Splunk4Ninjas - Data Onboarding

Randy Holloway, CISSP, MS
Staff Solutions Engineer

May, 2024

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>

#whoami

Randy Holloway

rholloway@splunk.com

Based in Houston, TX

25+ Years IT and Security Experience

16+ Years SIEM Experience

Came Over from ArcSight

Enjoys Michigan Football and Baseball

splunk>

Neville S. Farooq, CISSP

Sr. Solutions Engineer, Splunk

| **Career:** Deloitte > Lockheed Martin > Splunk
| **Location:** Arlington, Virginia
| **Hobbies:** Being Active/Outdoors, Reading, Sports

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

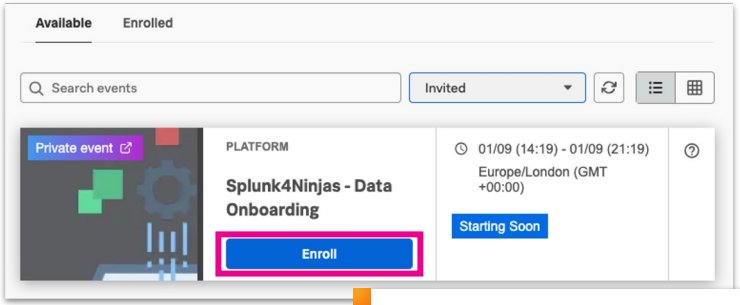
Tasks

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!

1. Download a copy of today's slide deck: https://splk.it/S4N-DataOnboarding

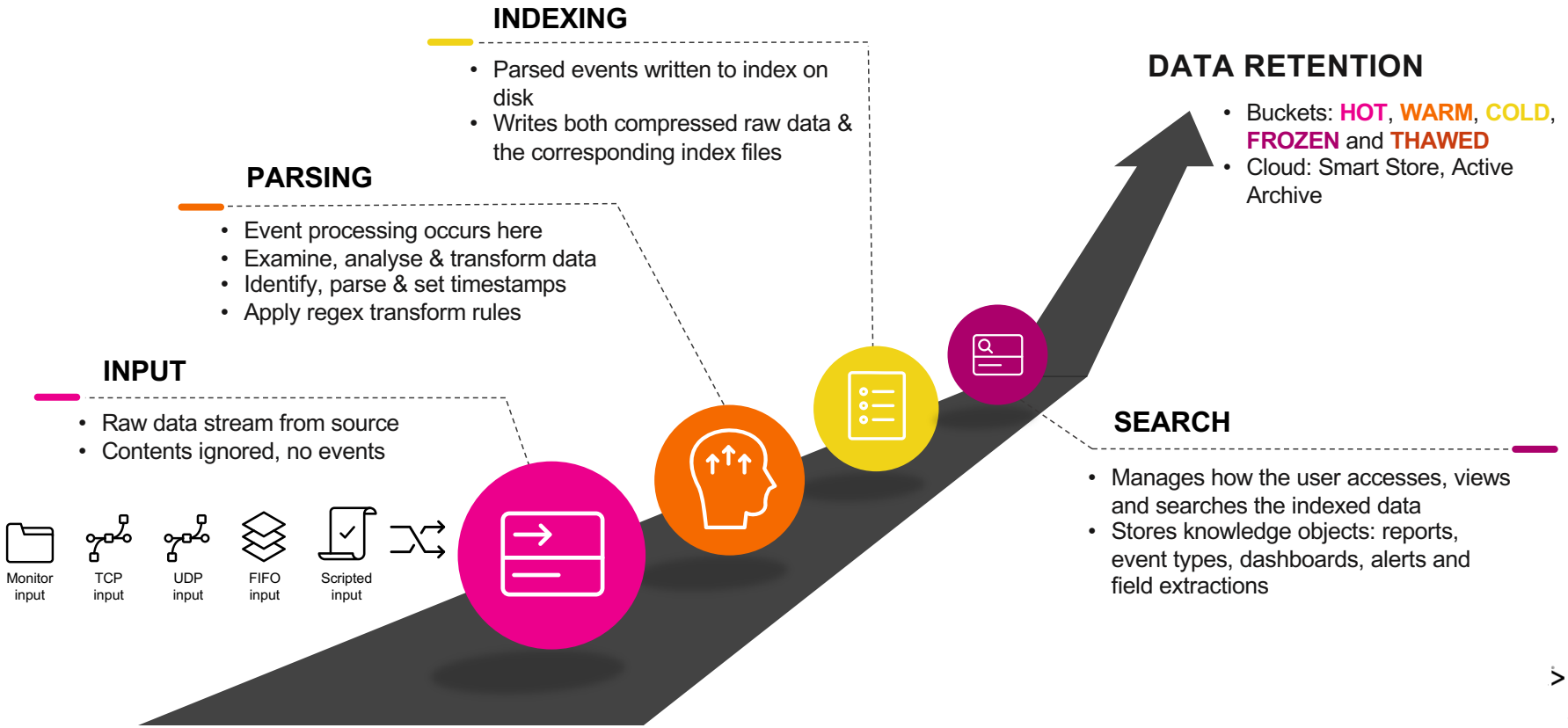
Goal



Enroll in today's event

splunk>

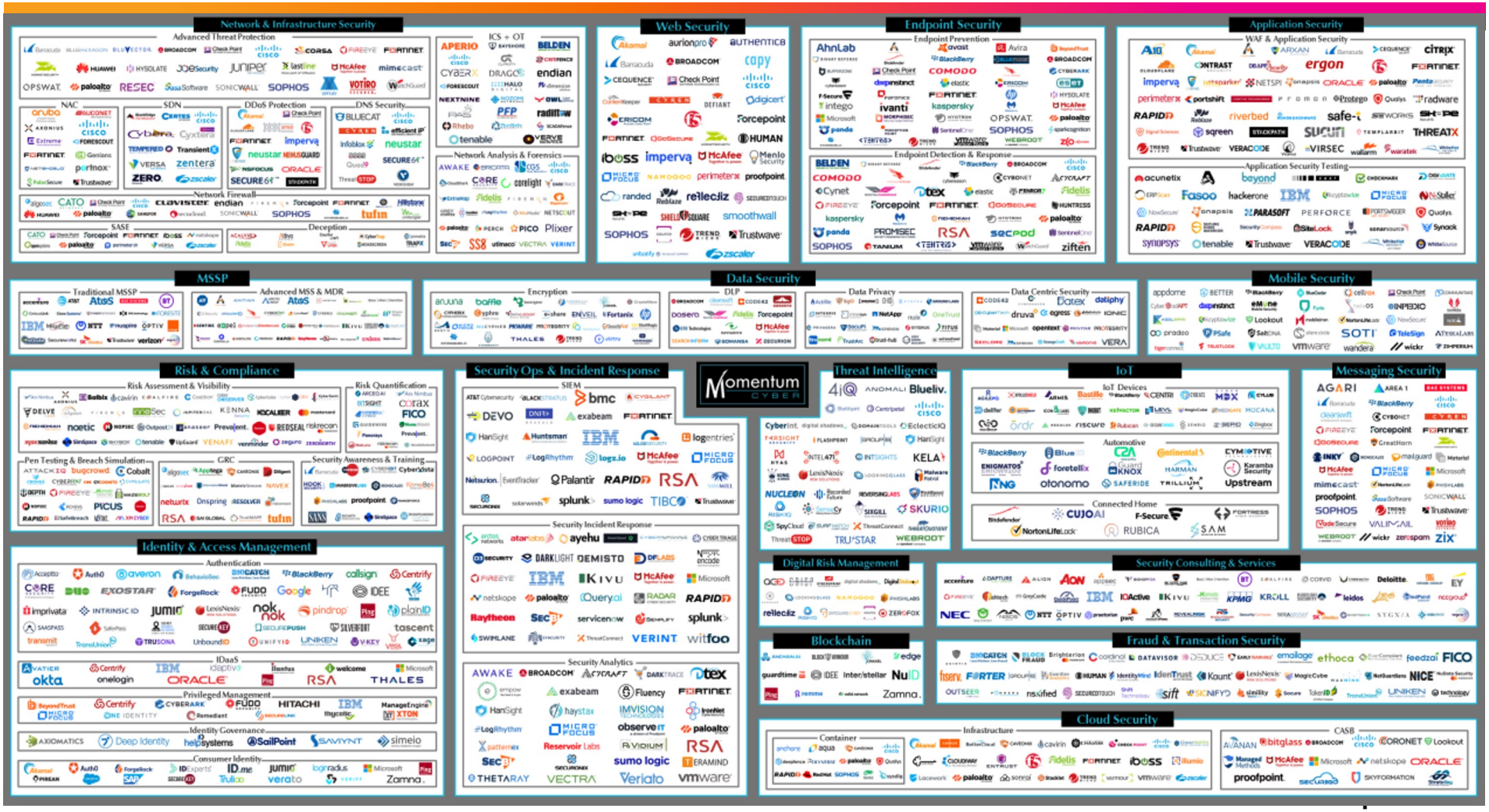
How Data Moves Through Splunk



Data Sources

and the importance of data quality

splunk>



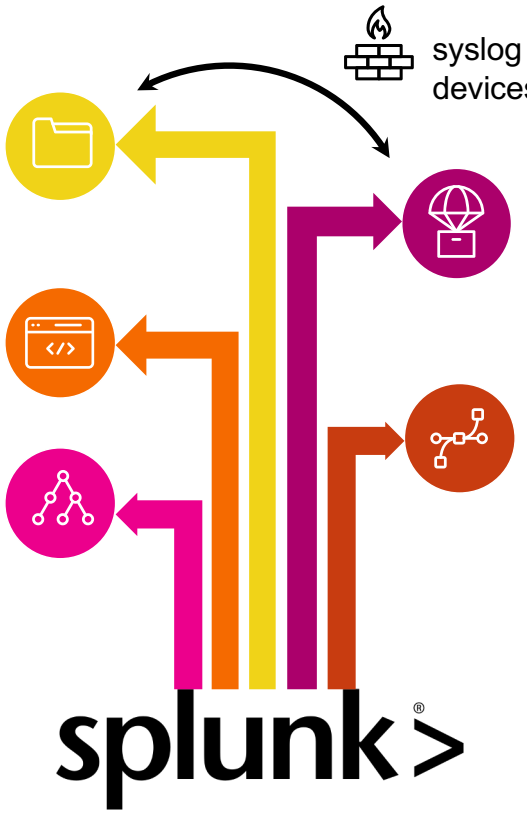
Source: Momentum Cyber

What Can Splunk Ingest?

Local File Monitoring
- Universal Forwarder

Aggregated / API Sources
- Heavy Forwarder

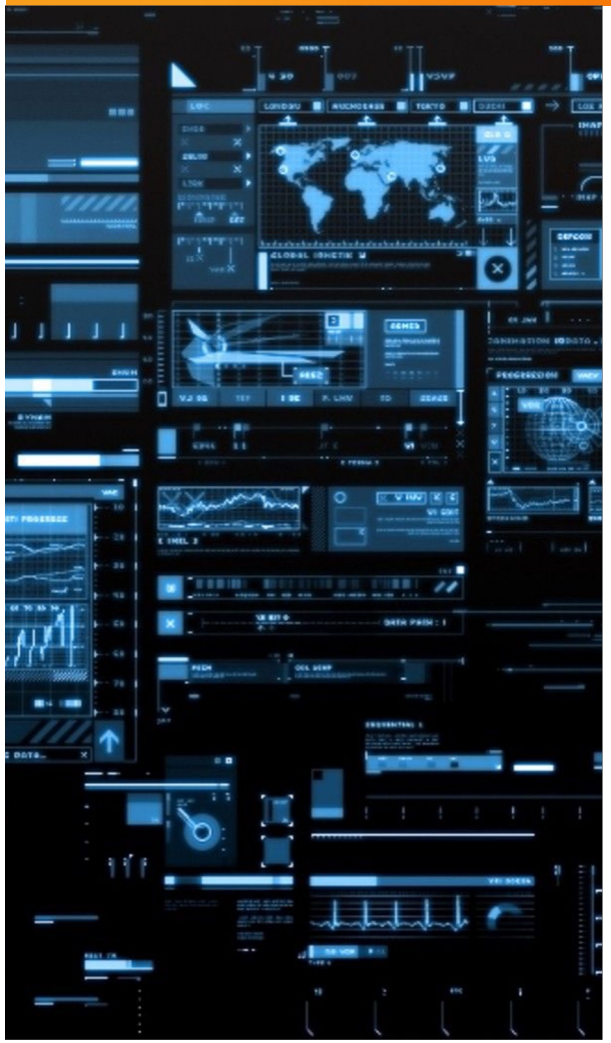
Event Logs, Active Directory, OS Stats, Unix/Linux/ Windows Hosts
- Universal Forwarder



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

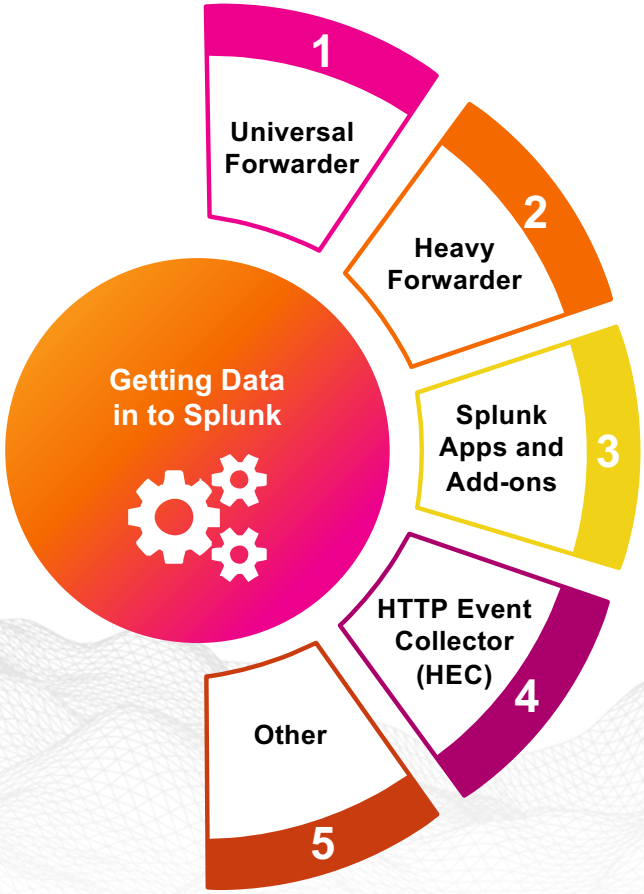
Timestamps
Improperly timestamped events can lead to early deletion

splunk>

SAGEFOX

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>

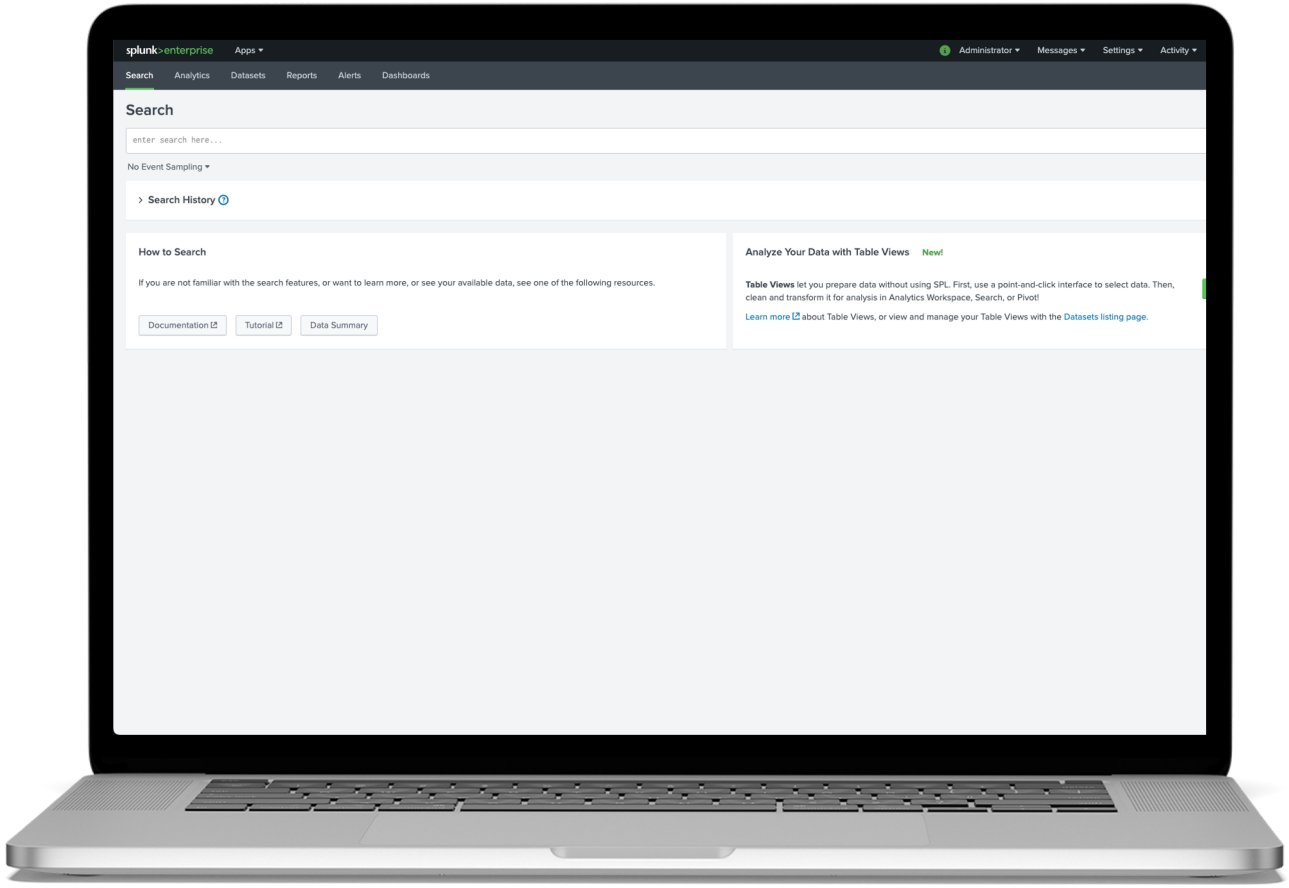
Universal vs Heavy Forwarder

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

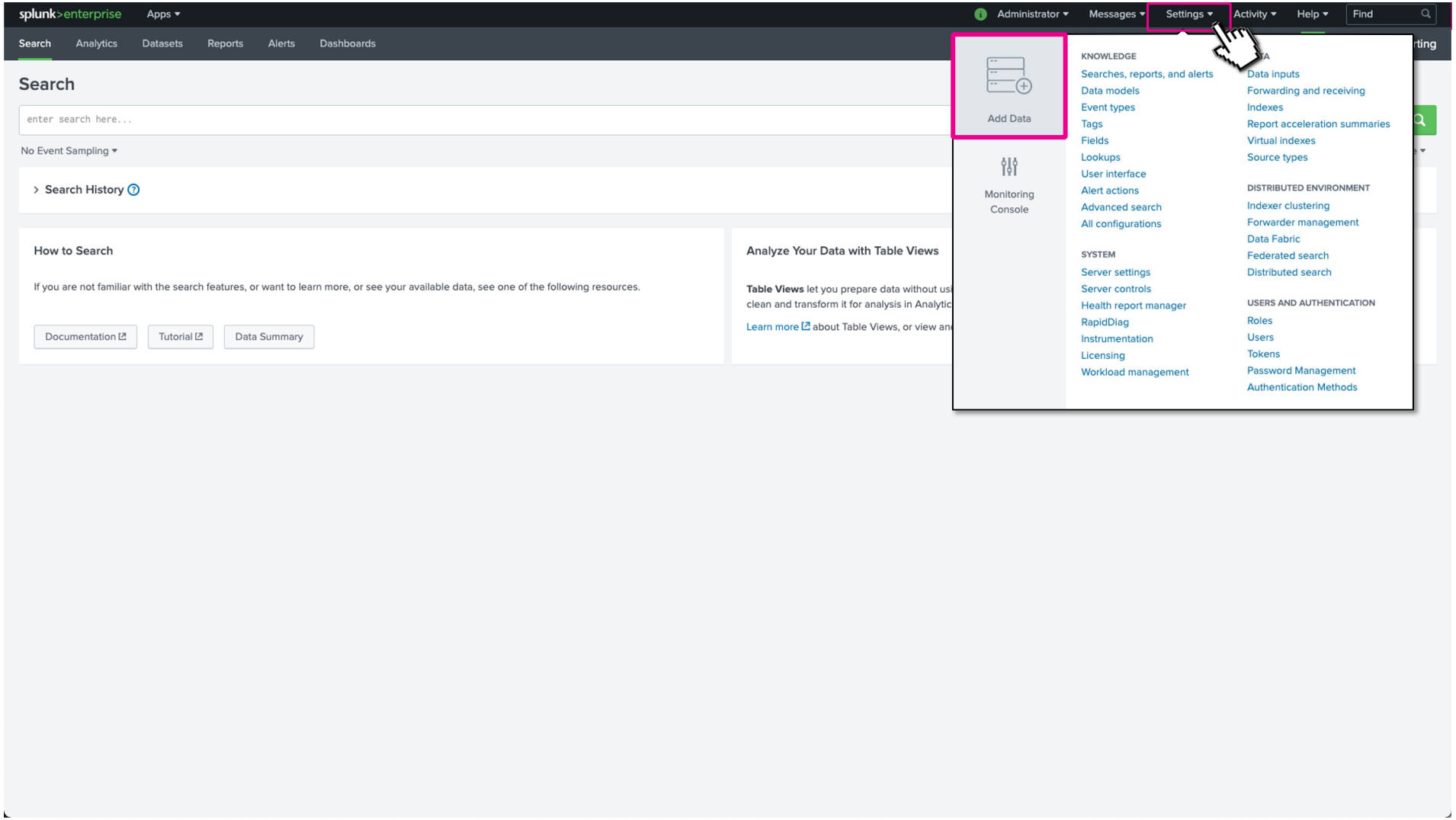
A full list of features and capabilities can be found at:
https://docs.splunk.com/Documentation/Splunk/latest/Forwarding/Typesofforwarders

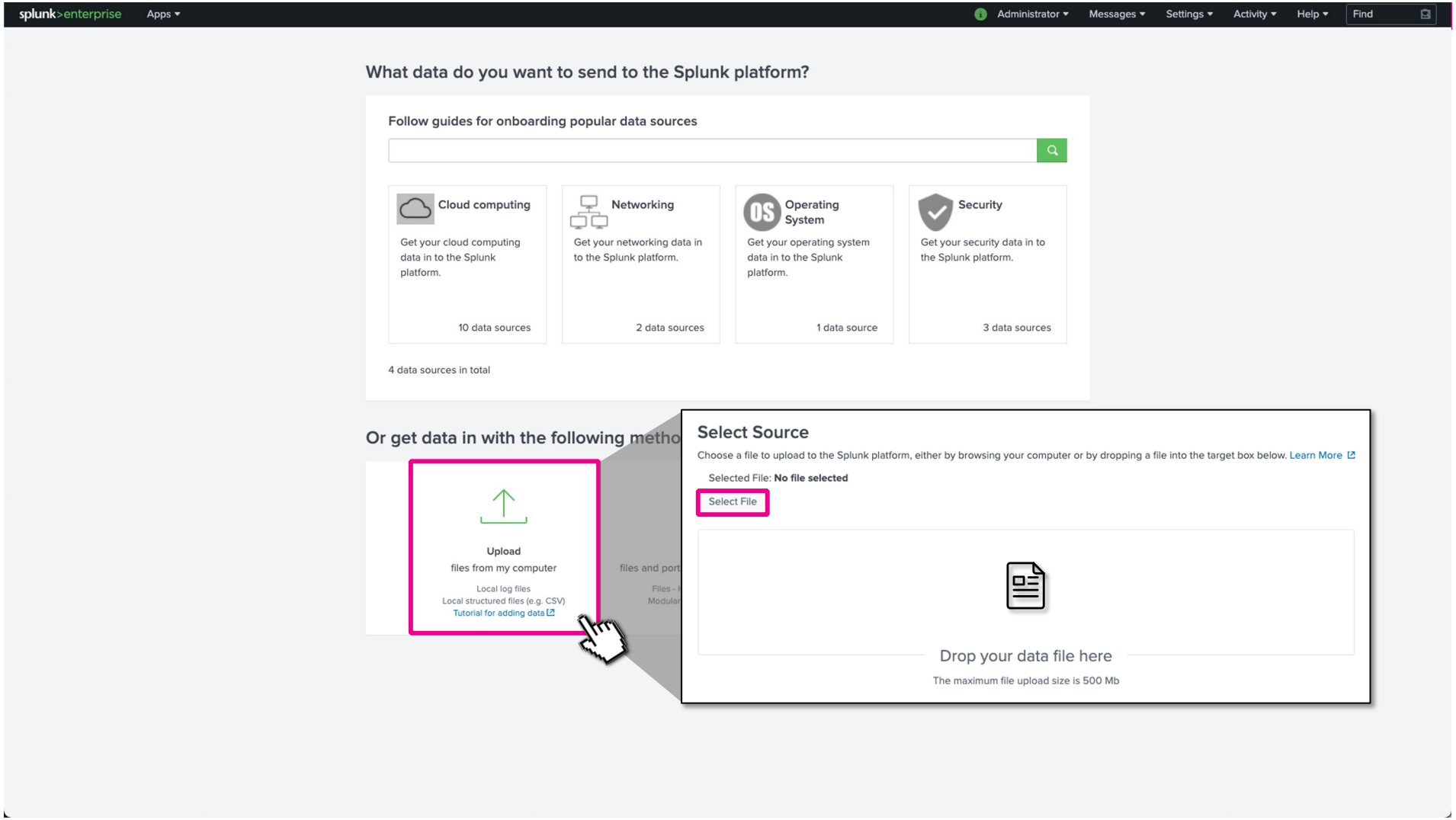
Getting Files into Splunk

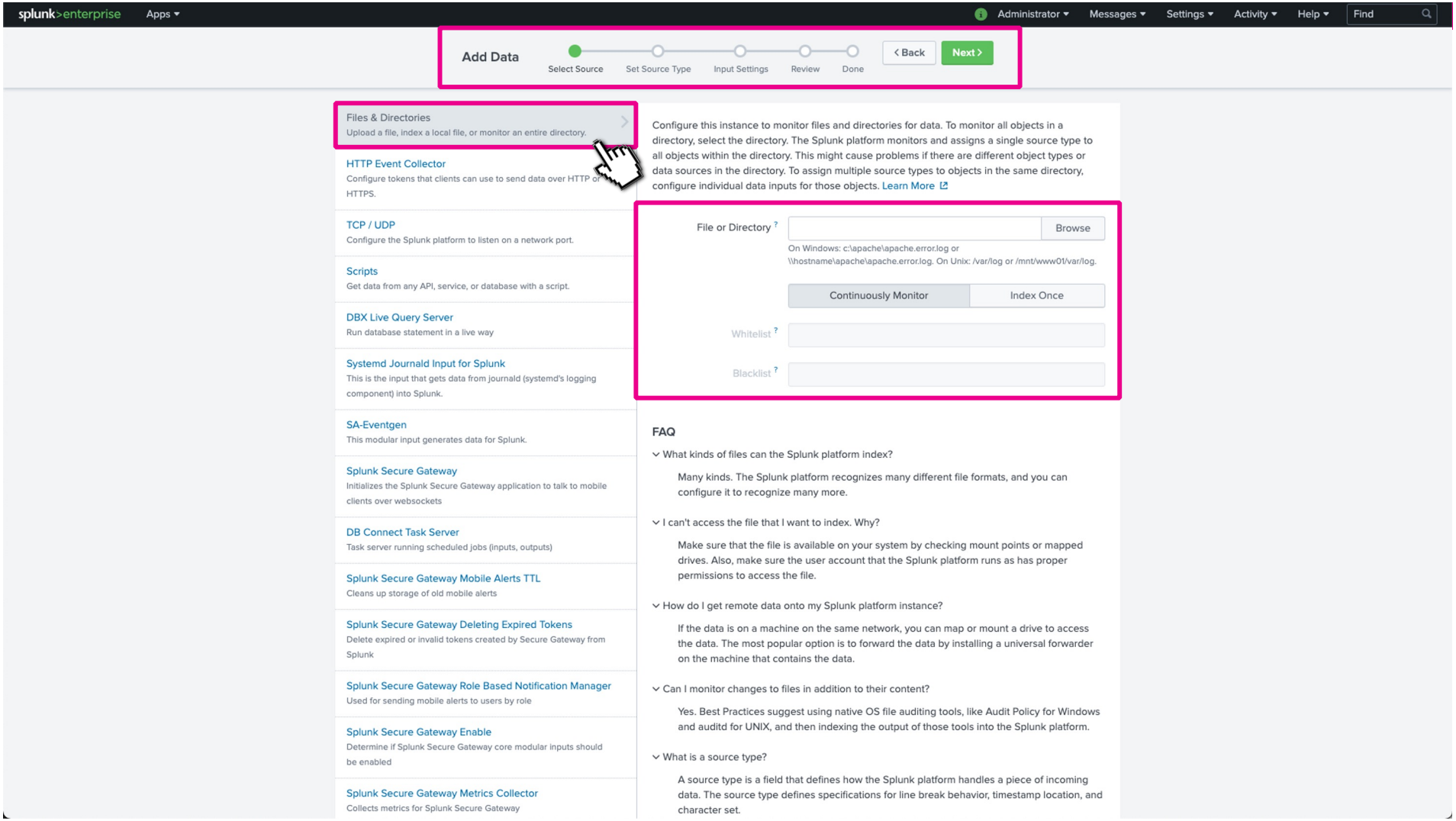
- Direct Upload
- File Monitor
- Custom scripts
- syslog and plenty more!

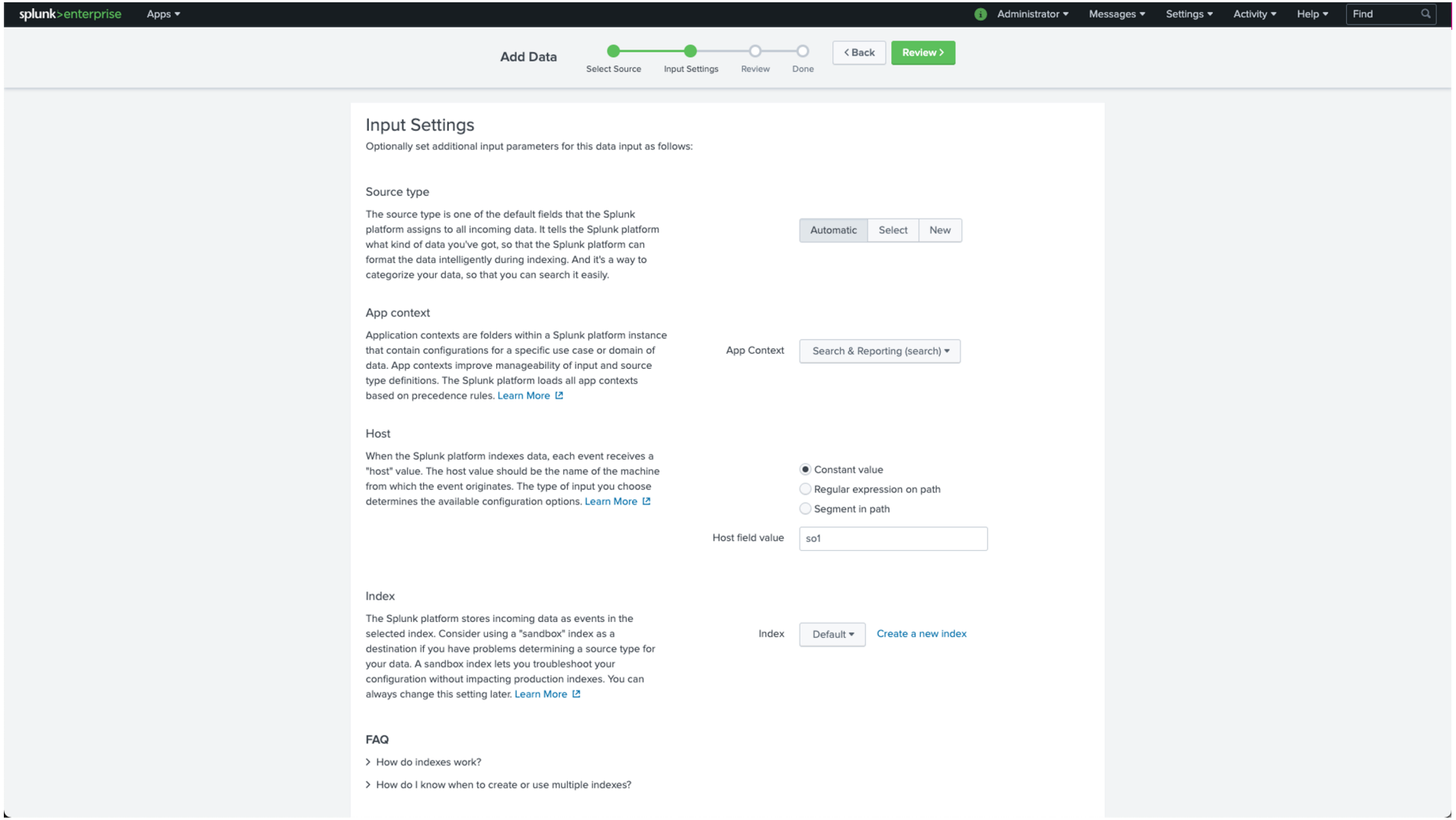


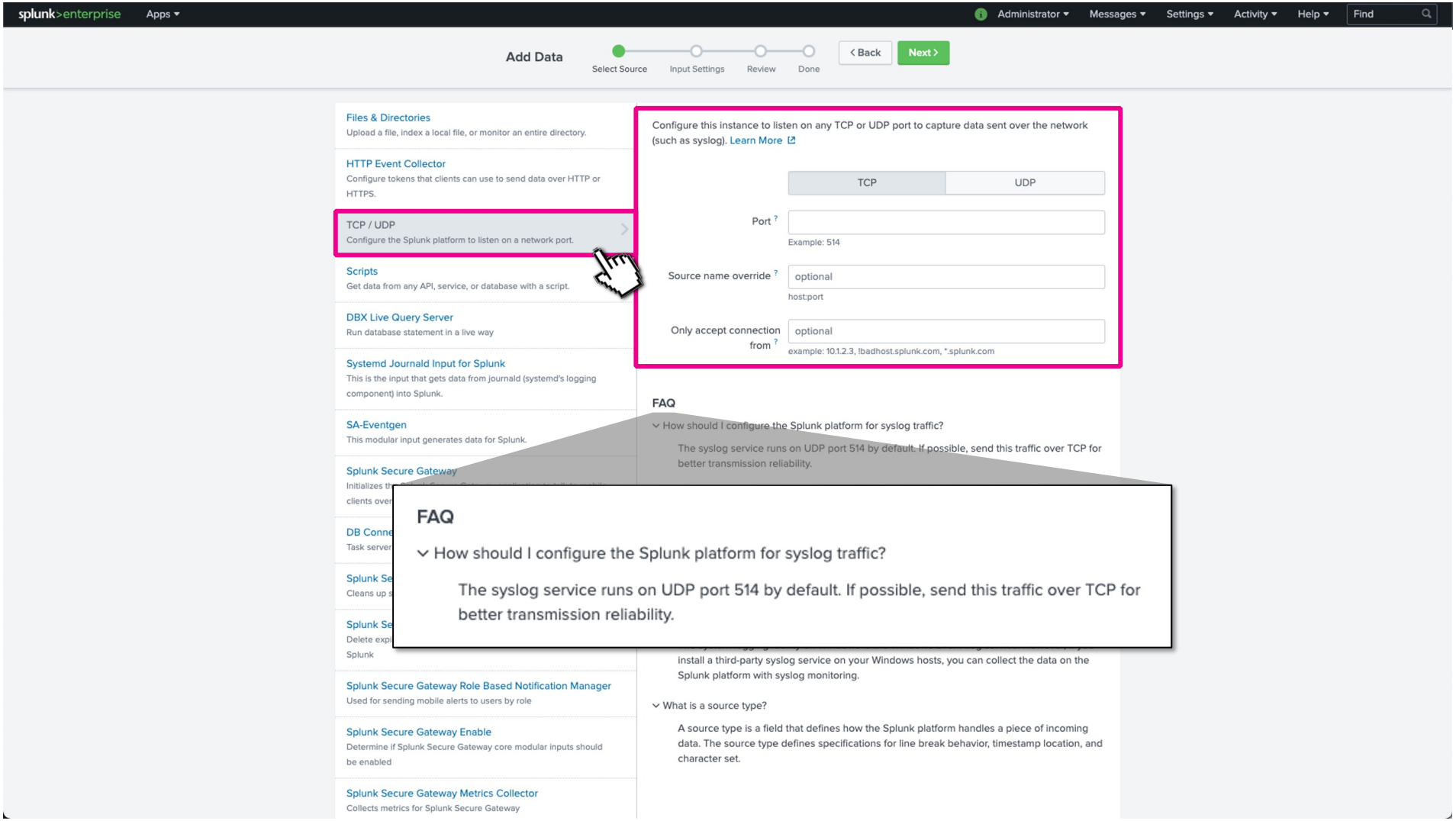
splunk>

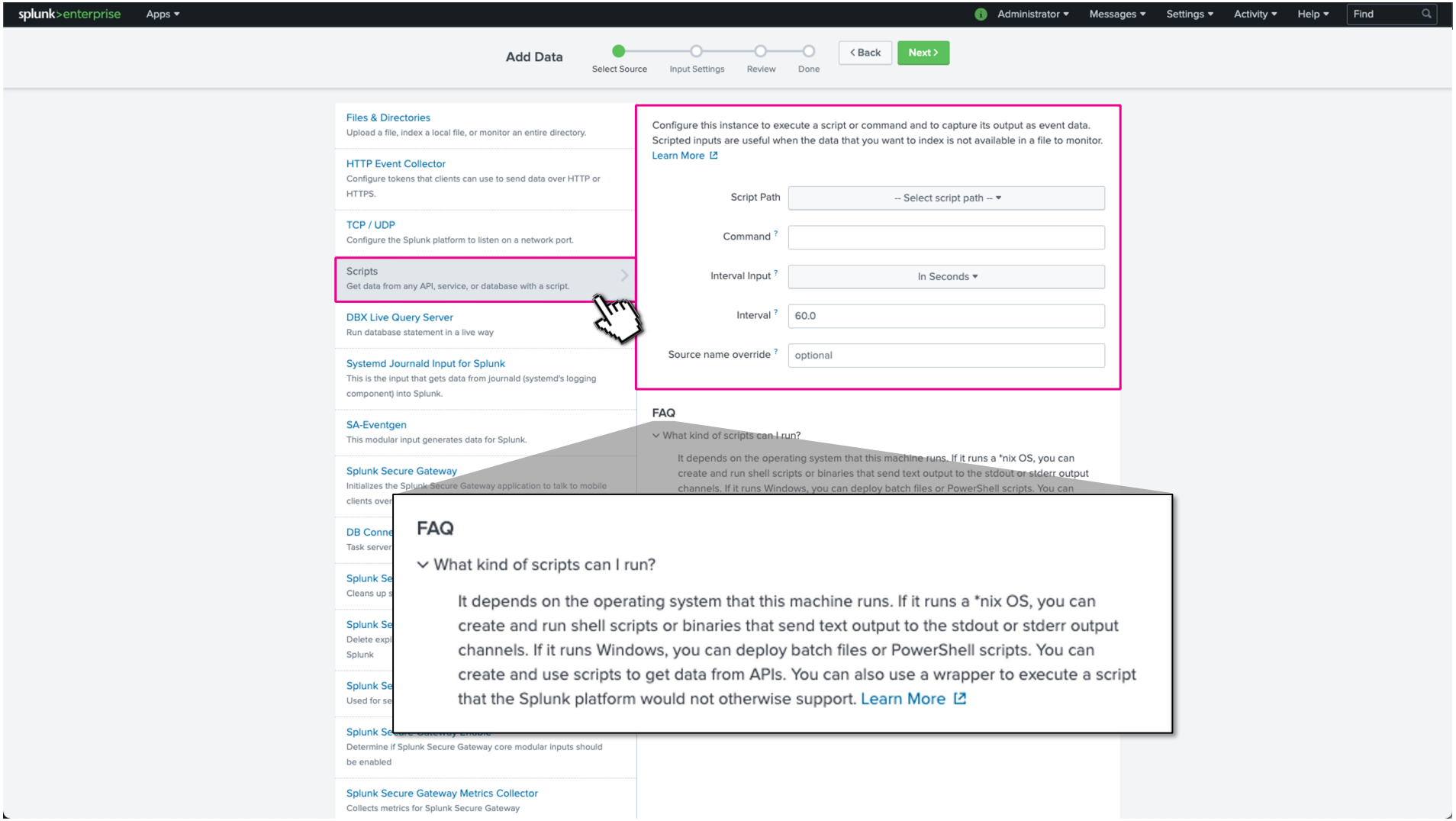

splunk>enterprise
Apps
Administrator
Messages
Settings
Activity
Help
Find
Search
Analytics
Datasets
Reports
Alerts
Dashboards
Search
enter search here...
No Event Sampling
Search History
How to Search
If you are not familiar with the search features, or want to learn more, or see your available data, 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 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
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


splunk>enterprise
Apps
Administrator
Messages
Settings
Activity
Help
Find
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.

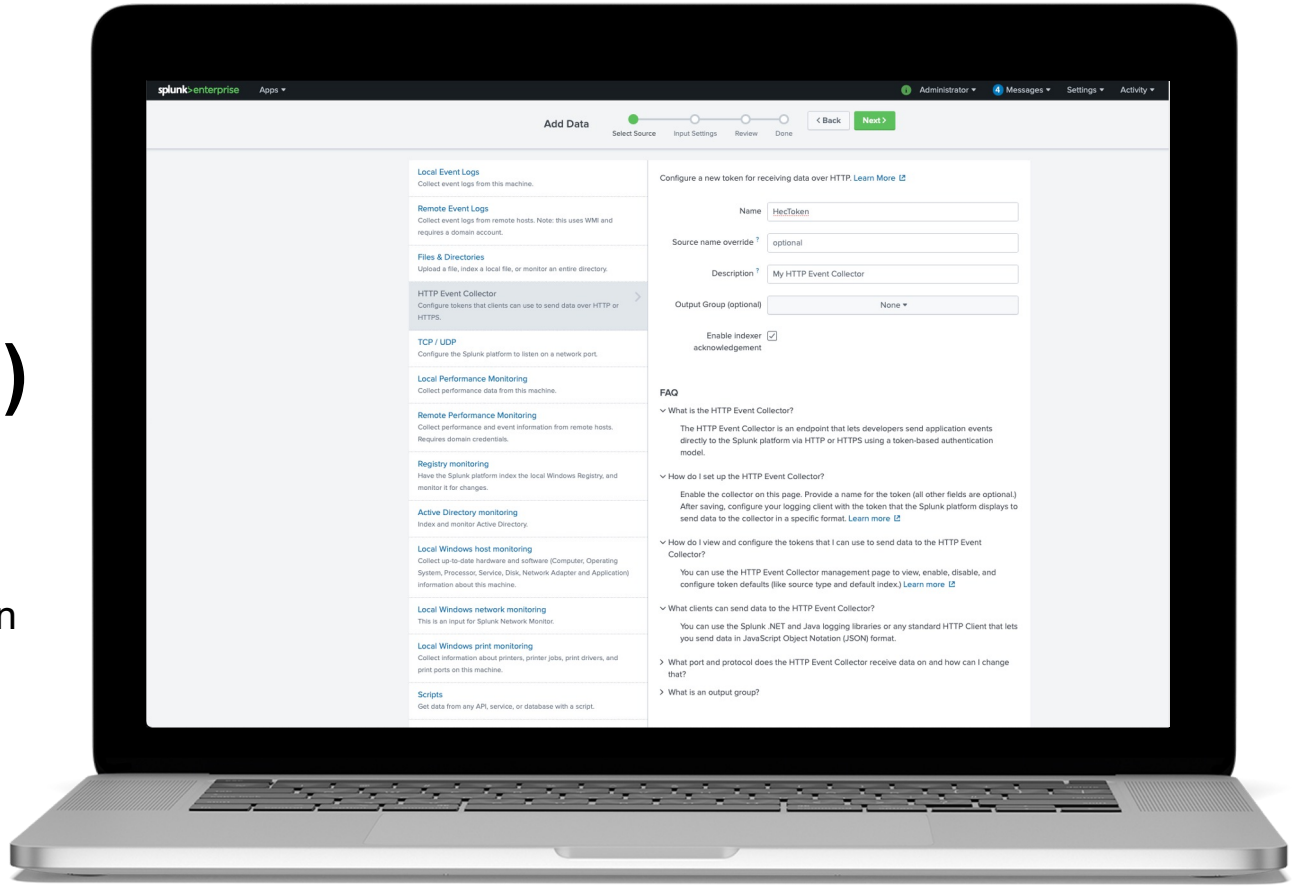

splunk>enterprise
Apps
Administrator
Messages
Settings
Activity
Help
Find
Add Data
Select Source
Input Settings
Review
Done
< Back
Review >
Input Settings
Optionally set additional input parameters for this data input as follows:
Source type
The source type is one of the default fields that the Splunk 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?


splunk>enterprise
Apps
Administrator
Messages
Settings
Activity
Help
Find
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
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.

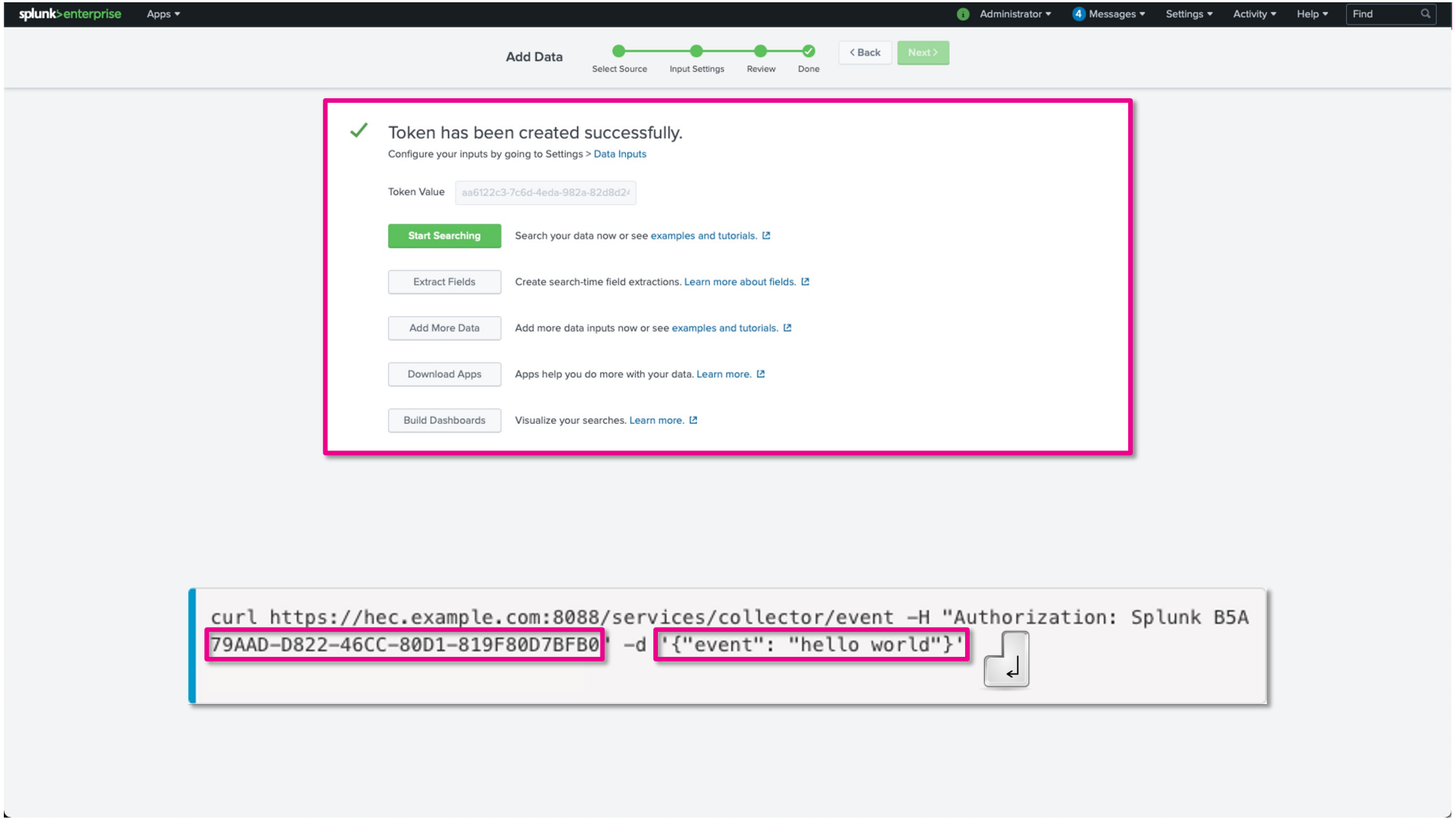

splunk>enterprise
Apps
Administrator
Messages
Settings
Activity
Help
Find
Add Data
Select Source
Input Settings
Review
Done
Back
Next
Files & Directories
Upload a file, index a local file, or monitor an entire directory.
HTTP Event Collector
Configure tokens that clients can use to send data over HTTP 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
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>


splunk>enterprise
Apps
Administrator
4 Messages
Settings
Activity
Help
Find
Add Data
Select Source
Input Settings
Review
Done
< Back
Next >
Token has been created successfully.
Configure your inputs by going to Settings > Data Inputs
Token Value
aa6122c3-7c6d-4eda-982a-82d8d2
Start Searching
Search your data now or see examples and tutorials.
Extract Fields
Create search-time field extractions. Learn more about fields.
Add More Data
Add more data inputs now or see examples and tutorials.
Download Apps
Apps help you do more with your data. Learn more.
Build Dashboards
Visualize your searches. Learn more.
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

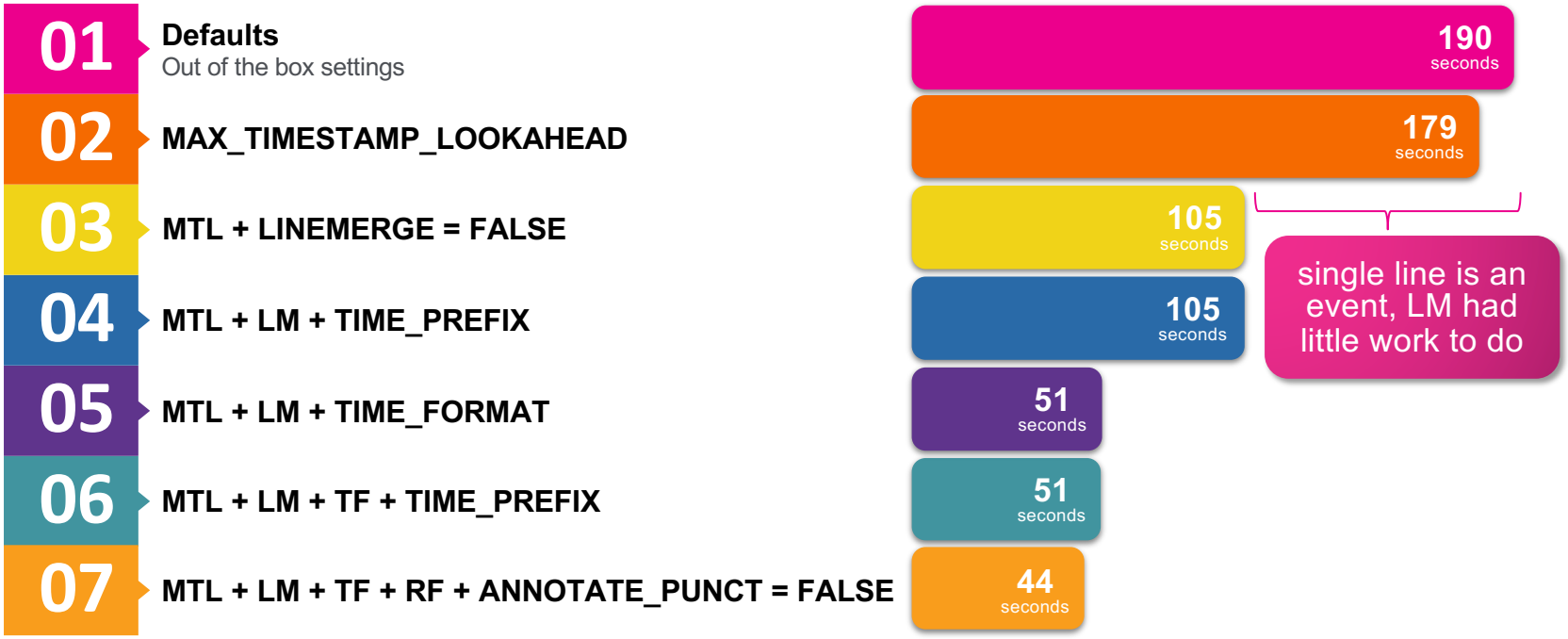
TIME_PREFIX:
regex of the text that leads up to the timestamp

MAX_TIMESTAMP_LOOKAHEAD:
how many characters for the timestamp

splunk>

Measurable Impact

Wall-Clock Seconds Ingesting 10M Events

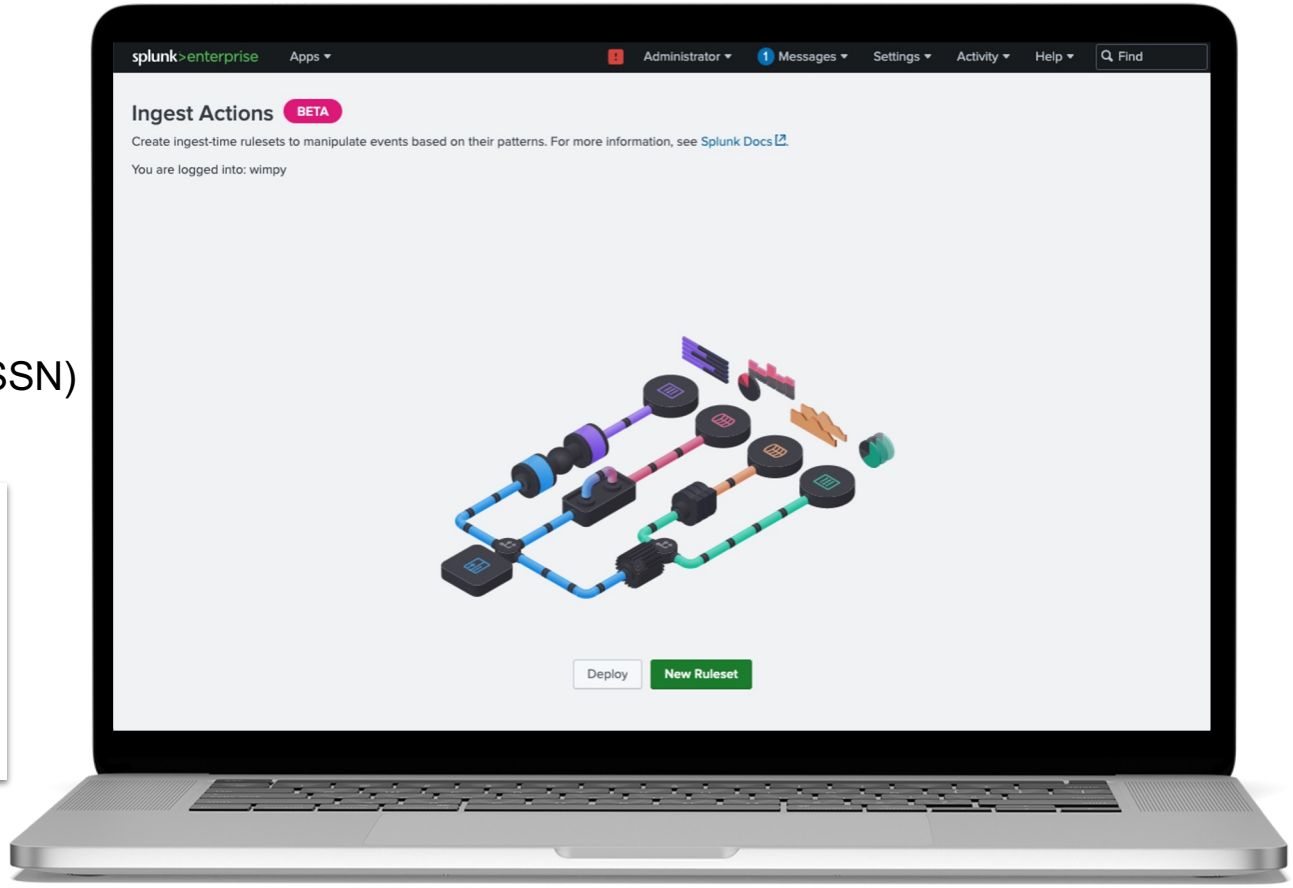


splunk>

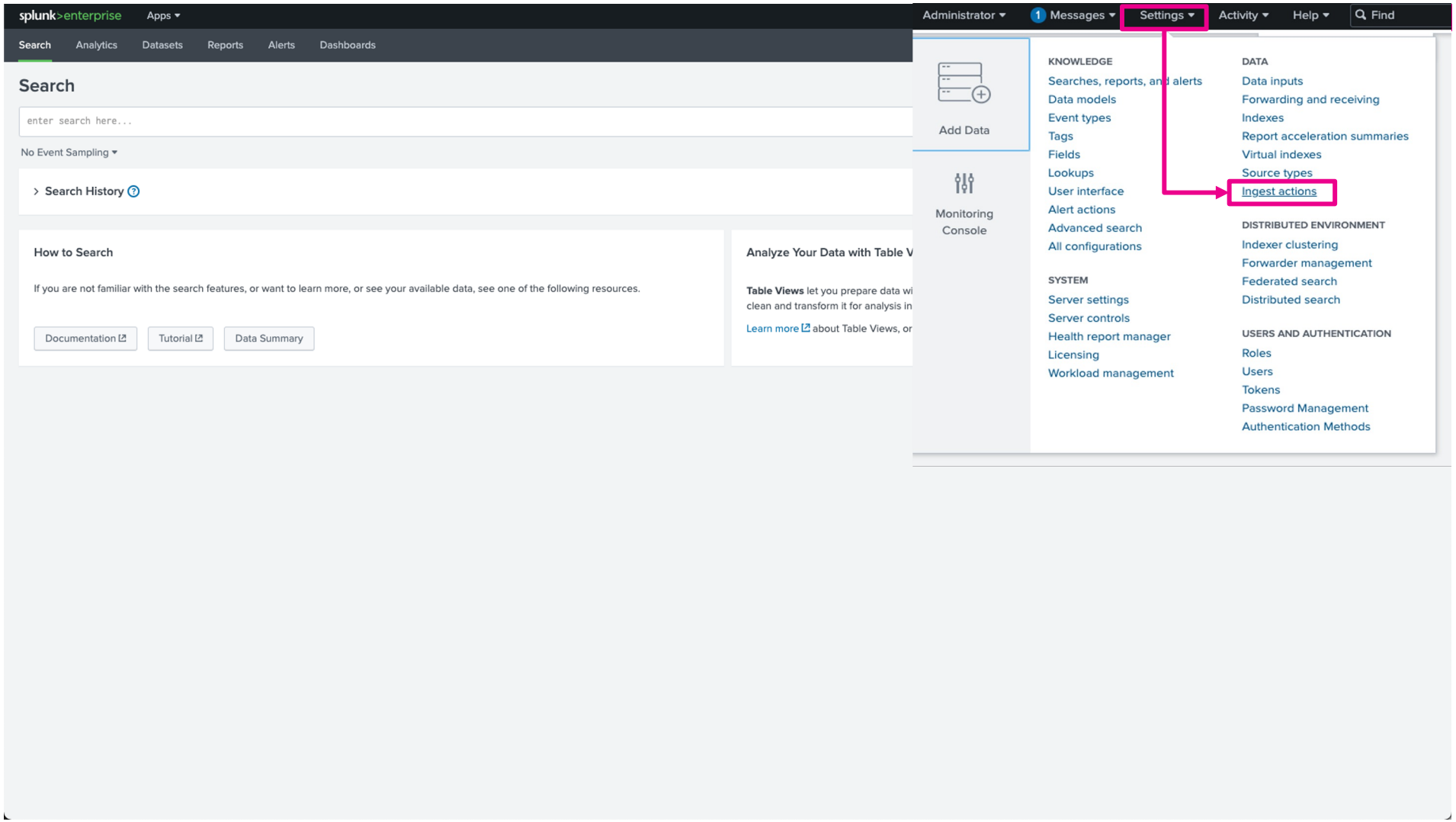
Ingest Actions

- 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>


splunk>enterprise
Apps
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
Administrator
1 Messages
Settings
Activity
Help
Find
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