and CIM: Guided Steps

Verify the Authentication Dashboard does not populate

Authentication Dashboard:

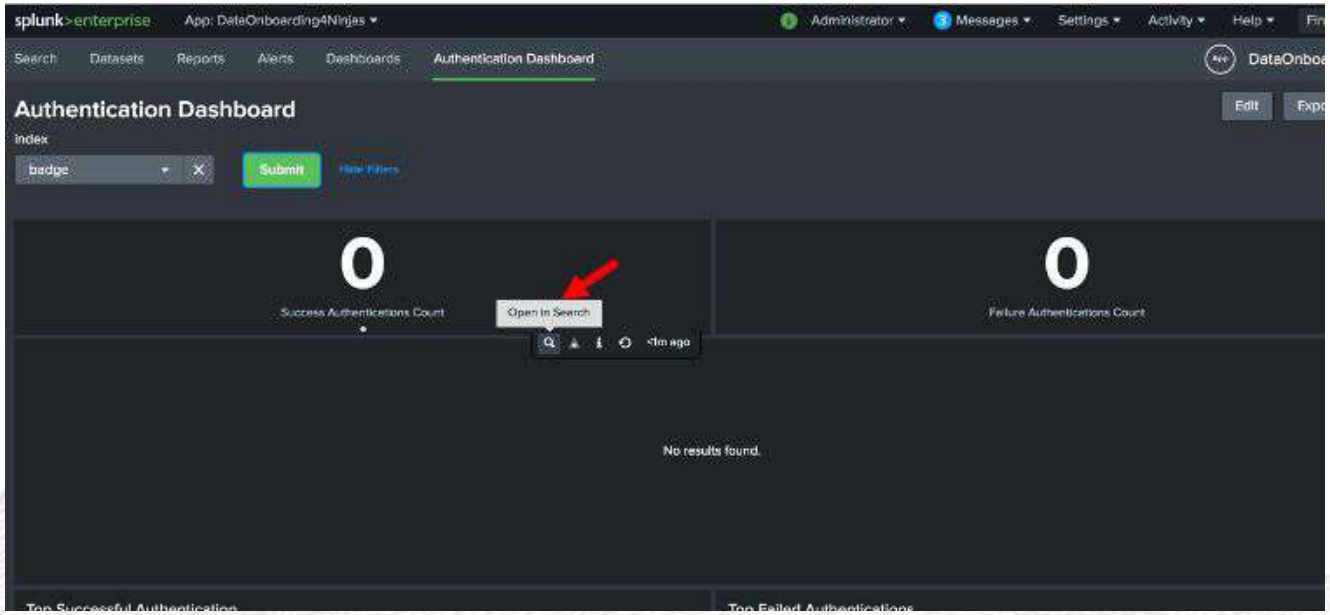
 # Check the Authentication Dashboard

It will not populate

 # WHY?

We need CIM compliant data!

Goal:



splunk>

Lab 3: Field Extraction and CIM: Guided Steps

Extract new fields

Add-on Builder GUI:

 # Extract Fields

Go to Extract Fields in the badge add-on

 # Structure your data

Choose Assisted Extraction

Table Format

Comma Separation

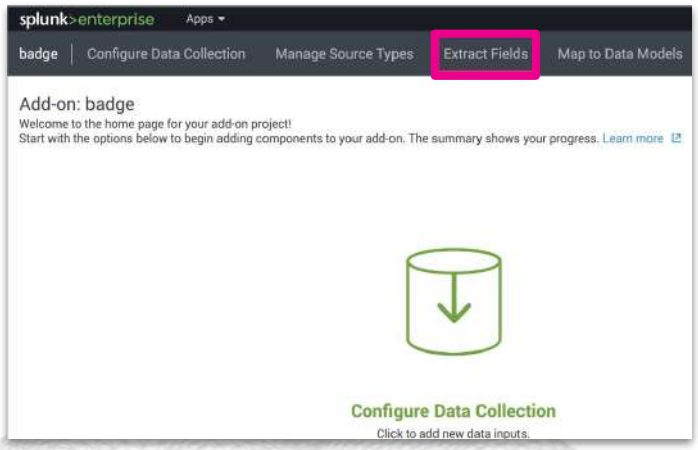
 # Name your fields

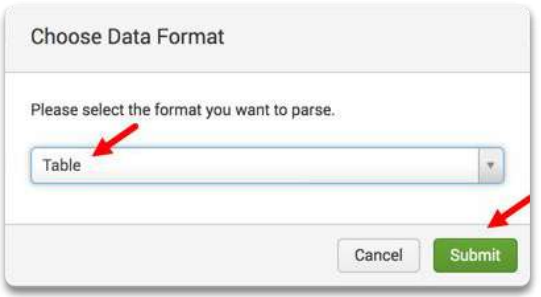
field_1 = employee

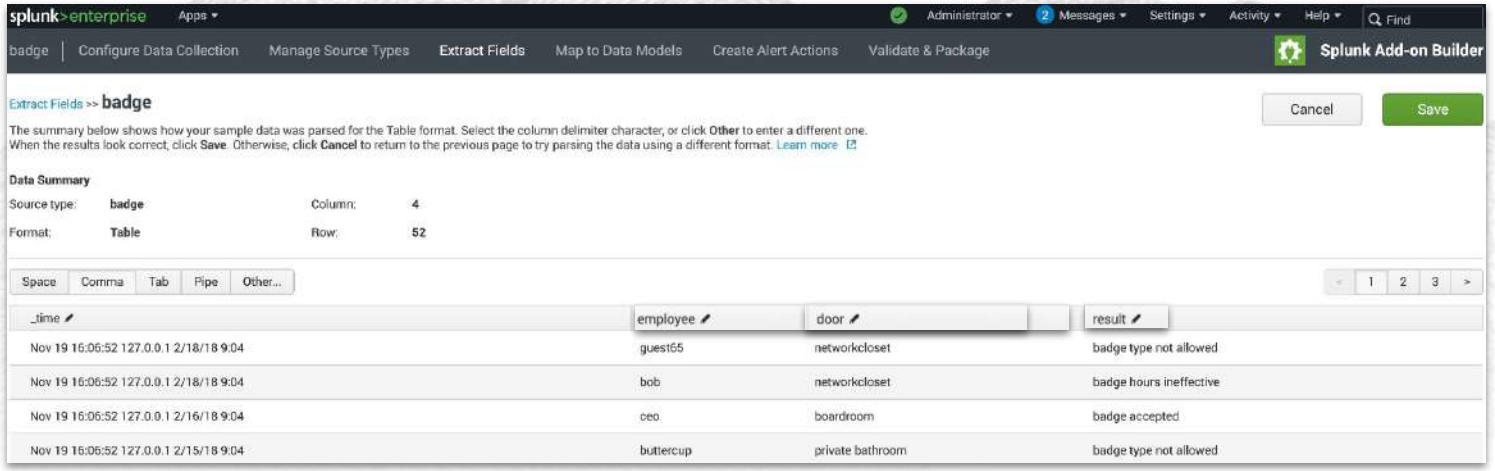
field_2 = door

field_3 = result

Goal:







Lab 3: Field Extraction and CIM: Guided Steps

Map to the Authentication Data Model

Add-on Builder GUI:

 # Map to Data Models

Go to Map to Data Models in the badge add-on

 # Define the Event Type

Event type name: badge_data

Source type: badge

Search: (sourcetype=badge)

 # Select your Data Model

Splunk_SA_CIM

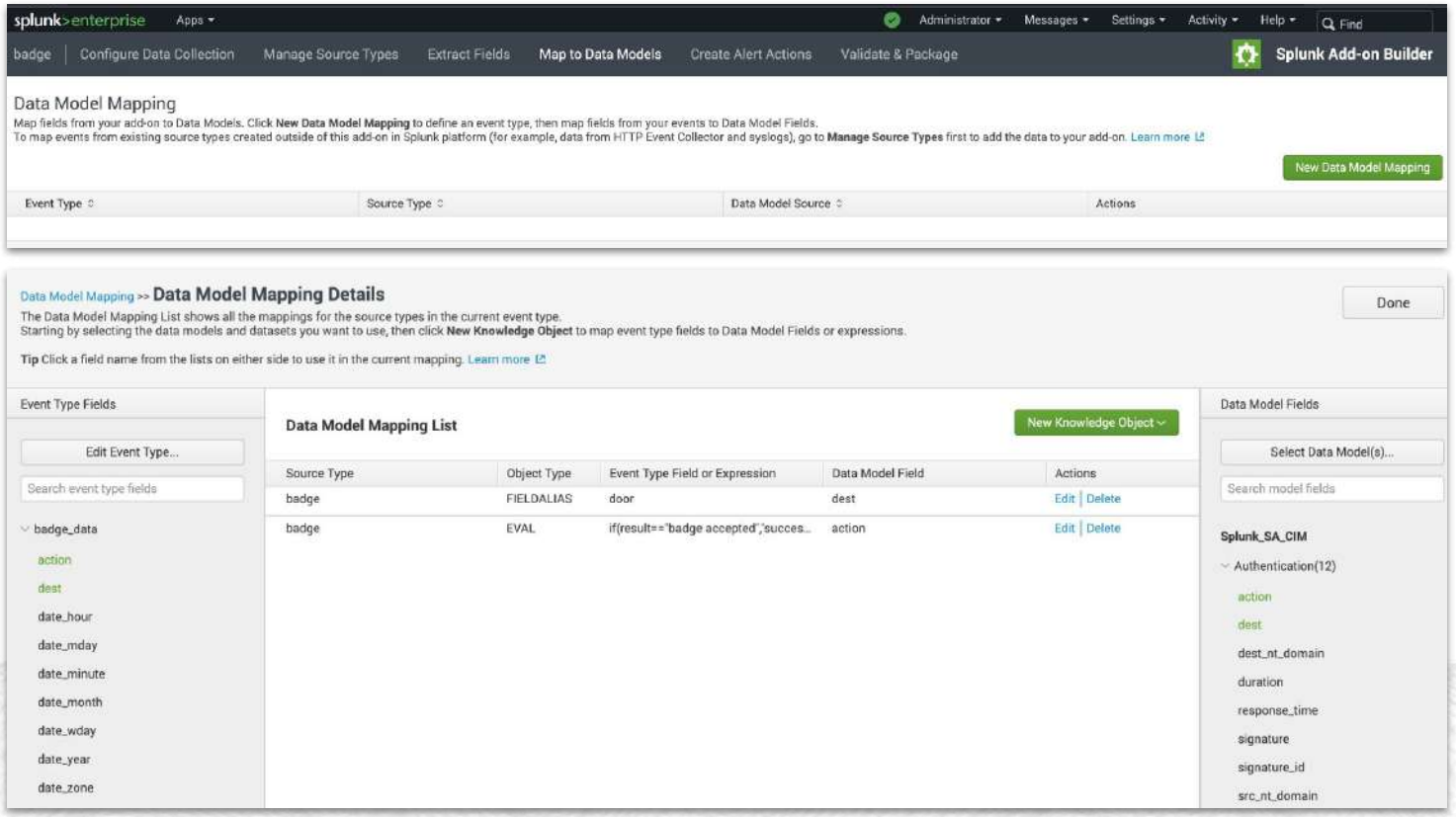
Authentication

 # Create a Field Alias and Eval

Field Alias mapping door to dest

Eval mapping result to success/failure

Goal:



Lab 3: Field Extraction and CIM: Guided Steps

Accelerate and view your dashboard

GUI:

 # Accelerate your Data Models

Navigate to Settings > Data Models

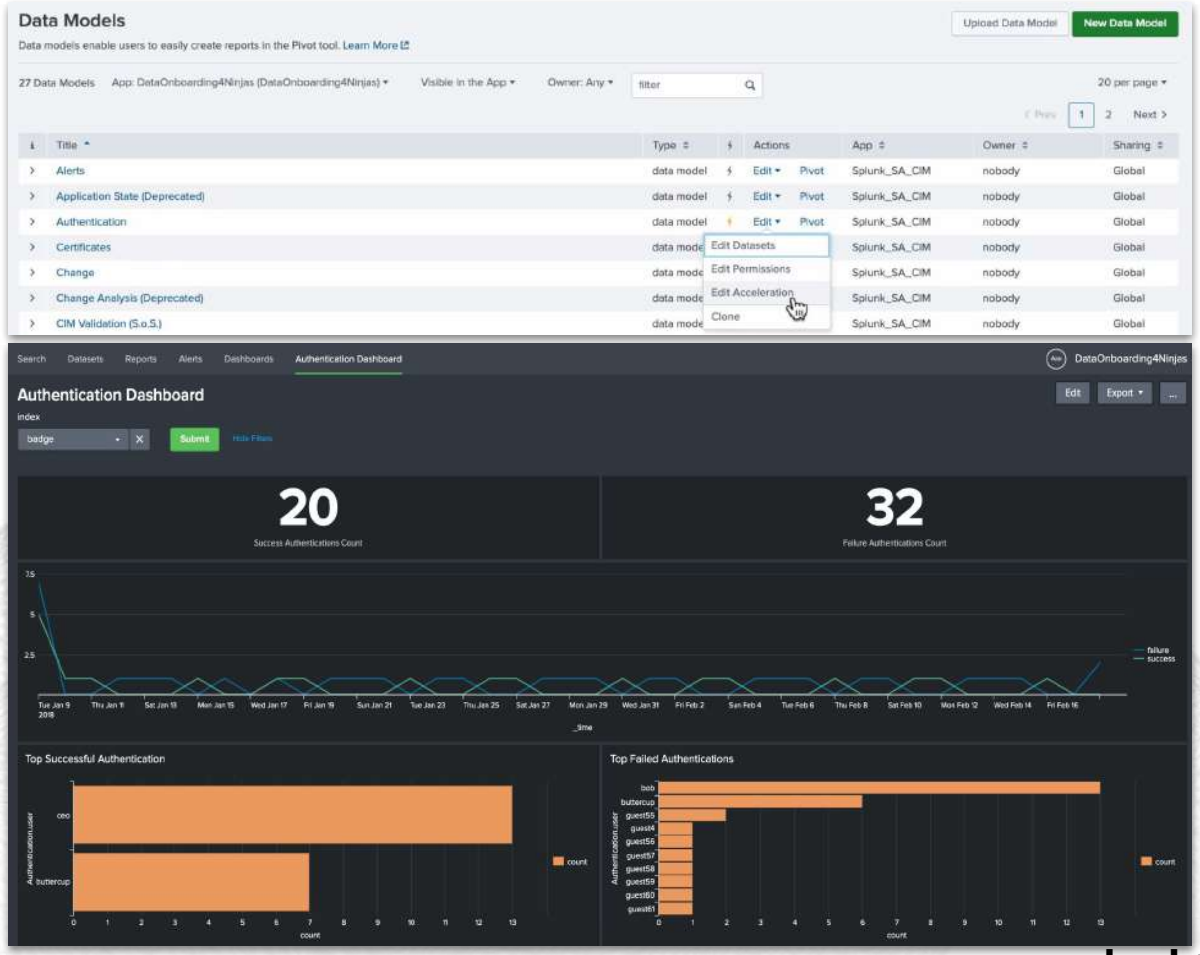
Edit the Authentication Data Model

Restart the Acceleration

 # View your dashboard

Back in the DataOnboarding4Ninjas App

Goal:



splunk>

What Happens Next?

splunk>

Additional Resources

Want more details?

Content

Splunk .conf session recordings: conf.splunk.com

- SEC1423A - To Data Model or Not to Data Model...
- FN1402 - Best Practises for Forwarder Hierarchies

Splunk Education and Training: education.splunk.com

- Free online courses
- Virtual instructor-led classes

This **slide deck**: splk.it/S4N-DataOnboarding

Help

Splunk Documentation: docs.splunk.com

- Getting Data In, Forwarding Data

Data Onboarding Cheat Sheet

- aplura.com/cheatsheets

Splunk Lantern: lantern.splunk.com

Community: community.splunk.com

Developer: dev.splunk.com

Splunk Events

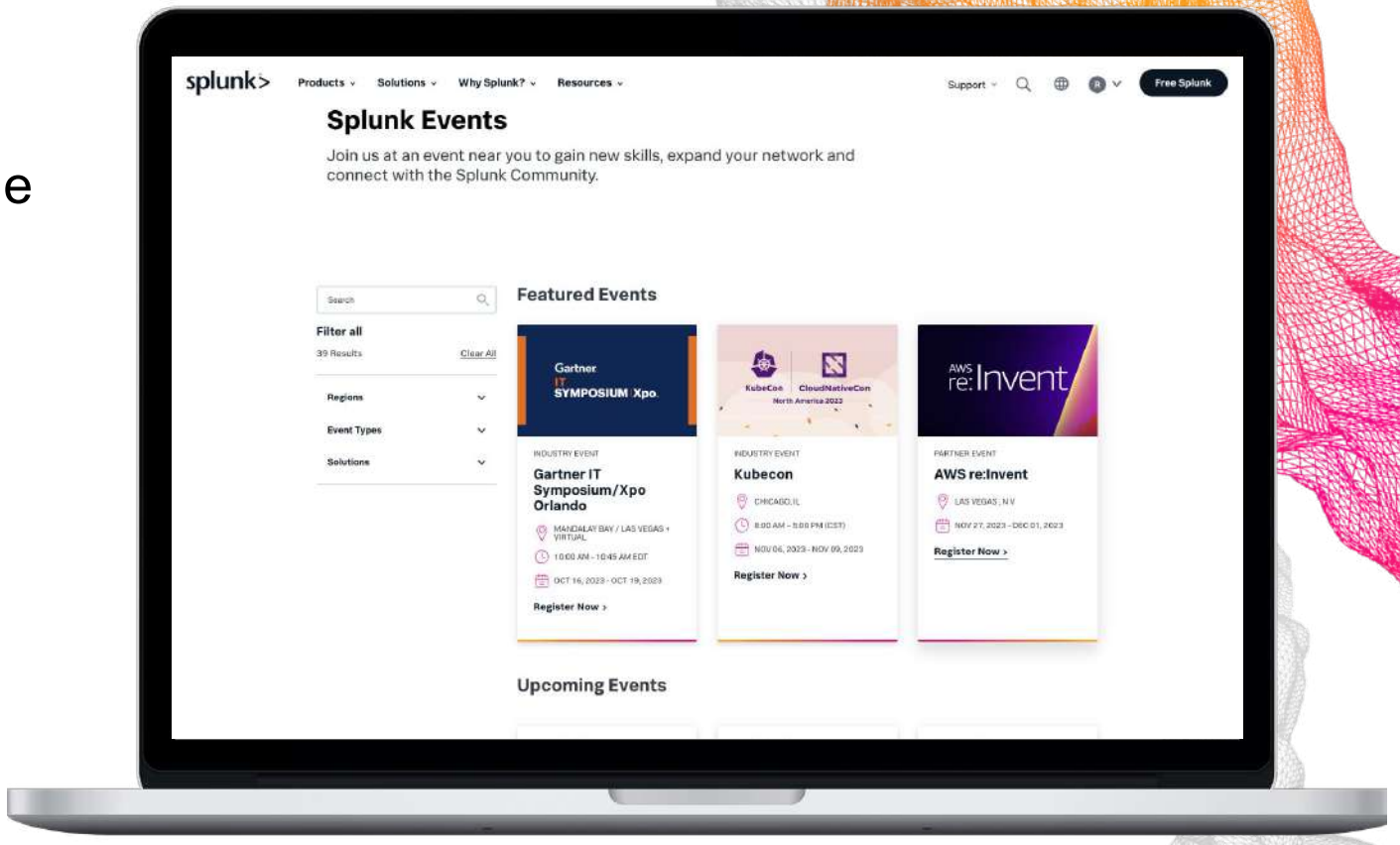
https://events.splunk.com

- Expand your network and connect with the global and local Splunk community

.conf

https://conf.splunk.com

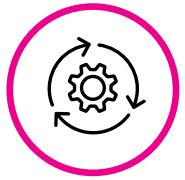
- Join us at .conf24!
- Hundreds of on-demand sessions from product updates to learning new Splunk skills!



splunk>

Thank You

Appendix



How to reset the input - If Things Go Wrong

Reset the input - Clean the fishbucket*:

1. `/opt/splunk/bin/splunk stop`
2. `/opt/splunk/bin/splunk cmd btprobe -d /opt/splunk/var/lib/splunk/fishbucket/splunk_private_db --file /opt/data/syslog/device1/firewall.log –reset`
3. `/opt/splunk/bin/splunk start`

You could 'clean' the index:

1. Make sure you have **can_delete** permission
2. `index="firewall" | delete` (this will mark your data non-searchable from this particular index)
3. Or use the **clean** command in the CLI: `/opt/splunk/bin/splunk clean eventdata -index firewall`

* **Fishbucket** - A subdirectory where Splunk software tracks how far into a file indexing has progressed, to enable the software to detect when data has been added to the file and resume indexing. The fishbucket subdirectory contains seek pointers and CRCs for indexed files.