



Splunk Enterprise Data Administration

Document Usage Guidelines

- Should be used only for enrolled students
- Not meant to be a self-paced document, an instructor is needed
- Do not distribute

Course Prerequisites

- Either the following courses:
 - What is Splunk?
 - Intro to Splunk
 - Using Fields
 - Introduction to Knowledge Objects
 - Creating Knowledge Objects
 - Creating Field Extractions
- Or the following courses:
 - Splunk Fundamentals 1 & 2
- And the following courses:
 - Splunk Enterprise System Administration

Course Goals

- Understand sourcetypes
- Manage and deploy forwarders with Forwarder Management
- Configure data inputs
 - File monitors
 - Network inputs (TCP/UDP)
 - Scripted inputs
 - HTTP inputs (via the HTTP Event Collector)
- Customize the input phase parsing process
- Define transformations to modify raw data before it is indexed
- Define search time knowledge object configurations

Course Outline

Module 1: Getting Data Into Splunk
Module 2: Config Files and Apps
Module 3: Configuring Forwarders
Module 4: Customizing Forwarders
Module 5: Managing Forwarders
Module 6: Monitor Inputs
Module 7: Network Inputs
Module 8: Scripted Inputs
Module 9: Agentless Inputs
Module 10: Operating System Inputs

Module 11: Fine-tuning Inputs
Module 12: Parsing Phase and Data Preview
Module 13: Manipulating Input Data
Module 14: Routing Input Data
Module 15: Supporting Knowledge Objects

System Administrator versus Data Administrator

Splunk System Administrator

System Management

- Install, configure, and manage Splunk components
- Monitor Splunk operations using the health report and Monitoring Console
- Manage Splunk licensing
- Manage Splunk configuration files
- Install and manage Splunk apps
- Create and manage Splunk indexes
- Manage Splunk user authentication
- Configure Distributed Search

Splunk Data Administrator

Data Onboarding and Management

- Work with users requesting new data sources
- Document existing and newly ingested data sources
- Design and manage inputs for UFs/HFs to capture data
- Manage parsing, event line breaking, timestamp extraction
- Move configuration through non-production testing as required
- Deploy changes to production
- Manage Splunk configuration files

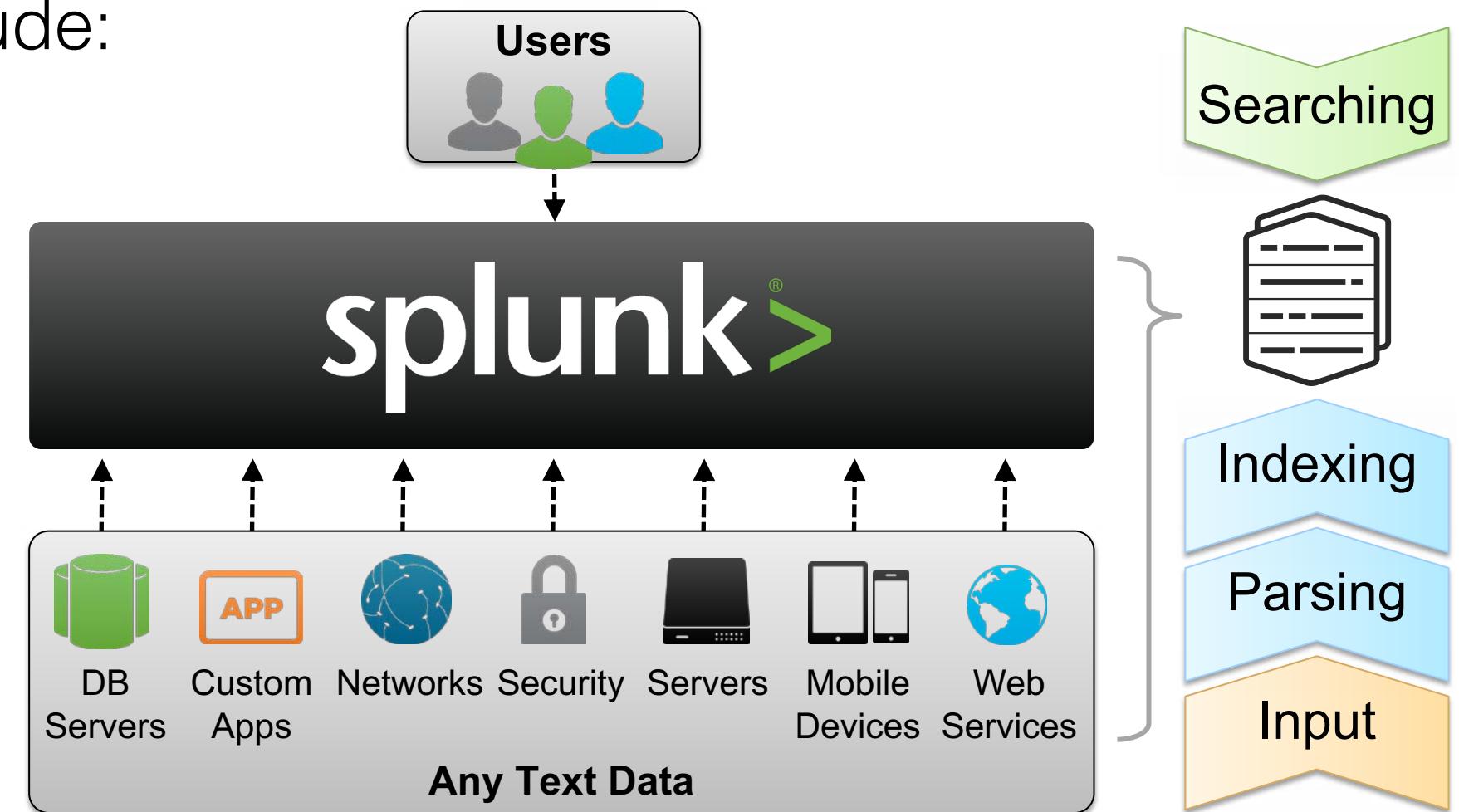
Module 1: Getting Data Into Splunk

Module Objectives

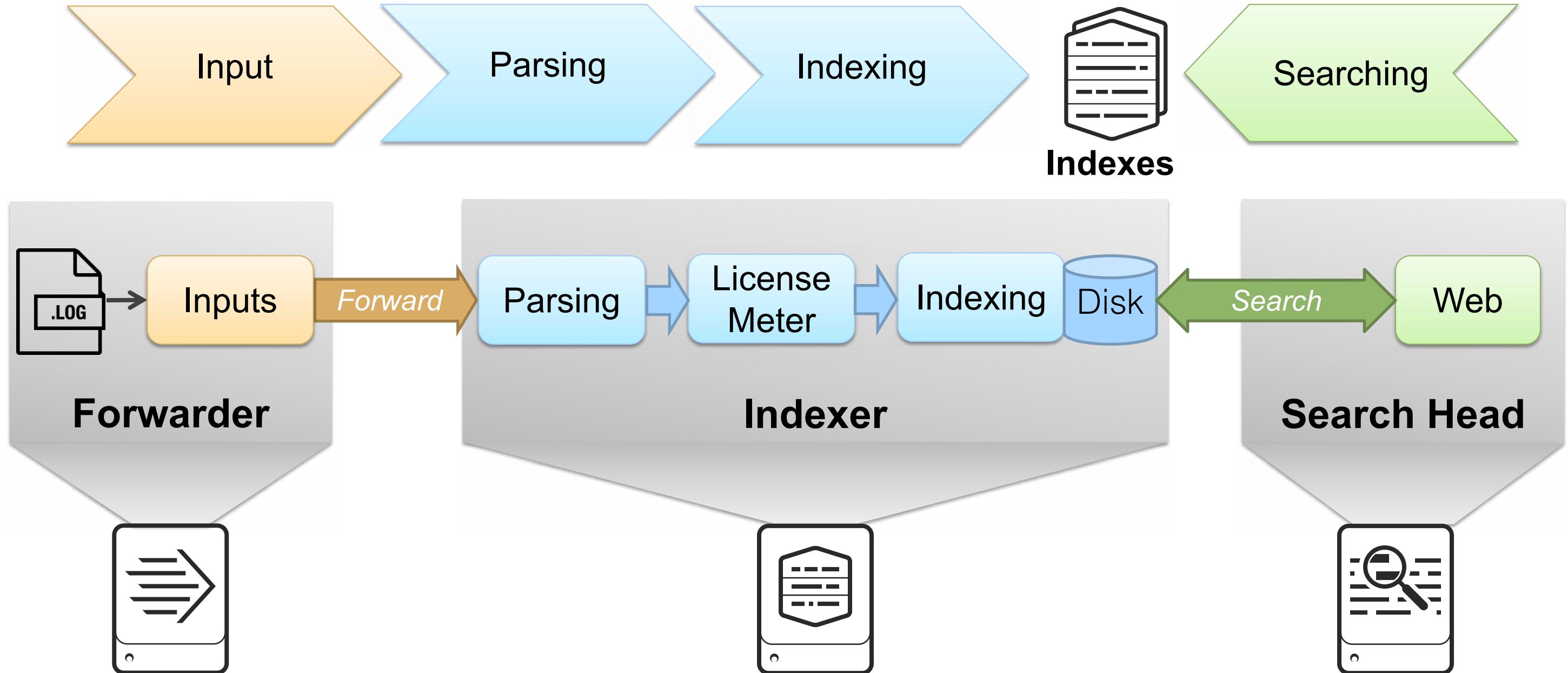
- Provide an overview of Splunk
- Describe the Splunk distributed model
- Describe data input types and metadata settings
- Configure initial input testing with Splunk Web
- Test indexes with input staging

Splunk Overview

- Splunk can be deployed in a variety of configurations
- Scales from a single server to a distributed infrastructure
- Four stages of Splunk include:
 - Input any text data
 - Parse the data into events
 - Index and store events
 - Search and report



The Splunk Distributed Model

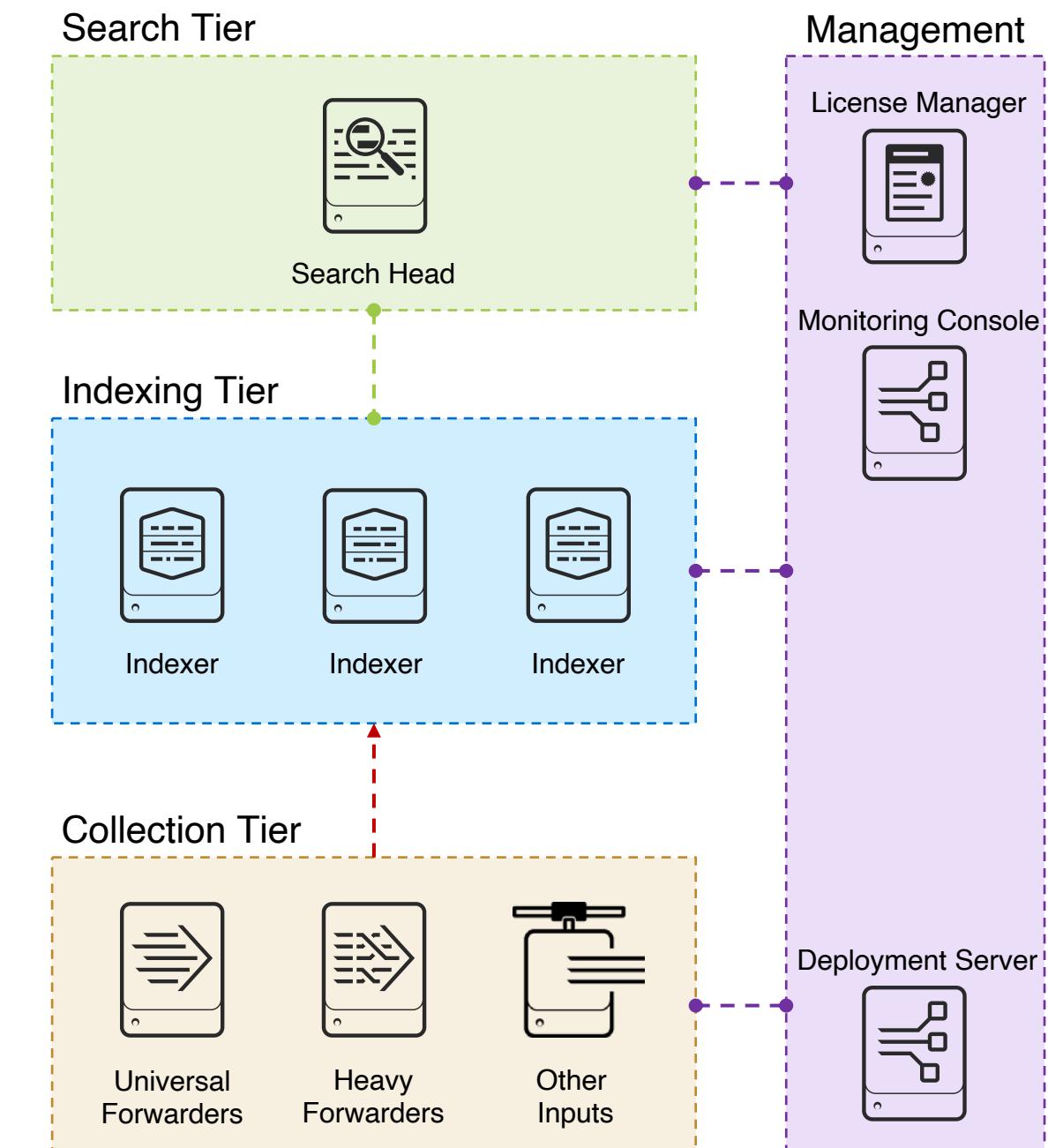


Distributed Non-Cluster Environment

- Scale Splunk in various ways
 - Add indexers to handle more inputs
 - Add indexers and search heads to handle more searching
- Centralize management using dedicated servers including:
 - Deployment Server
 - License Manager
 - Monitoring Console

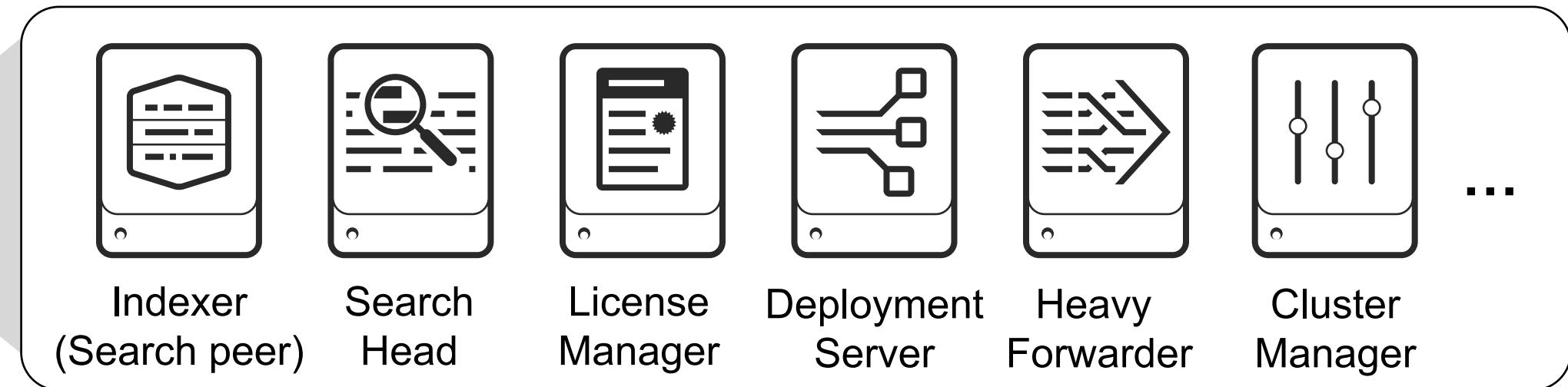
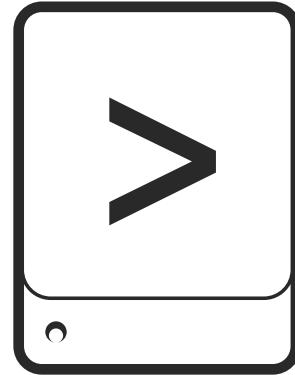
Note

You will configure a Deployment Server and different types of forwarders in later lab exercises.

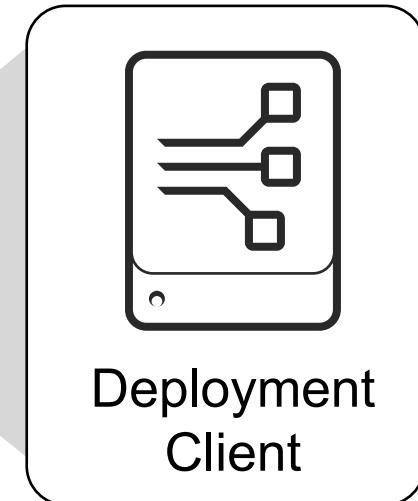
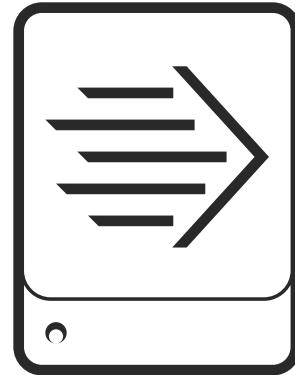


Software in Splunk Enterprise Packages

**Splunk
Enterprise
package**



**Universal
Forwarder
package**



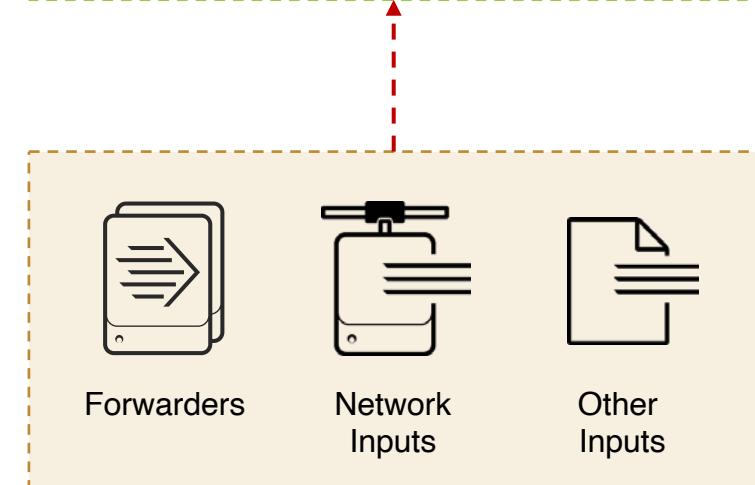
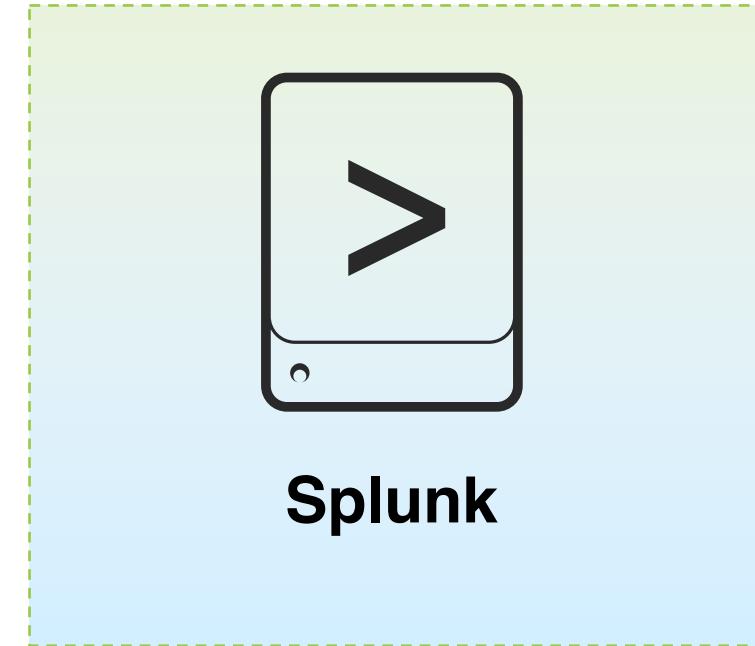
Note



The System Administrator is responsible for installing and configuring Splunk components.

Data Input Types

- Supported types of data input
 - **Files and directories**
 - **Network data**
 - **Script output**
 - **Linux and Windows logs**
 - **HTTP**
 - And more...
- You can add data inputs with:
 - Apps and add-ons
 - Splunk Web
 - CLI
 - Editing **inputs.conf**



Indexes any text data from any source

Metadata Settings

- Assigned when Splunk indexes event data
- Generally assigned to entire source during input phase
- Defaults are used if alternates are not specified
 - Overriding values can be performed at input time or later

Metadata	Description	Examples
host	Host where an event originates	<code>websvr1.example.com</code> <code>10.0.21.55</code>
source	Source file, stream or input of an event	<code>/var/log/messages</code> <code>UDP:514</code>
sourcetype	Format and category of the data input	<code>access_combined</code> <code>cisco_syslog</code>
index	Where data is stored by Splunk	<code>main</code> (default) <code>itops</code>

Adding an Input with Splunk Web

- Click the Add Data icon
 - On admin's Home page
 - On the Settings panel

- Or select:

1. Settings
2. Data inputs
3. Add new

The screenshot shows the Splunk Web interface with the following steps highlighted:

- Step 1:** Click the **Add Data** icon on the Home page. This icon is located in the top-left corner of the main content area, represented by a server icon with a plus sign.
- Step 2:** Select **Data inputs** from the **DATA** menu under the **Settings** panel. The **Data inputs** link is highlighted with a green box and a green arrow points from Step 1 to it.
- Step 3:** Click the **+ Add new** button in the **Actions** column for the **Files & Directories** input type. This button is highlighted with a green box and a green arrow points from Step 2 to it.

Data inputs
Set up data inputs from files and directories, network ports, and scripted inputs. If you want to set up forwarding and receiving between two Splunk instances, go to [Forwarding and receiving](#).

Type	Inputs	Actions
Files & Directories <small>Index a local file or monitor an entire directory.</small>	12	+ Add new
HTTP Event Collector	0	+ Add new

Add Data Menu

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

Guides for popular data sources

Or get data in with the following methods

Get data into Splunk

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

Monitor
files and ports on this Splunk platform instance
Files - HTTP - WMI - TCP/UDP - Scripts
Modular inputs for external data sources

Forward
data from a Splunk forwarder
Files - TCP/UDP - Scripts

Operating System

WIN Microsoft Windows
Windows event logs

Choose your deployment environment

Single instance
A single instance Splunk Enterprise deployment that combines indexing and search management functions.

Distributed
A distributed Splunk Enterprise deployment that separates indexing and search management into separate nodes.

Splunk Cloud
A cloud-based Splunk software service that performs all indexing and search management functions.

Overview of required configuration for your environment

Splunk Enterprise search head

Splunk Enterprise indexer cluster

High level steps

1. Configure security groups on the Windows hosts
2. Install a Splunk universal forwarder on each remote Windows host
3. Install and configure the Splunk Add-on for Windows on the universal forwarders
4. Install the Splunk Add-on for Windows across your Splunk platform deployment
5. Validate

[Full Configuration Documentation](#)

```
graph LR; OS[Operating System] --> WIN[WIN Microsoft Windows]; WIN --> DEPLOY[Choose your deployment environment]; DEPLOY --> DIST[Distributed]; DIST --> CONFIG[Overview of required configuration for your environment];
```

Add Data Menu (cont.)



Upload

files from my computer

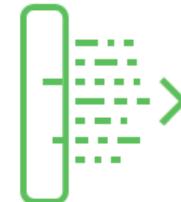
Local log files
Local structured files (e.g. CSV)
[Tutorial for adding data ↗](#)



Monitor

files and ports on this Splunk platform instance

Files - HTTP - WMI - TCP/UDP - Scripts
Modular inputs for external data sources



Forward

data from a Splunk forwarder

Files - TCP/UDP - Scripts

Upload

- Indexed once, for data that never gets updated
- Useful for testing
- File on the local machine
- Does not update **inputs.conf**

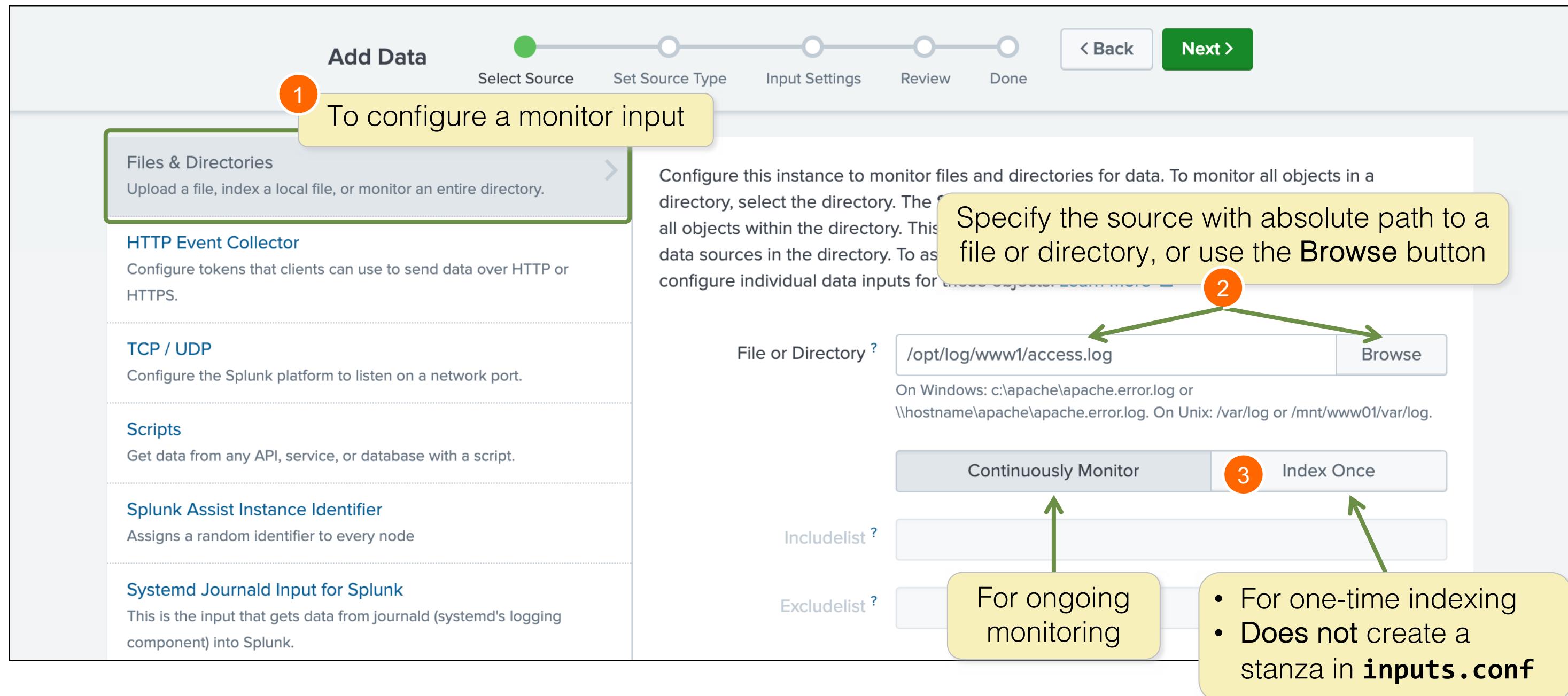
Monitor

- Indexed once or continuously
- Useful for testing or production
- File on the remote Splunk server
- Updates **inputs.conf**
- Supports files, directories, http events, network ports, and scripts

Forward

- Data from forwarders managed by this Deployment Server
- Sent to indexers' receiving port
- Main source of input in production
- Updates **inputs.conf**

Select Source



Select Source: Additional Information

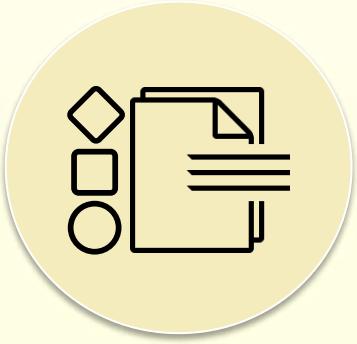
- To monitor a shared network drive, enter:

*nix:	<code><host>/<path></code>
Windows:	<code>\\"<host>\<path></code>

- Splunk requires read access to the share
- Additional sources on Linux Splunk instances
 - Systemd Journald Input
- Additional sources on Windows Splunk instances
 - Including Event Logs, Performance Monitoring, Registry monitoring, and Active Directory monitoring

Local Event Logs Collect event logs from this machine.
Remote Event Logs Collect event logs from remote hosts. Note: this uses WMI and requires a domain account.
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.
Local Performance Monitoring Collect performance data from this machine.
Remote Performance Monitoring Collect performance and event information from remote hosts. Requires domain credentials.
Registry monitoring Have the Splunk platform index the local Windows Registry, and monitor it for changes.
Active Directory monitoring <small>Index and monitor Active Directory</small>

Understanding Source Types



Source Types

- Splunk's way of categorizing data types
- Frequently used during index processes
- Used in searches, reports, apps, etc.
- Can be explicitly set with Splunk Web, CLI, or by modifying **inputs.conf**
- Assigned automatically when possible
- Can be set by administrators or apps

Source type: access_combined_wcookie ▾

> filter

> Default Settings
Splunk's default source type settings

> Application

Database

Email

Log to Metrics

Metrics

Miscellaneous

Network & Security

Operating System

Structured

Uncategorized

Web

access_combined
National Center for Supercomputing Applications (NCSA) combined format HTTP web server logs (can be generated by apache or other web servers)

apache_error
Error log format produced by the Apache web server (typically error_log on *nix systems)

iis
W3C Extended log format produced by the Microsoft Internet Information Services (IIS) web server

Pretrained Source Types

- Built-in source types shipped with Splunk
- Can be added to manually and defined by Splunk apps
- Listed in Splunk documentation:

docs.splunk.com/Documentation/Splunk/latest/Data/Listofpretrainedsourcetypes

Source type name	Origin	Examples
access_combined	NCSA combined format http web server logs (can be generated by apache or other web servers)	<code>10.1.1.43 - webdev [08/Aug/2005:13:18:16 -0700] "GET /HTTP/1.0" 200 0442 "-" "check_http/1.10 (nagios-plugins 1.4)"</code>
access_combined_wcookie	NCSA combined format http web server logs (can be generated by apache or other web servers), with cookie field added at end	<code>"66.249.66.102.1124471045570513" 59.92.110.121 -- [19/Aug/2005:10:04:07 -0700] "GET /themes/splunk_com/images/logo_splunk.png HTTP/1.1" 200 994 "http://www.splunk.org/index.php/docs" "Mozilla/5.0 (X11; U; Linux i686; en-US; rv:1.7.8) Gecko/20050524 Fedora/1.0.4-4 Firefox/1.0.4" "61.3.110.148.1124404439914689"</code>
access_common	NCSA common format http web server logs (can be generated by	<code>10.1.1.140 -- [16/May/2005:15:01:52 -0700] "GET</code>

Set Source Type: Data Preview

Automatically determined for major data types

1

Source: /opt/log/www1/access.log

Source type: access_combined_wcookie ▾

Save As

Event Breaks

Timestamp

Advanced

List ▾

Format

20 Per Page ▾

< Prev

1 2 3 4 5 6 7 8 ... Next >

	Time	Event
1	4/28/23 4:46:30.000 PM	211.166.11.101 - - [28/Apr/2023:16:46:30] "GET /product.screen?productId=MB-AG-G07&JSESSIONID=SD2SL9FF10ADFF4953 HTTP/1.1" 200 3426 "http://www.buttercupgames.com" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/536.5" 925
2	4/28/23 4:46:31.000 PM	211.166.11.101 - - [28/Apr/2023:16:46:31] "GET category.screen?categoryId=STRATEGY&JSESSIONID=SD2SL9FF10ADFF4953 HTTP/1.1" 200 2324 "http://www.buttercupgames.com/oldlink?itemId=EST-14" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 610
3	4/28/23 4:46:35.000 PM	211.166.11.101 - - [28/Apr/2023:16:46:35] "GET show.do?productId=SF-BVS-01&JSESSIONID=SD2SL9FF10ADFF4953 HTTP/1.1" 404 253 "http://www.buttercupgames.com/category.screen?categoryId=NULL" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 183

View Event Summary

2

Data Preview displays how processed events will be indexed

	Time	Event
1	4/28/23 4:46:30.000 PM	211.166.11.101 - - [28/Apr/2023:16:46:30] "GET /product.screen?productId=MB-AG-G07&JSESSIONID=SD2SL9FF10ADFF4953 HTTP/1.1" 200 3426 "http://www.buttercupgames.com" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/536.5" 925
2	4/28/23 4:46:31.000 PM	211.166.11.101 - - [28/Apr/2023:16:46:31] "GET category.screen?categoryId=STRATEGY&JSESSIONID=SD2SL9FF10ADFF4953 HTTP/1.1" 200 2324 "http://www.buttercupgames.com/oldlink?itemId=EST-14" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 610
3	4/28/23 4:46:35.000 PM	211.166.11.101 - - [28/Apr/2023:16:46:35] "GET show.do?productId=SF-BVS-01&JSESSIONID=SD2SL9FF10ADFF4953 HTTP/1.1" 404 253 "http://www.buttercupgames.com/category.screen?categoryId=NULL" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 183

Set Source Type: Data Preview (cont.)

Set Source Type

This page lets you see how the Splunk platform sees your data before indexing. If the events look correct and have the right timestamps, click "Next" to proceed. If not, use the options below to define proper event breaks and timestamps. If you cannot find an appropriate source type for your data, create a new one by clicking "Save As".

Source: /opt/log/www1/access.log [View Event Summary](#)

	Time	Event
1	4/28/23 4:46:30.000 PM	211.166.11.101 -- [28/Apr/2023:16:46:30] "GET /product.screen?productId=MB-AG-G07&JSESSIONID=SD2SL9FF10ADFF4953 HTTP 1.1" 200 3426 "http://www.buttercupgames.com" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 925
2	4/28/23 4:46:31.000 PM	211.166.11.101 -- [28/Apr/2023:16:46:31] "GET /category.screen?categoryId=STRATEGY&JSESSIONID=SD2SL9FF10ADFF4953 HTTP 1.1" 200 2324 "http://www.buttercupgames.com/oldlink?itemId=EST-14" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 610
3	4/28/23 4:46:35.000 PM	211.166.11.101 -- [28/Apr/2023:16:46:35] "GET show.do?productId=SF-BVS-01&JSESSIONID=SD2SL9FF10ADFF4953 HTTP 1.1" 404 253 "http://www.buttercupgames.com/category.screen?categoryId=NULL" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 183
4	4/28/23 4:46:40.000 PM	211.166.11.101 -- [28/Apr/2023:16:46:40] "GET /category.screen?categoryId=NULL&JSESSIONID=SD2SL9FF10ADFF4953 HTTP 1.1" 505 3503 "http://www.buttercupgames.com/category.screen?categoryId=NULL" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 490

Source type: access_combined_wcookie ▾ [Save As](#)

filter

Default Settings
Splunk's default source type settings

Application
Database
Email
Log to Metrics
Metrics
Miscellaneous
Network & Security
Operating System
Structured
Uncategorized
Web

Optional choose a different source type

Set Source Type: Data Preview Warning

Set Source Type

This page lets you see how the Splunk platform sees your data before indexing. If the events look correct and have the right timestamps, click "Next" to proceed. If not, use the options below to define proper event source type for your data, create a new one by clicking "Save As".

Source: /opt/log/www1/access.log

Source type: apache_error ▾

Save As

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

	Time	Event
1	4/28/23 4:46:30.000 PM	211.166.11.101 - - [28/Apr/2023:16:46:30] "GET /product.screen?productId=MB-AG-G07&JSESSIONID=SD2SL9FF10ADFF4953 HTTP 1.1" 200 3426 "http://www.buttercupgames.com" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 925
2	4/28/23 4:46:31.000 PM	211.166.11.101 - - [28/Apr/2023:16:46:31] "GET /category.screen?categoryId=STRATEGY&JSESSIONID=SD2SL9FF10ADFF4953 HTTP 1.1" 200 2324 "http://www.buttercupgames.com/oldlink?itemId=EST-14" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 610
3	4/28/23 4:46:35.000 PM	211.166.11.101 - - [28/Apr/2023:16:46:35] "GET /category.screen?categoryId=AC-1&JSESSIONID=SD2SL9FF10ADFF4953 HTTP 1.1" 200 2324 "http://www.buttercupgames.com/oldlink?itemId=EST-14" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/536.5 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 183

Allows creation of a new source type for a specific source data

Warning

If events are not separated correctly or have incorrect timestamps, select a different source type from the list or customize the source type settings.

splunk® turn data into doing™

24

Splunk Enterprise Data Administration
Copyright © 2023 Splunk, Inc. All rights reserved 2 October 2023

Input Settings

The screenshot shows the 'Add Data' wizard with five steps: 'Select Source' (green), 'Set Source Type' (green), 'Input Settings' (gray), 'Review' (white), and 'Done' (white). The 'Input Settings' step is active. The interface includes sections for 'App context', 'Host', and 'Index'. A yellow callout points to the 'App Context' dropdown set to 'Search & Reporting (search)'. Another yellow callout points to the 'Host field value' input field containing 'splunk08'. A third yellow callout points to the 'Index' dropdown set to 'Default'.

Input Settings
Optional set additional input parameters for this data input as follows:

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 ↗](#)

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 ↗](#)

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 ↗](#)

- Where input configuration is saved
- For Search & Reporting (search):
SPLUNK_HOME/etc/apps/search/local

By default, the **default host name** in **General settings** is used

Select index where input will be stored

Review

Review the input configuration summary and click Submit to finalize

The screenshot shows the 'Add Data' wizard in the 'Review' step. The top navigation bar includes 'Add Data' and five steps: 'Select Source' (green dot), 'Set Source Type' (green dot), 'Input Settings' (green dot), 'Review' (green dot), and 'Done' (grey dot). Below the navigation is a 'Review' section with the following configuration details:

Input Type	File Monitor
Source Path	/opt/log/www1/access.log
Continuously Monitor	Yes
Source Type	access_combined_wcookie
App Context	search
Host	splunk08
Index	default

To the right of the review section is a note box with a blue 'i' icon:

Note
Confirm settings before proceeding.
It is easier to use < Back and make changes than to rectify later.

What Happens Next?

- Indexed events are available for immediate search
 - Splunk may take a minute to start indexing the data
- You are given other options to do more with your data
- Input configuration is saved in:

SPLUNK_HOME/etc/apps/<app>/local/inputs.conf

```
[monitor:///opt/log/www1/access.log]
index = test
sourcetype = access_combined_wcookie
```

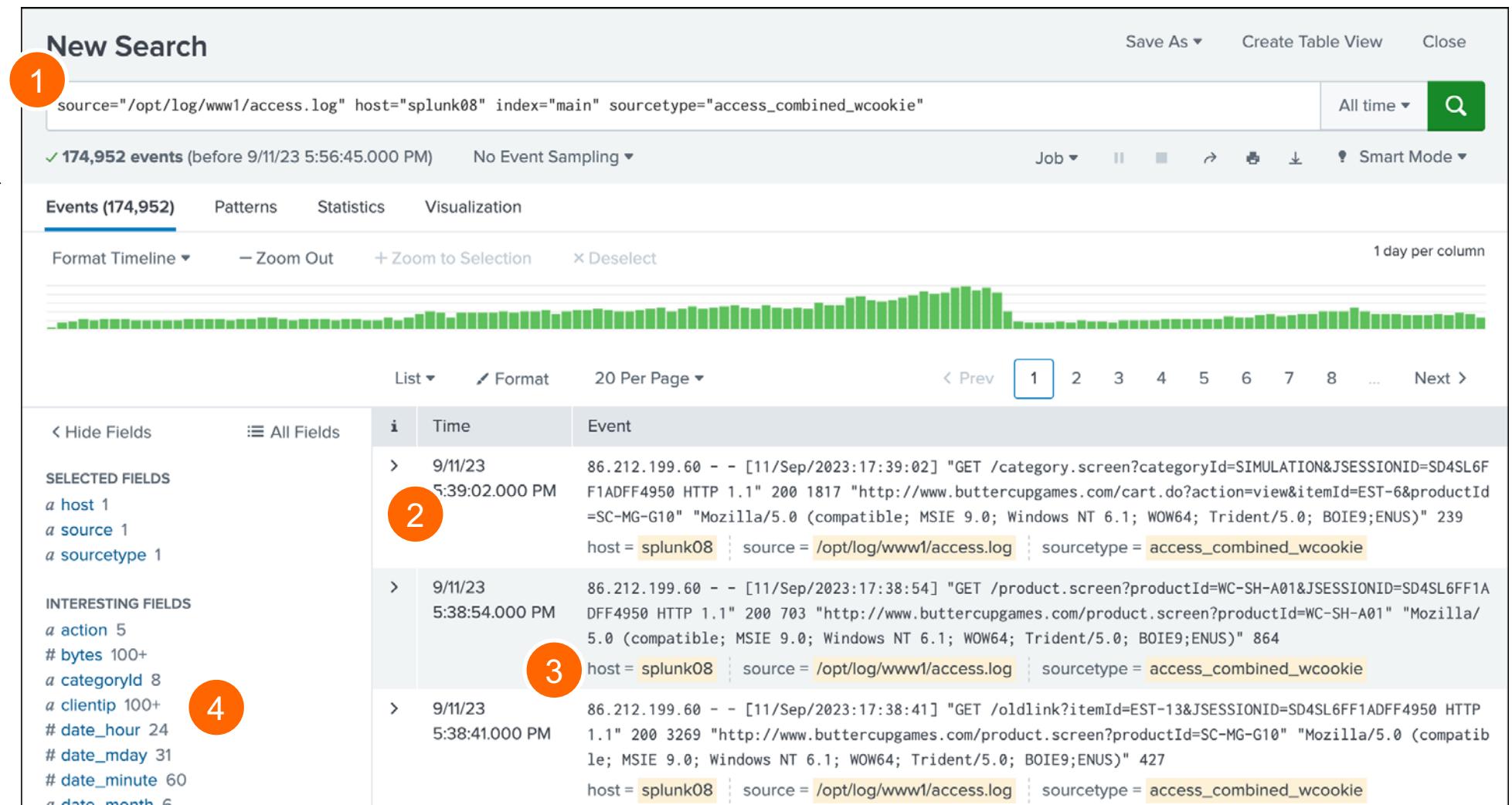
The screenshot shows the 'Add Data' wizard in Splunk. The progress bar at the top indicates the user is on the 'File input has been created successfully' step, which is the final step of the process. Below the progress bar, there is a success message: 'File input has been created successfully.' followed by a link to 'Configure your inputs by going to Settings > Data Inputs'. There are several buttons and links for further actions: 'Start Searching' (green button), 'Search your data now or see examples and tutorials.', 'Extract Fields' (button), 'Create search-time field extractions. Learn more about fields.', 'Add More Data' (button), 'Add more data inputs now or see examples and tutorials.', 'Download Apps' (button), 'Apps help you do more with your data. Learn more.', and 'Build Dashboards' (button), 'Visualize your searches. Learn more.'

Note

Entries in the **inputs.conf** file are not created when **Upload** or **Index Once** is selected.

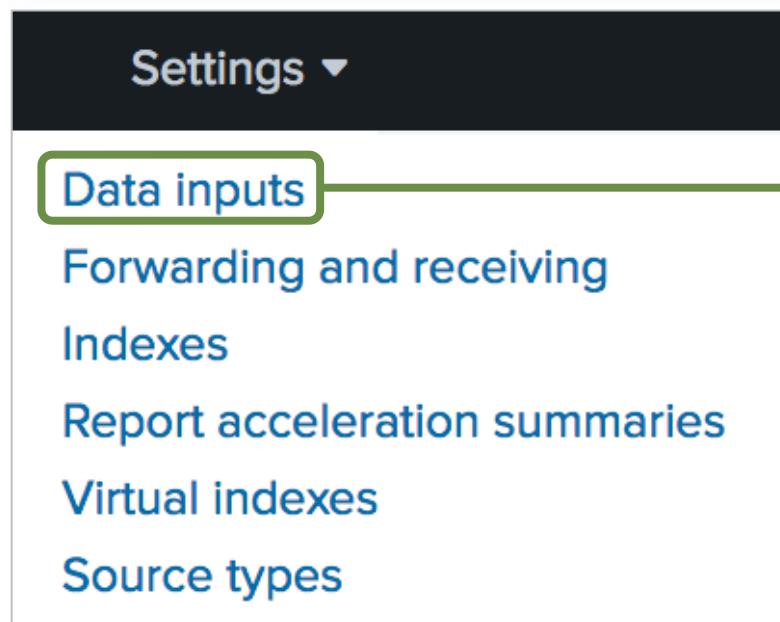
Verify your Input

1. Click Start Searching or search for **index=<indexname>**
2. Verify events and timestamps
3. Confirm the host, source, and sourcetype field values
4. Check the auto-extracted field names



Viewing Configured Inputs

Select Settings > Data Inputs



Data inputs
Set up data inputs from files and directories, network ports, and scripted inputs. If you want to set up forwarding and receiving between two Splunk instances, go to [Forwarding and receiving](#).

Local inputs Inputs handled by this server

Type	Inputs	Actions
Files & Directories Index a local file or monitor an entire directory.	10	+ Add new
HTTP Event Collector Receive data over HTTP or HTTPS.	0	+ Add new
TCP Listen on a TCP port for incoming data, e.g. syslog.	0	+ Add new
UDP Listen on a UDP port for incoming data, e.g. syslog.	0	+ Add new
...		

Forwarded inputs

Type	Actions
Windows Event Logs Collect event logs from forwarders.	0 + Add new
Files & Directories Monitor files or directories on forwarders.	0 + Add new

Viewing Configured Inputs: Files & Directories

Files & directories

Data inputs » Files & directories

Showing 1-12 of 12 items

filter 

Index Location of configuration (app context)

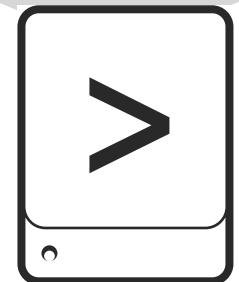
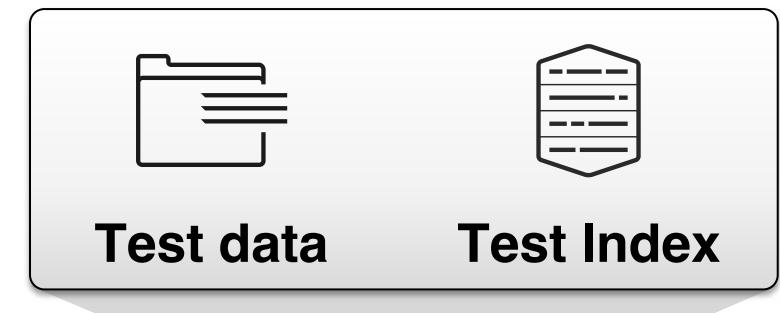
New Local File & Directory

Full path to your data	Set host	Source type	Index	Number of files	App	Status	Actions
\$SPLUNK_HOME/etc/splunk.version	Constant Value	splunk_version	_internal	1	system	Enabled Disable	
\$SPLUNK_HOME/var/log/introspection	Constant Value	Automatic	_introspection	7	introspection_generator_addon	Enabled Disable	
\$SPLUNK_HOME/var/log/splunk	Constant Value	Automatic	_internal	69	system	Enabled Disable	
\$SPLUNK_HOME/var/log/splunk/license_usage_summary.log	Constant Value	Automatic	_telemetry	1	system	Enabled Disable	
\$SPLUNK_HOME/var/log/splunk/splunk_instrumentation_cloud.log*	Constant Value	splunk_cloud_telemetry	_telemetry	1	system	Enabled Disable	
\$SPLUNK_HOME/var/log/watchdog/watchdog.log*	Constant Value	Automatic	_internal	1	system	Enabled Disable	
\$SPLUNK_HOME/var/run/splunk/search_telemetry/*search_telemetry.json	Constant Value	search_telemetry	_introspection	0	system	Enabled Disable	
\$SPLUNK_HOME/var/spool/splunk	Constant Value	Automatic	default		system	Disabled Enable	
\$SPLUNK_HOME/var/spool/splunk/...stash_hec	Constant Value	stash_hec	default		system	Disabled Enable	
\$SPLUNK_HOME/var/spool/splunk/...stash_new	Constant Value	stash_new	default		system	Disabled Enable	
\$SPLUNK_HOME/var/spool/splunk/tracker.log*	Constant Value	splunkd_latency_tracker	_internal	0	system	Enabled Disable	
/opt/log/www2/access.log	Constant Value	access_combined_wcookie	test	1	search	Enabled Disable	Delete

Click to edit existing input settings

Initial Data Input Testing

- Use a Splunk test server
 - Should be running same version as production
- Use test indexes
- Procedure:
 1. Copy production data to test server
 2. Use Splunk Web > Add Data
 3. Check to see if **sourcetype** and other settings are applied correctly
 4. Delete the test data, change your test configuration, and repeat as necessary



Test server

Question: Supported Splunk input types

Which of the following is *not* a supported Splunk data input type?

- A. Files and directories
- B. Network events
- C. Executable files
- D. Scripts

Answer: Supported Splunk input types

Which of the following is *not* a supported Splunk data input type?

- A. Files and directories
- B. Network events
- C. Executable files**
- D. Scripts

Splunk supports text-based data inputs. Executable files are not supported data input types.

Question: Inputs updating `inputs.conf`

Splunk will not update an `inputs.conf` file when you use which of the following:

- A. The Upload option in Settings > Add Data
- B. The Monitor option in Settings > Add Data
- C. The Forward option in Settings > Add Data
- D. Any Splunk command line when creating a data source

Answer: Inputs updating `inputs.conf`

Splunk will not update an `inputs.conf` file when you use which of the following:

- A. The Upload option in Settings > Add Data
- B. The Monitor option in Settings > Add Data
- C. The Forward option in Settings > Add Data
- D. Any Splunk command line when creating a data source

The Upload option ingests the data one time only, and therefore does not need to update the `inputs.conf` file with settings for that data input.

Question: Default index

What is the default index where Splunk will store event data, if an alternative is not specified?

- A. `_internal`
- B. `main`
- C. `summary`
- D. `default_index`

Answer: Default index

What is the default index where Splunk will store event data, if an alternative is not specified?

- A. `_internal`
- B. `main`**
- C. `summary`
- D. `default_index`

The default index for Splunk is **main**. You can customize the default database for your Splunk configuration in **indexes.conf** with **defaultDatabase = <index_name>**.

Question: Input Data Source Component

In most production Splunk environments, **most of** the input data comes from which Splunk components?

- A. Search Heads
- B. Indexers
- C. Deployment Servers
- D. Forwarders

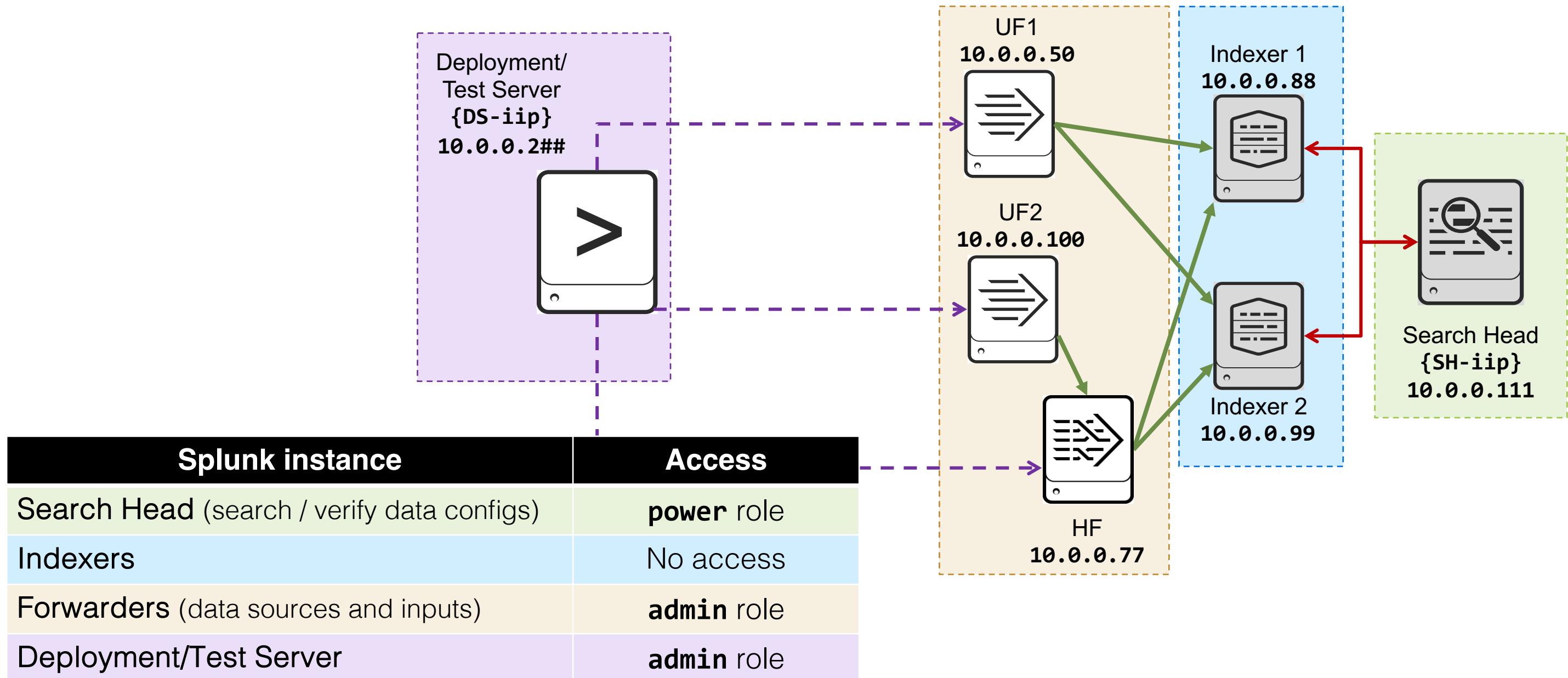
Answer: Input Data Source Component

In most production Splunk environments, **most of** the input data comes from which Splunk components?

- A. Search Heads
- B. Indexers
- C. Deployment Servers
- D. Forwarders**

In general, most data input comes from Splunk forwarders. The other Splunk components listed are used to manage the ingest of data, search, and Splunk management.

Access Scenario For Course Labs



Module 1 Lab

Time: 20 minutes

Description: Add a Local Data Input

Tasks:

- Discover Splunk Enterprise lab environment
- Log into search head and test/deployment server
- Create a test index on the deployment/test server
- Index a file on the deployment server
- Verify the indexed events with their metadata values

Module 2: Config Files and Apps

Module Objectives

- Identify Splunk configuration files and directories
- Describe index-time and search-time precedence
- Validate and update configuration files
- Explore Splunk apps and app installation

Note 

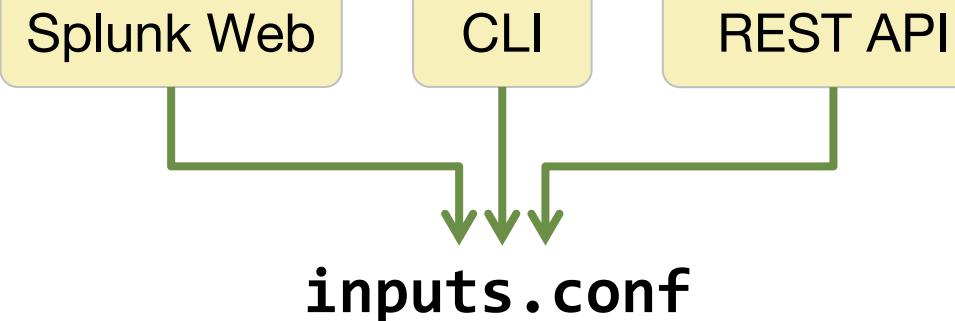
Configuration Files and Apps are covered in greater detail in the *Splunk Enterprise System Administration* course. This lesson and lab reviews both concepts.

Splunk Configuration Files



Configuration Files (.conf)

- Govern an aspect of Splunk functionality
- Text files are generally case sensitive with **[stanza]** and **attribute = value** format
- Modified using Splunk Web, CLI, REST API, app install, or directly editing
- Saved under **SPLUNK_HOME/etc**
- Come with documentation and examples under **SPLUNK_HOME/etc/system/README/**



```
[default]  
host=www
```

```
[monitor:///var/log/httpd]  
sourcetype = access_common  
ignoreOlderThan = 7d  
index = web
```

Note

For **.conf** file documentation and examples view **SPLUNK_HOME/etc/system/README/**:

- ***.conf.spec**
- ***.conf.example**

Methods for Modifying Splunk Configurations

- Splunk Web
- Splunk CLI

```
./splunk add monitor /opt/log/www1/access.log -index itops  
-sourcetype access_combined_wcookie -host splunk01
```

- Editing .conf files

```
[monitor:///opt/log/www1/access.log]  
disabled = false  
host = splunk01  
index = itops  
sourcetype = access_combined_wcookie
```

Host
Tell Splunk how to set the value of the host field in your events from this source.

Set host

Host field value

Source type
Tell Splunk what kind of data this is so you can group it with other sources. You can specify what you want if Splunk gets it wrong.

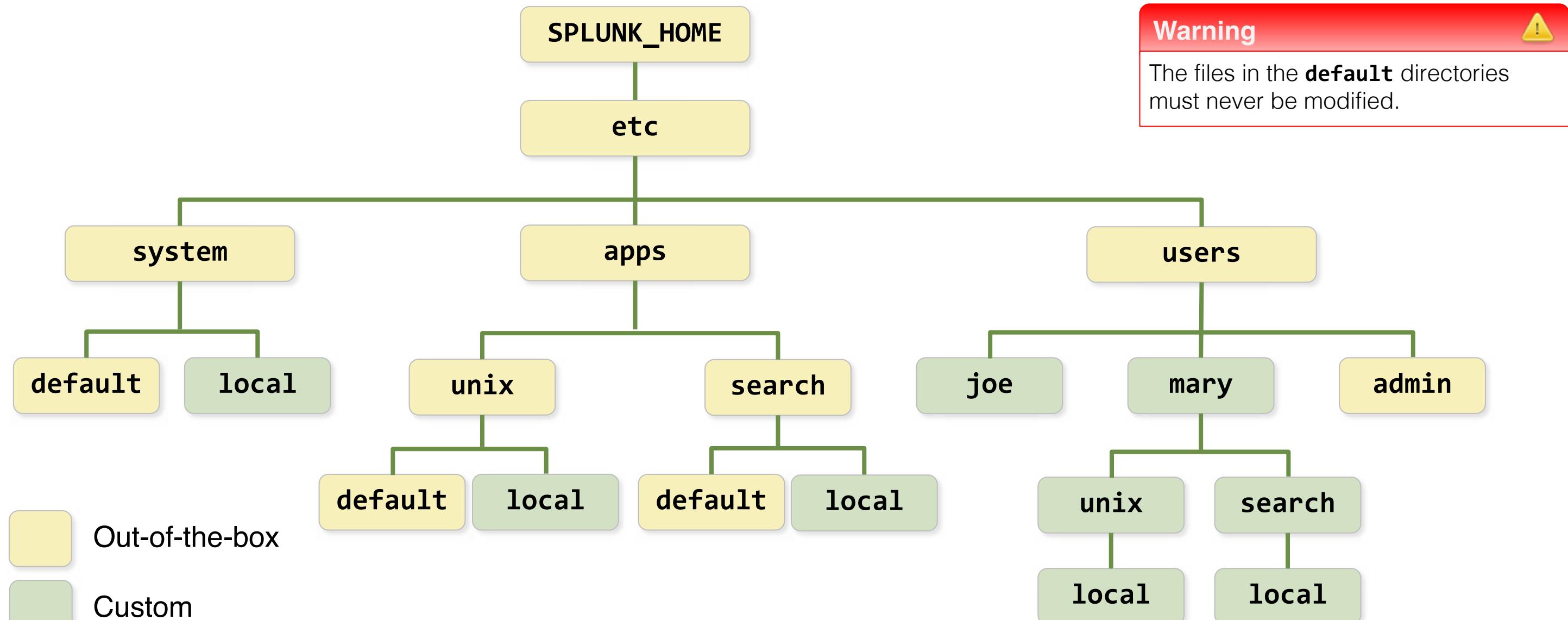
Set the source type
When this is set to automatic, Splunk uses the sourcetypes placeholder name to determine the source type.

Source type *

Index
Set the destination index for this source.

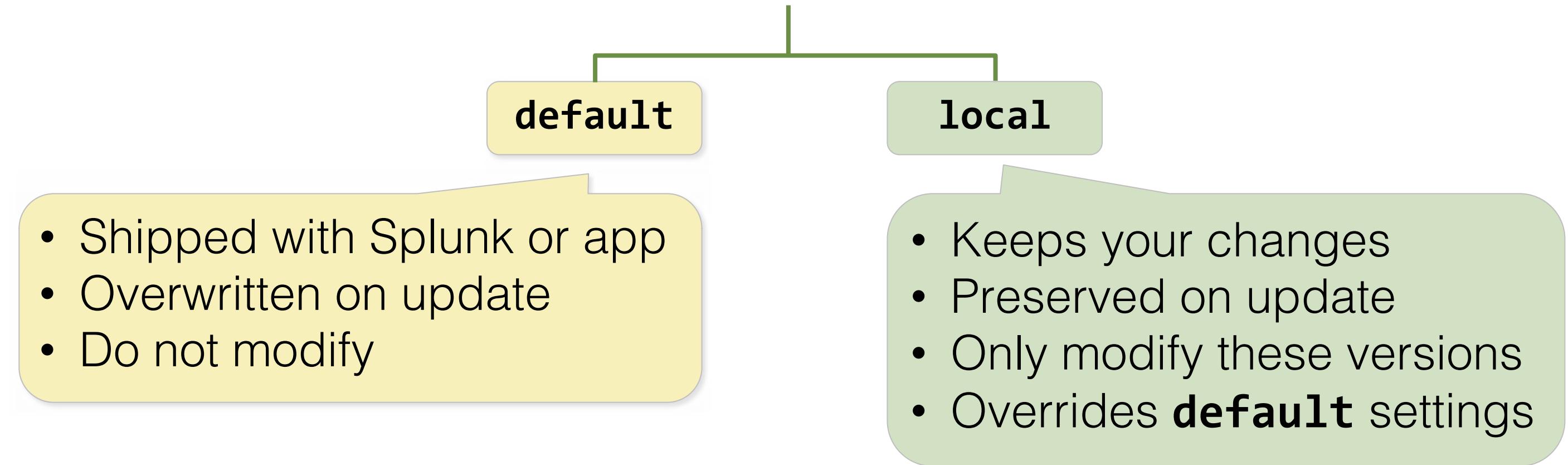
Index

Configuration Directories



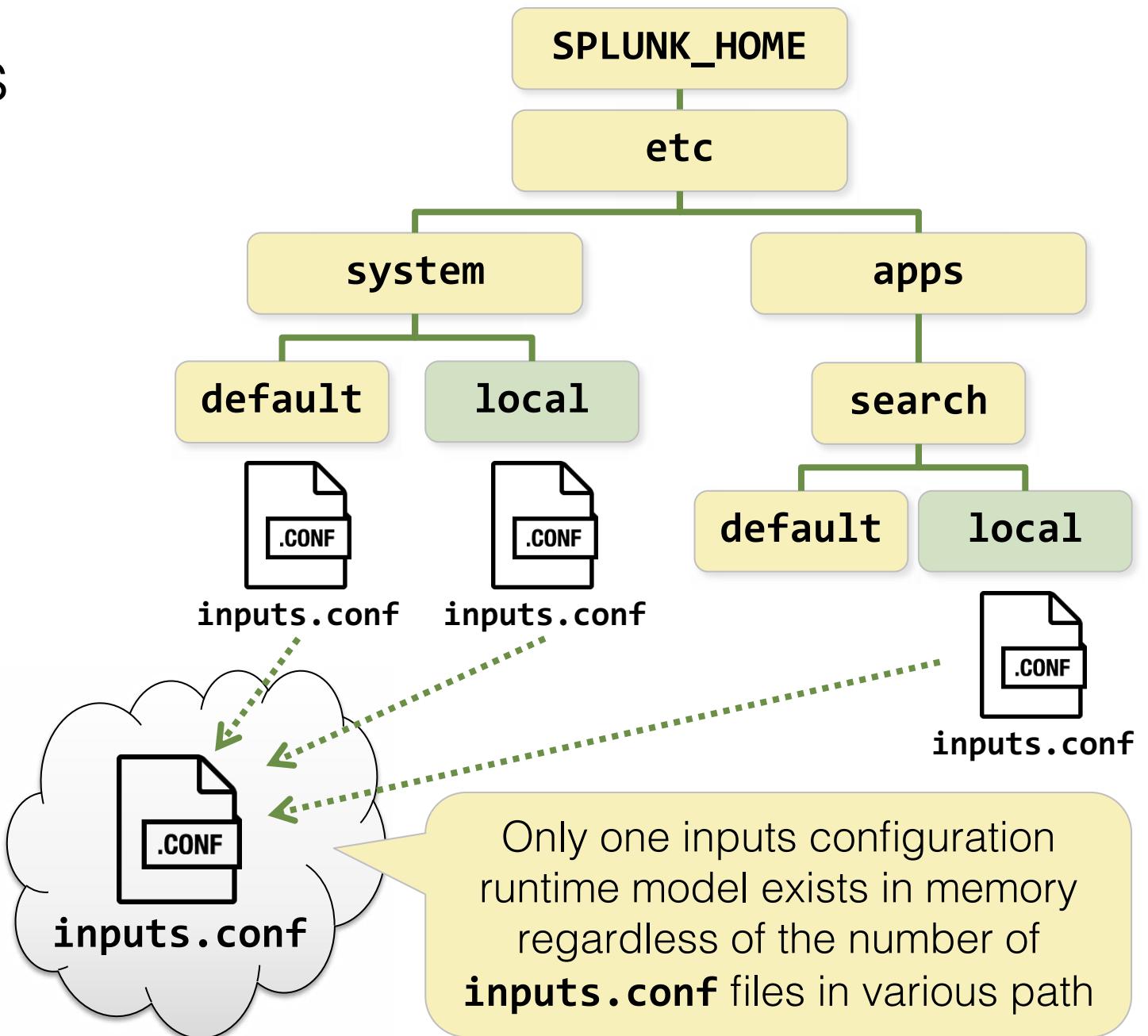
docs.splunk.com/Documentation/Splunk/latest/Admin>Listofconfigurationfiles

Default versus Local Configuration



Merging of Configuration Files

- Splunk merges configuration files
 - Generally, when Splunk starts, or when searches are run
 - Into a single run-time model for each file type
 - As a union of all files if no duplicates/conflicts exist
- In case of conflicts, priority is based on the context:
 - Global context (index-time)
 - App/User context (search-time)

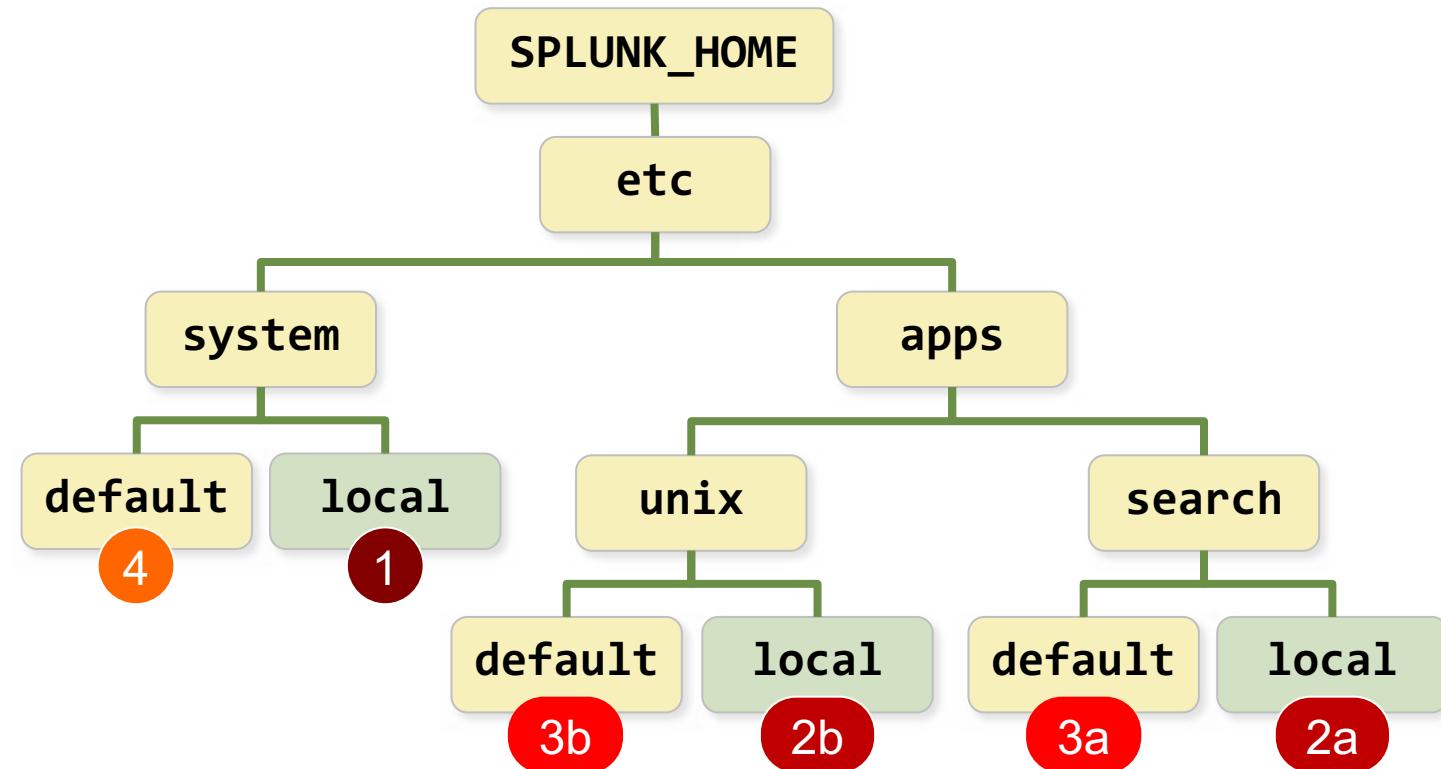


File Context and Index-time versus Search-time

	Global Context	App/User Context
<i>Used during:</i>	Index-time	Search-time
<i>Used by:</i>	<ul style="list-style-type: none">• User-independent tasks• Background tasks• Input, parsing, indexing	<ul style="list-style-type: none">• User-related activity• Searching• Search-time processing
<i>Example use-case:</i>	A network input to collect syslog data	Mary's private report in the Search app
<i>Example files:</i>	inputs.conf outputs.conf props.conf	macros.conf savedsearches.conf props.conf

docs.splunk.com/Documentation/Splunk/latest/Admin/Wheretofindtheconfigurationfiles

Index-Time Precedence (Global Context)



Precedence order

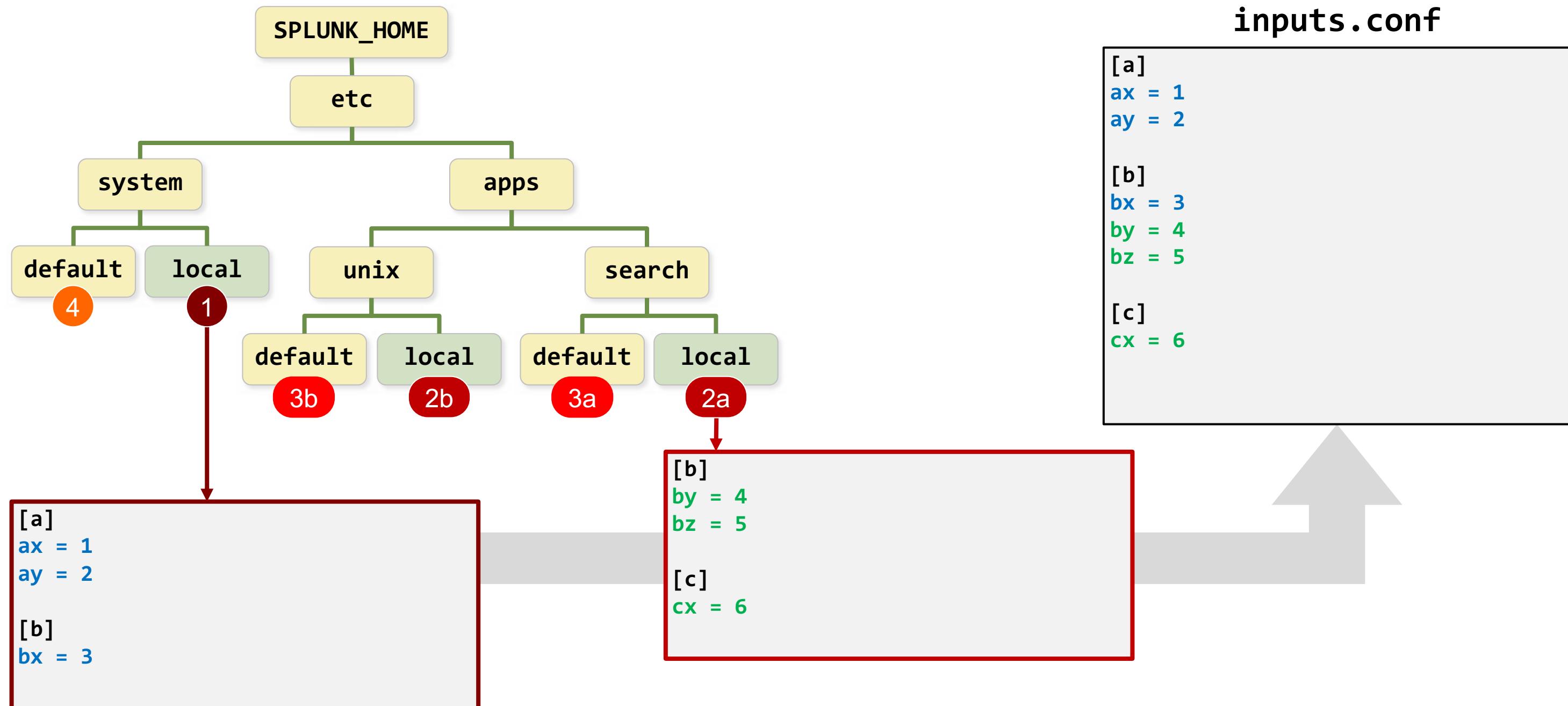
- 1 System local directory
`etc/system/local`
- 2 App local directories*
`etc/apps/appname/local`
- 3 App default directories*
`etc/apps/appname/default`
- 4 System default directory
`etc/system/default`

Note

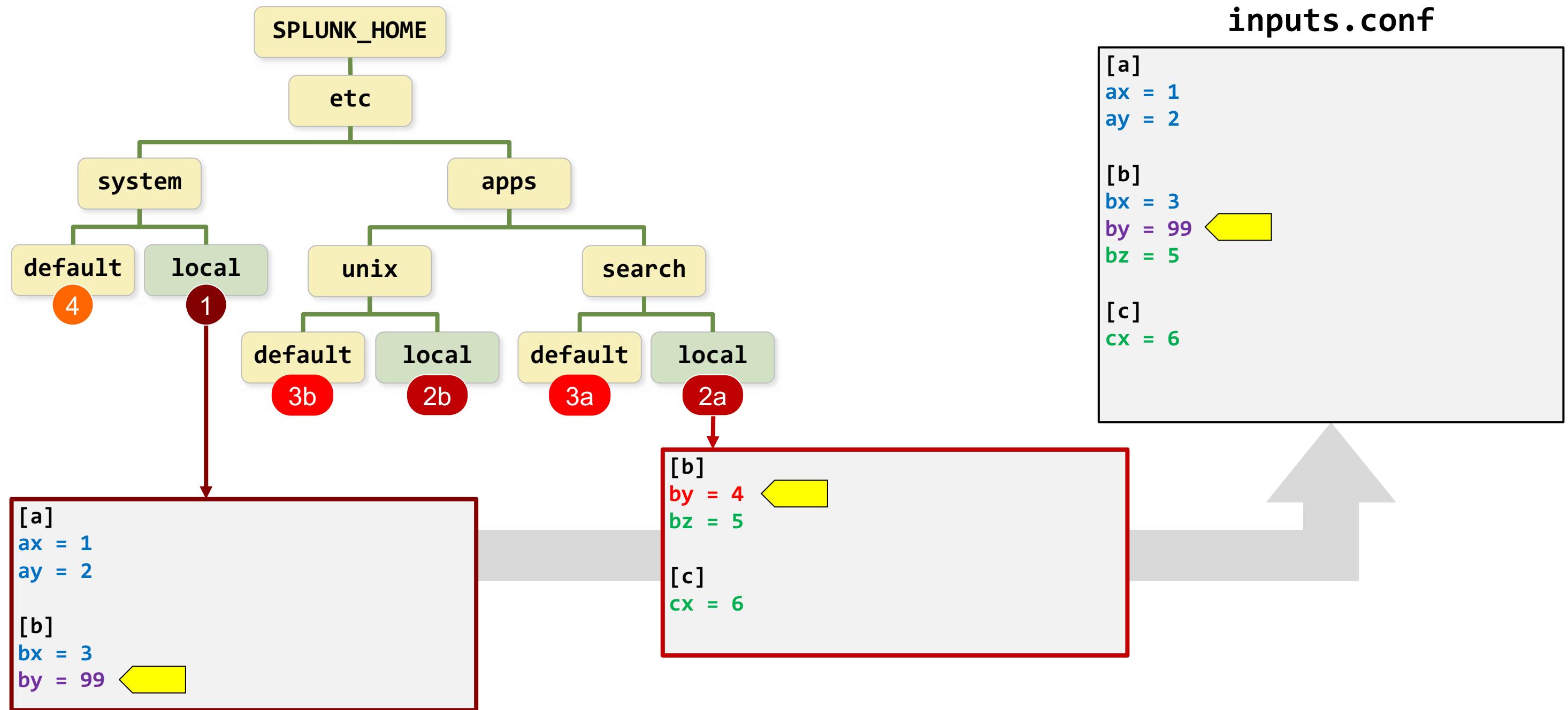


* When determining priority of app directories in **global** context (for steps 2 and 3), Splunk uses *lexicographical* order. (Files in apps directory "A" have higher priority than files in apps directory "B".)

Example of Index-Time Precedence: No Conflict

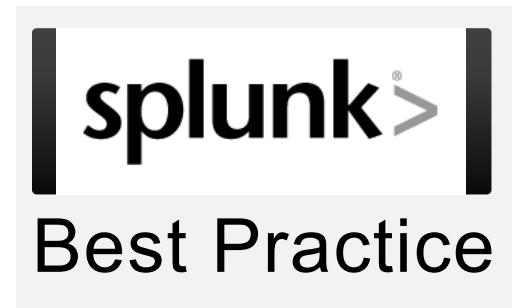


Example of Index-Time Precedence: Conflict



Configuration Best Practices

- Avoid storing configurations in **SPLUNK_HOME/etc/system/local**
 - Local context settings will *always* take precedence
 - Attempting to override index-time settings in an app will fail
 - Managing these settings with a deployment server is impossible
- Create an app to manage system settings
 - Allows you to manage settings with a deployment server
 - Manage system configurations in an app (e.g. **DC_app**) under **SPLUNK_HOME/etc/apps/<appname>/local**
 - Refer to the *Managing Forwarders* module



Validating the Splunk configuration

Validating the on-disk configuration

- Performed with **splunk btool** CLI
- Syntax: **splunk btool <conf_file> list**
- Example: **splunk btool inputs list**

Validating the in-memory configuration

- Performed with **splunk show config** CLI or REST API
- Syntax: **splunk show config <conf_file>**
- Example: **splunk show config inputs**

Reloading Configuration Files After Edit

- Changes made using Splunk Web or the CLI may not require restart
 - A message appears if restart is required (i.e. changing server settings)
- Changes made by editing **.conf** files are not automatically detected
- To force reload, go to **http://servername:webport/debug/refresh**
 - Reloads many of the configurations, including **inputs.conf**, but not all
- To reload all configurations, restart Splunk
 - Splunk Web: Settings > Server controls > Restart Splunk
 - CLI: **splunk restart**

Note



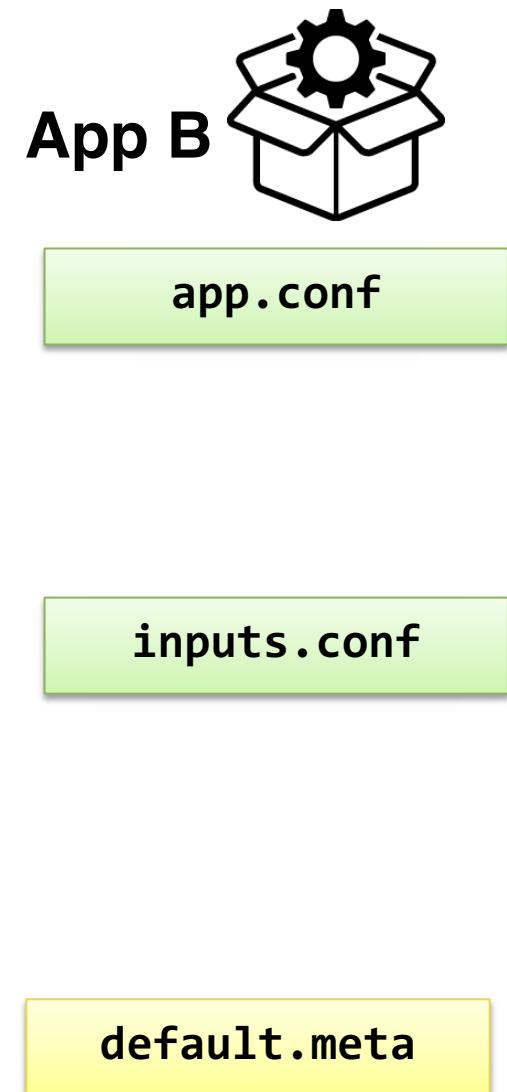
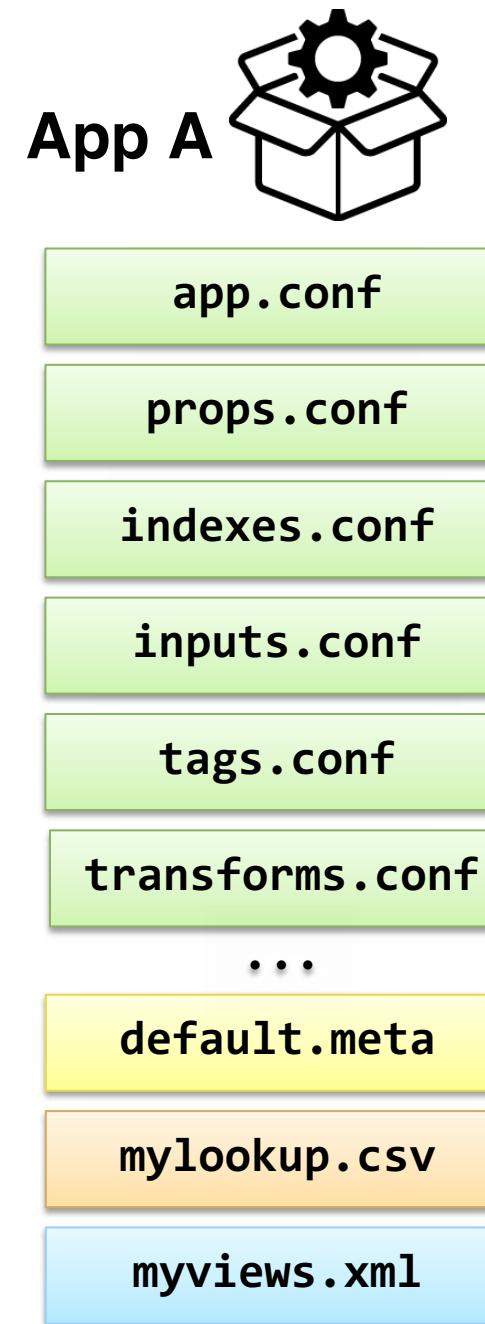
A Splunk refresh is only valid for standalone configuration or a search head.

What is an App?



Splunk App

- Collection of configuration files, scripts, web assets, and so on
- May be focused on specific type of data, vendor, OS, industry, or business need
- May be installed on any Splunk instance
- May be included with Splunk (as a default app)

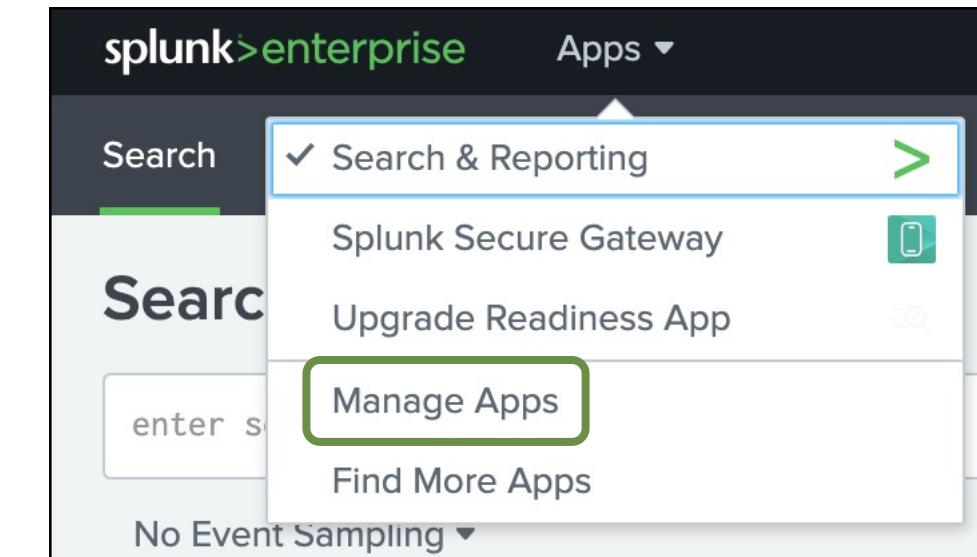
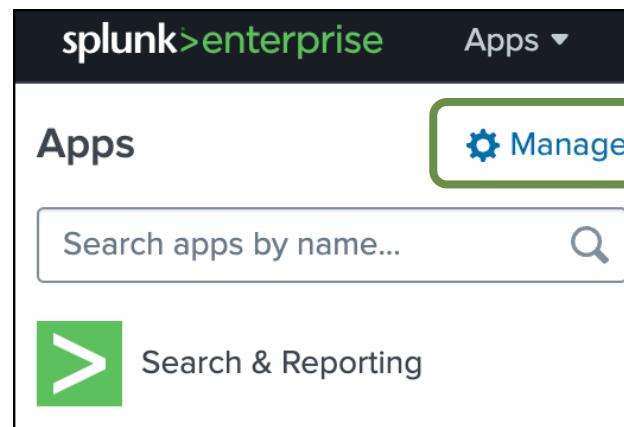


Viewing Installed Apps

- Apps are installed under **SPLUNK_HOME/etc/apps**
- Apps can be visible or hidden in Splunk Web
 - Several apps are installed by default
 - Internal apps used by Splunk should not be modified
- To manage apps in Splunk Web:

Within an app

On the Home view



Discovering Apps On Splunkbase

- Provides Splunk and community supplied apps and add-ons
 - Apps for Splunk Enterprise, Splunk Cloud, and Splunk SOAR
- Includes Splunk supported and developer supported apps

The screenshot shows the Splunkbase interface for discovering apps. At the top, there's a navigation bar with 'splunkbase' logo, 'Collections', 'Apps', a search bar ('Find an app'), and buttons for 'Submit an App' and 'Log In'. Below the navigation is a section titled 'Discover Apps'.

On the left, there are filtering options under 'PLATFORM':

- SPLUNK** (selected):
 - PRODUCT**:
 - Splunk Enterprise
 - Splunk Cloud
 - Splunk Enterprise Security
 - Splunk IT Service Intelligence
 - VERSION**:
 - 9.1

Below the filters, it says 'Showing 1-18 of 2084 Results' and 'Filtered by: Splunk > Product > Splunk Enterprise > Version > 9.1'. There are three app cards displayed:

- Splunk Add-on for Microsoft Windows** By Splunk Inc.
*** Important: Read upgrade instructions and test add-on update before deploying to production *** The Splunk Add-on f...
PLATFORM: Splunk Enterprise, Splunk Cloud, ...
RATING: ★★★★★ (40)
- Splunk Add-on for Unix and Linux** By Splunk Inc.
*** Important: Read upgrade Instructions and test add-on update before deploying to production *** There are changes to...
PLATFORM: Splunk Enterprise, Splunk Cloud, ...
RATING: ★★★★★ (48)
- Splunk Dashboard Examples** By Splunk Inc.
The Splunk Dashboard app delivers examples that give you a hands-on way to learn the basic concepts and tools...
PLATFORM: Splunk Enterprise, Splunk Cloud
RATING: ★★★★★ (71)

On the right, there's a 'Sort by' dropdown set to 'Popularity'.

Installing an App Using Splunk Web

The screenshot shows the Splunk Web interface for managing apps. At the top, there's a navigation bar with 'splunk>enterprise' and a dropdown menu for 'Administrator'. Below the navigation is a toolbar with buttons for 'Browse more apps' (highlighted with a green box), 'Install app from file' (also highlighted with a green box), and 'Create app'. A dropdown menu for '25 per page' is also visible. The main area is titled 'Apps' and shows a list of 25 items, with a 'filter' search bar above it. The list includes columns for 'Name', 'Folder name', 'Version', 'Update checking', 'Visible', 'Sharing', 'Status', and 'Actions'. Two arrows point down from the 'Install app from file' button to the 'SplunkForwarder' entry in the list, specifically pointing to the 'Actions' column where 'Permissions' and 'Enable' links are located.

Apps

Showing 1-25 of 25 items

filter

Name	Folder name	Version	Update checking	Visible	Sharing	Status	Actions
SplunkForwarder	SplunkForwarder	Yes	No	App	Permissions	Disabled	Enable

IT Essentials Work

IT Essentials Work helps you correlate logs and metrics for each entity, and then use that information to observe and understand the performance of your infrastructure. The app helps you get started monitoring and analyzing IT infrastructures such as *nix, Windows, virtualization with out-of-the-box dashboards, and pre-configured performance metric... [More](#)

Category: [IT Operations](#) | Author: [Splunk Inc.](#) | Downloads: 14891 | Released: 2 months ago | Last Updated: a month ago | [View on Splunkbase](#)

Splunk App for SOAR

The Splunk App for SOAR gets data from your Splunk SOAR instance for manipulation and display in Splunk. This app provides pre-built dashboards and enables you to use Splunk to power SOAR's search engine.

Install App From File

If you have a .spl or .tar.gz app file to install, you can upload it using this form.

You can replace an existing app via the Splunk CLI. [Learn more.](#)

File

No file chosen

Upgrade app. Checking this will overwrite the app if it already exists.

Apps on Forwarders

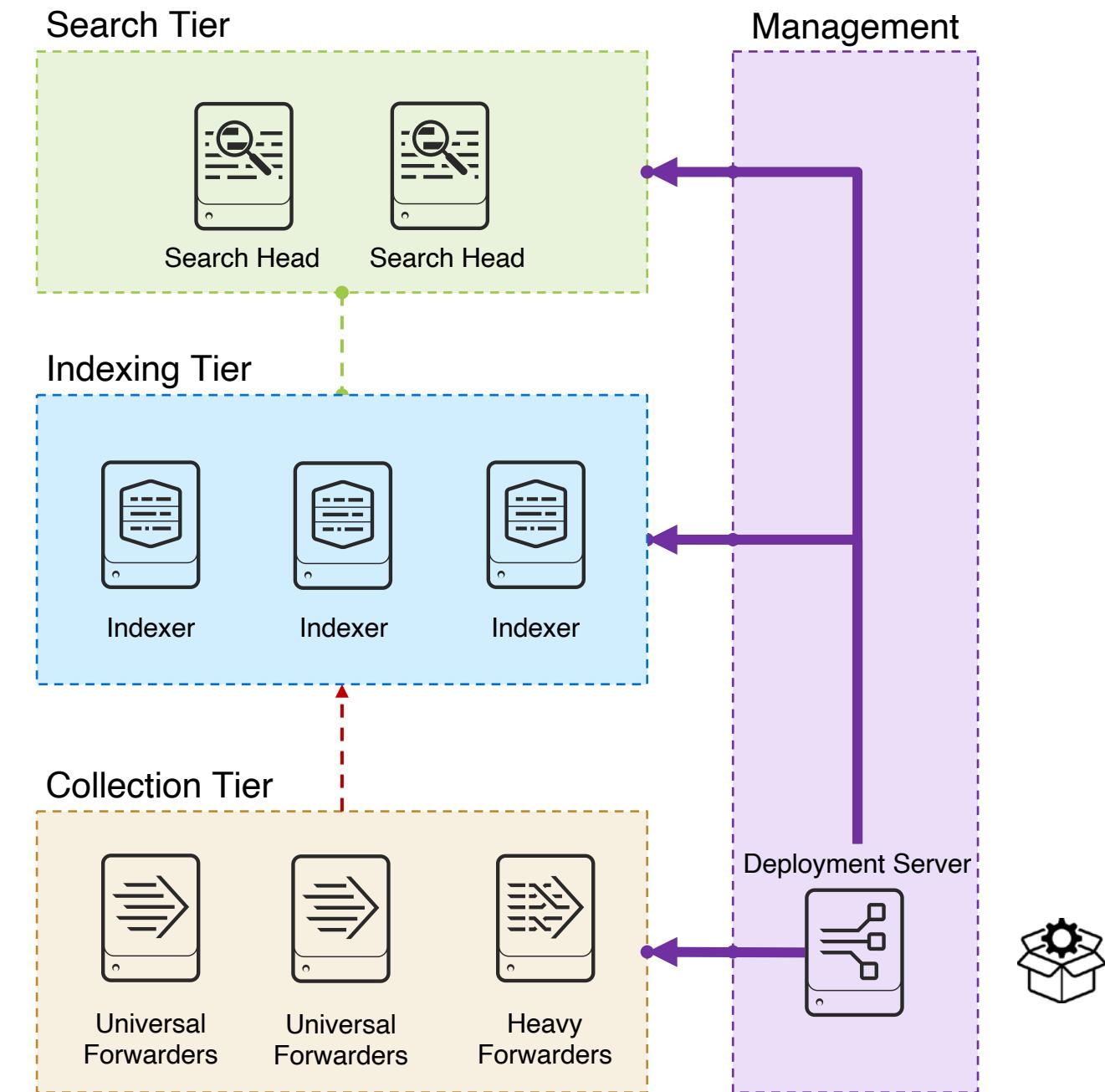
- Universal forwarders don't have a web interface, but may still benefit from an app
- Install the app on a forwarder using one of these methods:
 - Command line on the forwarder (**splunk install app**)
 - Extract app in proper location on the forwarder
 - Use a deployment server to deploy the app

Deploying Apps In Distributed Environments

- Without clustering, Deployment Server can deploy apps to all tiers
- Alternatively, deploy apps manually or using 3rd party software solutions

Note

Work with your Splunk administrator to properly deploy your app in a production environment. More information is provided in the *Forwarder Configuration and Management* course.

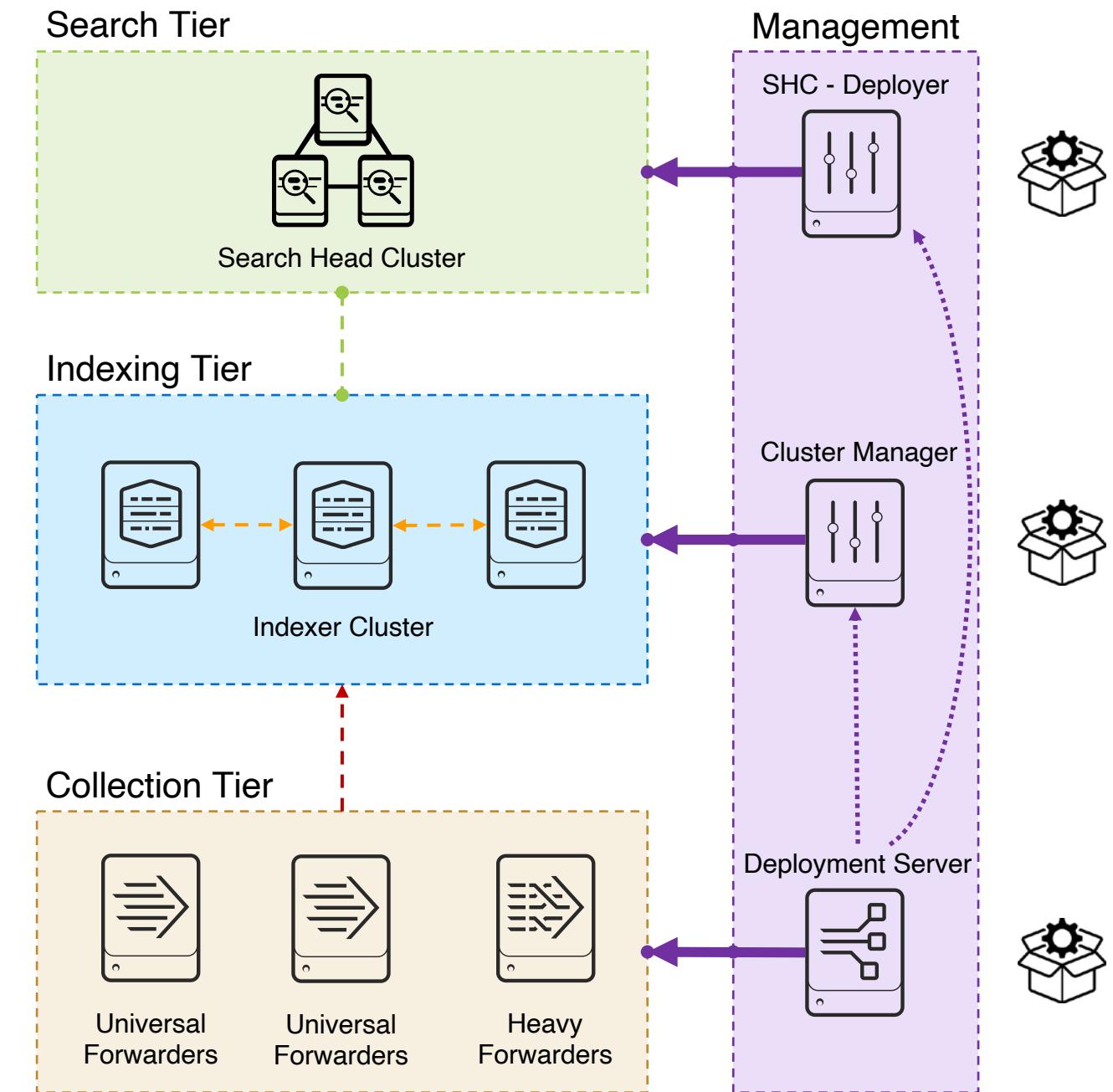


Deploying Apps In Clustered Environments

- With clustering use the deployment methods for the appropriate tier:
 - Deployer to search heads
 - Cluster Manager node to indexers
 - Deployment Server to forwarders
 - Deployment Server to Cluster Manager / Deployer (optional)

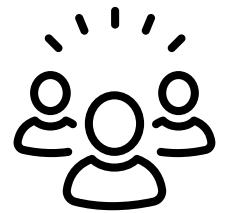
Note

More information is provided in the *Splunk Enterprise Cluster Administration* course.



Useful References

- Develop Splunk Apps
dev.splunk.com/enterprise/docs/developapps
- Splunk Web Framework
dev.splunk.com/enterprise/docs/developapps/visualizedata/usewebframework/
- Create a Splunk app and set properties
dev.splunk.com/enterprise/docs/developapps/createapps/
- Install Apps in Splunk Cloud
docs.splunk.com/Documentation/SplunkCloud/latest/User/SelfServiceAppInstall
- Documentation in **app.conf.spec**
docs.splunk.com/Documentation/Splunk/latest/admin/AppConf



Question: Configuration Files

Which configuration file tells a Splunk instance which data to ingest?

- A. inputs.conf
- B. outputs.conf
- C. server.conf
- D. props.conf

Answer: Configuration Files

Which configuration file tells a Splunk instance which data to ingest?

A. **inputs.conf**

B. **outputs.conf**

C. **server.conf**

D. **props.conf**

inputs.conf lists data inputs, or in some cases which network ports to listen on for data inputs.

outputs.conf lists where input data should be sent (for example from forwarders to indexers).

server.conf lists server settings, and **props.conf** lists processing properties for data input.

Question: App Context Location

If a data input is created in the Splunk Web interface with an App Context set to Search & Reporting (**search**), where will the **inputs.conf** configuration file be stored?

- A. SPLUNK_HOME/etc/apps/search/default
- B. SPLUNK_HOME/etc/apps/search/local
- C. SPLUNK_HOME/etc/system/default
- D. SPLUNK_HOME/etc/system/local

Answer: App Context Location

If a data input is created in the Splunk Web interface with an App Context set to Search & Reporting (**search**), where will the **inputs.conf** configuration file be stored?

- A. SPLUNK_HOME/etc/apps/search/default
- B. SPLUNK_HOME/etc/apps/search/local**
- C. SPLUNK_HOME/etc/system/default
- D. SPLUNK_HOME/etc/system/local

Question: The **btool** Utility

Which of the following about **btool** is true?

- A. btool is used to verify Splunk configuration in memory
- B. btool is used to verify Splunk configuration on disk
- C. btool is a Splunk Support utility you must download
- D. btool is unavailable on a Free license

Answer: The **btool** Utility

Which of the following about **btool** is true?

- A. **btool** is used to verify Splunk configuration in memory
- B. **btool** is used to verify Splunk configuration on disk**
- C. **btool** is a Splunk Support utility you must download
- D. **btool** is unavailable on a Free license

The **btool** command is used to verify Splunk configuration on disk. To verify Splunk configuration in memory, use the **splunk show config** command or REST API. It is available by default for all Splunk Enterprise installation.

Module 2 Lab

Time: 15 minutes

Description: Config Files and Apps

Tasks:

- Explore Splunk configuration files
- Use the **btool** and the **splunk show config** commands to investigate configuration files
- Explore Splunk apps on Splunkbase
- Explore Splunk **apps** and **deployment-apps** directories

Module 3: Configuring Forwarders

Module Objectives

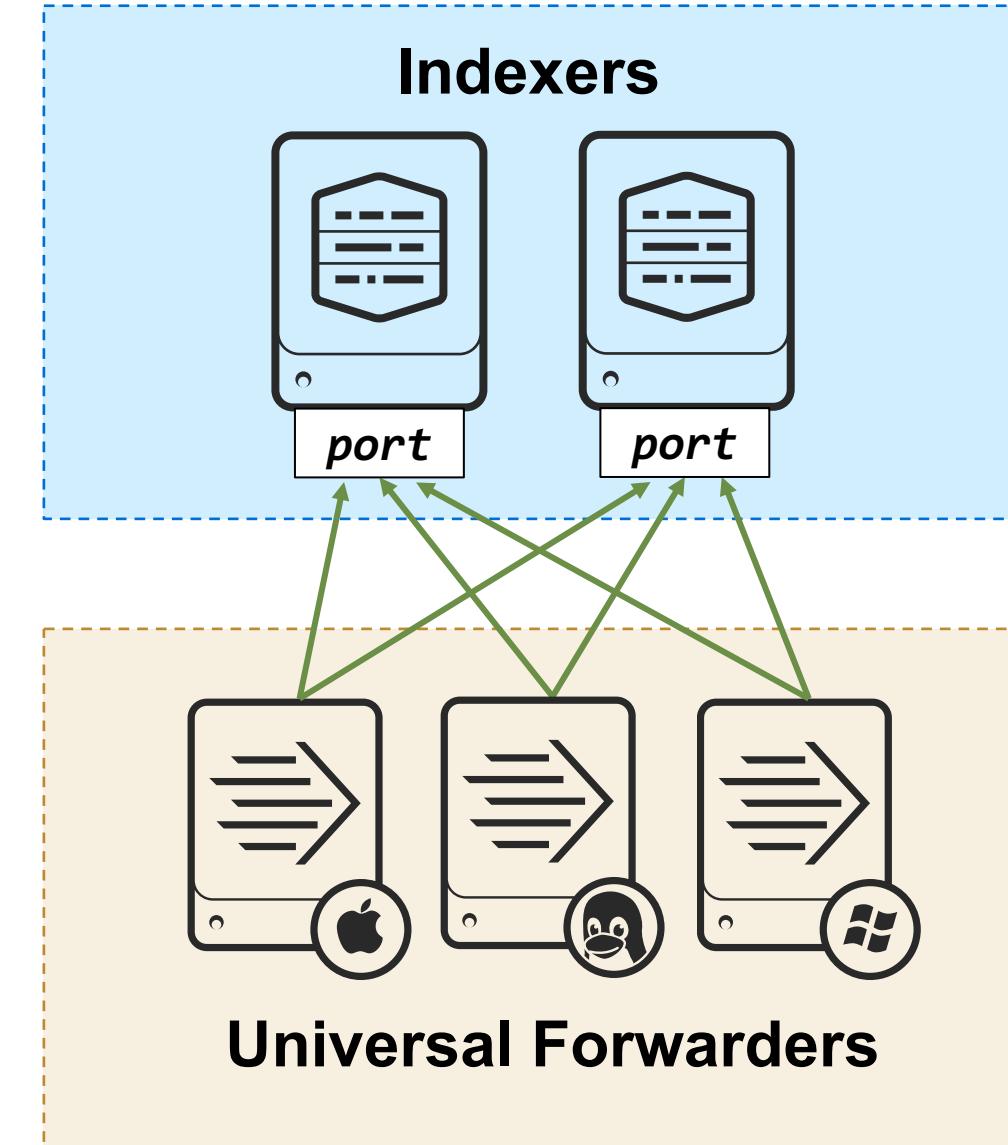
- Configure Universal Forwarders
- Configure Heavy Forwarders

Understanding Universal Forwarders



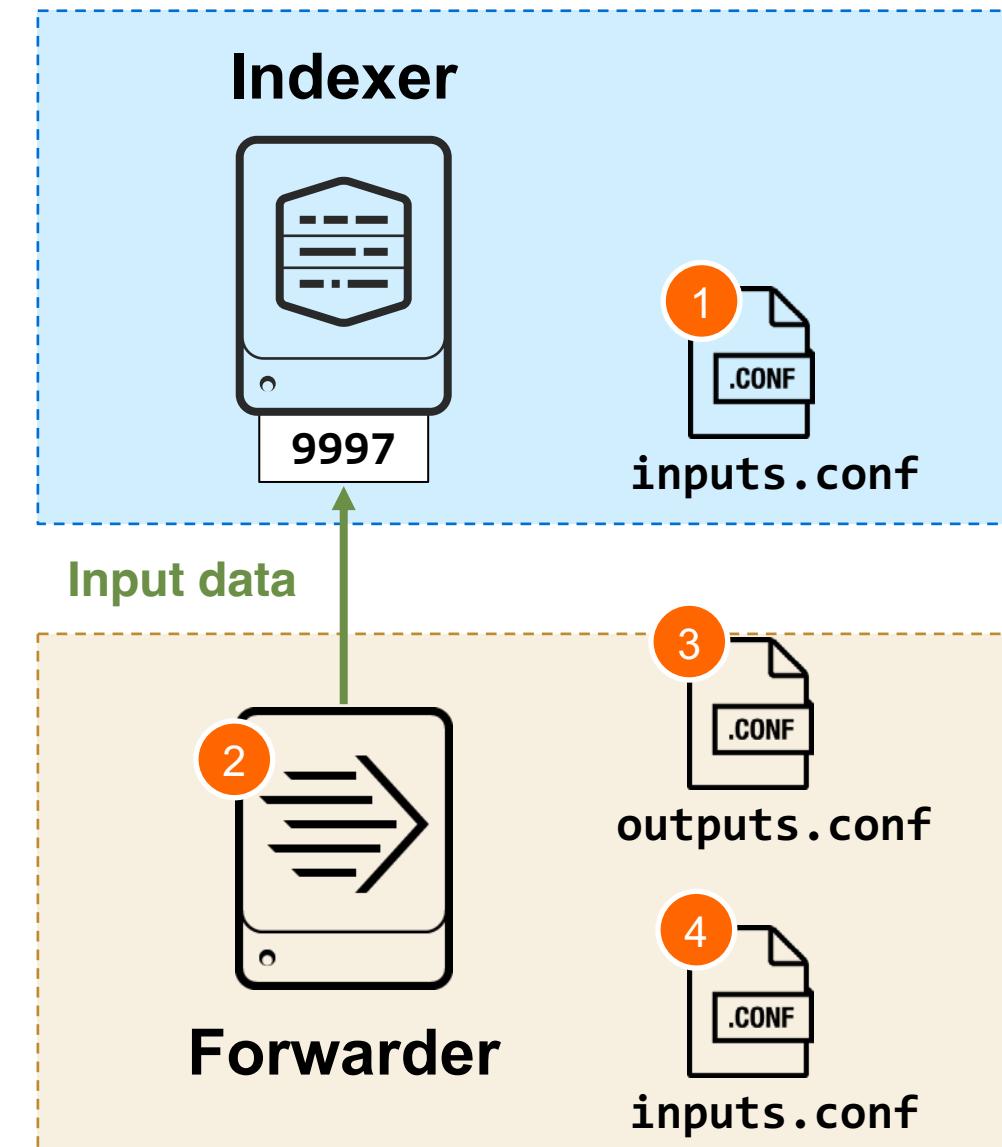
Universal Forwarders (UF)

- Gathers data from a host
- Sends data over the network to receiving ports on receivers (usually an indexer)
- Provided as separate installation binary with a built-in license (no limits)
- Designed to run on production servers
(minimal CPU / memory use, bandwidth constrained to 256 KBps by default, no web interface, cannot search or index)



Universal Forwarder Configuration Steps

1. Configure a receiving port on each indexer (one-time task)
2. Download and install Universal Forwarder
3. Configure forwarding on forwarders
4. Add inputs on forwarders



1. Configure Receiving Port on Each Receiver

- Using Splunk Web:
 1. Select Settings > Forwarding and receiving
 2. Next to Configure receiving, select Add new
 3. Enter a port number and click Save
 - Stored in most recently visited app:
SPLUNK_HOME/etc/apps/<app>/local
- Using CLI:
 - Run **splunk enable listen <port>**
 - Stored in **SPLUNK_HOME/etc/apps/search/local**
- Manually in **inputs.conf** as:
[splunktcp://port]

The screenshot shows the 'Forwarding and receiving' settings page. The main menu has two sections: 'Forward data' and 'Receive data'. Under 'Receive data', there is a 'Configure receiving' link. A green box highlights this link, and a green arrow points from it to the 'Listen on this port' input field in the 'Configure receiving' section below.

Forwarding and receiving

Forward data
Set up forwarding between two or more Splunk instances.

Forwarding defaults

Configure forwarding + Add new

Receive data
Configure this instance to receive data forwarded from other instances.

Configure receiving + Add new

Configure receiving

Set up this Splunk instance to receive data from forwarder(s).

Listen on this port * For example, 9997 will receive data on TCP port 9997.

2. Installing a Universal Forwarder

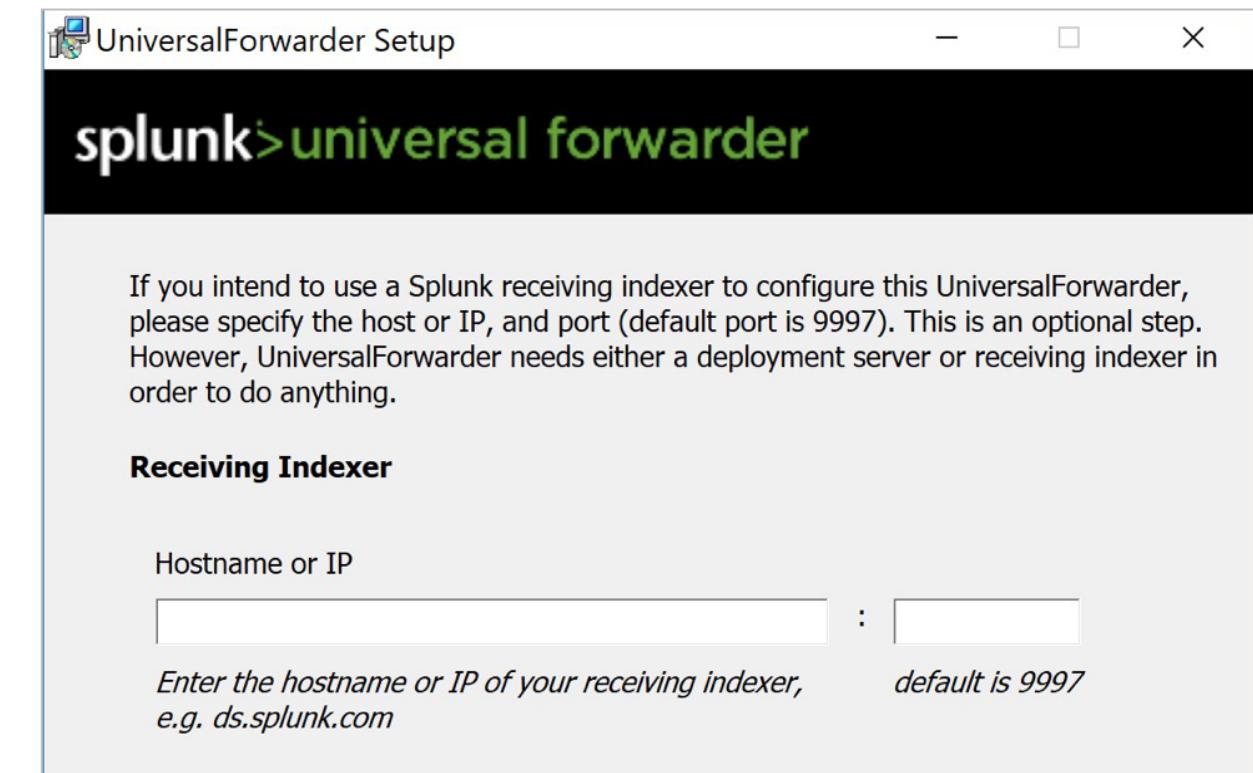
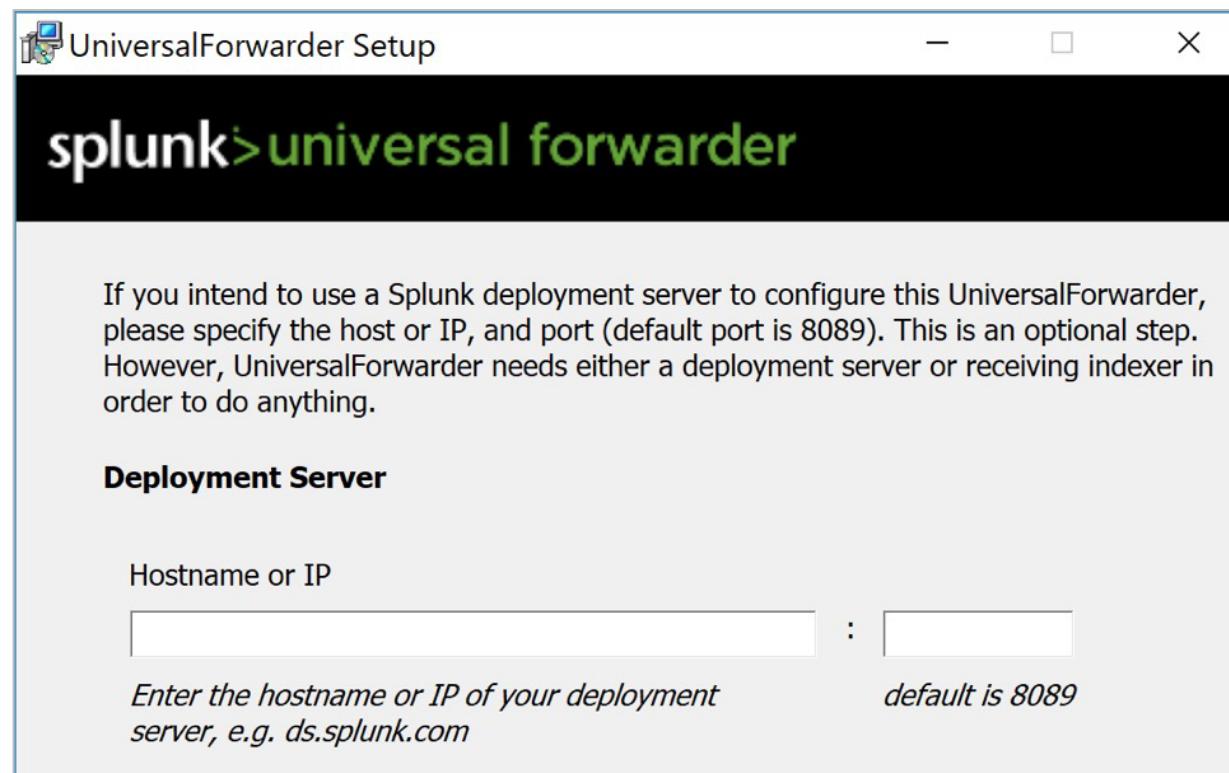
	*NIX	Windows
<i>Download</i>	www.splunk.com/en_us/download/universal-forwarder.html	
<i>Install</i>	<ul style="list-style-type: none">• Un-compress .tgz, .rpm, or .deb file in the path Splunk will run from• Default SPLUNK_HOME is: /opt/splunkforwarder	<ul style="list-style-type: none">• Execute .msi installer (or use the CLI)• Default SPLUNK_HOME is: C:\Program Files\SplunkUniversalForwarder

- Silent installation methods exist on all platforms
- Same **splunk** command-line interface in **SPLUNK_HOME/bin**
 - Same commands for start/stop, restart, etc.
 - An admin account and password are required

Using the Interactive Windows Installer

- Most forwarder settings can be configured using the installer wizard
 - Can run as a local or domain user without local administrator privileges
- CLI installation is available for scripted installations:

https://docs.splunk.com/Documentation/Forwarder/latest/Forwarder/Installawindowsuniversalforwarderfromaninstaller#Install_a_Windows_universal_forwarder_from_the_command_line



3. Configure Forwarding on Forwarders

- To configure target indexers on forwarders, either:
 - Run **splunk add forward-server <indexer:receiving_port>**
 - Modify **outputs.conf**
- Splunk logs are automatically sent to indexer's **_internal** index
- Example: **splunk add forward-server 10.1.2.3:9997f** configures **outputs.conf** as:

```
[tcpout]
defaultGroup = default-autolb-group

[tcpout-server://10.1.2.3:9997]

[tcpout:default-autolb-group]
disabled = false
server = 10.1.2.3:9997
```

docs.splunk.com/Documentation/Forwarder/latest/Forwarder/Configureforwardingwithoutputs.conf

4. Add inputs on forwarders

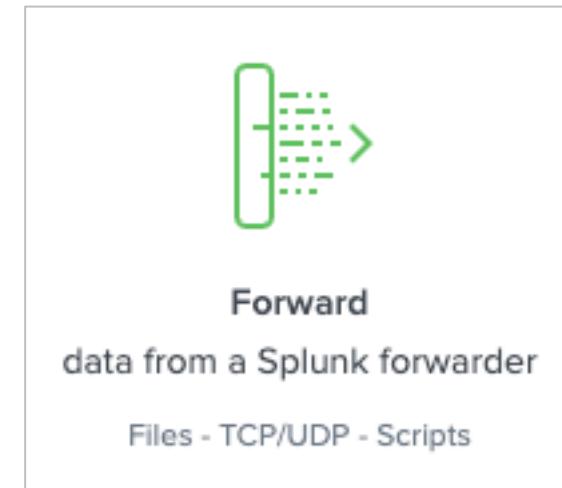
- Use Splunk CLI

```
./splunk add monitor /opt/log/www2/access.log -sourcetype access_combined_wcookie -index test
```

- Create Splunk config file (**inputs.conf**)

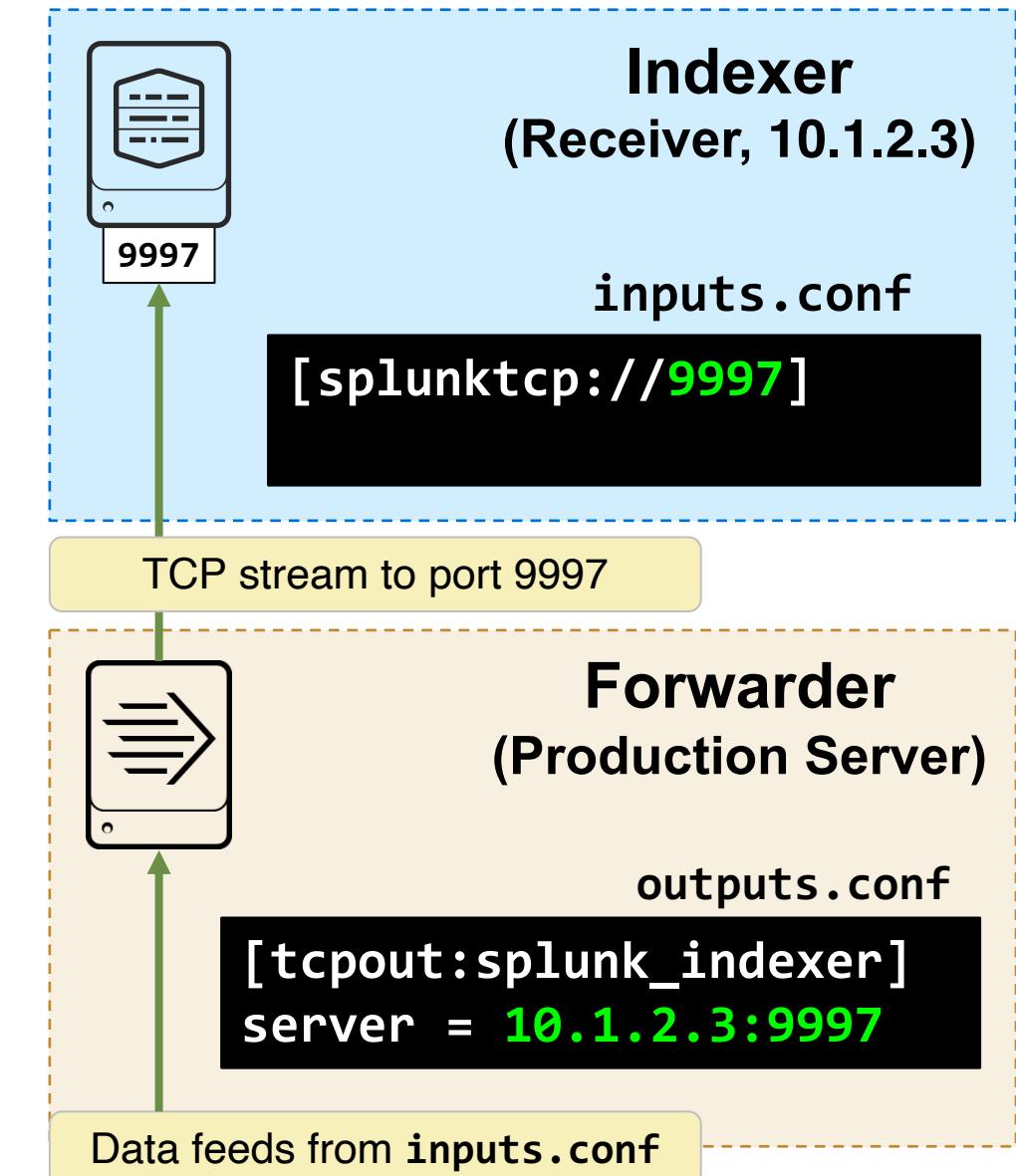
```
[monitor:///opt/log/www2/access.log]
disabled = false
index = test
sourcetype = access_combined_wcookie
```

- Use Spunk Web to Add Data > Forward
 - Performed on a Deployment Server
 - Requires deployment clients
 - Requires DS configurations such as server class
 - Discussed in later modules



Forwarder outputs.conf File

- Points the forwarder to the receivers
 - **splunktcp** stanza sets the indexer to listen on a port for feeds from Splunk forwarders
 - **server** sets a forwarder's destination to one or more receivers (IP or DNS name + receiver port), separated by commas
- Can specify additional options:
 - Load balancing
 - SSL
 - Compression
 - Alternate indexers
 - Indexer acknowledgement



Configuration Validation and Troubleshooting

- To verify the configuration:
 - On forwarder, run: **splunk list forward-server**
 - On indexer, run: **splunk display listen**
- To verify successful connection:
 - On search head, search: **index=_internal host=<forwarder_hostname>**
- Troubleshooting forwarder connection
 - Check **SPLUNK_HOME/var/log/splunk/splunkd.log** on forwarder:
tail -f splunkd.log | egrep 'TcpOutputProc|TcpOutputFd'

Selectively Forwarding Data to Indexers

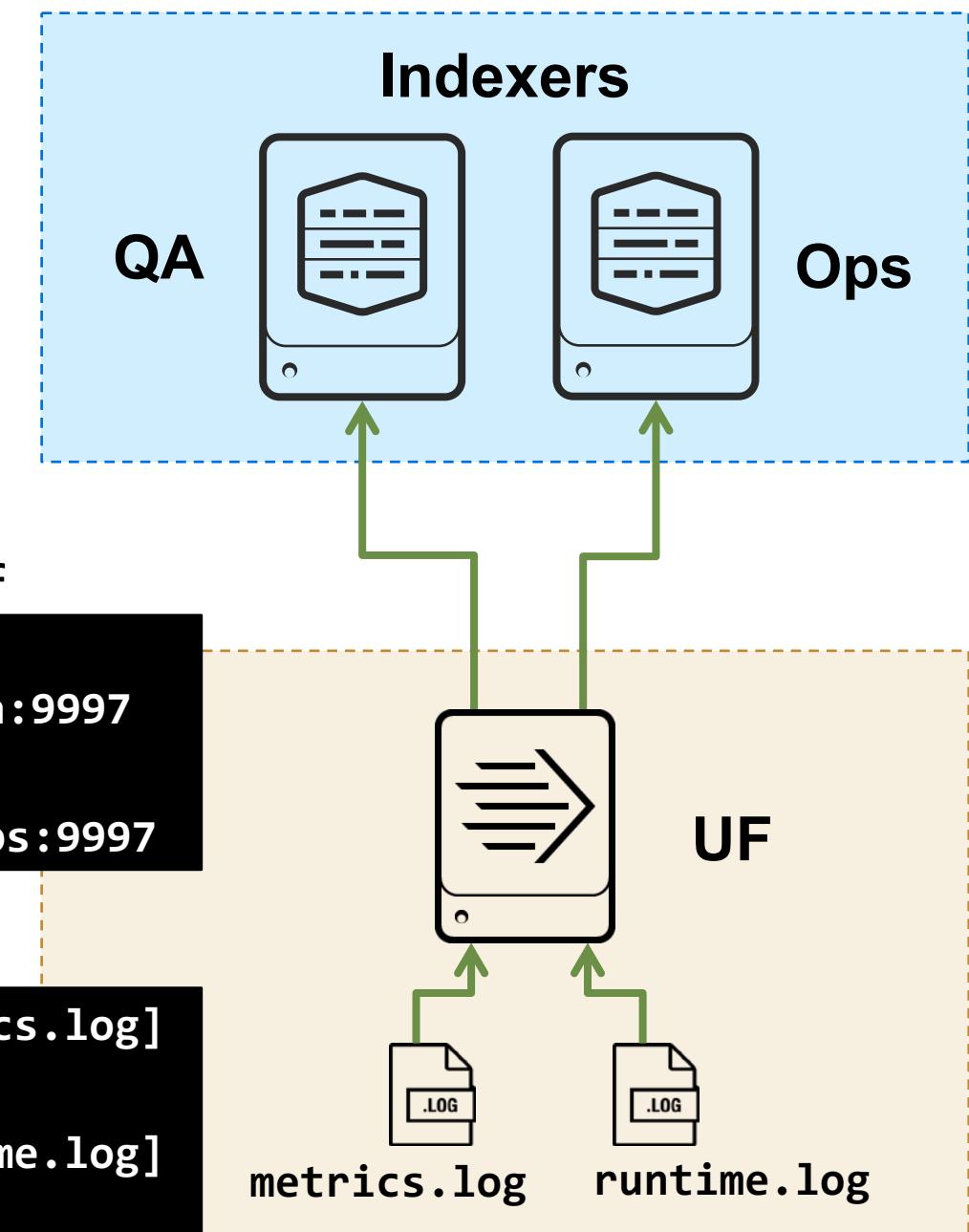
- Universal forwarder can route based on sources
- Example:
 - metrics.log** → QA indexer
 - runtime.log** → Ops indexer

Define multiple **tcpout** stanzas in **outputs.conf**

Specify **_TCP_ROUTING** for each source in **inputs.conf**

```
outputs.conf
[tcpout:QA]
server=srv.qa:9997
[tcpout:Ops]
server=srv.ops:9997

inputs.conf
[monitor://.../metrics.log]
_TCP_ROUTING = QA
[monitor://.../runtime.log]
_TCP_ROUTING = Ops
```

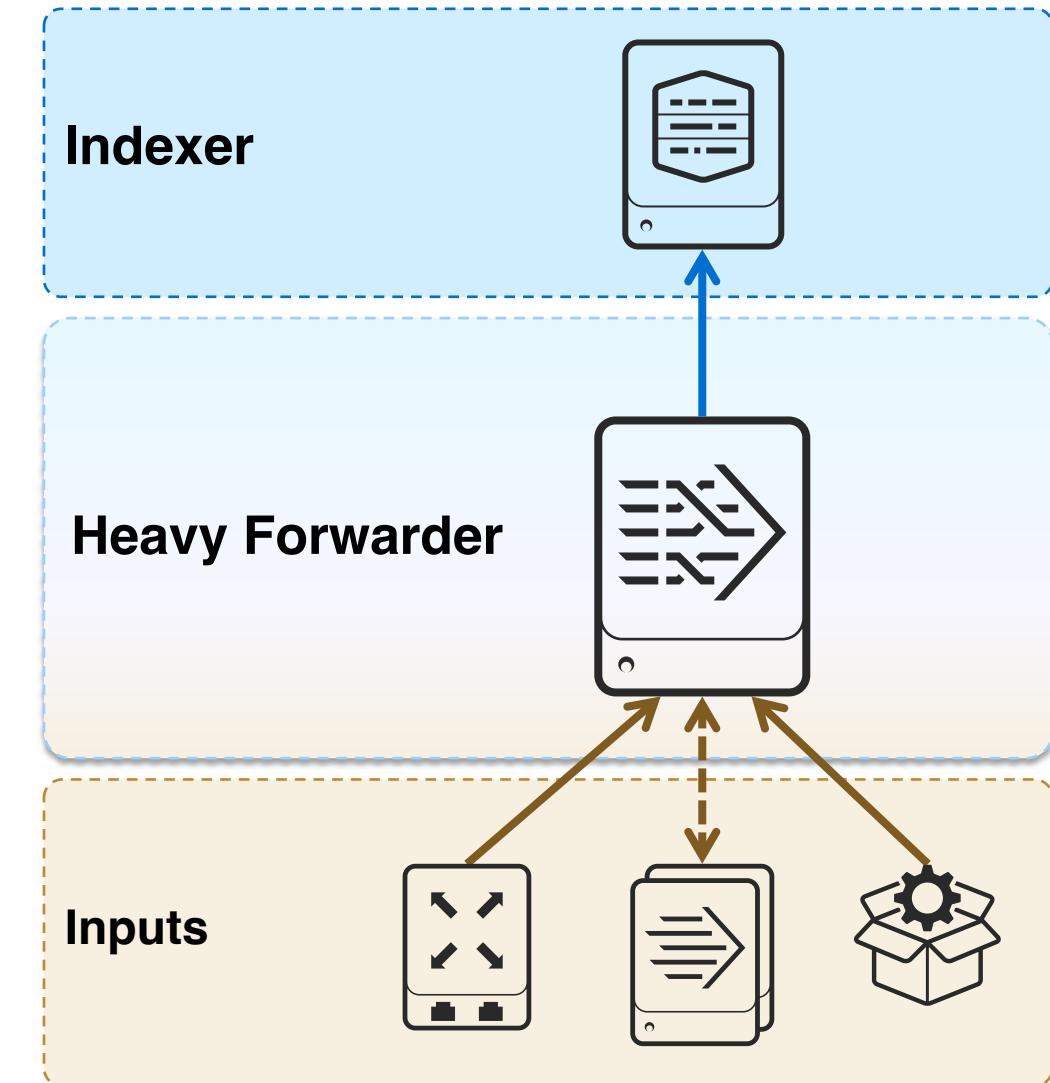


Understanding Heavy Forwarders (HF)



Heavy Forwarders (HF)

- Splunk Enterprise instance with the Forwarder License enabled
- Can parse data before forwarding it
- Can route data based on event criteria to different indexers or 3rd party receivers
- Supports some complex requirements
- Cannot perform distributed searches



Deciding Between UF and HF



Universal Forwarder

vs.



Heavy Forwarder

- Ideal for most circumstances, including collecting files or as intermediate forwarder
- Minimal footprint on production servers
- Generally, requires less bandwidth and has faster processing than same data on HF
- Supports simple routing or cloning data to separate indexers
- Does not support filtering based on regular expressions

- Generally runs on dedicated servers
- Required by some apps, add-ons, or input types (such as HEC, DBconnect)
- Supports complex, event-level routing
- Can anonymize or mask data before forwarding to an indexer
- Provides Splunk Web and predictable version of Python, if needed
- May increase network traffic

Optimizing the Heavy Forwarder

- Based on your use case
- Disable indexing data on the HF:

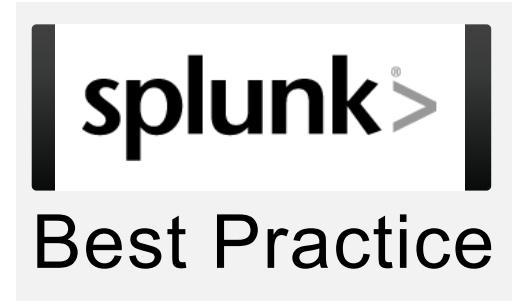
outputs.conf

```
[indexAndForward]  
index = false
```

- Disable Splunk Web on the HF:

web.conf

```
[settings]  
startwebserver = 0
```



Forwarding Resources

- Overview of forwarders

docs.splunk.com/Documentation/Splunk/latest/Data/Usingforwardingagents

- Forwarder deployment overview

docs.splunk.com/Documentation/Splunk/latest/Forwarding/Aboutforwardingandreceivingdata

- Splunk Blog: Universal or Heavy, that is the question?

www.splunk.com/en_us/blog/tips-and-tricks/universal-or-heavy-that-is-the-question.html

Useful Commands

Command	Operation
From the Forwarder:	
splunk add forward-server	Configures the forwarder to send data to the receiver
splunk list forward-server	Displays the current receivers
splunk remove forward-server	Removes the receiver from the forwarder
From the Receiver:	
splunk enable listen	Configures the Splunk receiving port number
splunk display listen	Displays the current Splunk receiving port number

Question: Describing UFs

Which of the following statements does NOT describe Universal Forwarders?

- A. Gathers data from a host
- B. Sends data over the network to receiving ports on receivers
- C. Provided as a separate install binary from Splunk Enterprise
- D. Must be manually licensed

Answer: Describing UFs

Which of the following statements does NOT describe Universal Forwarders?

- A. Gathers data from a host
- B. Sends data over the network to receiving ports on receivers
- C. Provided as a separate install binary from Splunk Enterprise
- D. Must be manually licensed**

Universal Forwarder comes with a built-in license.

Question: Defining Forwarder Destination

What configuration file on the forwarder defines where data is forwarded to?

- A. `server.conf`
- B. `indexes.conf`
- C. `inputs.conf`
- D. `outputs.conf`

Answer: Defining Forwarder Destination

What configuration file on the forwarder defines where data is forwarded to?

- A. `server.conf`
- B. `indexes.conf`
- C. `inputs.conf`
- D. `outputs.conf`

The **outputs.conf** file defines the forward destination. Example **outputs.conf** entries:

```
[tcpout:splunk_indexer]
server = 10.1.2.3:9997
```

Question: UF vs HF

Which of the following statements is true about Universal Forwarders (as compared with Heavy Forwarders)?

- A. Required by some apps or input types (such as HEC, DBconnect)
- B. Supports complex, event-level routing
- C. Provides Splunk Web, if needed
- D. Has a lighter CPU impact on the production server

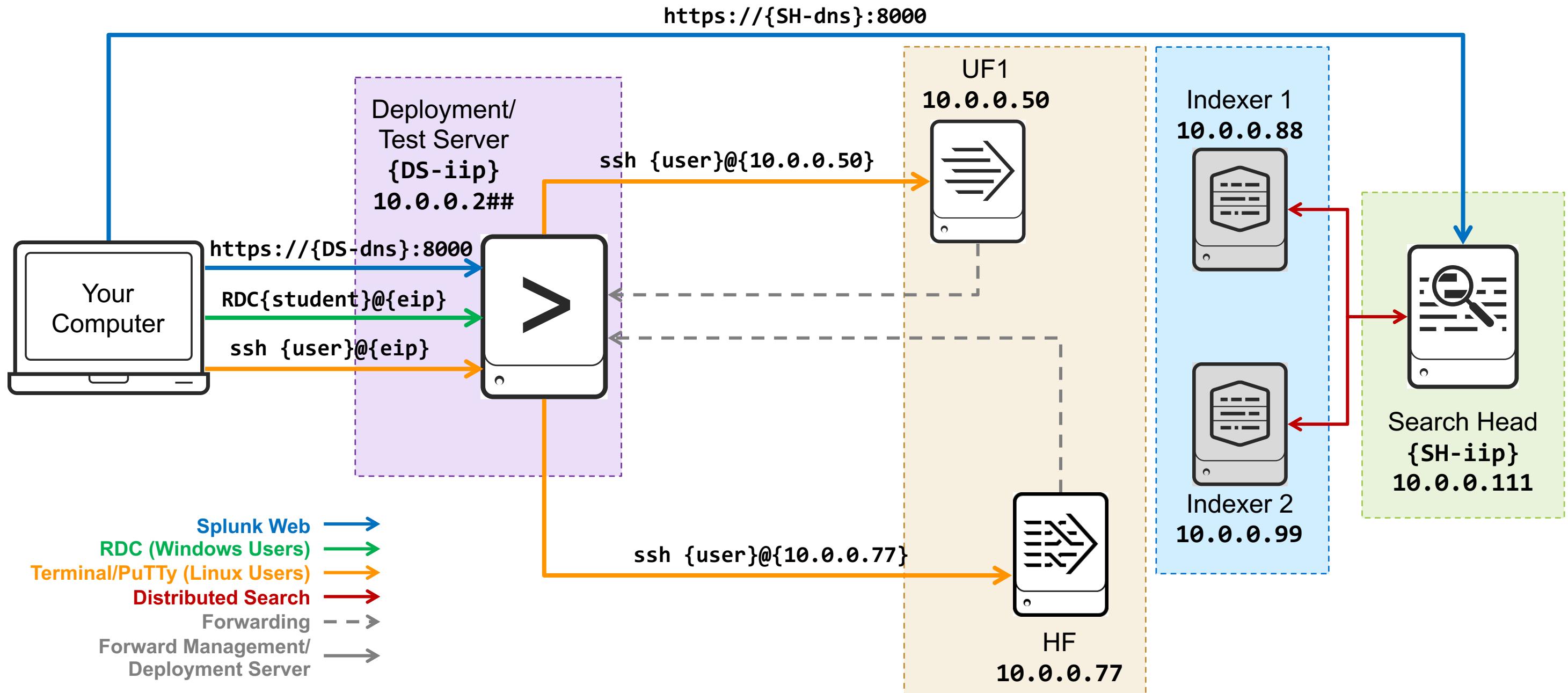
Answer: UF vs HF

Which of the following statements is true about Universal Forwarders (as compared with Heavy Forwarders)?

- A. Required by some apps or input types (such as HEC, DBconnect)
- B. Supports complex, event-level routing
- C. Provides Splunk Web, if needed
- D. Has a lighter CPU impact on the production server**

HFs can be used in cases where special requirements are needed, however one of the greatest advantages of UFs is that they provide a lighter impact on production servers.

Module 3 Lab – Environment Diagram



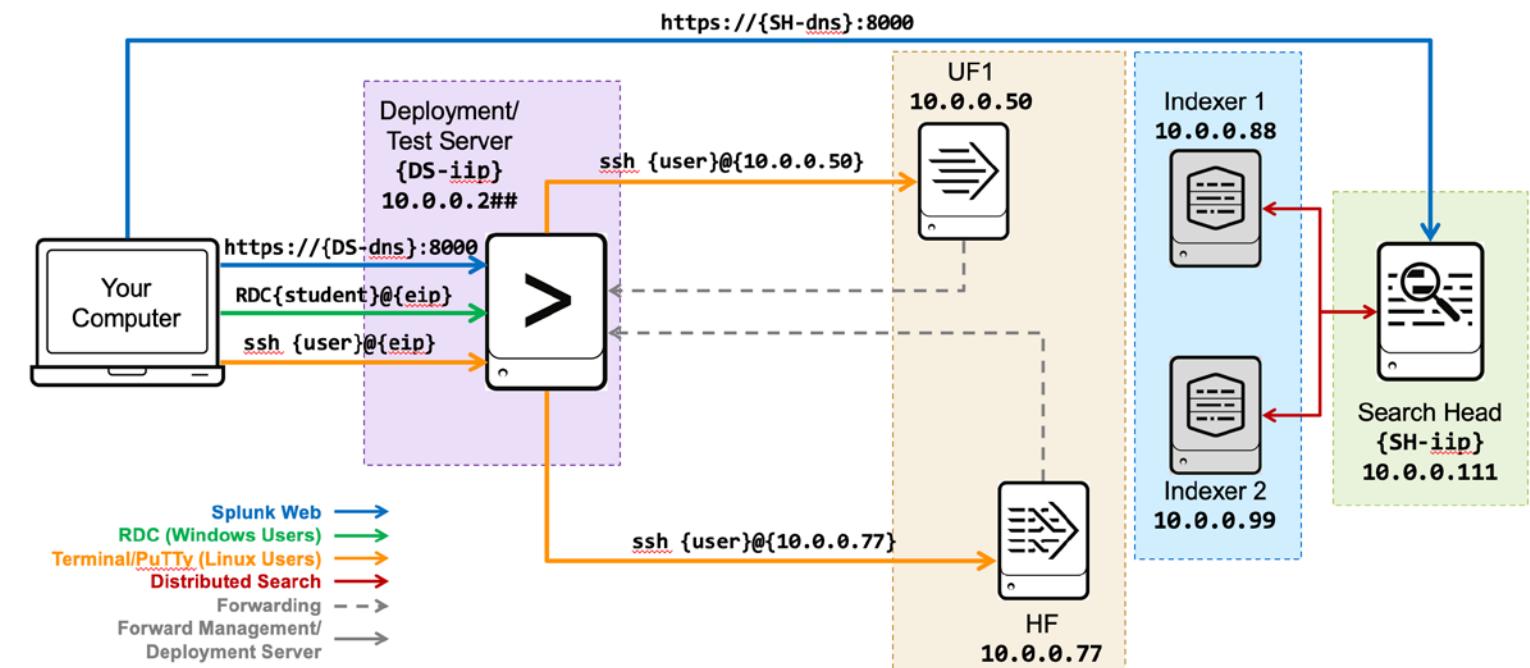
Module 3 Lab

Time: 20-25 minutes

Description: Configuring Forwarders

Tasks:

- Configure forwarders UF1 and HF to send data to Deployment/Test Server (**10.0.0.2##**)
- Add a monitor input on UF1
- Confirm forwarder connection from Deployment/Test Server

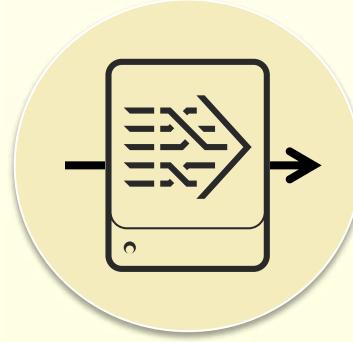


Module 4: Customizing Forwarders

Module Objectives

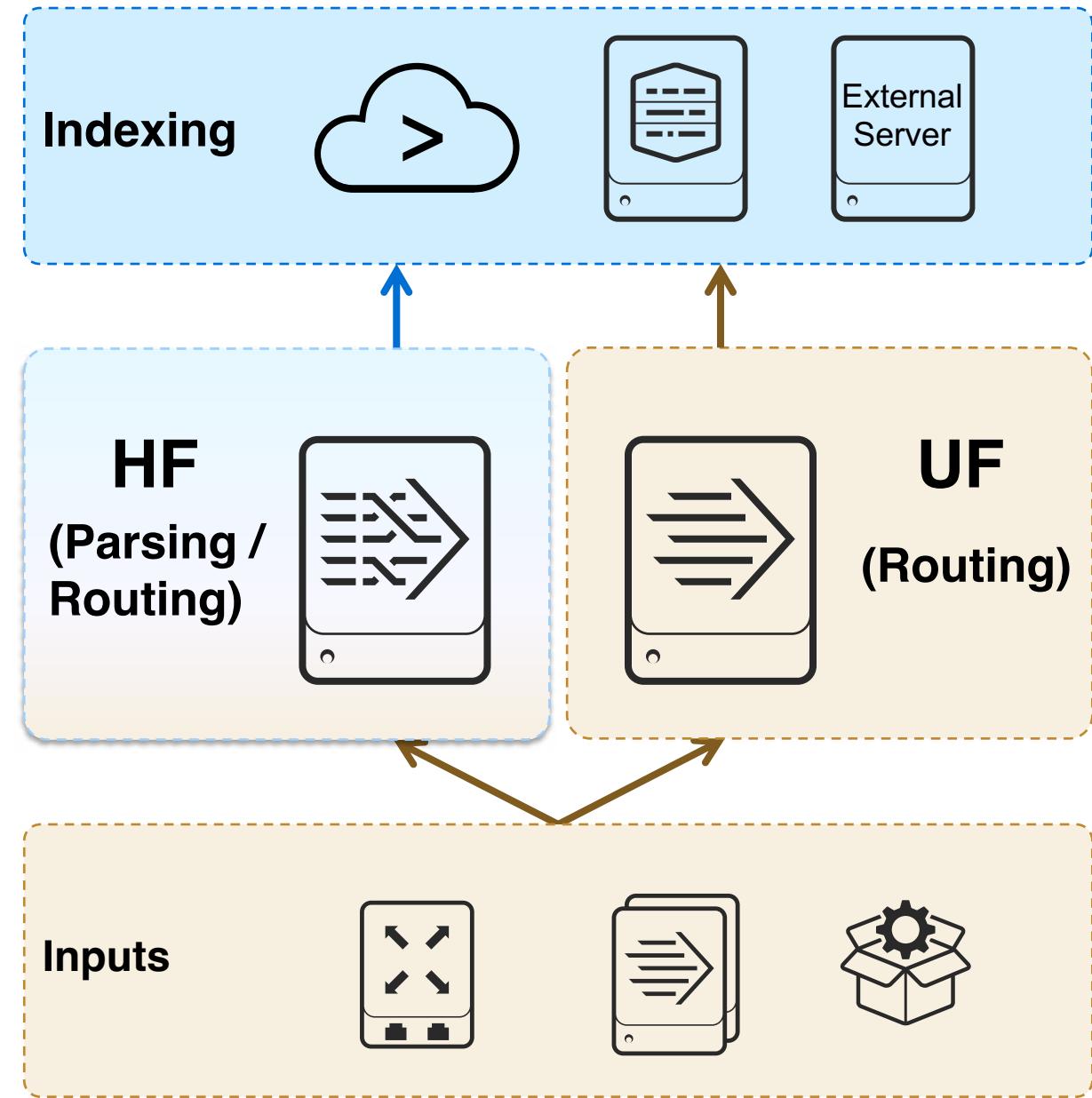
- Configure intermediate forwarders
- Identify additional forwarder options

Understanding Intermediate Forwarders



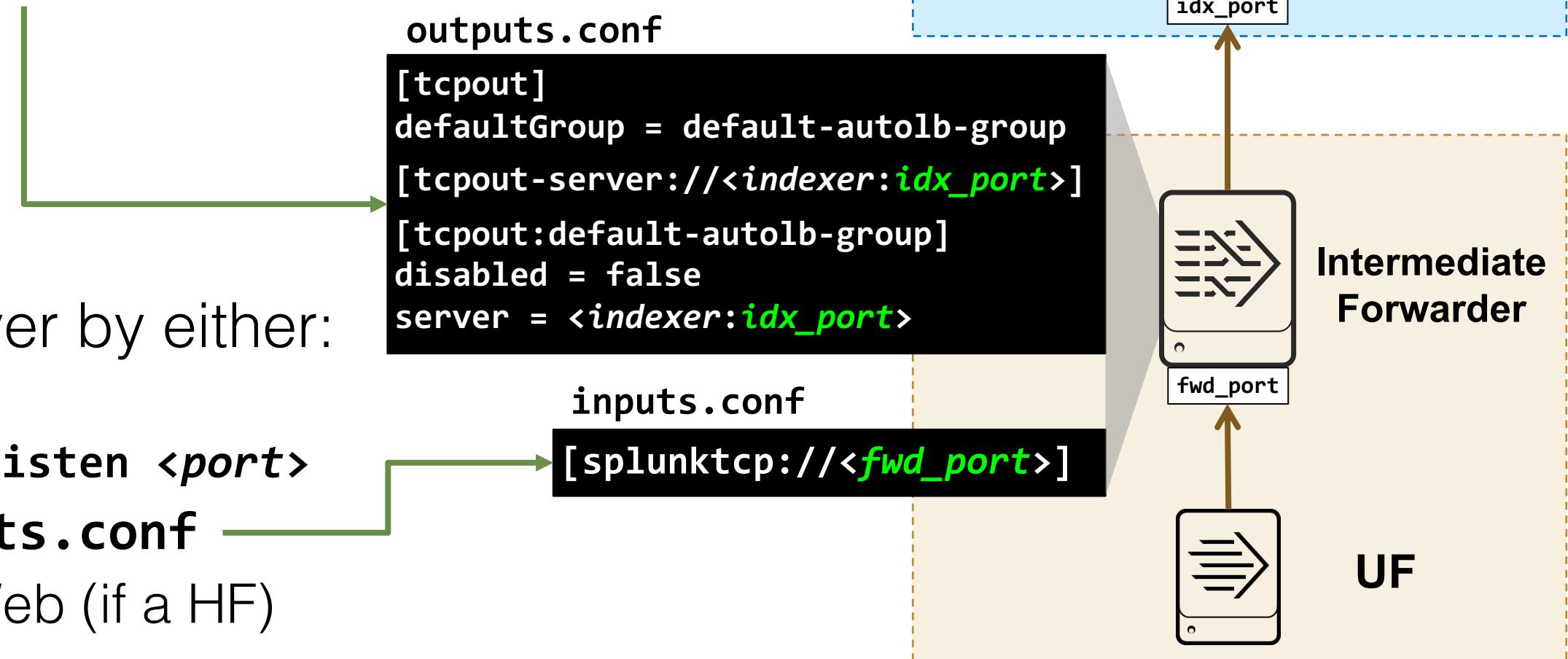
Intermediate Forwarders

- Generally Heavy Forwarders
- Route data from inputs to indexers or other intermediate forwarders
- Can reduce or limit bandwidth on specific network segments
- Can limit security concerns (DMZ, firewalls)
- Can parse, filter or index data if a HF



Configuring an Intermediate Forwarder

- Configure forwarding by either:
 - Running:
`splunk add forward-server:<idx:port>`
 - Modifying **outputs.conf**



- Configure receiver by either:
 - Running:
`splunk enable listen <port>`
 - Modifying **inputs.conf**
 - Using Splunk Web (if a HF)

Additional Forwarding Options



Compressing the feed



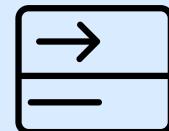
Securing the feed



Automatic load balancing to multiple indexers



Indexer acknowledgement to forwarder



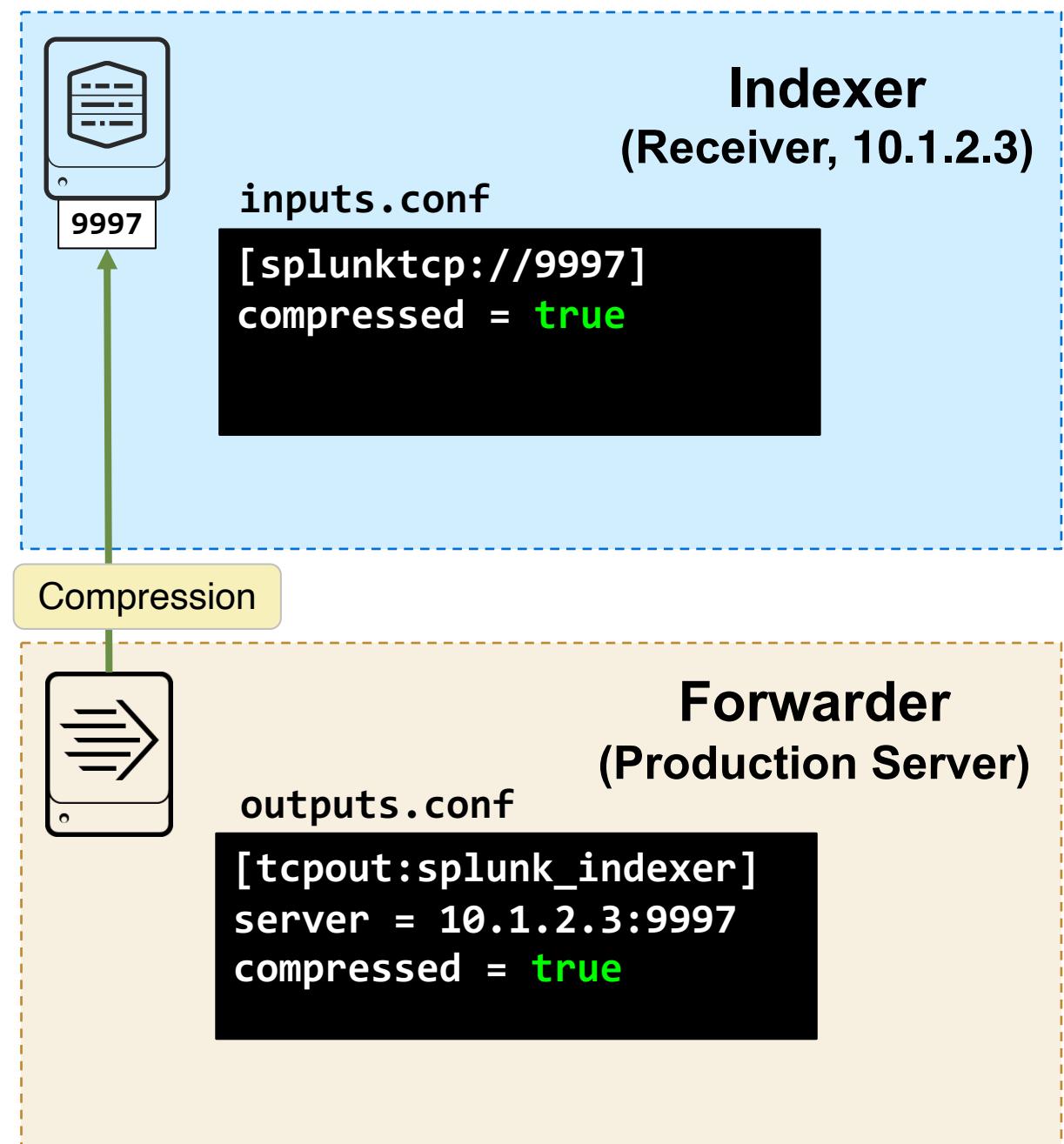
Forwarder queue size



Send the feed over HTTP

Compressing the Feed

- Reduces network utilization
- Increases CPU utilization slightly
- Set either at the forwarder or the indexer
 - Compress select feeds by setting on the forwarder
 - Compress all feeds by setting on the indexer



Securing the Feed with SSL

- Encrypts the feed
- Automatically compresses the feed
- Increases CPU utilization
- Requires the use of certificates
 - To configure with default root certificates:

- On a *nix indexer:

[sslConfig]

sslRootCAPath = SPLUNK_HOME/etc/auth/cacert.pem

- On a Windows indexer: Nothing required
- On a *nix forwarder:

[sslConfig]

sslRootCAPath = SPLUNK_HOME/etc/auth/cacert.pem

- On a Windows forwarder:

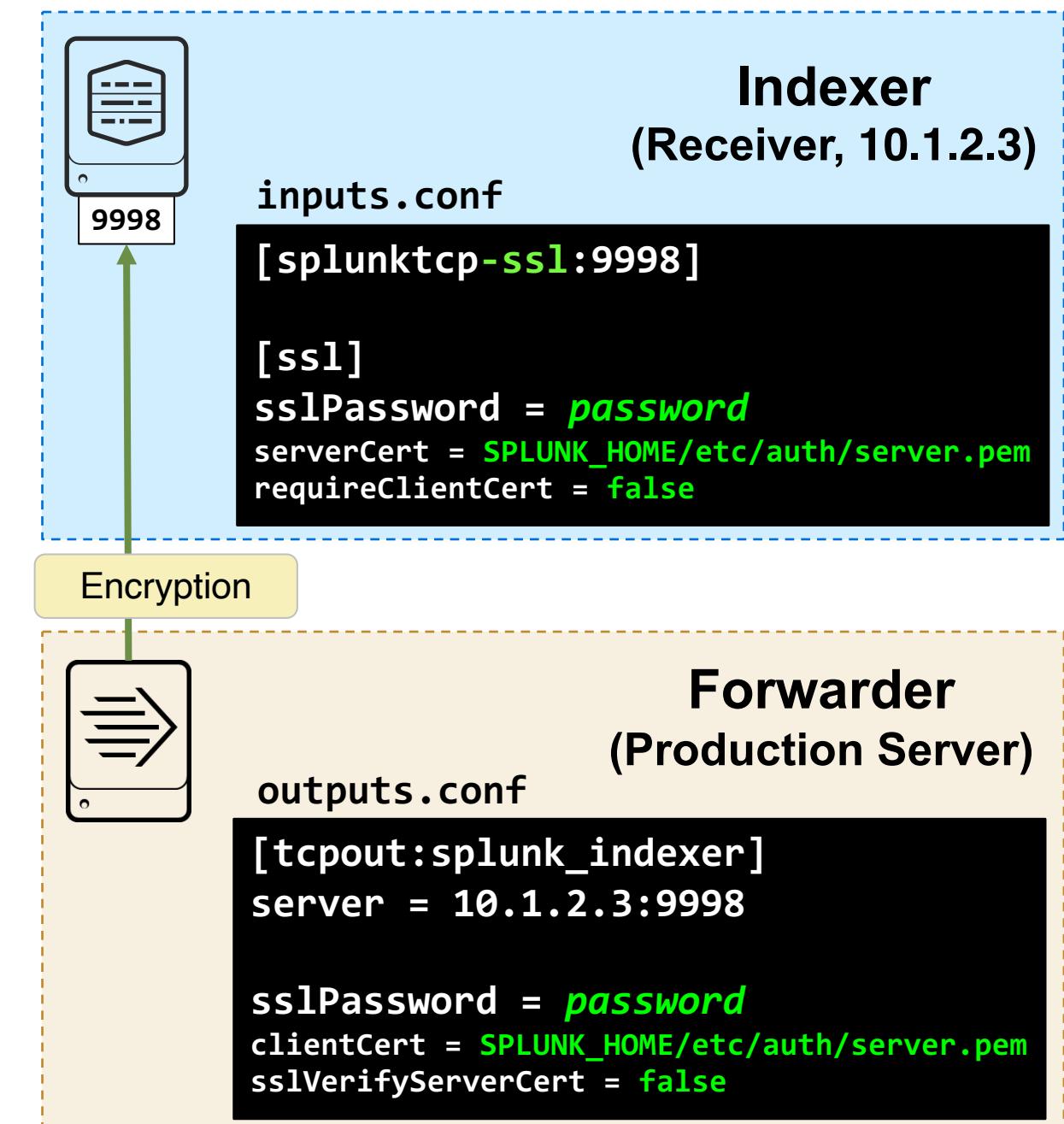
[sslConfig]

**caCertFile = cacert.pem
caPath = SPLUNK_HOME\etc\auth**

server.conf

server.conf

server.conf



Notes About SSL

- Splunk uses OpenSSL to generate its default certificates
 - Default certificate password is **password**
- Use external certs *or* create new ones using Splunk's OpenSSL
- Refer to:

docs.splunk.com/Documentation/Splunk/latest/Security/AboutsecuringyourSplunkconfigurationwithSSL

docs.splunk.com/Documentation/Splunk/8.2.9/Security/Aboutsecuringdatafromforwarders

docs.splunk.com/Documentation/Splunk/8.2.9/Security/ConfigureSplunkforwardingtousesthedefaultcertificate

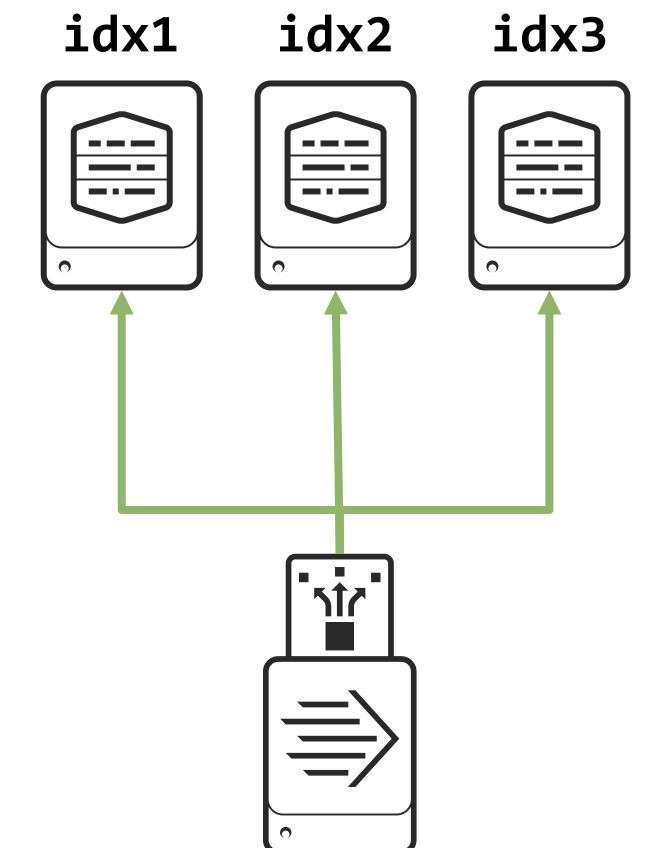
[docs.splunk.com/Documentation/Splunk/latest/Security/ConfigureSplunkforwardingtusesignedcertificates](https://docs.splunk.com/Documentation/Splunk/latest/Security/ConfigureSplunkforwardingtousesignedcertificates)

Automatic Load Balancing

- Configured in forwarder's **outputs.conf** using static list target:

```
[tcpout:my_LB_indexers]
server = idx1:9997, idx2:9997, idx3:9997
```

- Causes forwarder to split data between multiple indexers
- Switching indexers is performed:
 - By time, every **autoLBfrequency** seconds (default: 30 sec.)
 - By volume, every **autoLBvolume** bytes (default: 0 = disabled)
 - When it is safe for the data stream (e.g. an **EOF** is detected)
 - When a receiving indexer goes down



Load-balancing forwarder

docs.splunk.com/Documentation/Splunk/latest/Forwarding/Setuploadbalancinggd

Defining Event Boundary on UF

- Event boundaries
 - Detecting when one event ends and another starts
 - Normally determined during parsing (on indexer or HF)
- UF switches safely when:
 - An **EOF** (End of File) is detected
 - There is a short break in I/O activity
- Potential side effects
 - Streaming data (**syslog**) can prevent a UF from switching
 - A multi-line data (**log4j**) can result in event splits
 - Especially if the application has pauses in writing its log file
- Solution:
 - Enable event breaker on the UF per sourcetype

Defining Event Boundary on UF (cont.)

- Add event breaker settings on UF per sourcetype in **props.conf**
 - Single line event

```
[my_syslog]
EVENT_BREAKER_ENABLE = true
```

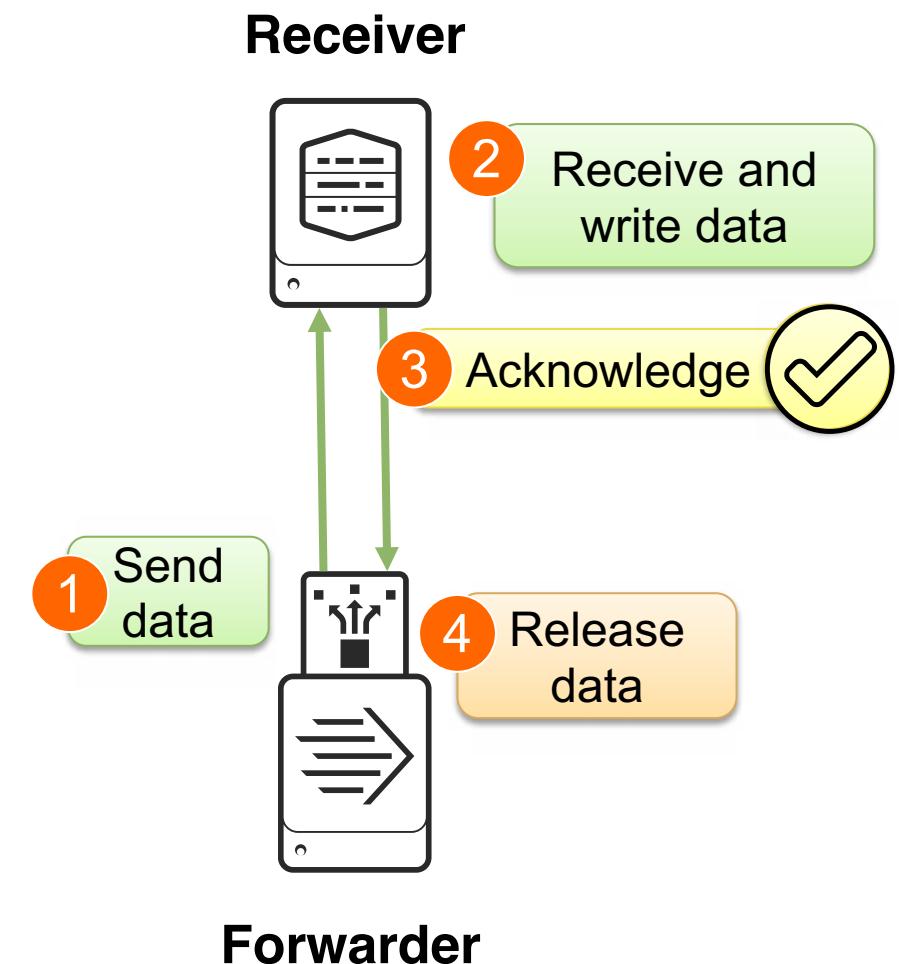
- Multi-line event

```
[my_log4j]
EVENT_BREAKER_ENABLE = true
EVENT_BREAKER = ([\r\n]+)\d\d\d\d-\d\d-\d\d
```

docs.splunk.com/Documentation/Forwarder/8.2.5/Forwarder/Configureloadbalancing

Indexer Acknowledgement

- Configured in **outputs.conf**
 - Disabled by default (**useACK=false**)
 - Enabled with **useACK=true**
- Guards against loss of forwarded data
 - If no acknowledgement is received, forwarder instead resends the data
- Enable along all segments of data path if using intermediate forwarders

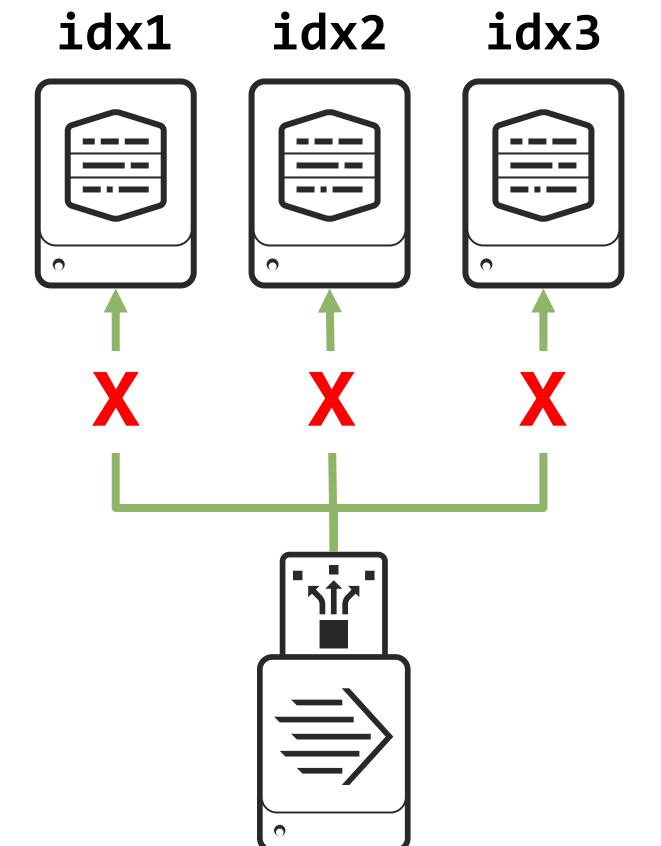


docs.splunk.com/Documentation/Splunk/latest/Forwarding/Protectagainstlossofin-flightdata

Forwarder Queue Size

- When forwarder can't reach an indexer, forwarder automatically switches to another indexer
- When forwarder can't reach any indexer, data is queued on the forwarder
- Output and wait queue sizes are affected by **maxQueueSize** and **useACK** in **outputs.conf**
 - Default: **maxQueueSize=auto**

maxQueueSize=	useACK=	Output queue	Wait queue
auto	false	500 KB	-
auto	true	7 MB	21 MB
20MB	true	20 MB	60 MB



Configuring a UF to Send Data over HTTP

- Use cases
 - Use existing network rules for HTTP
 - Easily supports off-the-shelf Load Balancers
- Limitations:
 - UF performs httpout or tcpout, but not both simultaneously
 - No support for indexer acknowledgements
- To break events on the UF for sending over HTTP:

outputs.conf

```
[httpout]
httpEventCollectorToken = <authToken>
uri = https://<ip>:8088
batchSize = 65536          (default: 64 KB)
batchTimeout = 30           (default: 30 sec)
```

props.conf

```
LB_CHUNK_BREAKER = ([\r\n]+)      (default)
```

Question: Automatic Load Balancing

Which of the following is true about automatic load balancing?

- A. Switching only occurs when an End of File (EOF) is detected
- B. Switching can be performed by time or by volume
- C. Configured in forwarder's **server.conf** file by listing targets
- D. Only supported on Heavy Forwarders

Answer: Automatic Load Balancing

Which of the following is true about automatic load balancing?

- A. Switching only occurs when an End of File (EOF) is detected
- B. Switching can be performed by time or by volume**
- C. Configured in forwarder's **server.conf** file by listing targets
- D. Only supported on Heavy Forwarders

Supported on any Splunk instance, load balancing switching can be performed by time (using **autoLBfrequency**; default is every 30 seconds) or by volume (using **autoLBvolume**; default is disabled) and occurs at EOF or a short break in I/O activity. Configuration is in the **outputs.conf** file.

Question: Configuring SSL for Data

Which is true about securing the data feed with SSL?

- A. Requires configuring the forwarder to send data over HTTP
- B. Requires additional settings to compress the feed
- C. Decreases CPU utilization
- D. Requires the use of certificates

Answer: Configuring SSL for Data

Which is true about securing the data feed with SSL?

- A. Requires configuring the forwarder to send data over HTTP
- B. Requires additional settings to compress the feed
- C. Decreases CPU utilization
- D. Requires the use of certificates**

Securing the feed with SSL encrypts and automatically compresses the standard Splunk data feed, but at the cost of possibly increasing CPU utilization. It requires the use of security certificates.

Question: Prevent Forwarded Data Loss

Which feature should be enabled to help prevent the loss of data from a forwarder to a receiver?

- A. Automatic load balancing
- B. Data encryption
- C. Indexer acknowledgement
- D. Configuring to send data over HTTP

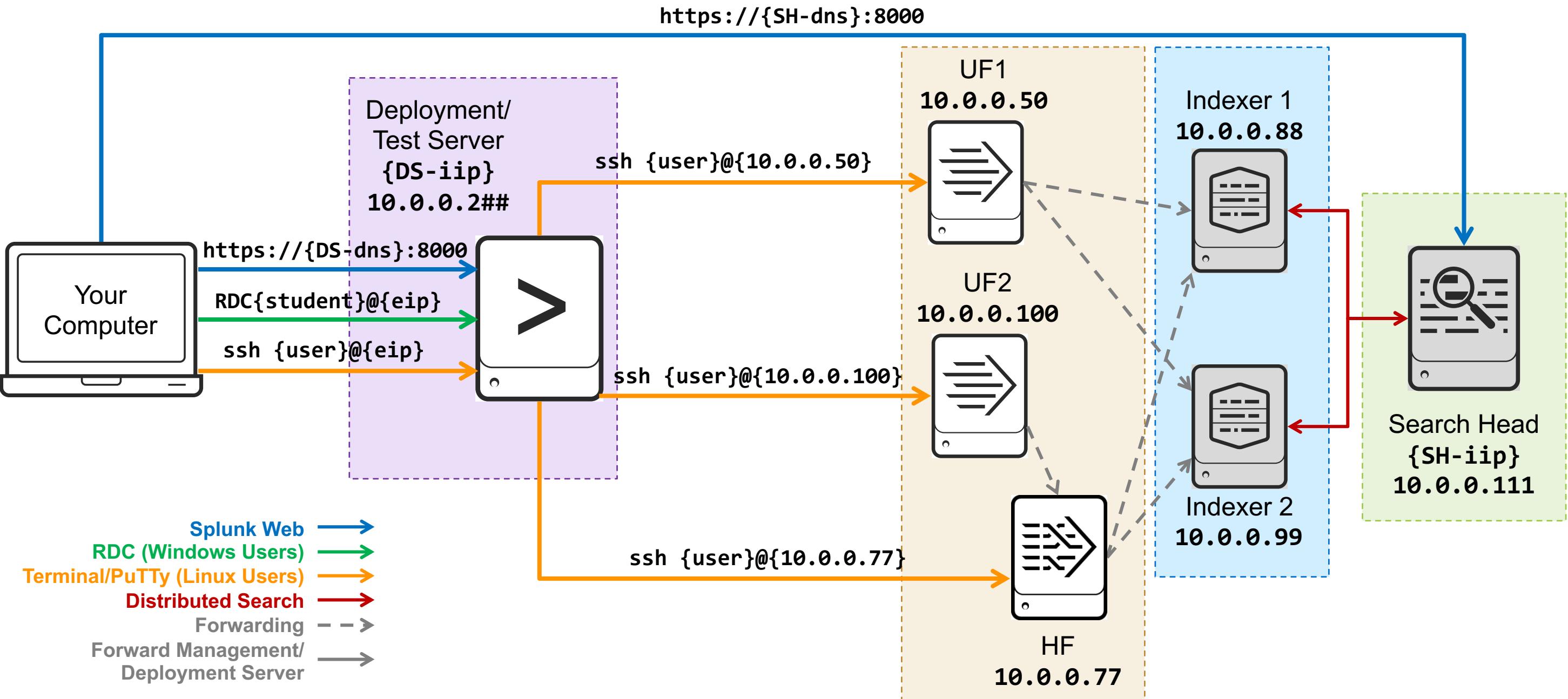
Answer: Prevent Forwarded Data Loss

Which feature should be enabled to help prevent the loss of data from a forwarder to a receiver?

- A. Automatic load balancing
- B. Data encryption
- C. Indexer acknowledgement**
- D. Configuring to send data over HTTP

With indexer acknowledgement, the forwarder will resend any data not acknowledged as "received" by the indexer or receiver.

Module 4 Lab – Environment Diagram



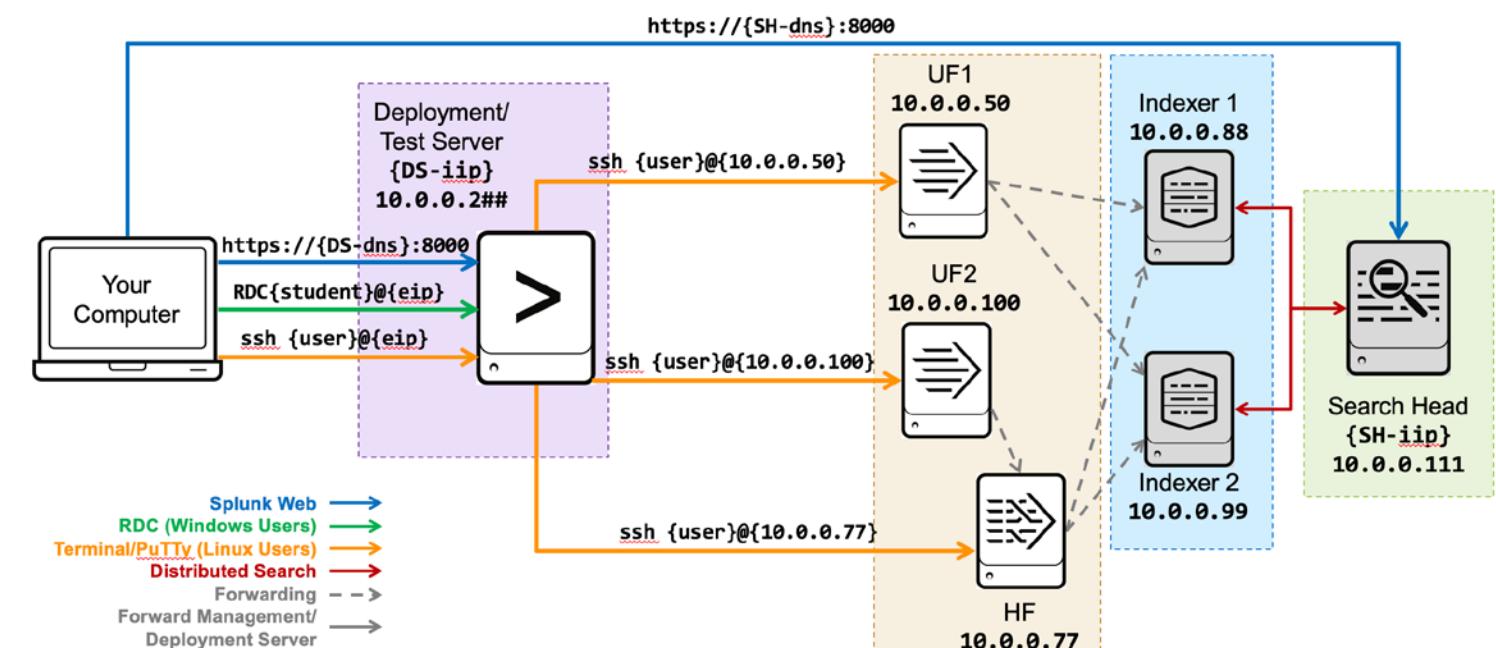
Module 4 Lab

Time: 25 minutes

Description: Customizing Forwarders

Tasks:

- Configure forwarders UF1 and HF to send data to Indexer 1 (**10.0.0.88**) and Indexer 2 (**10.0.0.99**)
- Confirm forwarder connection from your search head
- Configure HF to receive data as an intermediate forwarder from UF2

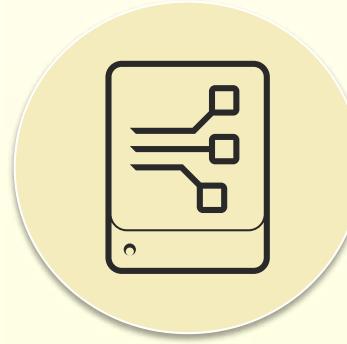


Module 5: Managing Forwarders

Module Objectives

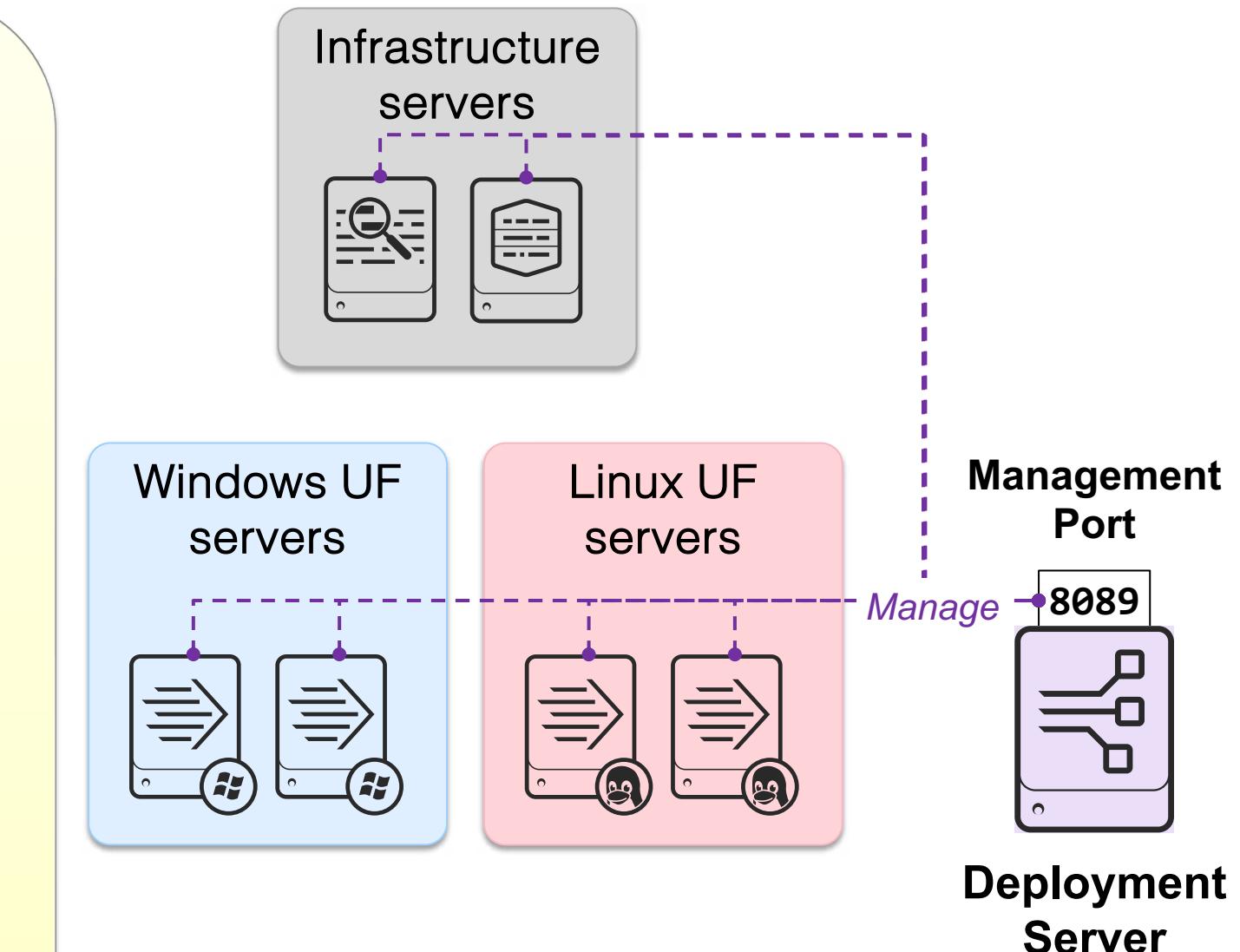
- Describe Splunk Deployment Server (DS)
- Manage forwarders using deployment apps
- Configure deployment clients and client groups
- Monitor forwarder management activities

Understanding the Deployment Server



Deployment Server (DS)

- Built-in tool for centrally managing configuration packages as apps for clients
- Includes **Forwarder Management** as the graphical user interface
- Can restart remote Splunk instances
- Requires an Enterprise license and should be on a dedicated server



Deployment Server Components

Deployment Apps

- Configuration files (such as **inputs.conf**) packaged as apps to be deployed to the deployment clients
- Reside in **SPLUNK_HOME/etc/deployment-apps/**

Deployment Clients

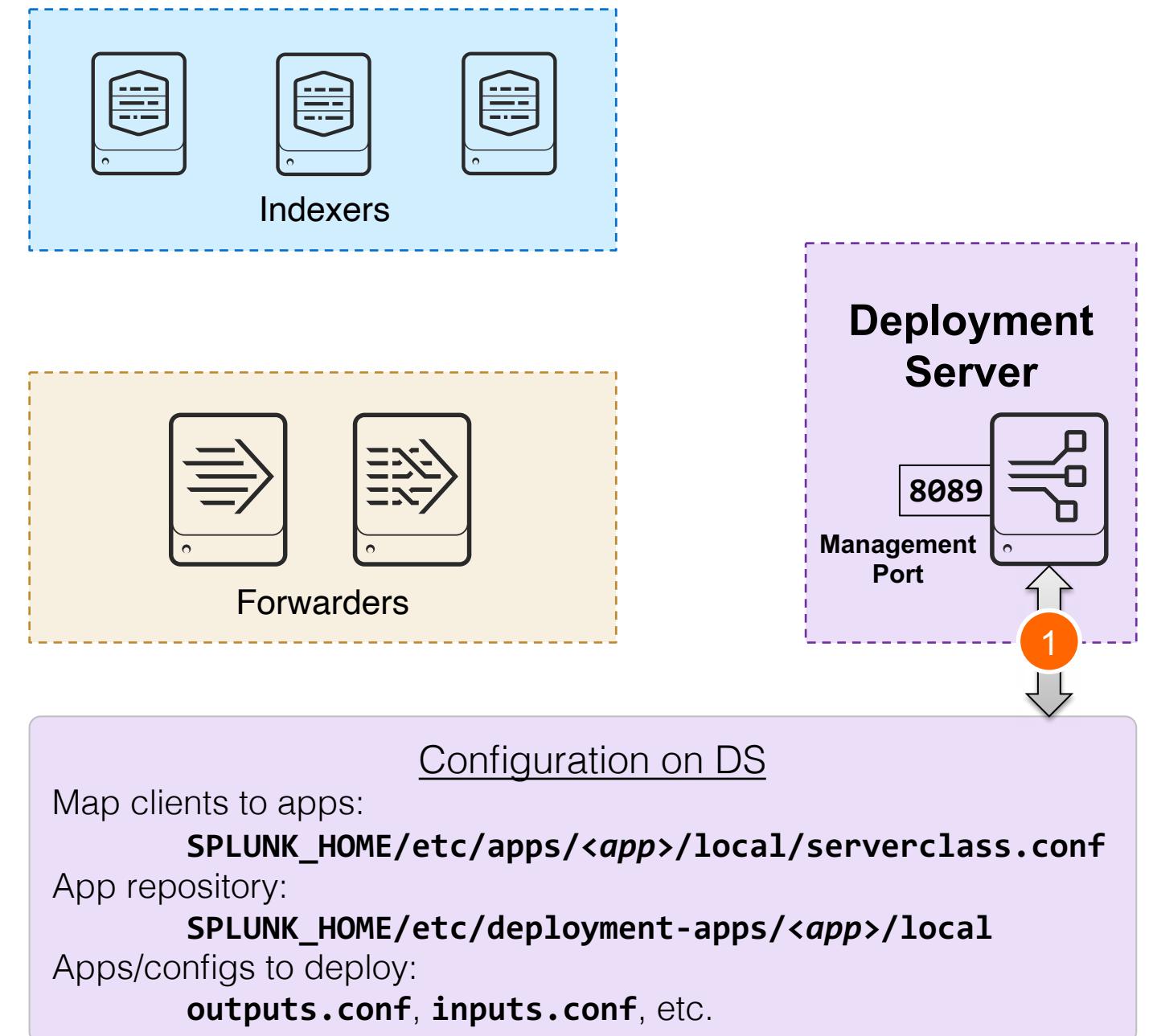
- Splunk instances (Enterprise or UF) that are connected to the Deployment Server (DS) and are phoning home
- Establish the connection from the Deployment Client

Server Classes

- Groupings of deployment clients
- Define what apps should be deployed to which clients
- Saved in **serverclass.conf**

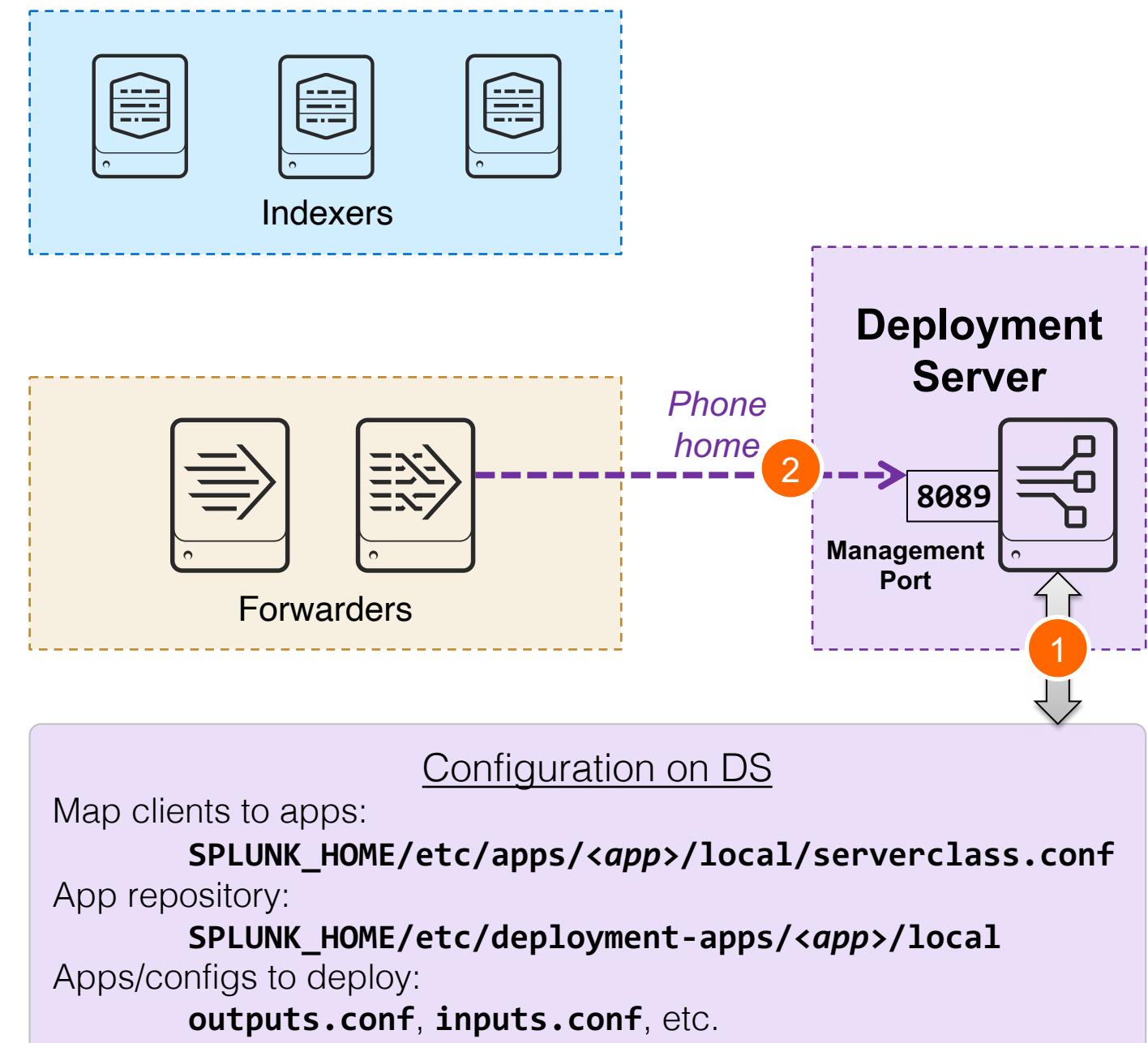
Deployment Server Configuration (1)

1. Configure DS, server classes, and app packages



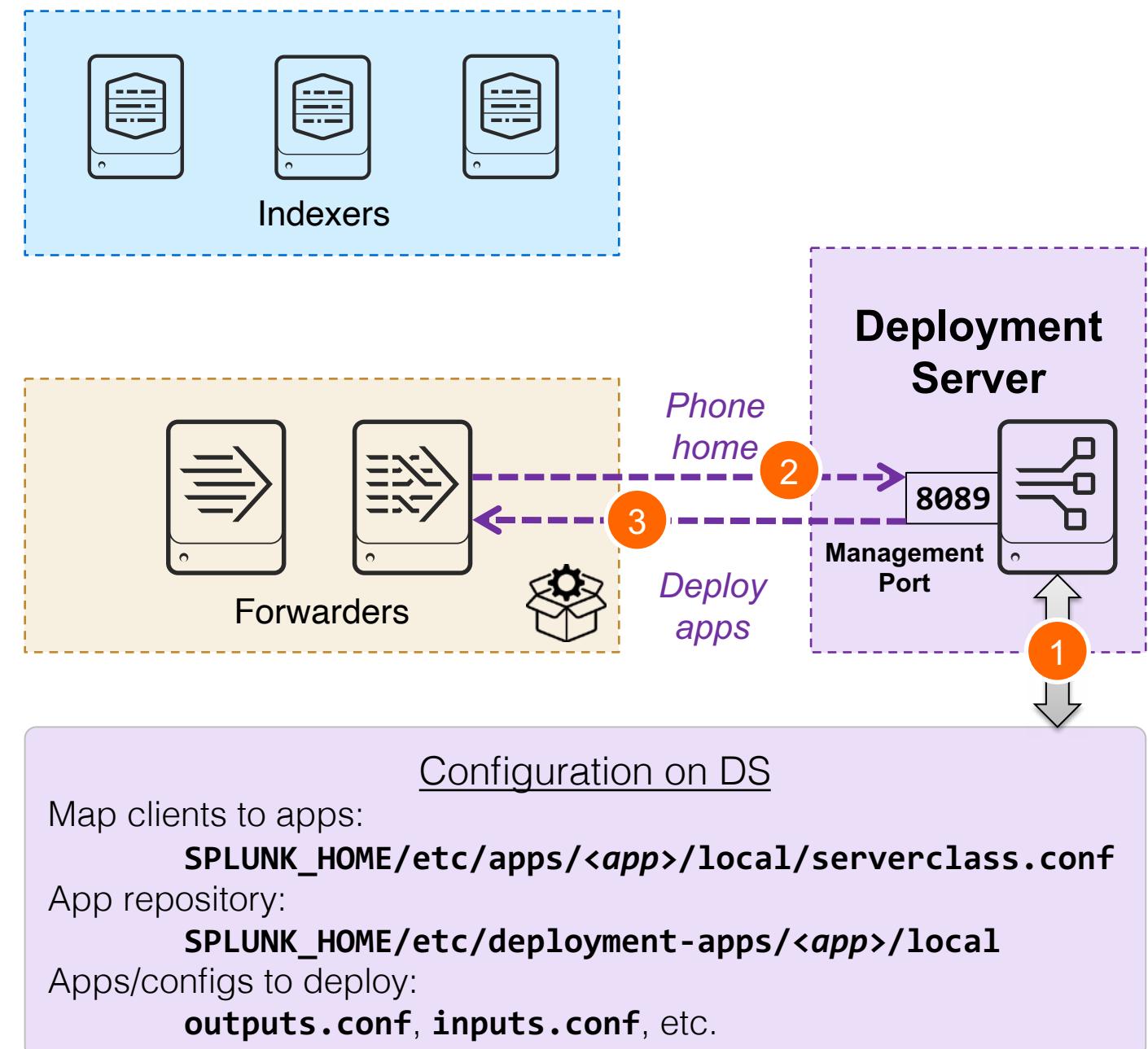
Deployment Server Configuration (2)

1. Configure DS, server classes, and app packages
2. Configure instances as deployment clients with **deploymentclient.conf**
 - Instances phone home to DS



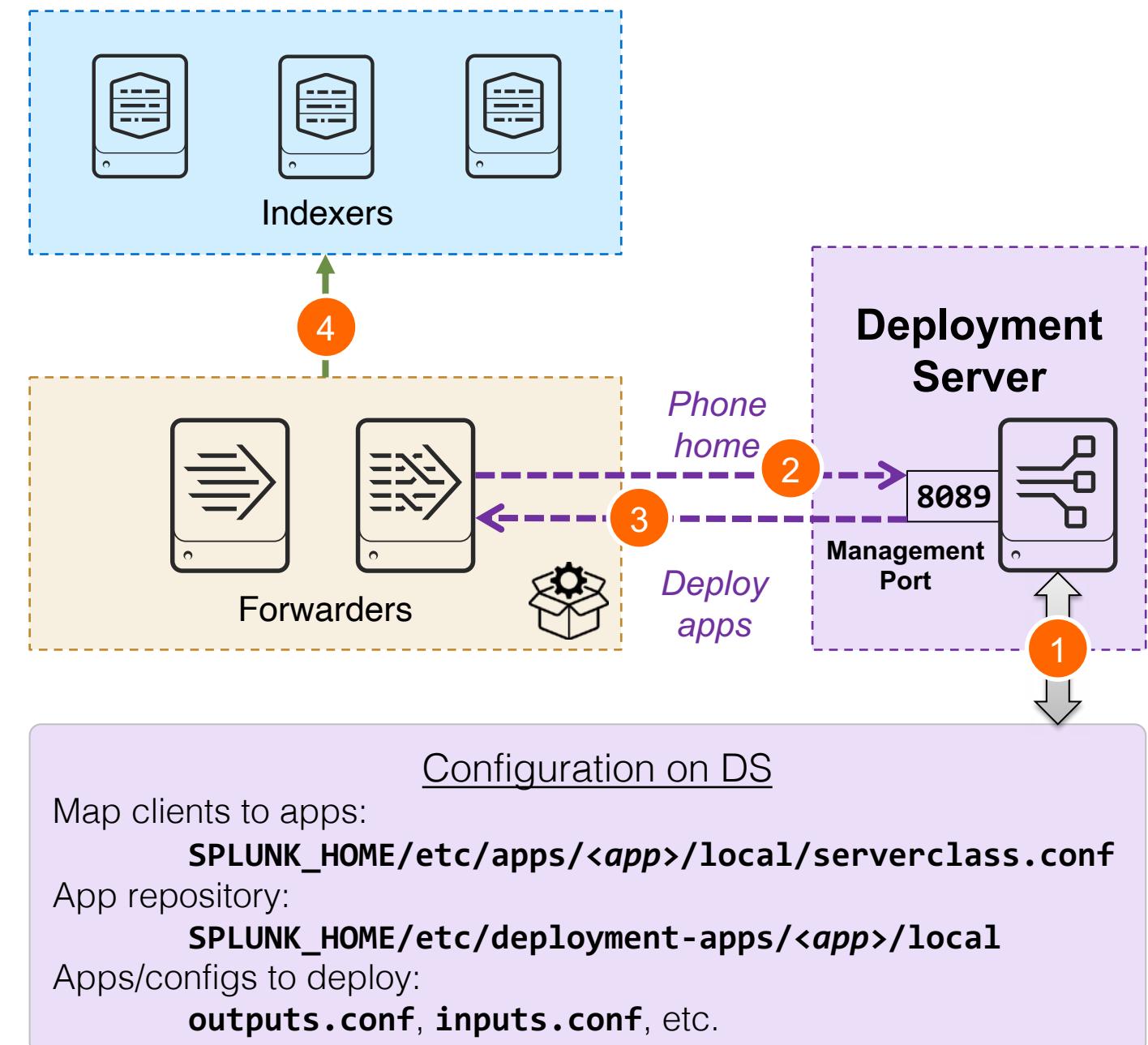
Deployment Server Configuration (3)

1. Configure DS, server classes, and app packages
2. Configure instances as deployment clients with **deploymentclient.conf**
 - Instances phone home to DS
3. Client downloads subscribed apps
 - As directed by server classes on DS



Deployment Server Configuration (4)

1. Configure DS, server classes, and app packages
2. Configure instances as deployment clients with **deploymentclient.conf**
 - Instances phone home to DS
3. Client downloads subscribed apps
 - As directed by server classes on DS
4. Client uses app configurations
 - For example: sending data to indexers

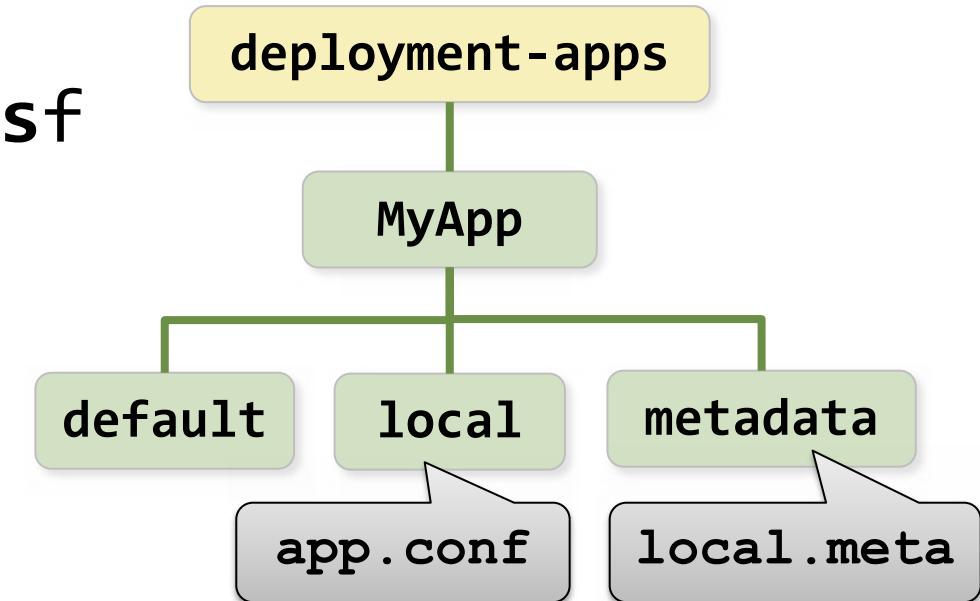


Enabling Forwarder Management

1. On deployment server: Add one or more apps in
 - Install an Enterprise license
 - Add one or more apps in **SPLUNK_HOME/etc/deployment-apps**
2. On forwarders: Set up the deployment client
 - Run **splunk set deploy-poll <deployment_server:splunkd_port>**
 - Run **splunk restart**
3. On deployment server: Create one or more server classes
 - Use forwarder management in Splunk Web
 - Modify **serverclasses.conf**

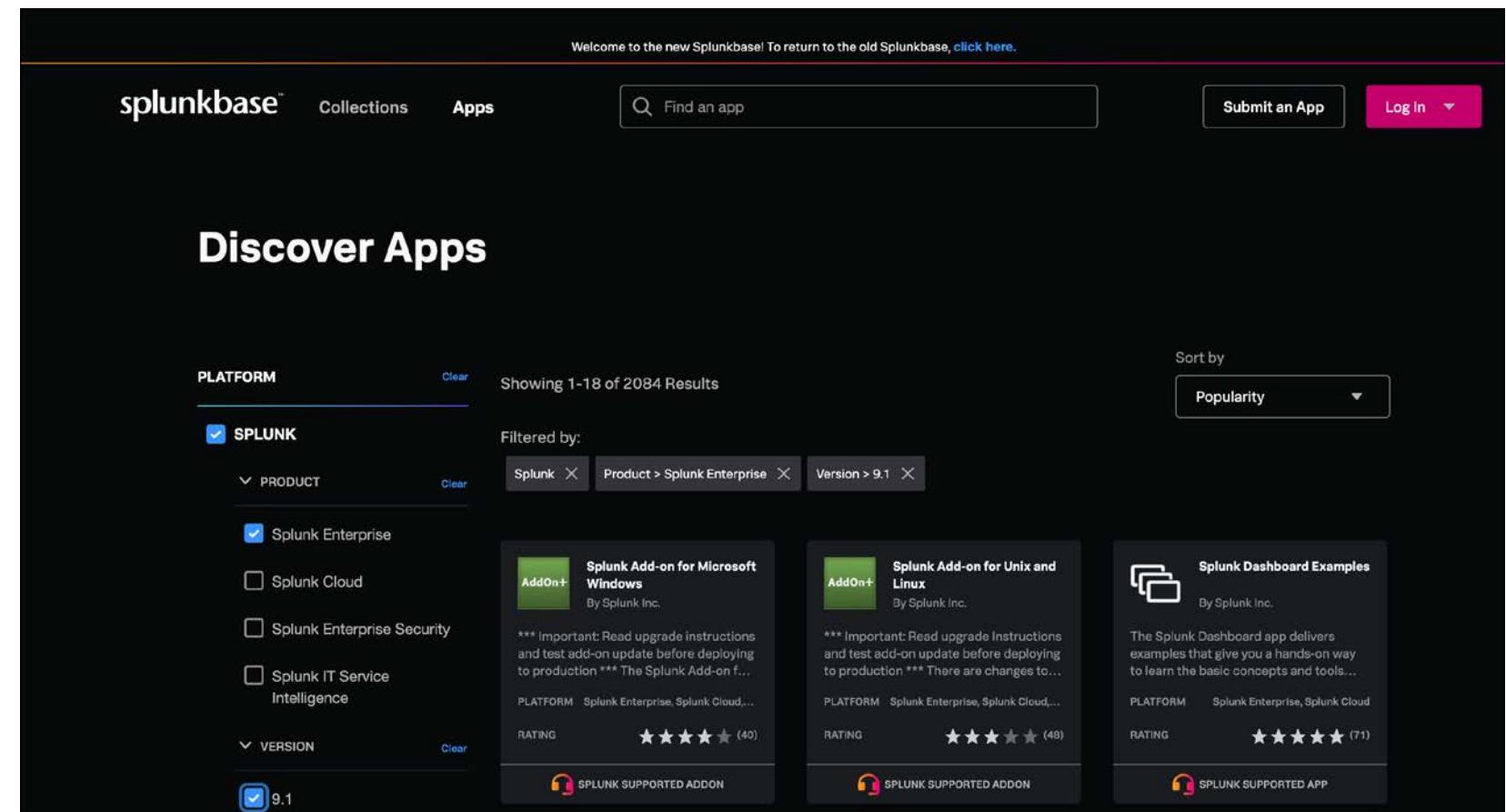
Configuring a Deployment App

- Follows app structure and rules
 - Place files in **SPLUNK_HOME/etc/deployment-apps**
 - Required files:
 - **app.conf** (in **default** or **local**)
 - **local.meta** (in **metadata**)
 - Optionally may contain configuration files, scripts, and other resources
- Files are deployed to client's **SPLUNK_HOME/etc/apps** folder by default
- Best practice
 - Create small and discrete deployment apps
 - Take advantage of **.conf** file layering
 - Use a consistent naming convention



Apps and Add-ons

- Can be downloaded from Splunkbase
- Installed on a Splunk instance:
 - Using the Deployment Server
 - Using CLI on the instance
 - Manually by installing the app
- Deploy to **SPLUNK_HOME/etc/apps**
- Comes with documentation for details about settings for **inputs.conf**, and so on



Configuring Deployment Clients

- On prospective deployment clients (usually forwarders):
 1. Run: **splunk set deploy-poll <deployment_server:splunkd_port>**
 - Creates **deploymentclient.conf** in **SPLUNK_HOME/etc/system/local**
 - Alternatively create **deploymentclient.conf** manually
 2. Restart the deployment clients:
splunk restart
- Edit **[deployment-client]** stanza to override defaults
 - Can be part of initial deployment app
 - Contains phone home setting (default: 60 seconds)

deploymentclient.conf

```
[target-broker:deploymentServer]
targetUri = splunk_server:8089
```

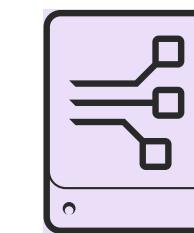
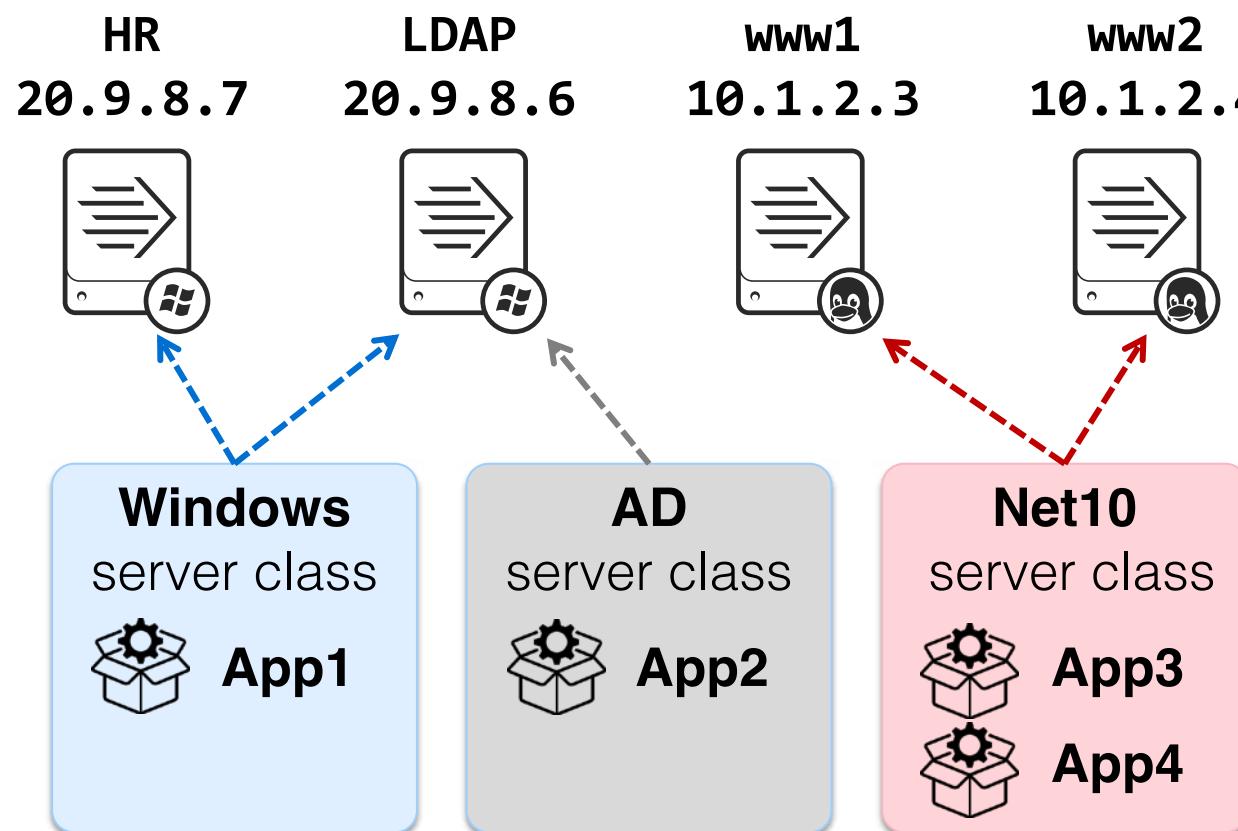
...

[deployment-client]

```
clientName = webserver_1
phoneHomeIntervalInSecs = 600
```

What's a Server Class?

- Maps groups of clients to deployment apps
 - Can be based on client name, host name, IP address, DNS name, or machine types



Deployment Server

Server class	Rules
Windows	<ul style="list-style-type: none">Assigned to Windows systemsInstalls App1
AD	<ul style="list-style-type: none">Assigned to Active Directory serversInstalls App2
Net10	<ul style="list-style-type: none">Assigned to hosts on 10.1.2.* subnetInstalls App3 and App4

Adding a Server Class

Forwarder Management

Repository Location: \$SPLUNK_HOME/etc/deployment-apps

0 Clients PHONED HOME IN THE LAST 24 HOURS

0 Clients DEPLOYMENT ERRORS

0 Total downloads IN THE LAST 1 HOUR

Apps (1) **Server Classes (0)** Clients (0)

No server classes. Learn more. [create one](#)

1 Client PHONED HOME IN THE LAST 24 HOURS

0 Clients DEPLOYMENT ERRORS

0 Total downloads IN THE LAST 1 HOUR

Apps (1) **Server Classes (1)** Clients (1)

All Server Classes filter

1 Server Classes 10 Per Page

Last Reload	Name	Actions	Apps	Clients
a few seconds ago	uf_base	Edit ▾	0	0 deployed

Select the Server Classes tab

Create New Server Class

Enter a name for the new server class

Save

Selecting Apps for the Server Class

Server Class: uf_base

[Back to Forwarder Management](#)

You haven't added any apps

1 [Add Apps](#)

You haven't added any clients

2 [Add Clients](#)

Edit Apps

Server Class: uf_base

1 Unselected App

filter

uf_base

hf_base

Select app to move it to Selected Apps

1 Selected App

filter

uf_base

3 [Save](#)

The screenshot shows the 'Edit Apps' page for the 'uf_base' server class. On the left, under 'Unselected Apps', there is a list with 'uf_base' highlighted and circled in orange. A callout bubble with the text 'Select app to move it to Selected Apps' points to this item. On the right, under 'Selected Apps', 'uf_base' is listed. At the top right of the 'Edit Apps' panel are 'Documentation' and 'Cancel' buttons, and at the bottom right is a green 'Save' button with the number '3' above it. The main header 'Server Class: uf_base' has an 'Edit' button and a 'Documentation' button above it.

Post Deployment Behavior Setting

Server Class: uf_base

[Edit](#) [Documentation](#)

[Back to Forwarder Management](#)

Apps [Edit](#)

Deployed Successfully [filter](#)

1 Apps 10 Per Page

Name	Actions	After Installation	Clients
uf_base	Edit Uninstall	Enable App	0 deployed

Edit App: uf_base

Documentation

Server Classes: [uf_base](#) [x](#) [+](#)

After Installation:

Enable App

Restart Splunkd

Ensure Restart Splunkd is enabled

[Cancel](#) [Save](#)

The diagram illustrates a three-step process for setting post-deployment behavior. Step 1 highlights the 'Edit' button for the 'uf_base' app in the main interface. Step 2 highlights the 'Restart Splunkd' checkbox in the 'Edit App' dialog. Step 3 highlights the 'Save' button in the dialog.

Selecting Clients for the Server Class

Server Class: uf_base

[Back to Forwarder Management](#)

Apps Edit

Deployed Successful

1 Apps 10 Per Page ▾

Name

uf_base

You haven't added any clients yet.

1 Add Clients

2 Enter Include, Exclude, and/or Machine Type filters

3 Save

Edit Clients
Server Class: uf_base

Include (whitelist)

ip-10*

Can be client name, host name.
Examples: 185.2.3.*, fwdr-*
[Learn more](#)

- Supports wildcards
- Exclude takes precedence over Include

Exclude (blacklist)

Optional

name.
Examples: ronnie, rarity
[Learn more](#)

Filter by Machine Type (machineTypesFilter)

+ Optional

All Matched Unmatched filter

1 10 Per Page ▾

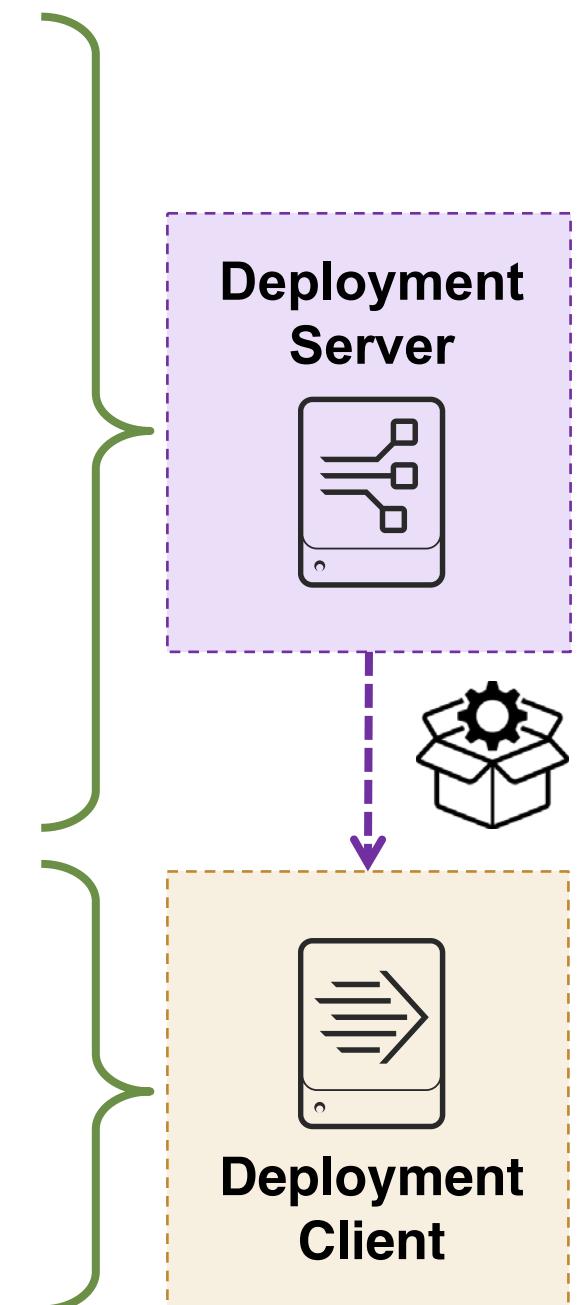
Matched	Host Name	DNS Name	Client Name	Instance Name	IP Address	Machine Type	Phone Home
	ip-10-0-0-100	10.0.0.100	E9DB9FFE-589E-4158-8B2F-77F26B4418A4	engdev203	10.0.0.100	linux-x86_64	a few seconds ago

Verify forwarder management

- On the deployment client:
 - Display the deployment server and management port:
splunk show deploy-poll
 - Confirm expected app directories and contents in
SPLUNK_HOME/etc/apps/app_name
 - ▶ Occurs at the next phone home interval
- On the deployment server:
 - Display information about the deployment clients:
splunk list deploy-clients

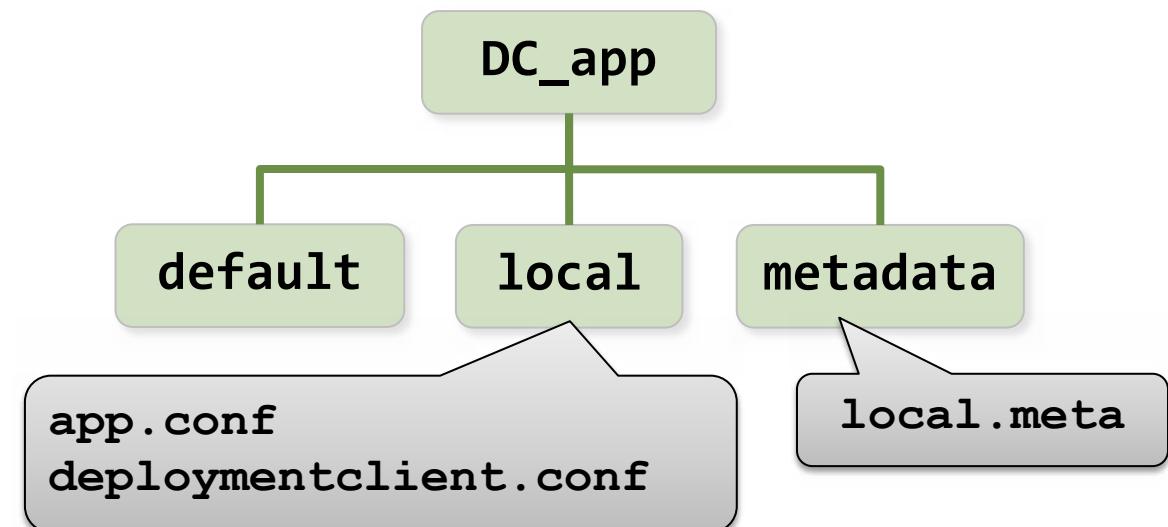
Updating deployed apps

1. Add new apps or change existing app in **deployment-apps**
2. Run **splunk reload deploy-server**
 - Detects changes to deployment apps on DS
 - Re-caches list of deployment apps
 - Re-calculates checksums used to uniquely identify apps by their contents
 - Eliminates need to restart Splunk
3. Verify the client downloads new/changed apps after next phone-home
 - Client downloads apps when checksums have changed



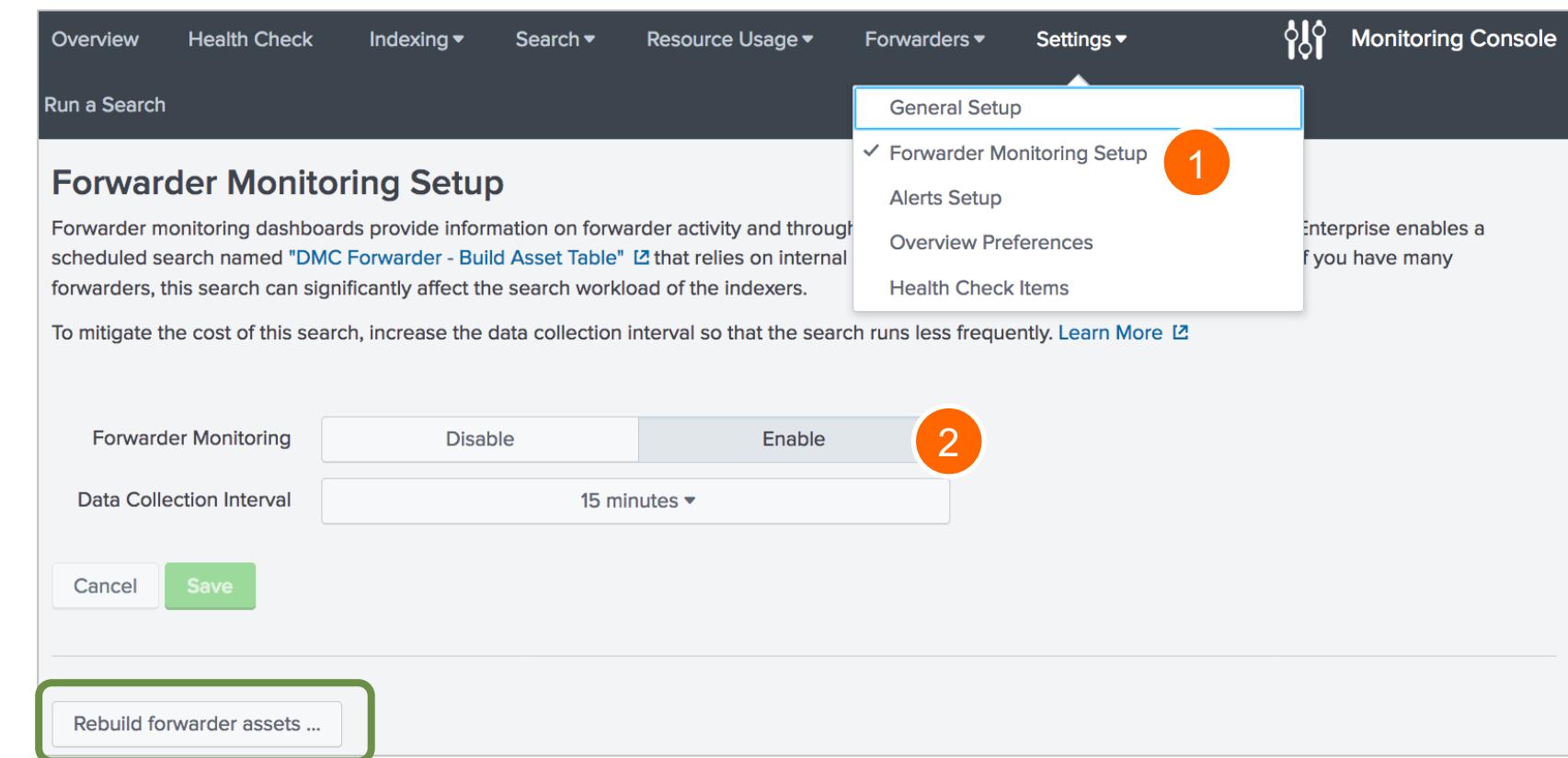
Manage Deployment Client Settings Centrally

- Use an app to manage deployment client settings
 - Create a deployment client settings app (example: **DC_app**)
 - Move **deploymentclient.conf** settings from **etc/system/local/** to **etc/apps/DC_app/local/**
 - Deploy **DC_app** to clients using a Server Class

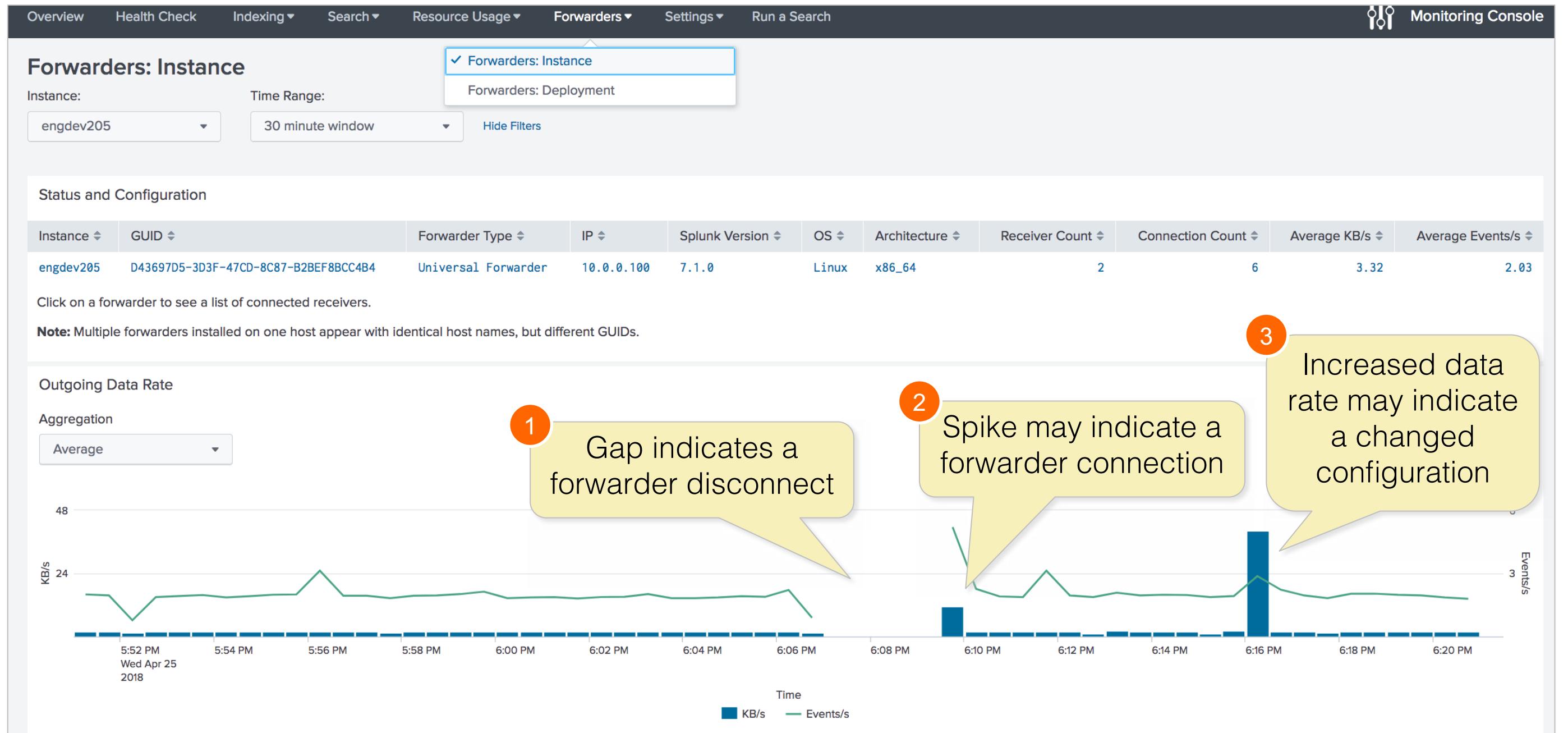


Forwarder Monitoring with Monitoring Console

- Provides valuable information on forwarder activity and throughput
- Runs a scheduled search that builds a forwarder asset table
 - Runs every 15 minutes by default
 - Relies on forwarder internal logs
 - Can affect search workload if you have many forwarders
 - Can be rebuilt manually
- Enabled with:
 1. MC: Settings > Forwarder Monitoring Setup
 2. Forwarder Monitoring: Enable



Forwarder Monitoring with MC



Useful Commands

Command	Operation
From the Deployment Client:	
splunk set deploy-poll	Connects the client to the deployment server and management port
splunk show deploy-poll	Displays the current deployment server and management port
splunk list forward-server	Displays the current forward server configuration
splunk disable deploy-client	Disables the deployment client
From the Deployment Server (DS):	
splunk reload deploy-server	Checks all apps for changes and notifies the relevant clients the next time they phone home
splunk list deploy-clients	Displays information about the deployment clients

Question: Deployed Apps on Clients

When an app is deployed from the Deployment Server to the client, where will you find that app on the client by default?

- A. SPLUNK_HOME/etc/system
- B. SPLUNK_HOME/etc/apps
- C. SPLUNK_HOME/etc/deployment-apps
- D. SPLUNK_HOME/var/lib/splunk

Answer: Deployed Apps on Clients

When an app is deployed from the Deployment Server to the client, where will you find that app on the client by default?

- A. `SPLUNK_HOME/etc/system`
- B. `SPLUNK_HOME/etc/apps`**
- C. `SPLUNK_HOME/etc/deployment-apps`
- D. `SPLUNK_HOME/var/lib/splunk`

The apps in the **`SPLUNK_HOME/etc/deployment-apps`** on the Deployment Server are apps for deployment to a client. The apps get deployed to the **`SPLUNK_HOME/etc/apps`** folder on the client.

Question: DS Polling Port

On which port do clients poll the Deployment Server by default?

- A. 8000
- B. 8089
- C. 8191
- D. 9997

Answer: DS Polling Port

On which port do clients poll the Deployment Server by default?

- A. 8000
- B. 8089**
- C. 8191
- D. 9997

Clients poll the Deployment Server on its management (**splunkd**) port, which is **8089** by default.

Question: Connect client to DS

Which command is used to connect the deployment client to the deployment server and management port?

- A. `splunk set deploy-poll`
- B. `splunk set deploy-client`
- C. `splunk set deploy-server`
- D. `splunk set forward-server`

Answer: Connect client to DS

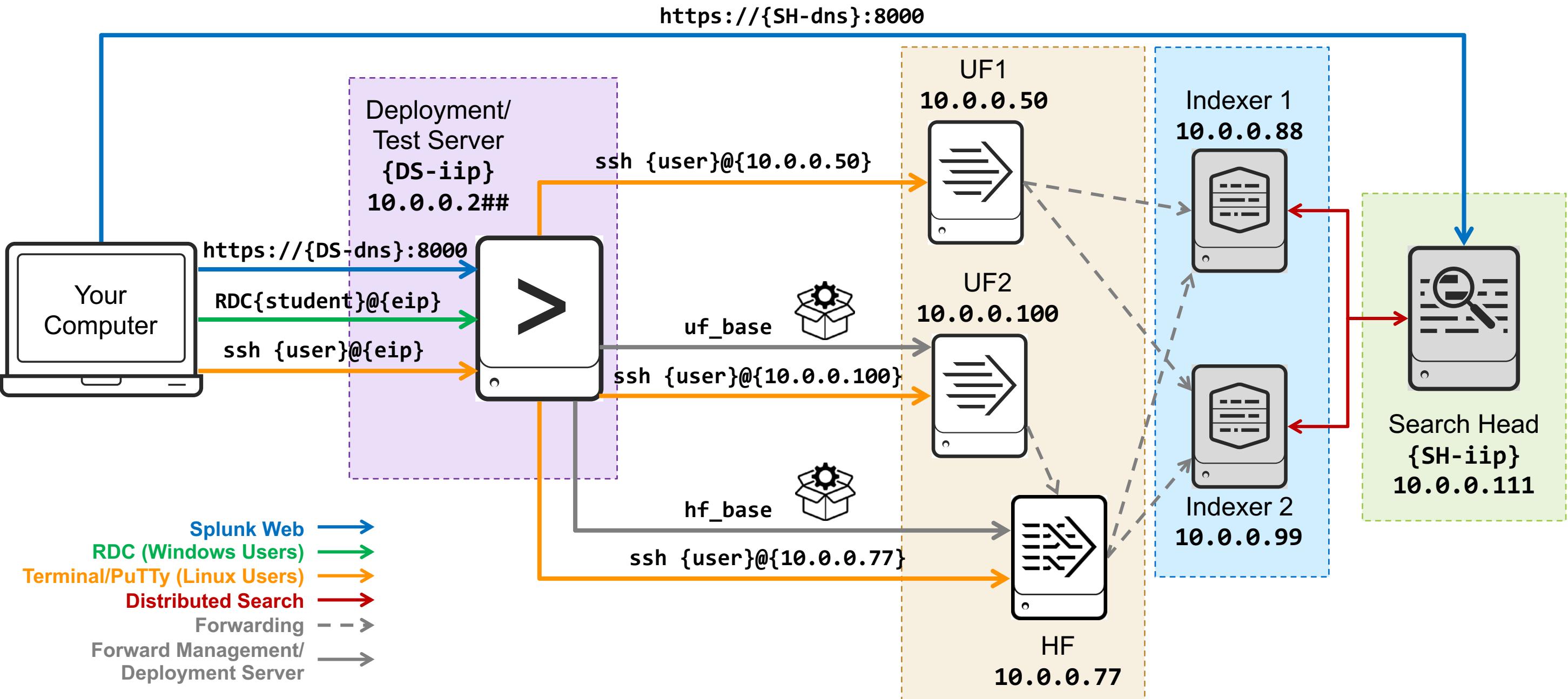
Which command is used to connect the deployment client to the deployment server and management port?

- A. **splunk set deploy-poll**
- B. **splunk set deploy-client**
- C. **splunk set deploy-server**
- D. **splunk set forward-server**

To connect a client to the deployment server:

1. Run **splunk set deploy-poll <deployment_server:splunkd_port>**
2. Restart the deployment client

Module 5 Lab – Environment Diagram



Module 5 Lab

Time: 30 minutes

Description: Managing Forwarders

Tasks:

- Remove forwarding configuration from UF2 and HF
- Copy deployment apps to the DS folders
- Configure UF2 and HF as deployment clients
- Create two server classes to manage UF2 and the HF from the DS
- Confirm deployment of deployment apps on UF2 and HF

Module 6: Monitor Inputs

Module Objectives

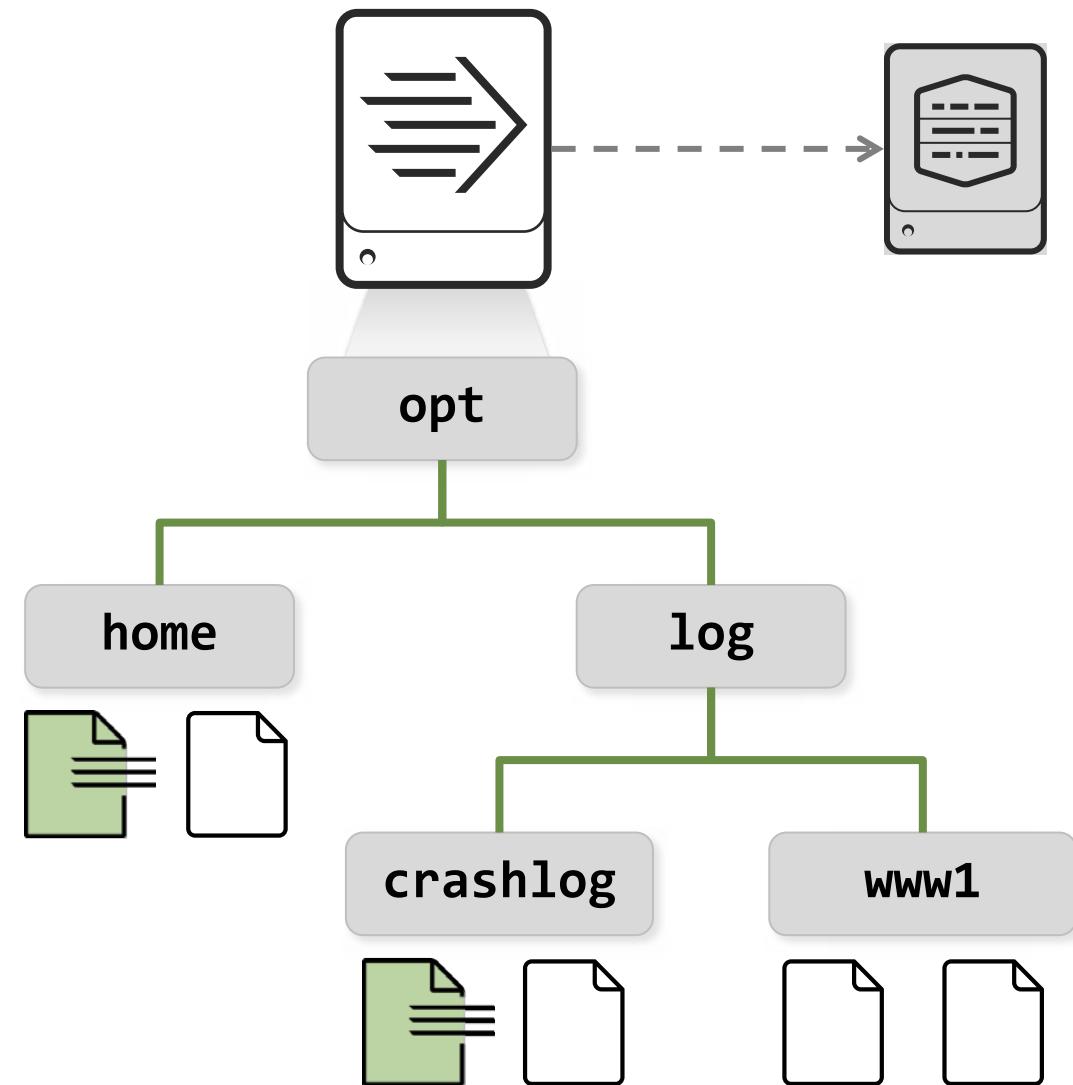
- Create file and directory monitor inputs
- Use optional settings for monitor inputs
- Deploy a remote monitor input

Monitoring Input Files

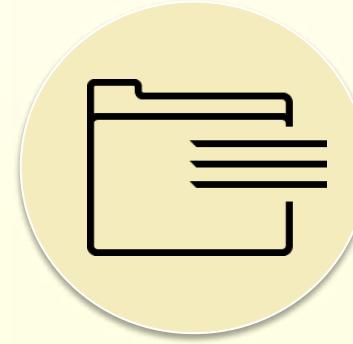


Monitoring Files

- Defines a single file as the source, with input settings (**sourcetype**, **index**, **host**, etc.)
- Ingests current contents of the file
- Continuously monitors for new content using the Splunk Fishbucket to keep a checkpoint
- Supports any text format, such as: plain text, structured text (**CSV**, **XML**, **JSON**), multi-line logs (**Log4J**), and files compressed with **gzip**

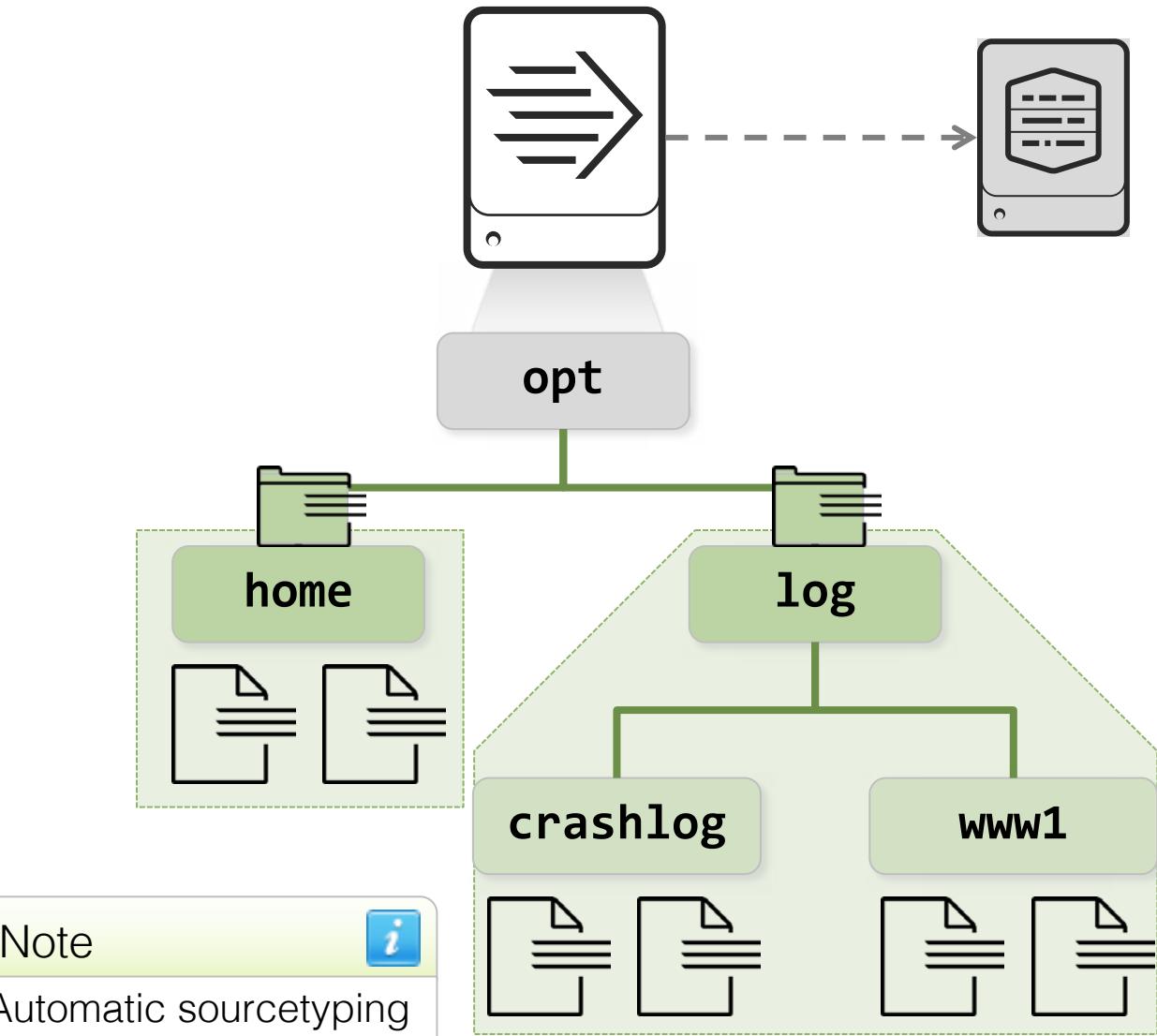


Monitoring Input Directories



Monitoring Directories

- Defines a directory tree as data source
- Recursively traverses directory and monitors all discovered text files
- Unzips compressed files automatically before ingesting them, one at a time
- Includes new files added to the directories
- Detects and handles log file rotation
- Input settings applied to all contained files



Monitor Input Options in `inputs.conf`

- Defining the source
 - Place after `monitor://` in stanza header
 - Absolute path to a file or directory
 - Can contain wildcards
- Defining attributes
 - All attributes are optional
 - Default `host` is defined in **SPLUNK_HOME/etc/system/local/inputs.conf**
 - Omitting `sourcetype` causes Splunk to try to determine it automatically
- For more attributes and documentation
 - See `inputs.conf.spec` in **SPLUNK_HOME/etc/system/README**

`inputs.conf` format:

```
[monitor://<path>]
disabled=[0|1|false|true]
sourcetype=<string>
host=<string>
index=<string>
blacklist=<regular expression>
whitelist=<regular expression>
```

Example `monitor` path entries:

```
[monitor:///var/log/secure]
[monitor:///var/log/]
[monitor://C:\logs\system.log]
[monitor://C:\logs\]
```

File Pathname Wildcards in `inputs.conf`

Wildcard	Description
...	The ellipsis wildcard recurses through directories and subdirectories to match.
*	The asterisk wildcard matches anything in that specific directory path segment but does not go beyond that segment in the path. Normally it should be used at the end of a path.

File and Directory Matching

```
[monitor:///var/log/www1/secure.log]
sourcetype = linux_secure
```

- ✓ /var/log/www1/secure.log
- ✗ /var/log/www1/secure.1
- ✗ /var/log/www1/logs/secure.log
- ✗ /var/log/www2/secure.log

```
[monitor:///var/log/www1/secure.*]
sourcetype = linux_secure
```

- ✓ /var/log/www1/secure.log
- ✓ /var/log/www1/secure.1
- ✗ /var/log/www1/logs/secure.log
- ✗ /var/log/www2/secure.log

```
[monitor:///var/log/www*/secure.*]
sourcetype = linux_secure
```

- ✓ /var/log/www1/secure.log
- ✓ /var/log/www1/secure.1
- ✗ /var/log/www1/logs/secure.log
- ✓ /var/log/www2/secure.log

```
[monitor:///var/log/.../secure.*]
sourcetype = linux_secure
```

- ✓ /var/log/www1/secure.log
- ✓ /var/log/www1/secure.1
- ✓ /var/log/www1/logs/secure.log
- ✓ /var/log/www2/secure.log

✓ Matches
✗ Doesn't match

Additional Options

Follow tail (**followTail**)

- Splunk ignores file's existing content, indexing new data as it arrives
- DO NOT leave enabled indefinitely

Ignore older than (**ignoreOlderThan**)

- Only index events after the time window (such as only events within last 60 days with **ignoreOlderThan = 60d**)
- Completely ignores files with modification time outside the time window (even if the file is updated later)

Include and exclude list (**whitelist**, **blacklist**)

- Use regular expressions to filter files or directories from the input
- In case of a conflict, the exclude list prevails

Example: Using Include List (**whitelist**)

- Files/directories that match the regular expression are indexed
- The syntax for exclude list (**blacklist**) is identical

```
[monitor:///var/log/www1/]
whitelist = \.log$
```

✓ /var/log/www1/access.log
✓ /var/log/www1/dbaccess.log
✓ /var/log/www1/access.1.log
✗ /var/log/www1/access.log.2

```
[monitor:///var/log/www1/]
whitelist = query\.log$|my\.log$
```

✓ /var/log/www1/query.log
✓ /var/log/www1/dbquery.log
✓ /var/log/www1/my.log
✗ /var/log/www1/my.log4j

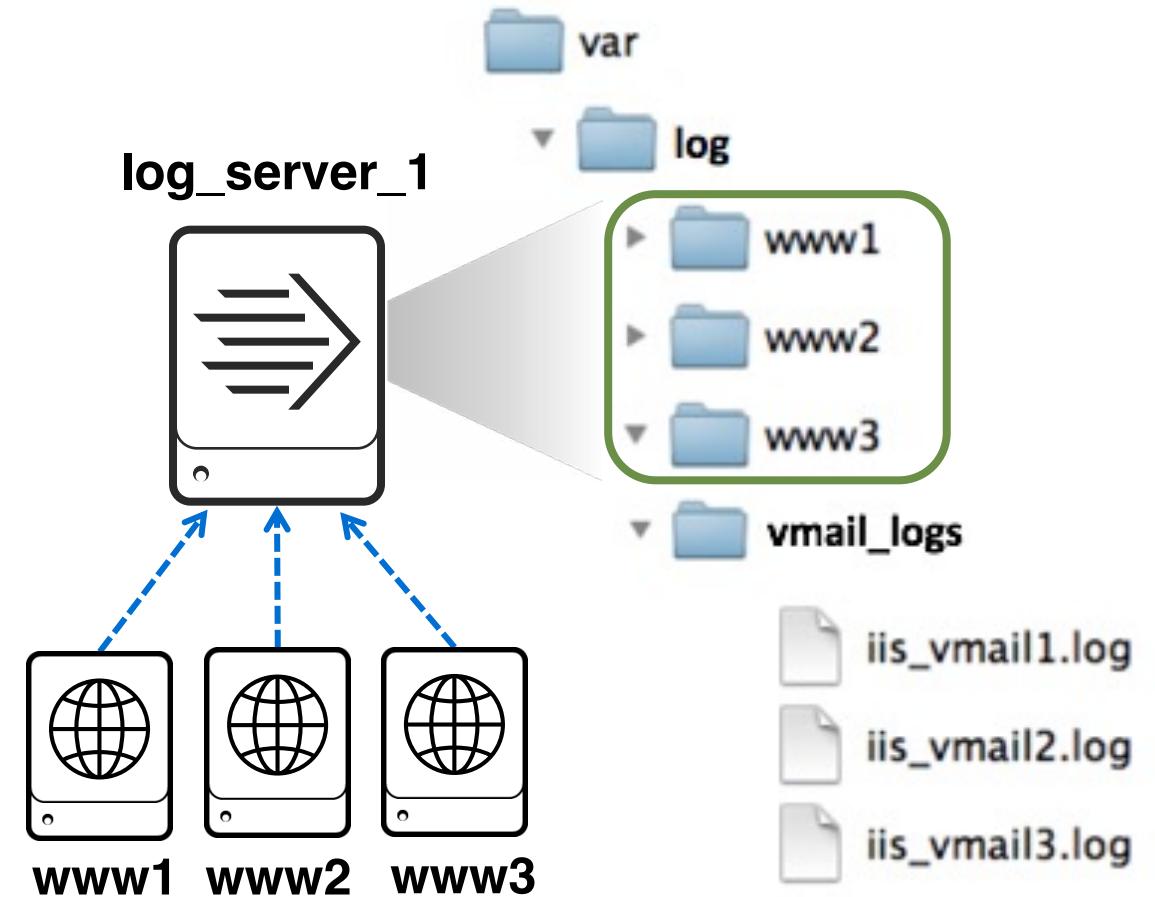
```
[monitor:///var/log/www1/]
whitelist = /query\.log$|/my\.log$
```

✓ /var/log/www1/query.log
✓ /var/log/www1/my.log
✗ /var/log/www1/dbquery.log
✗ /var/log/www1/my.log4j

✓ Matches
✗ Doesn't
match

Overriding the Host Field

- When data is stored on a different server than its origin
 - Example: A web farm where each web server stores its log file on a centralized file server
- By explicitly setting the host
 - Using a specified value
 - Using a directory name
 - Using a regular expression



Setting the Host With a Directory Name

- Used with **host_segment = <integer>**

Example: Setting **host_segment** to **3** uses the 3rd segment of the directory path as the host name for files in that directory

The screenshot shows the "Add Data" wizard in Splunk. The steps are: Select Source (green dot), Set Source Type (green dot), Input Settings (green dot), Review (gray dot), and Done (gray dot). The "Review" and "Done" buttons are grayed out. On the left, there's a file tree: "var" contains "log", which contains "www1", "www2", and "www3". The "www3" folder is highlighted with a green background. A green arrow points from the "www3" folder to a text input field labeled "Segment number?". This field contains the value "3". To the right of the input field is a list of options: "Constant value" (radio button), "Regular expression on path" (radio button), and "Segment in path" (radio button, which is selected and highlighted with a red border). Below the input field is a black box containing the configuration command:

```
[monitor:///var/log/]
host_segment=3
```

Setting the Host With a Regular Expression

- Used with **host_regex = <regular expression>**

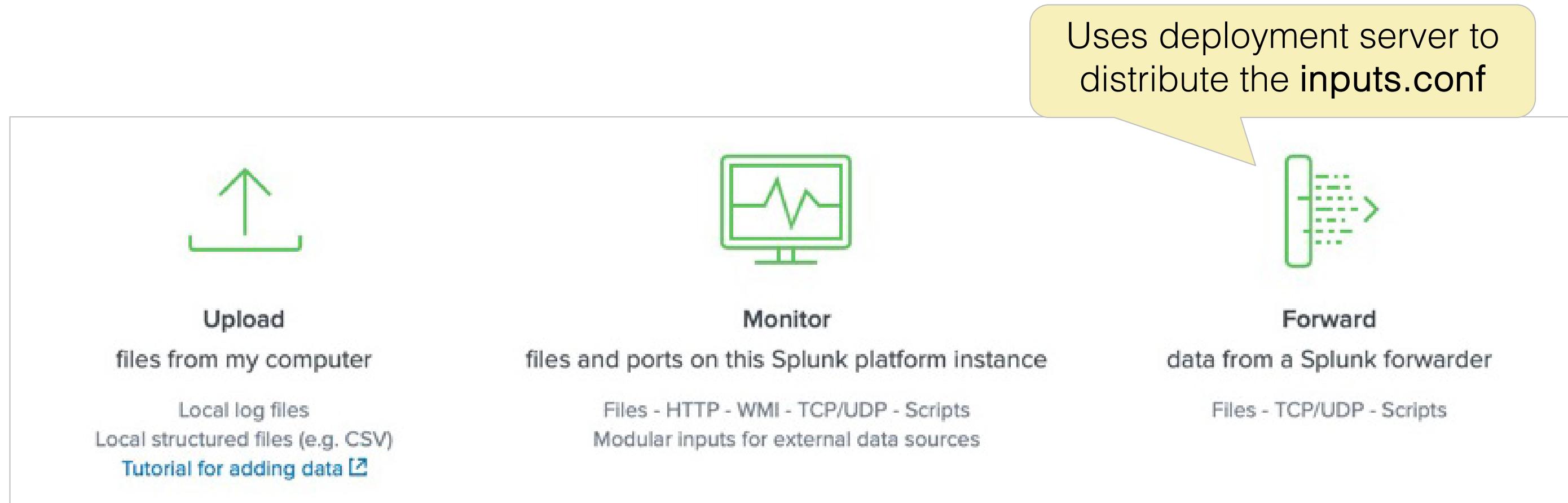
Example: Setting **host_regex** to `\w+(vmail.+)\.log$` selects the second part of the log file name as its host name

The screenshot shows the 'Add Data' wizard in Splunk, specifically the 'Input Settings' step. On the left, a file system tree shows a path: var/log/vmail_logs containing files iis_vmail1.log, iis_vmail2.log, and iis_vmail3.log. A green arrow points from the vmail_logs folder to a dropdown menu labeled 'Regular expression?'. This menu contains three options: 'Constant value', 'Regular expression on path' (which is selected and highlighted with a red box), and 'Segment in path'. Below the menu is a text input field containing the regular expression `\w+(vmail.+)\.log$`. To the right of the input field is a large black box containing the configuration command:

```
[monitor://C:\var\log\vmail_logs]
host_regex=\w+(vmail+)\.log$
```

Creating Forwarded Inputs

- Use the deployment server to create forwarded inputs
- Optionally create deployment apps for configuring inputs on deployment clients



Creating Forwarded Inputs (cont.)

The screenshot shows the 'Add Data' wizard with five steps: 'Select Forwarders' (highlighted with a green dot), 'Select Source', 'Input Settings', 'Review', and 'Done'. The 'Select Forwarders' step is currently active. It includes instructions to create or select a server class for data inputs. Below this, there's a note about enabling forwarding of data from deployment clients. A 'Select Server Class' section has 'New' selected, showing 'Available host(s)' (LINUX ip-10-0-0-100, LINUX ip-10-0-0-77) and an empty 'Selected host(s)' list. A 'New Server Class Name' field contains 'eng_webservers'. A yellow callout box lists two bullet points:

- Creates new server class or uses existing one
- Creates a new app for this input (or updates existing)

On the right side of the wizard, there's a detailed configuration panel for 'File & Directories' input type. It shows a configuration for '/opt/log/www2' with 'optional' included and excluded lists. The configuration text describes monitoring files or directories, setting up TCP/UDP ports, using scripts, and using Splunk Assist Instance Identifier.

Editing Forwarded Inputs

The screenshot shows the 'Forwarded inputs' configuration page in Splunk. The left sidebar lists 'Type' (Windows Event Logs, Files & Directories), 'Filebeat' (disabled), and 'Logstash' (disabled). The 'Files & directories' section is selected, highlighted by a red circle labeled '1'. It shows two items: '/opt/log' and '/opt/www2/access.log', both with 'Host' set to 'None' and 'Source type' set to 'Automatic'. A red circle labeled '2' highlights the second item. The main panel on the right is titled 'Forwarded inputs' and contains the following sections:

- Type**: Describes how to collect event logs from forwarders.
- Host**: Sets the host field value to 'constant value' with 'www-03' specified.
- Source type**: Sets the source type to 'Automatic'.
- Index**: Sets the destination index to 'sales'.
- Advanced options**: Includes 'Allowrule' (set to 'www') and 'Denyrule' (set to 'secure').

A green arrow points to the 'Forwarded inputs' title, and a red circle labeled '3' is positioned above the 'Host' section.

Verifying Forwarder Management

Forwarder Management

Repository Location: \$SPLUNK_HOME/etc/deployment-apps

2 Clients PHONED HOME IN THE LAST 24 HOURS **0** Clients DEPLOYMENT ERRORS

Apps (13) Server Classes (14) Clients (2)

All Server Classes ▾ filter

14 Server Classes 10 Per Page ▾

Last Reload	Name
3 days ago	100_IngestAction_AutoGenerated
3 days ago	dcrusher_tcp
3 days ago	devserver_vmail
3 days ago	devserver_vmstat
3 days ago	eng_badge_access
3 days ago	eng_crashlog
3 days ago	eng_dreamcrusherXML
3 days ago	eng_hf
3 days ago	eng_sales_entries
3 days ago	eng_sysmonitor



Server Class: eng_crashlog

Back to Forwarder Management

1 App IN THE SERVER CLASS **1** Client IN THE SERVER CLASS **100%** Clients DEPLOYED APPS SUCCESSFULLY

Apps Edit

Deployed Successfully ▾ filter

1 Apps 10 Per Page ▾

Name	Actions	After Installation	Clients
_server_app_eng_crashlog	Edit ▾	Enable App	1 deployed

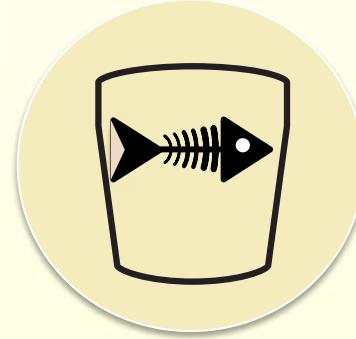
Clients Edit

Phone Home: All ▾ All Clients ▾ filter

1 Clients 10 Per Page ▾

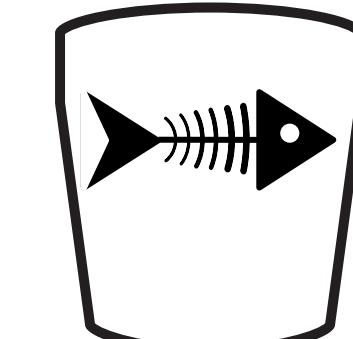
i	Host Name	Client Name	Instance Name	IP Address	Actions	Machine Type	Deployed Apps	Phone Home
>	ip-10-0-0-100	7EB2AAEA-1C69-44E7-B6F1- XXXXXXXXXX	engdev203	10.0.0.100	Delete Record	linux-x86_64	11 deployed	a few seconds ago

What is the Fishbucket?

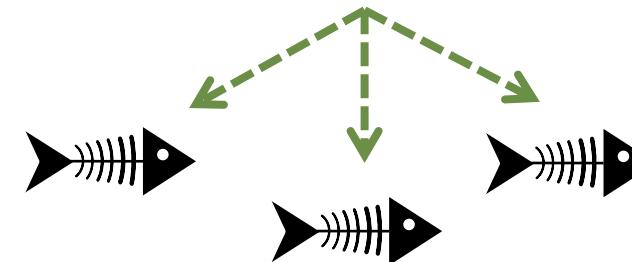


Fishbucket

- Allows Splunk to track monitored input files
- Contains file metadata which identifies a pointer to the file, and a pointer to where Splunk last read the file
- Exists on all Splunk instances
- Stored in a special subdirectory found at **SPLUNK_DB/fishbucket**



Fishbucket index



One record per monitored file

Includes:

- **Head:** Pointer to the file
- **Tail:** Pointer showing where Splunk last left off indexing in the file

Editing Inputs and Re-indexing Data

- Editing the **inputs.conf**
 - Only applies changes to new data
 - Does not change or cause re-indexing of existing ingested data
- To re-index:
 1. Delete the old, erroneous data on the indexers
 - ▶ May require assistance from the system administrator
 2. Change the **inputs.conf** on the deployment server (or forwarders)
 3. Reset the fishbucket checkpoint on the involved forwarders
 4. Restart Splunk forwarders

Resetting Input File Monitors

1. Stop Splunk
2. Reset applicable file monitors on the source system
 - Individually for each source:

```
splunk cmd btprobe -d SPLUNK_DB/fishbucket/splunk_private_db  
--file <source> --reset
```

- All sources (use only on test systems / with extreme caution):

```
splunk clean eventdata -index _thefishbucket
```

Or

```
rm -r SPLUNK_DB/fishbucket
```

3. Start Splunk

Warning

Resetting the fishbucket forces re-indexing of all file monitors affected. The re-indexing results in more license usage.

Question: Directory Data Source

Which of the following is not true (false) when defining a directory as a data source?

- A. Traverses directory and monitors all discovered text files
- B. Unzips compressed files automatically before ingestion
- C. Includes new files added to the directories
- D. Does not handle log file rotation

Answer: Directory Data Source

Which of the following is not true (false) when defining a directory as a data source?

- A. Traverses directory and monitors all discovered text files
- B. Unzips compressed files automatically before ingestion
- C. Includes new files added to the directories
- D. Does not handle log file rotation**

Log file rotation is handled when defining a directory as a data source.

Question: File and Directory Matching

Which of the following is not a valid match for this **inputs.conf** stanza entry?

[monitor:///var/log/www*/secure.*]

- A. /var/log/www1/secure.log
- B. /var/log/www1/secure.1
- C. /var/log/www1/logs/secure.log
- D. /var/log/www2/secure.log

Answer: File and Directory Matching

Which of the following is not a valid match for this **inputs.conf** stanza entry?

[monitor:///var/log/www*/secure.*]

- A. /var/log/www1/secure.log
- B. /var/log/www1/secure.1
- C. /var/log/www1/logs/secure.log**
- D. /var/log/www2/secure.log

To match all of these – including subdirectories – use the ellipsis with this stanza entry:

[monitor:///var/log/.../secure.*]

Question: Changing the Host Value

If the host value is changed, what must be done to reflect this new host value for already ingested data?

- A. Already ingested data will automatically reflect this host value
- B. Delete the data and re-ingest
- C. Delete the data, reset the fishbucket and re-ingest
- D. There is no procedure for handling this

Log file rotation is handled when defining a directory as a data source.

Answer: Changing the Host Value

If the host value is changed, what must be done to reflect this new host value for already ingested data?

- A. Already ingested data will automatically reflect this host value
- B. Delete the data and re-ingest
- C. Delete the data, reset the fishbucket and re-ingest**
- D. There is no procedure for handling this

Already ingested data cannot be modified. To change the host value for this data, it must first be deleted. The fishbucket must be reset to convince Splunk that the data hasn't been ingested yet, and then the data should be re-ingested with the new host value.

Question: Resetting Individual Source

Which of the following steps is *not* used when resetting an individual file monitor in the fishbucket?

- A. `splunk stop`
- B. `splunk cmd btprobe -d <fb_path> --file <source> --reset`
- C. `splunk clean eventdata -index _thefishbucket`
- D. `splunk start`

Answer: Resetting Individual Source

Which of the following steps is *not* used when resetting an individual file monitor in the fishbucket?

- A. `splunk stop`
- B. `splunk cmd btprobe -d <fb_path> --file <source> --reset`
- C. `splunk clean eventdata -index _thefishbucket`**
- D. `splunk start`

Reset applicable file monitors on the source system by stopping Splunk, resetting an individual source with the **splunk cmd btprobe** command, and then starting Splunk.

The **splunk clean eventdata** command is used to reset *all* sources for that client.

Module 6 Lab

Time: 20-25 minutes

Description: File Monitor Input

Tasks:

- Add a monitor input for a remote directory on UF2 to the **test** index
- Modify the **inputs.conf** file using the following caveats
 - Send the source logs to the **sales** index
 - Override the **default-host** name value
 - Monitor only the **www.*** sub-directories
 - Exclude the indexing of the **secure.log** files
- Re-deploy the **inputs.conf** file

Module 7: Network Inputs

Module Objectives

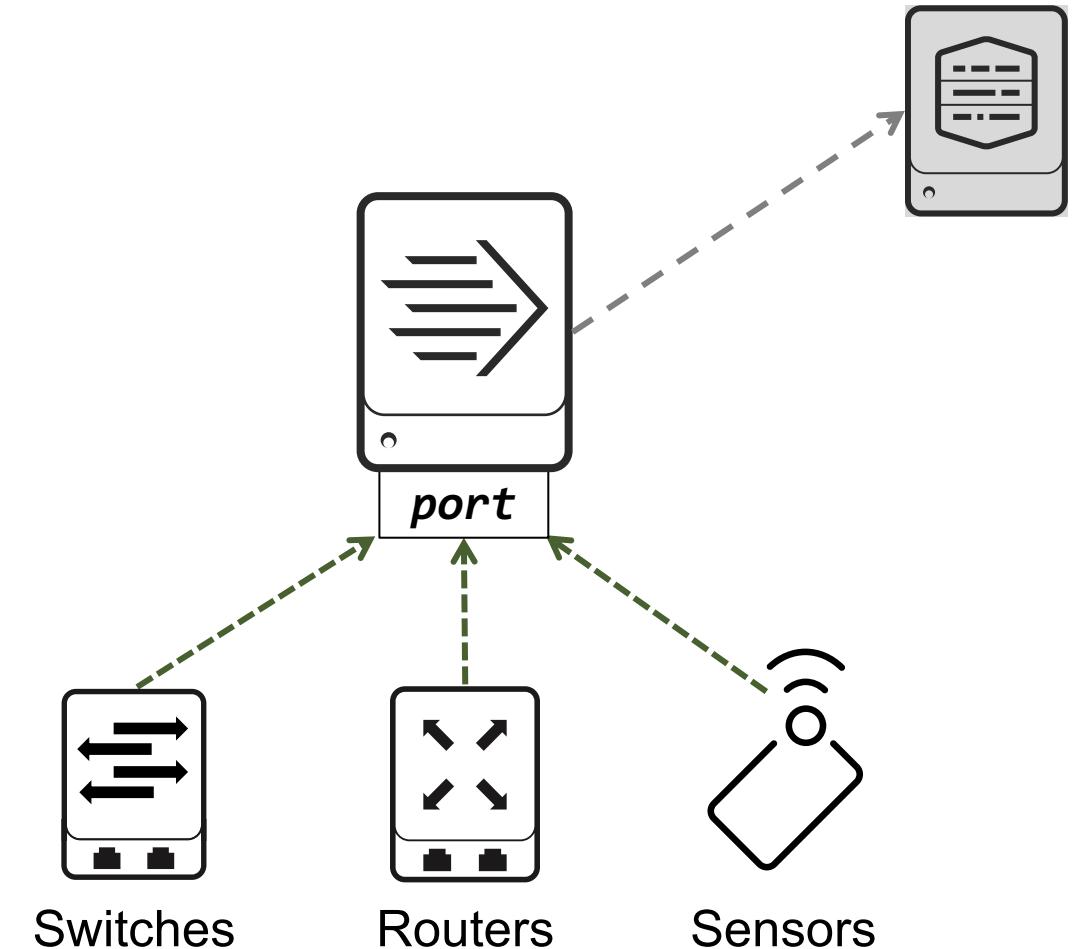
- Create network (TCP and UDP) inputs
- Describe optional settings for network inputs

Network Inputs



Network Inputs

- Input data sent to a Splunk instance on a TCP/UDP port (for example: Syslog)
- Adds a layer of resiliency (buffering, load balancing, cloning, indexer restarts)
- Can minimize indexer workload by managing network connections on the forwarder (which can additionally bridge network segments)



Adding Network Input

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.

Source name override ? **Port ?** **Only accept connection from ?**

TCP **UDP**

9001
Example: 514

dns_10-1-2-3
host:port

10.1.2.3
example: 10.1.2.3, !badhost.splunk.com, *.splunk.com

If not specified, default:
• TCP: **tcp:<port>**
• UDP: **udp:<port>**

• If specified, only accepts connections from this host
• If unspecified: all hosts are allowed

Optional Network Input Settings

- Edit the stanza directly to fine-tune input settings:
 - Metadata override
 - Sender filtering options
 - Network input queues
 - Memory queues
 - Persistent queues

```
[udp://<[host:]port>]  
connection_host = dns  
sourcetype=<string>
```

```
[tcp://<[host:]port>]  
connection_host = dns  
source=<string>
```

Examples:

```
[udp://514]  
connection_host = dns  
sourcetype=syslog
```

```
[tcp://10.1.2.3:9001]  
connection_host = dns  
source = dns_10-1-2-3
```

Network Input: Host Field

- Set in **inputs.conf** with the **connection_host** attribute:
 - **dns** (default for TCP inputs)
 - The host is set to a DNS name using reverse IP lookup
 - **ip** (default for UDP inputs)
 - The host is set to the originating host's IP address
 - **none** (Custom in the UI)
 - Requires explicit setting of the **host** value

```
[tcp://9002]
sourcetype=auth-data
connection_host=dns

[tcp://9003]
sourcetype=ops-data
connection_host=ip

[tcp://9001]
sourcetype=dnslog
connection_host=none
host=dnsserver
```

The screenshot shows the 'Source' configuration page in Splunk. The 'Host' section is highlighted with a green rounded rectangle. It contains fields for 'Set host' (radio buttons for 'IP', 'DNS', and 'Custom'), where 'Custom' is selected and the value 'dnsserver' is entered. Below this is the 'Index' section, which has a field for 'Index' with the value 'test'.

Source

Source name override If set, overrides the default source v

Source type

Set sourcetype for all events from this source.

Set sourcetype

Source type If this field is left blank, the default v

More settings

Host

Set the host with this value.

Set host IP DNS Custom

Index

Set the destination index for this source.

Index

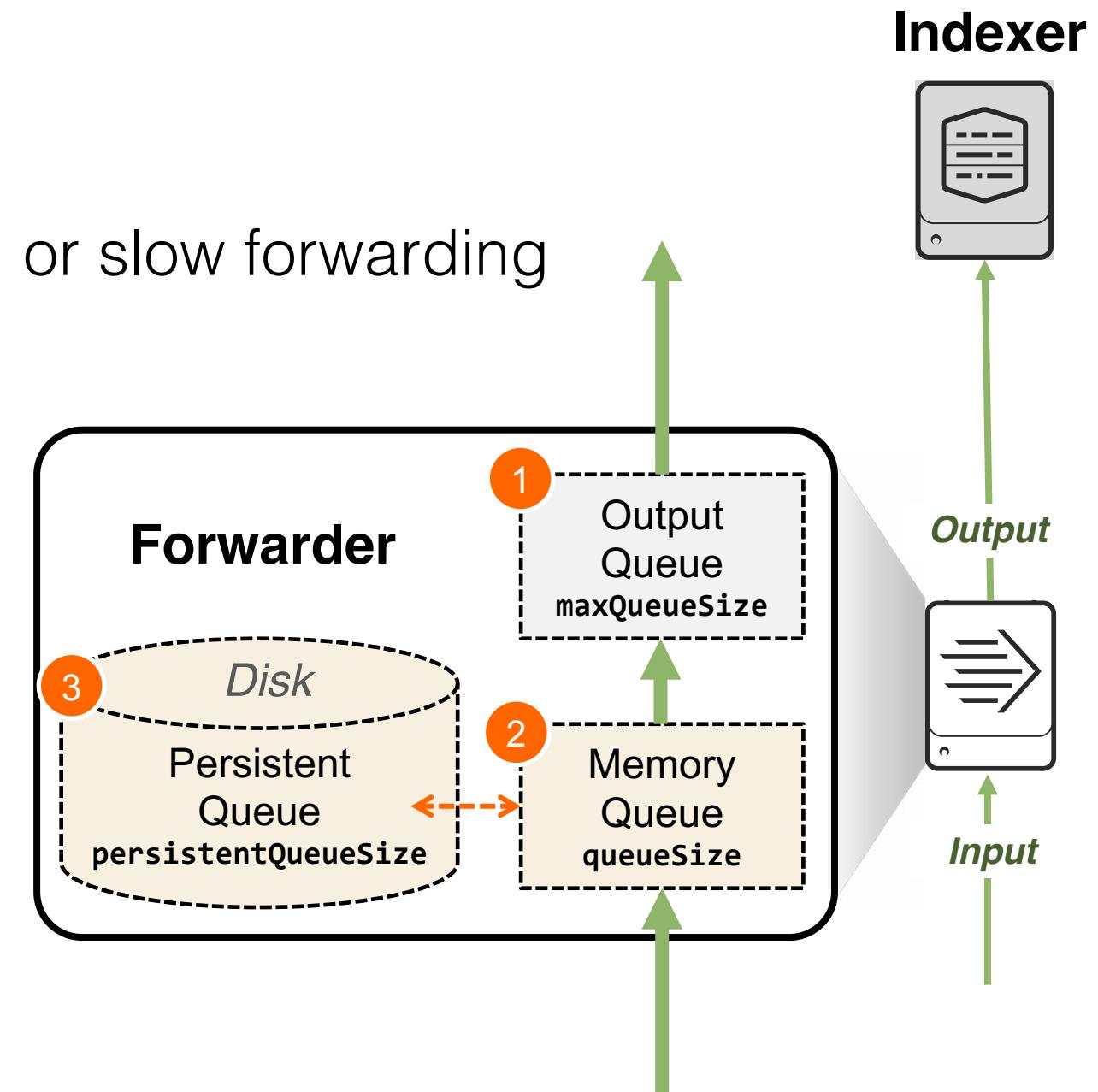
Network Input: Sender Filtering Options

- Specify which input streams are accepted by Splunk
- Example:
 - Network devices are sending syslog reports (UDP 514) to the Splunk network input, but want to accept UDP inputs selectively
- Use **acceptFrom = <network_acl>**
 - List address rules separated by commas or spaces
 - Available formats include:
 - ▶ Single IPv4 or IPv6 address
 - ▶ CIDR block of addresses
 - ▶ DNS name
 - ▶ Wildcards: * (any), ! (not)

```
[udp://514]
sourcetype=syslog
connection_host=ip
acceptFrom=!10.1/16, 10/8
```

Network Input: Queues

- Provide input flow control
- Apply to TCP, UDP, scripted input
- Control network data bursts, slow resources, or slow forwarding
 1. If indexers can't be reached:
 - Data is stored in the output queue
 2. If the output queue is full:
 - Data is stored in the memory queue
 3. If the memory queue is full:
 - Data is stored in the persistent queue
- Persistent queue preserves across restarts
 - Not a solution for input failure



Network Input: Setting Queue Attributes

- Memory queue
 - Set with **queueSize** (default = 500 KB)
 - Memory-resident queue that buffers data before forwarding
 - Useful if indexer receives data slower than forwarder is acquiring it
 - Independent of forwarder's **maxQueueSize** attribute
- Persistent queue
 - Set with **persistentQueueSize** (doesn't exist by default)
 - Provides additional, file-system buffering of data
 - Written to **SPLUNK_HOME/var/run/splunk/...**
 - Useful for high-volume data and in the case of network outage to indexers

inputs.conf

```
[tcp://9001]
queueSize=10MB
persistentQueueSize=5GB
```

Special Handling and Best Practices

UDP

- Splunk merges UDP data until it finds a timestamp by default
- Default behavior can be overridden during the parsing phase

Syslog

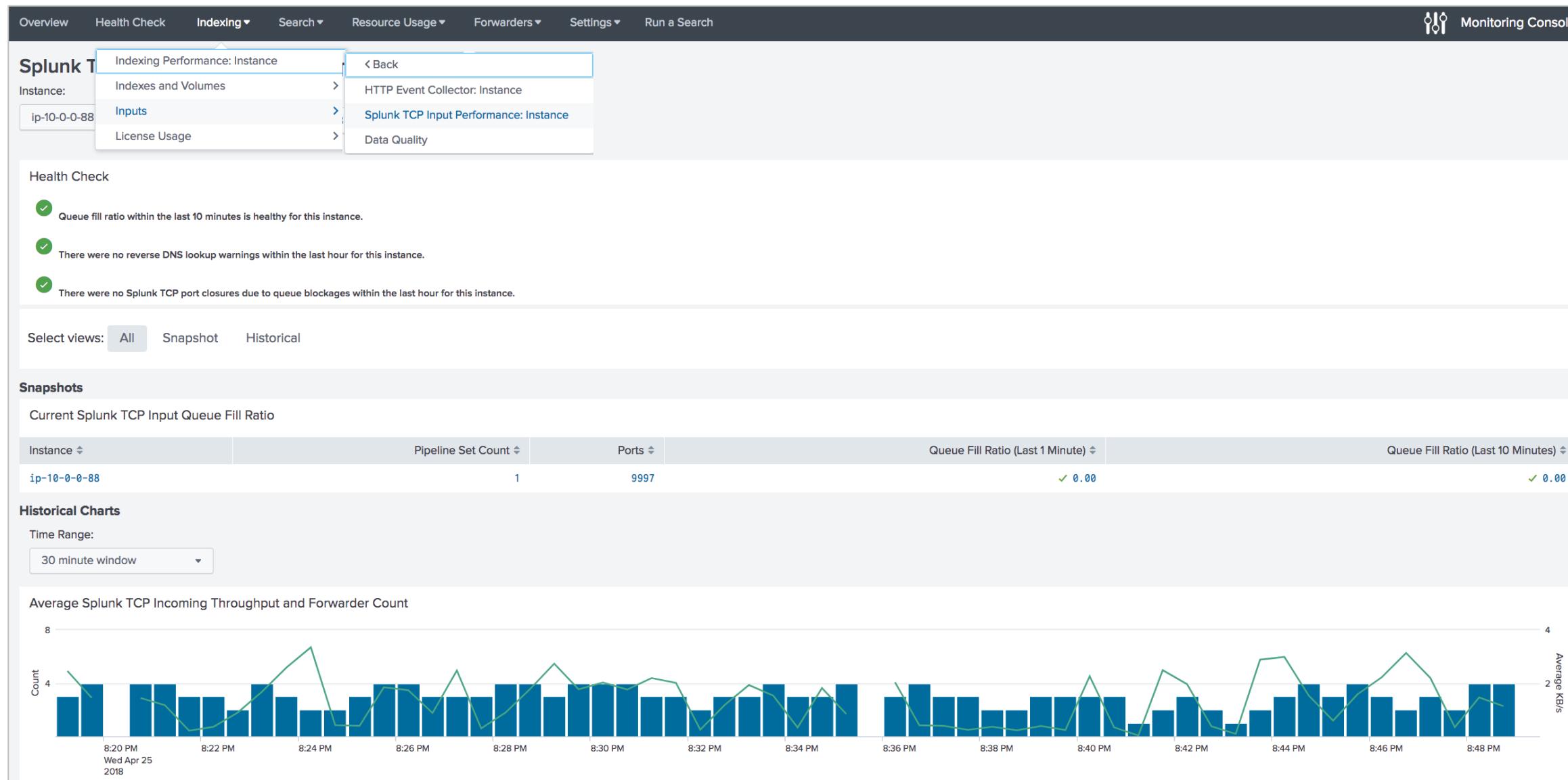
- Send data to a syslog collector that writes into a directory structure (for example: `/var/log/syslog/hostname/filename.txt`)
- Monitor the directory and use **host_segment**
- docs.splunk.com/Documentation/Splunk/latest/Data/HowSplunkEnterprisehandlessyslogdata

SNMP traps

- Write the traps to a file and use the monitor input
- docs.splunk.com/Documentation/Splunk/latest/Data/SendSNMPEventstoSplunk

Monitoring with MC: Splunk TCP Inputs

For remote input monitoring, click Indexing > Inputs > Splunk TCP Input Performance

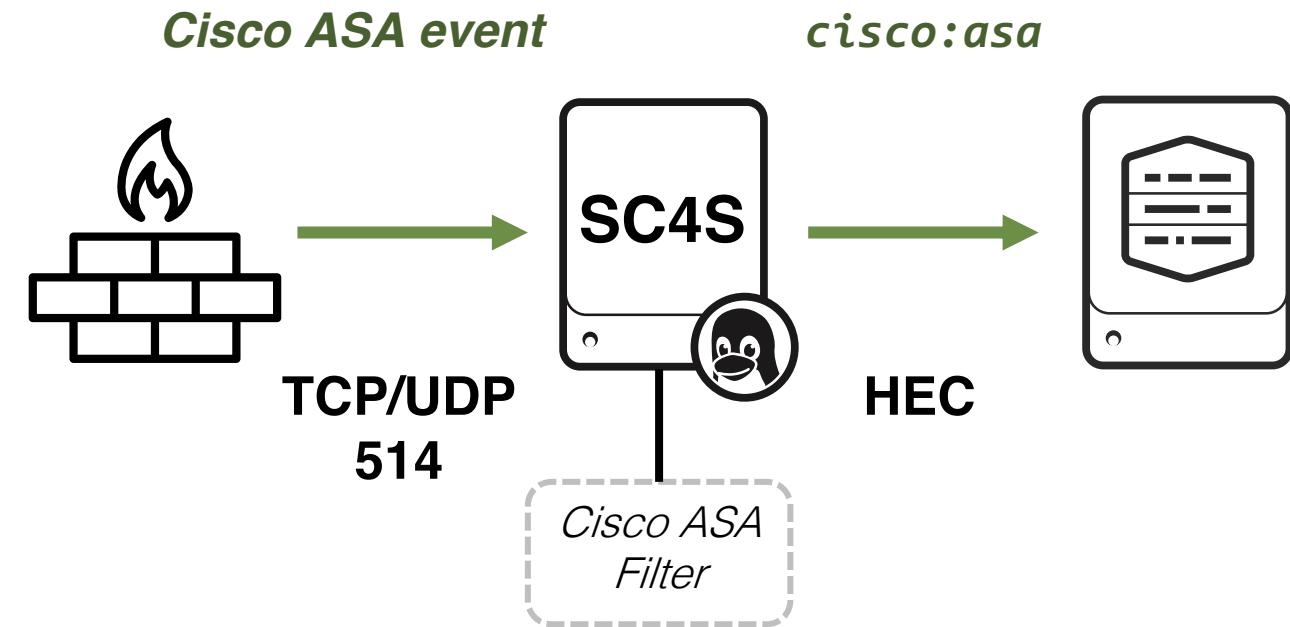


Understanding Splunk Connect for Syslog (SC4S)



Splunk Connect for Syslog

- Lower burden of getting syslog into Splunk
- Consistent, documented, repeatable
- Turnkey data ingestion for common source types
- Lower Splunk overhead for improved scaling and data distribution
- Containerized Syslog appliance



Identify / Parse / Format

Question: Specifying a Host Value

For a network input how can you specify a host value instead of using what is automatically assigned?

- A. `inputs.conf`: Set `connection_host=dns` and `host=<hostname>`
- B. `inputs.conf`: Set `connection_host=ip` and `host=<hostname>`
- C. `inputs.conf`: Set `connection_host=none` and `host=<hostname>`
- D. This can only be performed in Splunk Web

Answer: Specifying a Host Value

For a network input how can you specify a host value instead of using what is automatically assigned?

- A. `inputs.conf`: Set `connection_host=dns` and `host=<hostname>`
- B. `inputs.conf`: Set `connection_host=ip` and `host=<hostname>`
- C. `inputs.conf`: Set `connection_host=none` and `host=<hostname>`**
- D. This can only be performed in Splunk Web

Under the stanza in **`inputs.conf`** set the **`connection_host=none`** and specify the **`host`** value. This can also be performed in Splunk Web by selecting a Host value of Custom (instead of IP or DNS) and then specifying the host value.

Question: Network Input Queues

Which of the following is incorrect about network input queues?

- A. If indexers can't be reached, data is stored in output queue
- B. If output queue is full, data is stored in memory queue
- C. If memory queue is full, data is stored in persistent queue
- D. If persistent queue is full, data is stored in parsing queue

Answer: Network Input Queues

Which of the following is incorrect about network input queues?

- A. If indexers can't be reached, data is stored in output queue
- B. If output queue is full, data is stored in memory queue
- C. If memory queue is full, data is stored in persistent queue
- D. If persistent queue is full, data is stored in parsing queue**

If the persistent queue is full, additional network data must be dropped. To prevent dropped network event data: create a larger persistent queue using the **persistentQueueSize** value.

Module 7 Lab

Time: 15 minutes

Description: Network Inputs

Tasks:

- Create and test a simple TCP-based network input
- On the deployment/test server, add a test network input
- Modify the host value for the test network input

Note 

Your instructor will run a script to send TCP data ports on the forwarder.

Use your assigned port to listen for the TCP data.

Module 8: Scripted Inputs

Module Objectives

Create a basic scripted input

Scripted Inputs



Scripted Inputs

- Schedules script execution and indexes the output
- Used to collect diagnostic data from OS commands (such as **top**, **netstat**, **vmstat**, **ps** etc.)
- Used by many Splunk apps to gather information from the OS or other server applications
- Can gather transient data that cannot be collected with Monitor or Network inputs (Examples: APIs, message queues, Web services, custom transactions)
- Supports Shell (**.sh**), Batch (**.bat**), PowerShell (**.ps1**) and Python (**.py**) scripts

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 Splunk to listen on a network port.

Scripts

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

Warning



Splunk only executes scripts from:

- **SPLUNK_HOME/etc/apps/<app_name>/bin**
- **SPLUNK_HOME/bin/scripts**
- **SPLUNK_HOME/etc/system/bin**

Defining a Scripted Input

1. Develop and test the script
2. Test your script from the context of a Splunk app
 - Copy the script to the app's **bin** directory on a test/dev server
 - Run script using the **splunk cmd scriptname** command
Example: **splunk cmd SPLUNK_HOME/etc/apps/<app>/bin/myscript.sh**
3. Deploy the script to production servers, for example if using a deployment server:
 - Copy script to **SPLUNK_HOME/etc/deployment-apps/<app>/bin/**
 - Deploy script using Add Data > Forward from Splunk Web
4. Verify the output of the script is being indexed

Scripted Input Stanza

```
[script://<cmd>]  
passAuth = <username>  
host = <as indicated>  
source = <defaults to script name>  
sourcetype = <defaults to script name>  
interval = <number in seconds or cron syntax>
```

inputs.conf

Use **passAuth** to run the script as a specified OS user; Splunk passes an authorization token via stdin to the script

Interval is the time period between script executions (default: 60 seconds)

Warning



- Splunk only executes scripts from:
- **SPLUNK_HOME/etc/apps/<app_name>/bin**
 - **SPLUNK_HOME/bin/scripts**
 - **SPLUNK_HOME/etc/system/bin**

Scripted Inputs Example

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.

```
[script://./bin/myvmstat.sh]
disabled = false
interval = 60.0
source = vmstat
sourcetype = myvmstat
```

inputs.conf

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

\$SPLUNK_HOME/bin/scripts ▾

Script Name

myvmstat.sh ▾

Command ?

\$SPLUNK_HOME/bin/scripts/myvmstat.sh

Interval Input ?

In Seconds ▾

Interval ?

60.0

In Seconds

Cron Schedule

Source name override ?

vmstat

Editing Scripted Inputs

The screenshot illustrates the process of editing a scripted input in Splunk. It consists of two main panels: a left panel showing the list of scripts and a right panel showing the configuration details for a specific script.

Left Panel (Script List):

- Section:** Script
- Path:** Data inputs > Script
- Count:** Showing 1-1 of 1 item
- Filter:** A search bar with a magnifying glass icon.

Right Panel (Configuration):

- Path:** \$SPLUNK_HOME/etc/apps/_server_app_devserver_vmstat/bin/myvmstat.sh
Data inputs > Script > \$SPLUNK_HOME/etc/apps/_server_app_devserver_vmstat/bin/myvmstat.sh
- Source:**
 - Interval:** 120.0 (Number of seconds to wait before running the command again, or a valid cron schedule.)
 - Source name override:** (If set, overrides the default source value for your script entry (script:path_to_script).)
- Source type:**
 - Set sourcetype:** Manual
 - Source type *:** vmstat (If this field is left blank, the default value of script will be used for the source type.)
- Host:** Host field value (empty field)
- Index:**
 - Set the destination index for this source:** (empty field)
 - Index:** A dropdown menu showing available indices:
 - default
 - history
 - itops** (selected)
 - main
 - summary
 - test

Scripted Input Buffering

- Potential loss of data
 - Forwarder running the script is not able to connect to the indexer due to networking problems
- Workaround
 - The **queueSize** and **persistentQueueSize** attributes can be set for scripted input (in the **[script://...]** stanza)
 - Buffers data on the forwarder when the network or indexer is unavailable

Alternates to Using Scripted Input

Monitor a file containing the output of the script

- Allows the use of Splunk's simple configuration of monitoring files
- Takes advantage of the file system and Splunk's robust file monitoring capabilities
- Can easily recover even when forwarder goes down
- Configured with a scripted log file:
 1. Schedule the script to run using an external scheduler (such as cron)
 2. Append script output to a log file
 3. Set up a monitor input to ingest the log file

Use Splunk's modular input

- Simple UI for configuring a scripted input
- Appears as its own type of input
- docs.splunk.com/Documentation/Splunk/latest/AdvancedDev/ModInputsScripts

Question: Script Interval Settings

An administrator wants to ingest the output from a custom script **splk.sh**. Which method would not provide the results desired?

- A. Create a scripted input using a [script://splk.sh] stanza
- B. Ingest the script as a file input
- C. Send the script output to a file and ingest that output file
- D. Use Splunk modular inputs

Answer: Script Interval Settings

An administrator wants to ingest the output from a custom script **splk.sh**. Which method would not provide the results desired?

- A. Create a scripted input using a [script://splk.sh] stanza
- B. Ingest the script as a file input**
- C. Send the script output to a file and ingest that output file
- D. Use Splunk modular inputs

Ingesting the script as a file input would ingest the script itself, and not the output generated by the script.

Question: Script Interval Settings

For a script configured under stanza **[script://myscript.sh]**, which of the following is an invalid **interval** field?

- A. 30
- B. 30.5
- C. 6-7
- D. 0/30 6-7 * * *

Answer: Script Interval Settings

For a script configured under stanza **[script://myscript.sh]**, which of the following is an invalid **interval** field?

- A. 30
- B. 30.5
- C. **6-7**
- D. 0/30 6-7 * * *

The interval field specifies how often to run the script. It may be a fractional number (run every **30** or **30.5** seconds) or a **cron** schedule in format "**<minute> <hour> <day_of_month> <month> <day_of_week>**" (**0/30 6-7 * * *** runs any day at 6:00 AM, 6:30 AM, 7:00 AM, and 7:30 AM).

Module 8 Lab

Time: 10 minutes

Description: Scripted Inputs

Tasks:

- Add a scripted input on your deployment server
- Deploy the scripted input to your forwarder
- Disable the forward scripted input

Module 9:

Agentless Inputs

Module Objectives

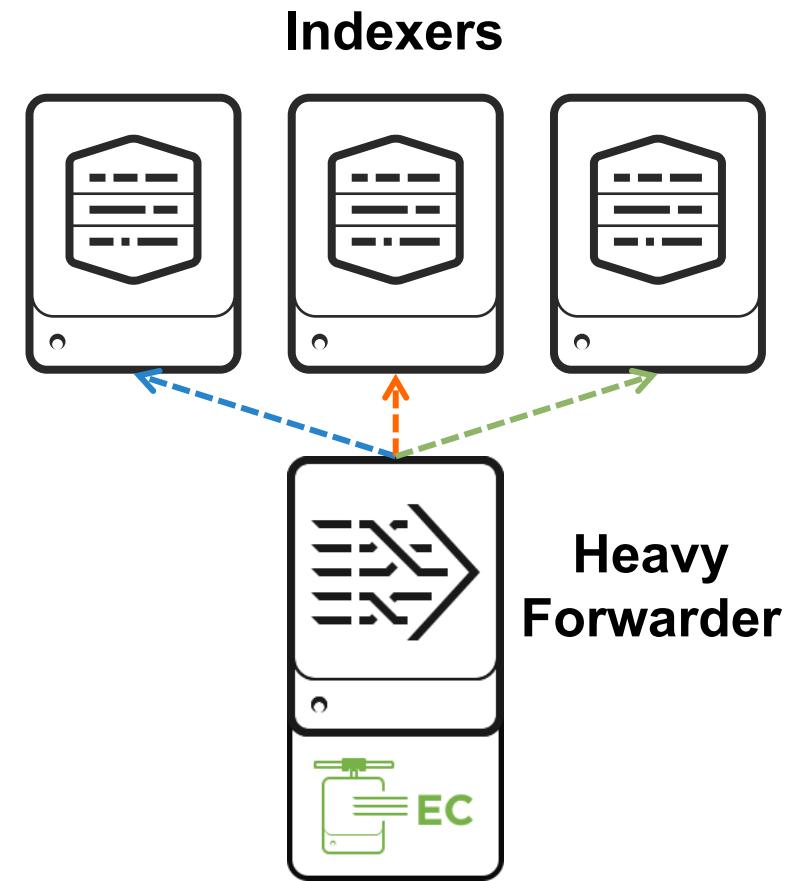
- Configure Splunk HTTP Event Collector (HEC) agentless input
- Describe Splunk App for Stream

HTTP Event Collector (HEC) Agentless Inputs



HTTP Event Collector (HEC)

- A token-based HTTP input that is secure and scalable
- Sends events to Splunk without the use of forwarders (such as log data from a web browser, automation scripts, or mobile apps)
- Can facilitate logging from distributed, multi-modal, and/or legacy environments

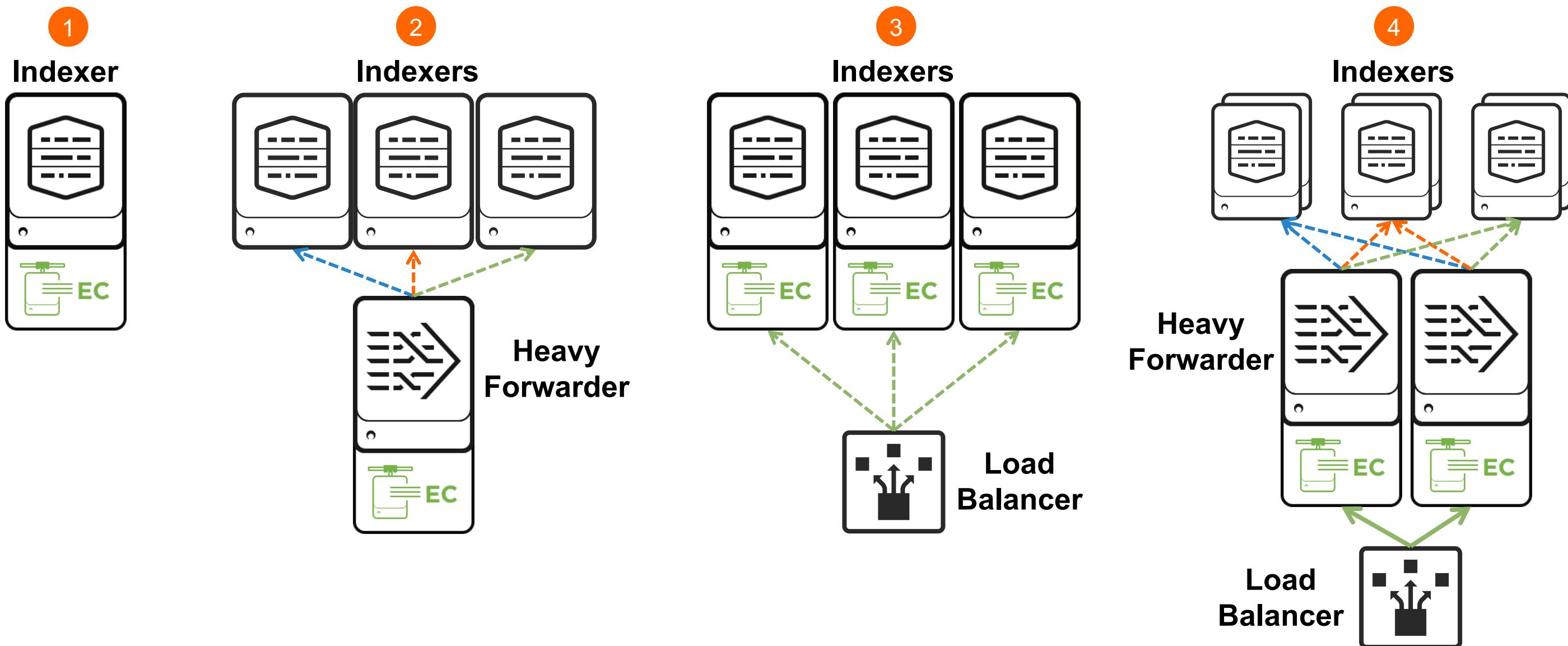


Event collector enabled
to receive HTTP events



Distributed HEC Deployment Options

HEC can scale by taking advantage of Splunk distributed deployment



Configuring HTTP Event Collector

1. Enable the HTTP event collector (disabled by default)
 - Navigate to Settings > Data inputs > HTTP Event Collector
 - Click Global Settings > Enabled
2. Generate a HTTP-input token by clicking New Token
 - The Add Data workflow starts
 - Name the input token and optionally set the default source type and index

The screenshot shows the Splunk 'HTTP Event Collector' configuration page. At the top, there's a navigation bar with 'Data Inputs」 > HTTP Event Collector'. Below it, a summary shows '1 Tokens' and filtering options. A 'Global Settings' button and a 'New Token' button are visible. The main table lists a single token: 'iot_sensors' with a value of 'af58d9a4-4df6-4fda-a209-1c3988e1ceaf'. The token is highlighted with a green box and labeled '2'. The table has columns for Name, Actions, Token Value, Source Type, Index, and Status.

Name	Actions	Token Value	Source Type	Index	Status
iot_sensors	Edit Disable Delete	af58d9a4-4df6-4fda-a209-1c3988e1ceaf	test		Disabled

Sending HTTP Events from a Device

- Create a request with its authentication header to include the input token
 - Can send data from any client
 - Simplify the process by using the Splunk logging libraries
 - ▶ Supports JavaScript, Java and .NET
- POST data in JSON format to the token receiver

```
curl "http[s]://<splunk_server>:8088/services/collector"
-H "Authorization: Splunk <generated_token>"
-d '{
    "host": "xyz",
    "sourcetype": "f101_S2",
    "source": "sensor125",
    "event": {"message": "ERR", "code": "401"}
}'
```

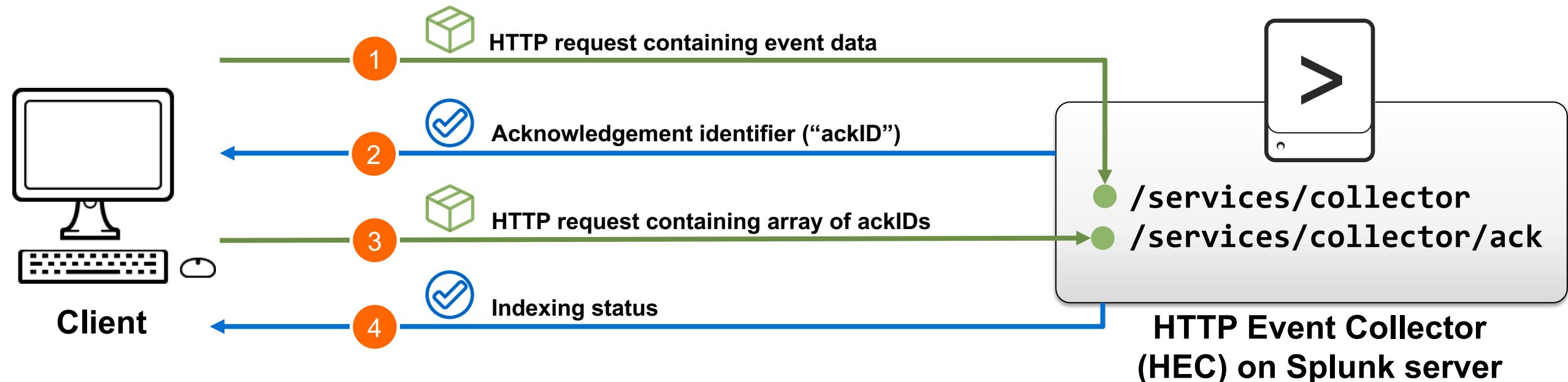
HTTP Event Collector Options

- Enable HEC acknowledgments
- Send *raw* payloads
- Configure dedicated HTTP settings

docs.splunk.com/Documentation/Splunk/latest/Data/UseHECusingconffiles

HEC Indexer Acknowledgement

1. Request sent from client to the HEC endpoint using a token, with indexer acknowledgement enabled
2. Server returns an acknowledgement identifier (**ackID**) to client
3. Client can query the Splunk server with the identifier to verify if all events in the send request have been indexed (HTTP request containing array of **ackID**'s)
4. Splunk server responds with status information of each queried request



HEC Indexer Acknowledgement Notes

- **ACK** is configured at the token level
- Each client request must provide a *channel* (a unique identifier created by the client)
- When an event is indexed, the channel gets the **ackID**
- Client polls a separate endpoint using one or more **ackID**'s
- After an **ACK** has been received, it is released from memory
- Client polling functionality is not built into Splunk and requires custom programming

docs.splunk.com/Documentation/Splunk/latest/Data/AboutHECIDXAck

Configure a new token for receiving data over HTTP. [Learn More ↗](#)

Name	mainframe
Source name override ?	optional
Description ?	optional
Output Group (optional)	None ▾ None
<input checked="" type="checkbox"/> Enable indexer acknowledgement	

Sending Raw Payloads to HEC

- Example:
 - Application developers want to send data in a proprietary format
- Solution:
 - HEC allows any arbitrary payloads, not just JSON
- Configuration Notes:
 - No special configuration required
 - Must use channels similar to ACK
 - Supports ACK as well
 - Events MUST be bounded within a request

```
curl "http[s]://<splunk_server>:8088/services/collector/raw?channel=<client_provided_channel>"  
-H "Authorization: Splunk <generated_token>"  
-d 'ERR,401,-23,15,36'
```

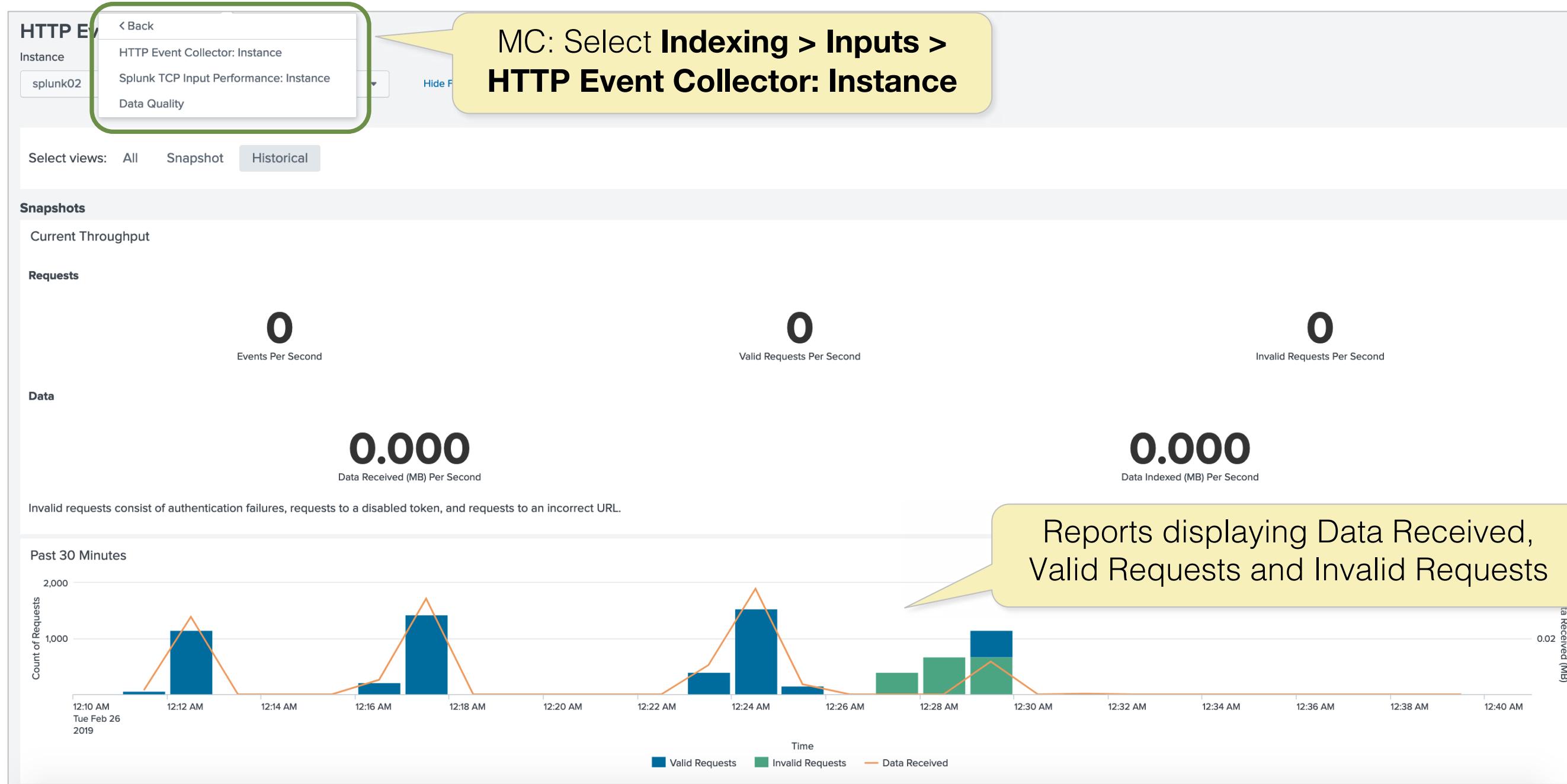
Configuring Dedicated HTTP Settings

- Example:
 - Splunk admins want to limit who can access the HEC endpoints
- Solution:
 - Manually add the dedicated server settings in **inputs.conf**
- Configuration Notes:
 - Available attributes under the **[http]** stanza
 - Configure a specific SSL cert for HEC and client certs
 - Enable cross-origin resource sharing (CORS) for HEC
 - Restrict based on network, hostnames, etc.

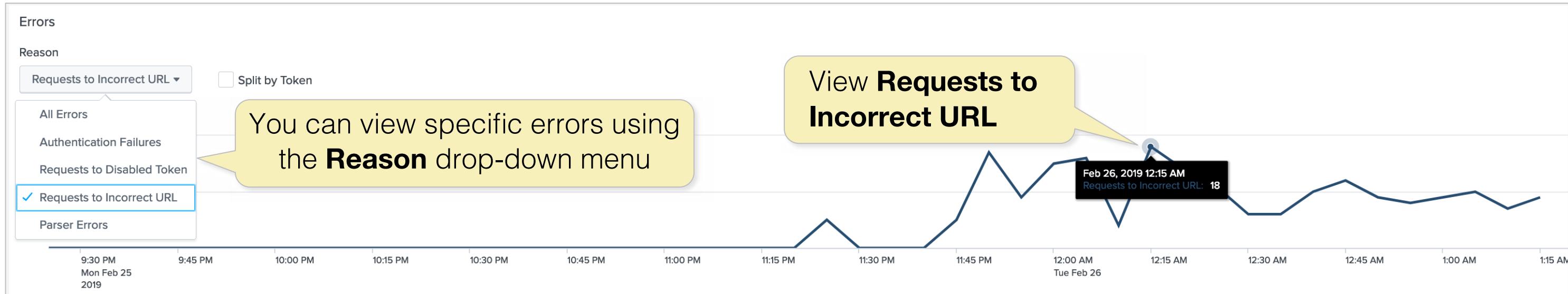
inputs.conf

```
[http]
enableSSL = 1
crossOriginSharingPolicy = *.splunk.com
acceptFrom = "!45.42.151/24, !57.73.224/19, *"
```

Monitoring HEC with MC



Monitoring HEC with MC – Viewing Errors



HTTP Event Collector (HEC) Documentation

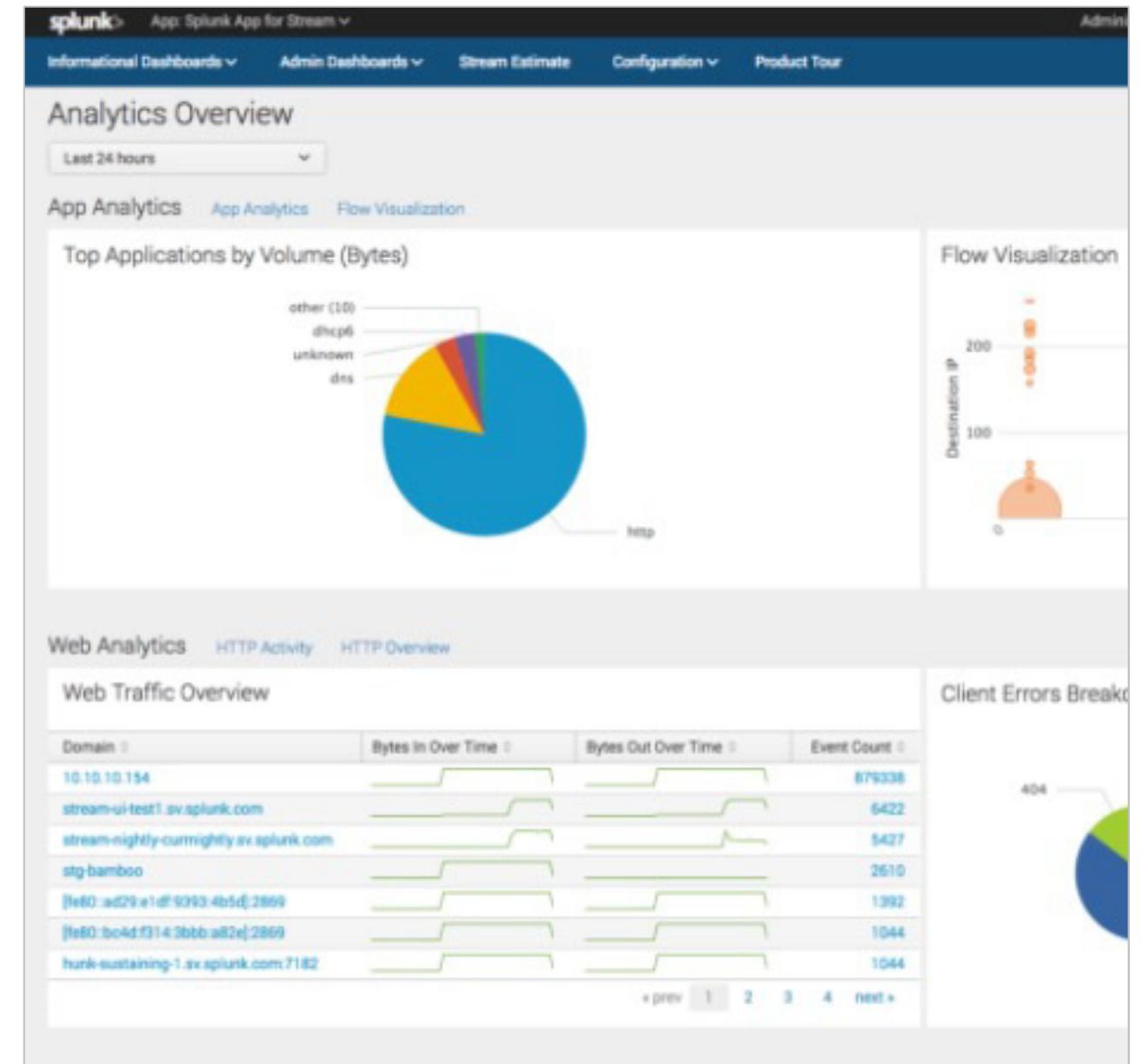
- Refer to:
 - Introduction to Splunk HTTP Event Collector
dev.splunk.com/enterprise/docs/devtools/httpeventcollector/
 - Blogs: Tips & Tricks on HTTP Event Collector
www.splunk.com/en_us/blog/tips-and-tricks/http-event-collector-your-direct-event-pipe-to-splunk-6-3.html

Understanding Splunk App for Stream



Splunk App for Stream

- Part of purpose-built wire data collection and analytics solution from Splunk
- An alternative way to collect “difficult” inputs
 - Database servers without forwarders
 - Network traffic not visible to web logs
- Able to read data off the wire
- Supports Windows, Mac, and Linux



Question: Configuring HEC

Where is the HTTP event collector configured?

- A. On any Splunk instances
- B. On an indexer or search head
- C. On an indexer or heavy forwarder
- D. On a heavy or universal forwarder

Answer: Configuring HEC

Where is the HTTP event collector configured?

- A. On any Splunk instances
- B. On an indexer or search head
- C. On an indexer or heavy forwarder**
- D. On a heavy or universal forwarder

The HTTP event collector is configured on the indexer or heavy forwarder. It is not configured on universal forwarders.

Question: Understanding HEC

Which of the following is true about the HTTP Event Collector (HEC)?

- A. HEC is enabled by default.
- B. Use the default token when possible.
- C. Provides the ability to have indexer acknowledgement
- D. Must be monitored using command line

Answer: Understanding HEC

Which of the following is true about the HTTP Event Collector (HEC)?

- A. HEC is enabled by default.
- B. Use the default token when possible.
- C. Provides the ability to have indexer acknowledgement**
- D. Must be monitored using command line

HEC is disabled by default, and you must create tokens manually. HEC can be monitored in the MC under Indexing > Inputs > HTTP Event Collector: Instance.

Module 9 Lab

Time: 15 minutes

Description: Agentless Inputs with HTTP Event Collector

Tasks:

- Enable HTTP event collector on the deployment/test server
- Create a HTTP event collector token
- Send HTTP events from your UF1 (**10.0.0.50**)

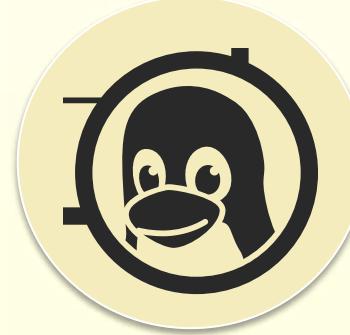
Module 10:

Operating System Inputs

Module Objectives

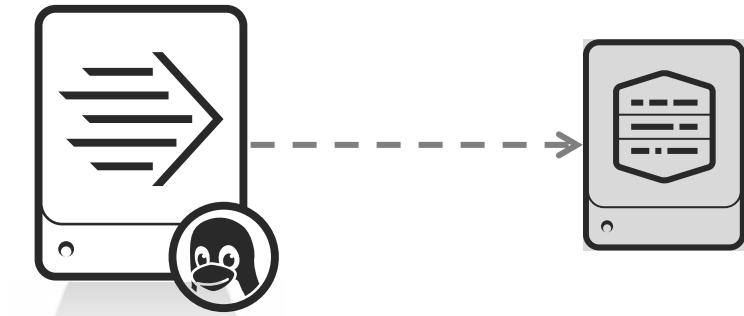
- Identify Linux-specific inputs
- Identify Windows-specific inputs

Identifying JournalID Inputs For UF



JournalD Inputs on Linux

- Natively supports **journalctl** command for viewing logs collected by **systemd**
- Collects thousands of events per second with minimal impact
- Only requires **inputs.conf** configuration
- Supported in Splunk 8.1 and later



inputs.conf

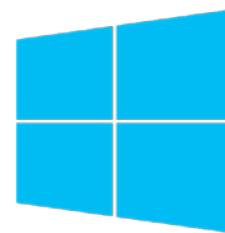
```
[journald://my-stanza]
journalctl-include-list = PRIORITY, CMD, EXE
journalctl-exclude-list =
journalctl-filter = _SYSTEMD_UNIT=my.service
    _PID=232 + _SYSTEMD_UNIT=sshd
journalctl-grep = ^WARN.*disk,
    .*errno=\d+\$+restarting
journalctl-user-unit = unit1, unit2
```

Windows-Specific Inputs

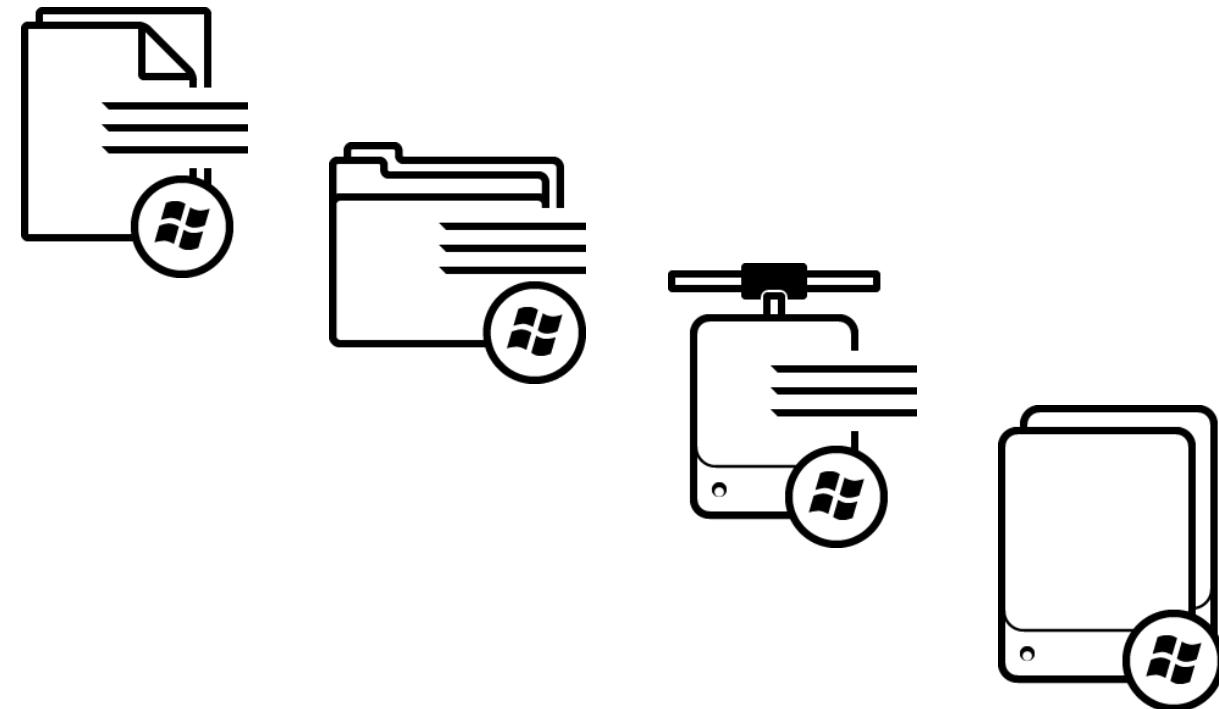


Windows-Specific Inputs

- Generally stored in binary format (for example some state data and logs)
- Accessed using Microsoft APIs
- Use special Splunk input types
- Can be forwarded to an indexer running any OS platform
- May require that Windows Universal Forwarder run as a domain user



Windows



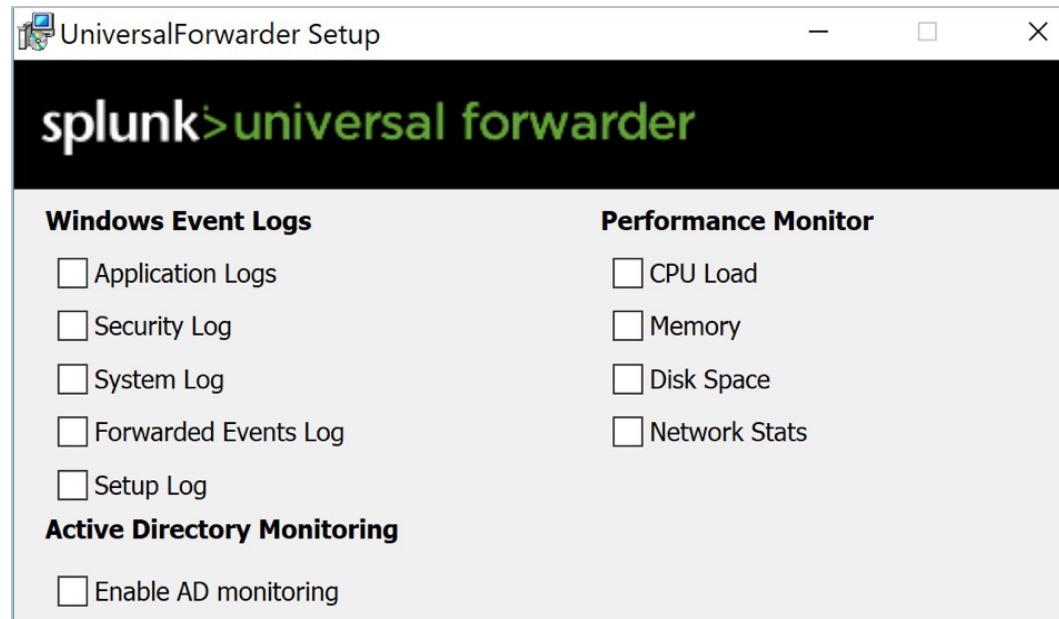
Windows-Specific Input Types

Input Type	Description
Event Log*	Consumes data from the Windows OS logs
Performance*	Consumes performance monitor data
Active Directory	Monitors changes in an Active Directory server
Registry	Monitors changes in a Windows registry
Host	Collects data about a Windows server
Network	Monitors network activity on a Windows server
Print	Monitors print server activity

* Supports both local and remote (WMI) data collection

Options for Configuring Local Windows Inputs

- During the Windows forwarder install
 - Easy to use for testing and proof of concept (PoC)
 - Entries created in the app **SplunkUniversalForwarder**
 - Presents issues when centrally managing configuration with Deployment Server (DS)



- Manually (Best Practice)
 - Create entries in custom app or use Splunk Add-on for MS Windows:
splunkbase.splunk.com/app/742/
 - Easy to manage using a DS
 - For details refer to:
 - **inputs.conf.spec**
 - **inputs.conf.example**

```
[admon://name]
[perfmon://name]
[WinEventLog://name]
[WinHostMon://name]
[WinNetMon://name]
[WinPrintMon://name]
[WinRegMon://name]
```

Configuring Local Windows Inputs Using Add Data

The screenshot shows the 'Add Data' wizard with four steps: 'Select Source', 'Input Settings', 'Review', and 'Done'. The 'Select Source' step is active, indicated by a green dot above it. The 'Local Event Logs' option is selected and highlighted with a green border. To the right, there is descriptive text about monitoring local Windows Event Log channels. Below the options, there is a configuration interface for selecting event logs. It includes a 'Select Event Logs' section, a list of 'Available item(s)' (Application, Security, Setup, System, ForwardedEvents, Els_Hyphenation/Analytic, EndpointMapper, FirstUXPerf-Analytic, Analytic), and a 'Selected item' section containing 'Security'. At the bottom, a note says 'Select the Windows Event Logs you want to index from the list.'

Add Data

Select Source Input Settings Review Done

Local Event Logs
Collect event logs from this machine.

Remote Event Logs
Collect event logs from remote hosts. Note: this uses WMI and requires a domain account.

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.

inputs.conf

```
[WinEventLog://Security]
checkpointInterval = 5
current_only = 0
disabled = 0
start_from = oldest
```

Select Event Logs Available item(s) add all » Selected item

Application
Security
Setup
System
ForwardedEvents
Els_Hyphenation/Analytic
EndpointMapper
FirstUXPerf-Analytic
Analytic

Select the Windows Event Logs you want to index from the list.

Windows Input Filtering Options

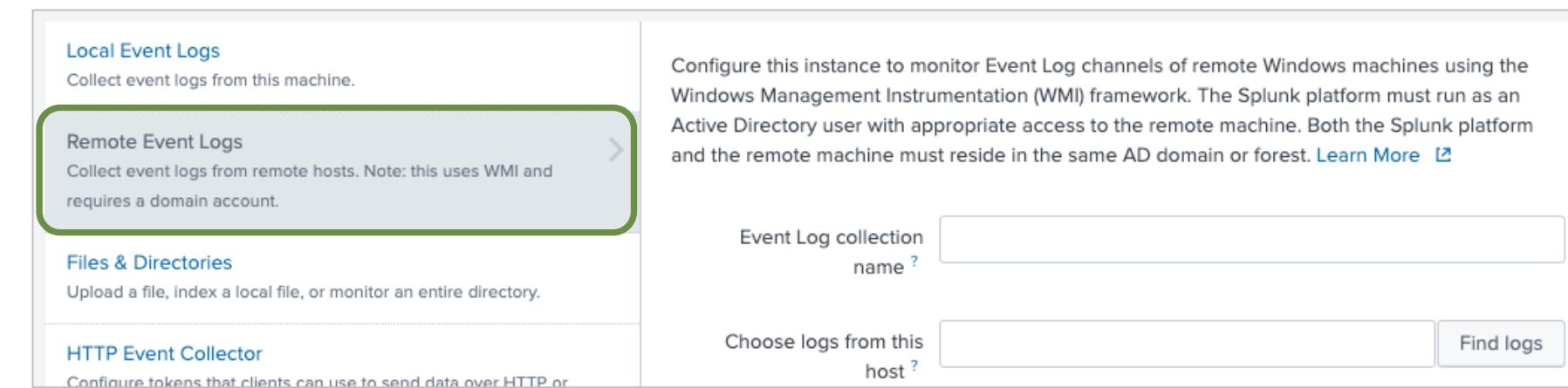
- Filter out non-essential events
 - Use include lists (**whitelist**) and exclude lists (**blacklist**)
 - Configure up to 10 entries for each list per stanza
 - Set entries based on event field names and regex:
 - **whitelist[1-9]** = <List> | *key=regex* [*key=regex*]
 - **blacklist[1-9]** = <List> | *key=regex* [*key=regex*]
 - In case of a conflict, the exclude lists (**blacklist**) prevails

`inputs.conf`

```
[WinEventLog://Security]
disabled=0
whitelist1= EventCode=/^4|5.*$/ Type=Error|Warning/
whitelist2= TaskCategory=%^Log.*%
blacklist = 540
```

Windows Remote Inputs With WMI

- Available for two types of Windows inputs:
 - Event logs
 - Performance monitor
- Advantage:
 - Collect input without a forwarder
- Disadvantage:
 - Uses WMI as a transport protocol
 - Not recommended in high latency networks
 - Requires Splunk to run as a domain account



Configuring WMI Inputs

- Remote inputs are configured in **wmi.conf**
- See **wmi.conf.spec** and **wmi.conf.example** for full details

wmi.conf

```
[WMI:remote-logs]
interval = 5
server = server1, server2, server3
event_log_file = Application, Security, System

[WMI:remote-perfmon]
interval = 5
server = server1, server2, server3
wql = Select DatagramsPersec
```

Special Field Extractions

- Several Microsoft products use a special multi-line header log format
 - Examples: IIS/W3C, JSON, and other delimited/structured sources
- Challenges:
 - These logs often get re-configured by the product administrator
 - Requires coordination between source administrator and Splunk administrator to sync the field extraction
- Solution:
 - Use indexed field extraction on the Windows forwarder
 - Normally the field extraction magic happens on the index/search tier

Powershell Input

- Uses built-in **powershell.exe** scripting facility in Windows
 - No custom external library dependencies

PowerShell v3 or higher

Command or a script file

Blank field executes once only

inputs.conf

```
[powershell://<name>]
script = <command>
schedule = [<number>|<cron>]
```

Windows Inputs Resources

- Monitoring Windows data with Splunk Enterprise
docs.splunk.com/Documentation/Splunk/latest/Data/AboutWindowsdataandSplunk
- Microsoft: Diagnostics - Windows Event Log
docs.microsoft.com/en-us/windows/desktop/wes/windows-event-log
- Microsoft: Diagnostics - Performance Counters
docs.microsoft.com/en-us/windows/desktop/PerfCtrs/performance-counters-portal
- Microsoft: Diagnostics - Performance Counters Reference
docs.microsoft.com/en-us/windows/desktop/PerfCtrs/performance-counters-reference

Question: JournalD Input Requirements

Which of the following is not a requirement for JournalD inputs?

- A. Indexer with Linux platform
- B. Source system with Linux platform running `systemd`
- C. Configuration settings in `inputs.conf`
- D. Splunk 8.1 and later

Answer: JournalD Input Requirements

Which of the following is not a requirement for JournalD inputs?

- A. Indexer with Linux platform
- B. Source system with Linux platform running `systemd`
- C. Configuration settings in `inputs.conf`
- D. Splunk 8.1 and later

Any indexer can ingest JournalD data. Other than a source system with JournalD inputs, the only requirements are configuration settings in `inputs.conf`, and Splunk 8.1 and later.

Question: Windows Inputs

Which of the following is true about Windows inputs?

- A. Configure remote collection of Active Directory using **wmi.conf**
- B. Windows input from a Windows UF requires a Windows indexer
- C. Windows events can be excluded using exclude lists
- D. Windows inputs must be configured in Splunk Web

Answer: Windows Inputs

Which of the following is true about Windows inputs?

- A. Configure remote collection of Active Directory using **wmi.conf**
- B. Windows input from a Windows UF requires a Windows indexer
- C. Windows events can be excluded using exclude lists**
- D. Windows inputs must be configured in Splunk Web

WMI remote inputs are available for Event logs and Performance monitoring. Windows inputs from Windows UFs can be ingested by any indexer, and can be configured in Splunk Web, using configuration files, and even during Windows UF installation.

Module 11:

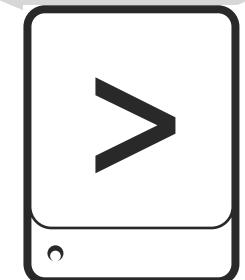
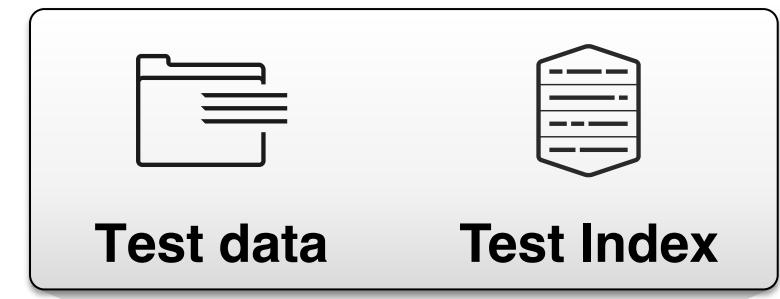
Fine-tuning Inputs

Module Objectives

- Understand the default processing that occurs during input phase
- Configure input phase options, such as source type fine-tuning and character set encoding

Review: Initial Data Input Testing

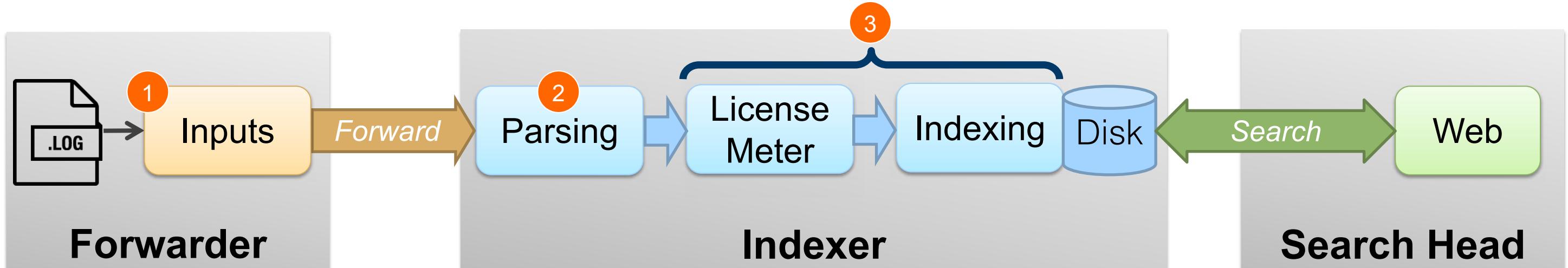
- Use a Splunk test server
 - Should be running same version as production
- Use test indexes
- Procedure:
 1. Copy production data to test server
 2. Use Splunk Web > Add Data
 3. Check to see if **sourcetype** and other settings are applied correctly
 4. Delete the test data, reset fishbucket if needed, change test configuration, and repeat as necessary



Test server

Index-Time Process

1. **Input phase:** Handled at the source (usually a forwarder)
 - The data sources are being opened and read
 - Data is handled as streams; configuration settings are applied to the entire stream
2. **Parsing phase:** Handled by indexers (or heavy forwarders)
 - Data is broken up into events and advanced processing can be performed
3. **Indexing phase:** Handled by indexers
 - License meter runs as data is initially written to disk, prior to compression
 - After data is written to disk, it cannot be changed



Things to Get Right at Index Time

Input phase

- Host
- Source type
- Source
- Index

Parsing phase

- Line breaking (event boundary)
- Date/timestamp extraction
- Adjust meta fields*
- Mask raw data*
- Eliminate events*

* Optional

What if I Don't Get It Right?

On a testing / development server

- This is what a test/dev server is for!
- Clean or delete+recreate test index, change configuration, try again
- May need to reset the fishbucket

On a production server

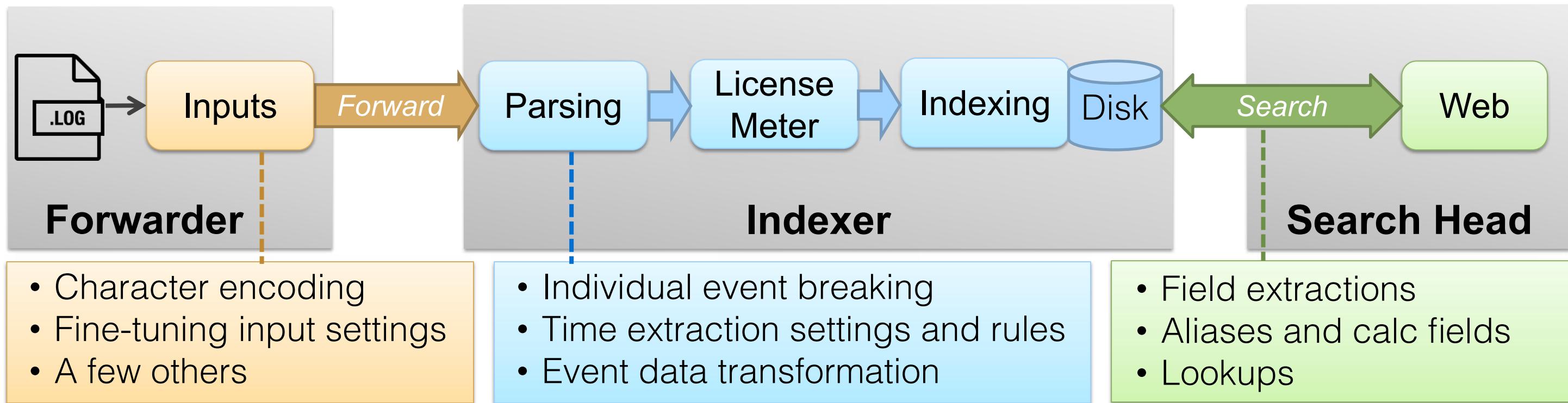
- Leave erroneous data in the system until it naturally “ages out” (reaches the index size or retention time limits)
- Attempt to delete the erroneous data
- Only re-index when it is absolutely necessary

The props.conf File

- Config file referenced during all phases of Splunk data processing (inputs, indexing, parsing and searching)
- Documentation:
 - The **props.conf.spec** and **props.conf.example** files in **SPLUNK_HOME/etc/system/README**
 - docs.splunk.com/Documentation/Splunk/latest/admin/Propsconf

Phases and `props.conf`

- Settings from `props.conf` applied during phases:



- Configure `props.conf` on the appropriate Splunk instances

docs.splunk.com/Documentation/Splunk/latest/Admin/Configurationparametersandthedatapipeline

Stanzas in props.conf

- All data modifications in **props.conf** are based on either source, sourcetype, or host

syntax

```
[source::source_name]  
attribute = value
```

```
[host::host_name]  
attribute = value
```

```
[sourcetype_name]  
attribute = value
```

example

```
[source::/var/log/secure*]  
sourcetype = linux_secure
```

```
[host::nyc*]  
TZ = US/Eastern
```

```
[sales_entries]  
CHARSET = UTF-8
```

- You can use wildcards (*) and regex in the **source::** and **host::** stanzas

Character Encoding

- During the input phase, Splunk sets all input data to UTF-8 encoding by default
 - Can be overridden, if needed, by setting the **CHARSET** attribute

```
[source:::/var/log/locale/korea/*]
```

```
CHARSET=EUC-KR
```

```
[sendmail]
```

```
CHARSET=AUTO
```

- Use **AUTO** to attempt automatic encoding based on language

docs.splunk.com/Documentation/Splunk/latest/Data/Configurecharacterencoding

Fine-tuning Directory Monitor Source Types

- When you add a directory monitor:
 - Specify a **sourcetype** to apply it to all files (contained recursively under that directory)
 - Omitting the **sourcetype** causes Splunk to try to use automatic pre-trained rules
- Override specific source types selectively in **props.conf**
 - Identify input with a **[source::<source>]** stanza and set the **sourcetype** attribute
 - Place this configuration on the source server, as this is an input phase process

inputs.conf

```
[monitor:///var/log/]
```

props.conf

```
[source::/var/log/mail.log]
sourcetype=sendmail
```

```
[source::/var/log/secure/]
sourcetype=secure
```

...

Note

If you explicitly set the source type in **inputs.conf** for a given source, you cannot override the source type value for the source in **props.conf**

Question : `props.conf` Stanza

In the `props.conf` example below, what is `sendmail`?

```
[sendmail]  
CHARSET=AUTO
```

- A. Source
- B. Source type
- C. Event
- D. Input

Answer: `props.conf` Stanza

In the `props.conf` example below, what is `sendmail`?

```
[sendmail]  
CHARSET=AUTO
```

- A. Source
- B. Source type
- C. Event
- D. Input

It is a source type in `props.conf`. Source types are specified as a string value in the stanza without the `sourcetype::` prefix.

Question: Invalid `props.conf` Entry

What is wrong with the `props.conf` example shown?

```
[source:::/var/.../korea/*]  
CHARSET=EUC-KR
```

```
[sendm*]  
CHARSET=AUTO
```

- A. The source cannot use the wildcard ...
- B. The source cannot use the wildcard *
- C. The source is using an invalid CHARSET
- D. The source type cannot use the wildcard *

Answer: Invalid `props.conf` Entry

What is wrong with the `props.conf` example shown?

```
[source:::/var/.../korea/*]  
CHARSET=EUC-KR
```

```
[sendm*]  
CHARSET=AUTO
```

- A. The source cannot use the wildcard ...
- B. The source cannot use the wildcard *
- C. The source is using an invalid CHARSET
- D. The source type cannot use the wildcard *

You cannot use a wildcard with source types in `props.conf`.

Module 11 Lab

Time: 10-15 minutes

Description: Fine-tuning Inputs

Tasks:

- Add a test directory monitor to sample the auto-sourcetype behavior
 - Make note of the source type value
- Override the auto-sourcetyping of a specific source by adding a source type declaration in **props.conf**
- Deploy it to your forwarder and check again

Note



These input files are not being updated. Therefore, you must reset the file pointer and re-index the files.

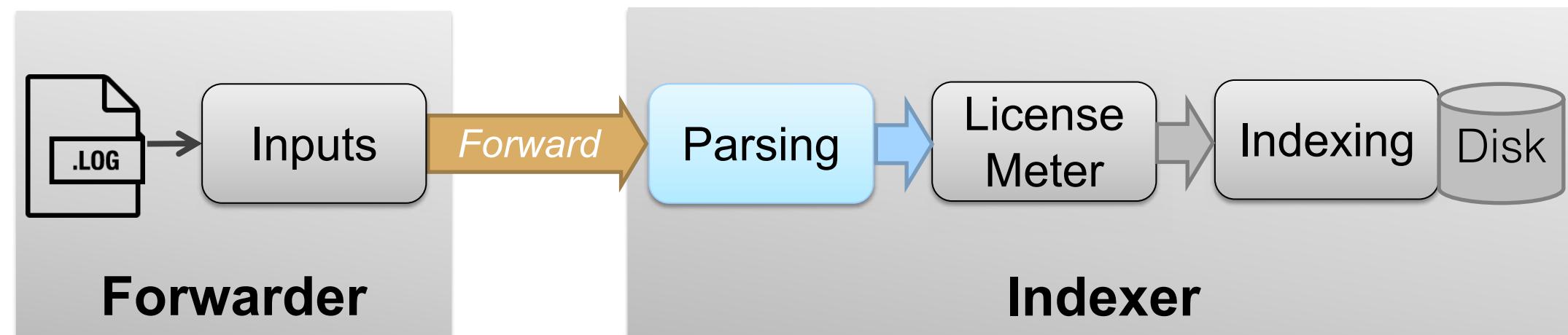
Module 12: Parsing Phase and Data Preview

Module Objectives

- Understand the default processing that occurs during parsing
- Optimize and configure event line breaking
- Explain how timestamps and time zones are extracted or assigned to events
- Use Data Preview to validate event creation during the parsing phase

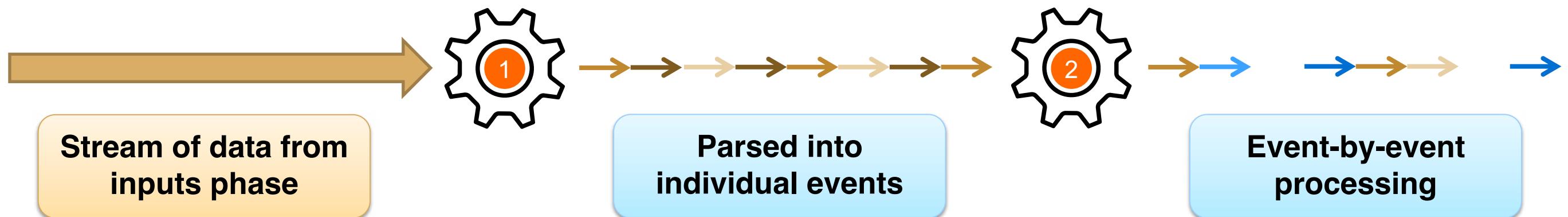
The Parsing Phase

- Occurs as data arrives at the indexer (or heavy forwarder)
- Breaks up input data stream into discrete **events**, each with a timestamp and time zone
- Creates, modifies, and redirects events
 - Applies additional transformation steps to modify the metadata fields or modify raw data



Event Creation

- Occurs during the parsing phase
 1. Data from input phase is broken up into individual events
 2. Event-level processing is performed



- Relies on **event boundaries**: distinguishing where events begin and end
 - Usually determined by line breaks
 - May be determined by other settings in **props.conf**
- Should be verified using **Data Preview**, with new source types

Determining Event Boundaries

Step 1: Line breaking

- Splits the incoming stream of bytes into separate lines
- Configured with **LINE_BREAKER** = *<regular_expression>*
- Default is any sequence of new lines and carriage returns: `([\r\n]+)`

Step 2: Line merging

- Merges separate lines to make individual events
- Configured with **SHOULD_LINEMERGE** = **true** (default)
- Uses additional settings to determine how to merge lines (such as **BREAK_ONLY_BEFORE**, **BREAK_ONLY_BEFORE_DATE**, and **MUST_BREAK_AFTER**)
- If each event is a separate line, disable (set to **false**) to improve performance

docs.splunk.com/Documentation/Splunk/latest/Data/Configureeventlinebreaking

Event Boundary Examples

Monitored input: Single line input with 3 events

```
[19/Sep/2020:18:22:32] VendorID=7033 Code=E AcctID=4390644811207834 ↵
[19/Sep/2020:18:22:48] VendorID=1239 Code=K AcctID=5822351159954740 ↵
[19/Sep/2020:18:22:59] VendorID=1243 Code=F AcctID=8768831614147676 ↵
```

`props.conf`

```
[sourcetype1]
LINE_BREAKER = ([\r\n]+)
SHOULD_LINEMERGE = false
```

Monitored input: Multi-line input with 3 events

```
Sep 12 06:11:58 host1.example.com storeagent[49597] <Critical>: Starting update scan ↵
Sep 12 06:11:58 host1.example.com storeagent[49597] <Critical>: UpdateController: Message tracing {
    "power_source" = ac; ↵
    "start_date" = "2018-08-21 20:10:39 +0000"; ↵
} ↵
Sep 12 06:11:58 host1.example.com storeagent[49597] <Critical>: Asserted BackgroundTask power ↵
```

`props.conf`

```
[sourcetype2]
LINE_BREAKER = ([\r\n]+)
SHOULD_LINEMERGE = true
BREAK_ONLY_BEFORE_DATE = true
```

Using Splunk Data Preview

- Splunk attempts to auto-detect a source type
 - Alternatively select from a list or define your own source type
 - Supports both unstructured and structured data sources
 - CSV, JSON, W3C/IIS, XML, etc.
- Event breaking and date/timestamp settings are evaluated
 - Use test environment to determine settings before taking a new data input into production
- Use Data Preview configuration settings to create new source types

Setting Event Breaks in Data Preview

Set Source Type

This page lets you see how the Splunk platform sees your data before indexing. If the events look correct and have the right timestamps, click "Next" to proceed. If not, use the options below to define proper event breaks and timestamps. If you cannot find an appropriate source type for your data, create a new one by clicking "Save As".

Source: /opt/log/crashlog/dreamcrusher.xml

[View Event Summary](#)

Source type: default ▾

Save As

Event Breaks

Define event boundaries for incoming data.

Event-breaking Policy: Auto Every Line Regex

Pattern: `([\r\n]+)\s*<Interceptor>`

* Specifies a regular expression that determines how the raw text stream is broken into initial events, before line endings.

= false and used regular expressions.

is broken into any number of lines.

ing group -- a pair of identified subcomponents of the match.

* Wherever the regex matches, Splunk considers the start of the first capturing group to be the end of the previous event, and considers the end of the first capturing group to be the start of the next event.

* The contents of the first capturing group are discarded, and will not be present in any event. You

List ▾ Format 20 Per Page ▾

< Prev 1 2 3 4 5 6 7 8 ... Next >

	Time	Event
1	12/3/19 4:49:04.000 AM	<?xml version="1.0" encoding="UTF-8" ?> <dataroot> timestamp = none
2	12/3/19 4:49:04.000 AM	<Interceptor> <AttackCoords>-80.33100097073213,25.10742916222947</AttackCoords> <Outcome>Interdiction</Outcome> <Infiltrators>23</Infiltrators> <Enforcer>Ironwood</Enforcer> <ActionDate>2019-11-20</ActionDate> <RecordNotes></RecordNotes> <NumEscaped>0</NumEscaped> <LaunchCoords>-80.23429525620114,24.08680387475695</LaunchCoords> <AttackVessel>Rustic</AttackVessel> </Interceptor> Collapse timestamp = none
3	12/3/19 4:49:04.000 AM	<Interceptor> <AttackCoords>-80.1462234 <Outcome>Interdiction</Outcome> <Infiltrators>6</Infiltrators> <Enforcer>Cunningham</Enforcer>

Show all 11 lines

Note i

Although **Event Breaks** have now been set correctly, notice that the timestamp is not yet properly captured for this input.

Date/timestamp Extraction

- Correct date/timestamp extraction is essential
 - Splunk works well with standard date/time formats and well-known data types
- Always verify timestamps when setting up new data types
 - Pay close attention to timestamps during testing/staging of new data
 - Check UNIX time or other non-human readable timestamps
- Custom timestamp extraction is specified in **props.conf**

Incorrectly Determined Timestamps

The screenshot shows a Splunk interface for a crash log. On the left, a detailed stack trace and system information are displayed. On the right, a timeline of events is shown.

Stack Trace (Left):

```
[167154] 2019-03-06 00:46:26
Received fatal signal 6 (Aborted).
Cause:
Signal sent by PID 6241 running under UID 5
Crashing thread: Main Thread
Registers
RDI: [0x00000B0500000C09]
RSI: [0x0F0097000009A300]
RBP: [0x0000000000002000]
RSP: [0x004B00000000D000]
RAX: [0x00042000010D0000]
RBX: [0x3005000000100000]
RCX: [0xE0E00000C010000]
RDX: [0x0000000A00000C00]
EFL: [0x0000000000002000]

OS: Linux
Arch: x86-64

Backtrace:
[0x04050A000000D000] gsignal + 53 (/lib64/libc.so.6)
[0x0600000000000000] abort + 373 (/lib64/libc.so.6)
[0x000C000000000000] ? (/lib64/libc.so.6)
[0x8000000090300B0] __assert_perror_fail +
[0x0F000000E00B000] _ZN11XmlDocument8addChildERK7XmlNode + 61 (dcrusherd)
[0x0800000070500C00] _Z18getSearchConfigXMLR11XmlDocumentPKPKc + 544 (dcrusherd)
[0x0000100000000000] _Z22do_search_process_impliPKPKcP12BundlesSetupb + 6141 (dcrusherd)
Linux /usr13.eng.buttermcupgames.com / 2.6.32-279.5.2.el6.x86_64 / #1 SMP Fri Aug 24 01:07:11 UTC 2018 / x86_64
/etc/redhat-release: CentOS release 6.3 (Final)
glibc version: 2.12
glibc release: stable
Last errno: 2
```

Event Timeline (Right):

Add Data < Back Next >

Set Source Type

This page lets you see how the Splunk platform sees your data. Click "Next" to proceed. If not, use the options below to define your data, create a new one by clicking "Save As".

Source: /opt/log/crashlog/crash-2019-03-06-00_46_26.log

Source type: Select Source Type Save As

	Time	Event
1	3/6/19 12:46:26.000 AM	[167154] 2019-03-06 00:46:26 Received fatal signal 6 (Aborted). Cause: Signal sent by PID 6241 running under UID 5 Crashing thread: Main Thread Show all 25 lines
2	8/24/18 1:07:11.000 AM	Linux /usr13.eng.buttermcupgames.com / 2.6.32-279.5.2.el6.x86_64 / #1 SMP Fri Aug 24 01:07:11 UTC 2018 / x86_64 /etc/redhat-release: CentOS release 6.3 (Final) glibc version: 2.12 glibc release: stable Last errno: 2 Show all 20 lines

A callout bubble points from the text "Splunk makes its best attempt to identify event boundaries and timestamps; however, if you are more familiar with the data, provide more info" to the timestamp "2019-03-06 00:46:26" in the first event row.

Annotations:

- An orange circle with the number 1 is positioned over the timestamp in the first event row.
- An orange circle with the number 2 is positioned over the timestamp in the second event row.

Failed To Parse Timestamps

Add Data

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

Set Source Type

This page lets you see how Splunk sees your data before indexing. If the events look correct and have the right timestamps, click "Next" to proceed. If not, use the options below to define proper event breaks and timestamps.

Source: /opt/log/crashlog/dreamcrusher.xml

Source type: default ▾ Save As List ▾

View Event Summary

Event Breaks

Timestamp

When an event is not being parsed correctly, use the warning indicator to help identify possible solutions

1 !

MAX_EVENTS (256) was exceeded without a single event break. Will set BREAK_ONLY_BEFORE_DATE to False, and unset any MUST_NOT_BREAK_BEFORE or MUST_NOT_BREAK_AFTER rules. Typically this will amount to treating this data as single-line only. Show all 257 lines

2 !

Failed to parse timestamp. timestamp = none Defaulting to file modtime.

3/6/18 8:16:05.000 PM timestamp = none

<?xml version="1.0" encoding="UTF-8" ?>

<Interceptor>

<AttackCoords>-80.33100097073213,25.10742916222947</AttackCoords>

<Outcome>Interdiction</Outcome>

Sebastian Jiménez,

< Prev 1 2 3 4 5 6 7 8 ... Next >

Using TIME_PREFIX

- Syntax: **TIME_PREFIX = <REGEX>**
- Matches characters right BEFORE the date/timestamp
 - Use this syntax to specify where the timestamp is located in the event

[167154] 2019-03-06 00:46:26
Received fatal signal 6 (Aborted).
Cause:
Signal sent by PID 6241 running under UID 5898.

Event

props.conf

```
[my_custom_source_or_sourcetype]
TIME_PREFIX = [\d+]\s+
```

Using MAX_TIMESTAMP_LOOKAHEAD

- Syntax: **MAX_TIMESTAMP_LOOKAHEAD = <integer>**
- Specifies how many characters to look for a timestamp
 - Generally, starts from beginning of the event
 - If **TIME_PREFIX** is set, starts from the point the **TIME_PREFIX** indicates
 - Improves efficiency of timestamp extraction

→ [167154] 2019-03-06 00:46:26
Received fatal signal 6 (Aborted).
Cause:
Signal sent by PID 6241 running under UID 5898.

props.conf

```
[my_custom_source_or_sourcetype]
TIME_PREFIX = [\d+]\s+
MAX_TIMESTAMP_LOOKAHEAD = 30
```

Event

Note



The complete timestamp string must be present within the specified range.

Using Timestamp Lookahead In Splunk Web

The screenshot shows the 'Add Data' workflow in Splunk Web, specifically the 'Set Source Type' step. A yellow callout box labeled 'Timestamp > Advanced' points to the 'Advanced' button in the extraction section. Another yellow callout box points to the 'Lookahead' field set to 30. The main pane displays event details from a log file, with one event highlighted.

Source: /opt/log/crashlog/crash-2019-03-06-00_46_26.log

View Event Summary

Source type: Select Source Type

Event Breaks

Timestamp

Determine how timestamps for the incoming data are defined.

Extraction

- Auto
- Curr...
- Adva...** (highlighted)
- Conf...

Time Zone -- Default System Timezone --

Timestamp format A string in strftime() format that helps Splunk recognize timestamps. [Learn More](#)

Timestamp prefix Timestamp is always prefaced by a regex pattern eg: \d+abc123\d[2,4]

Lookahead 30

Timestamp never extends more than this number of events

Timestamp > Advanced

Time 3/6/19 12:46:26.000 AM

Event [167154] 2019-03-06 00:46:26
Received fatal signal 6 (Aborted).
Cause:
Signal sent by PID 6241 running under UID 5898.
Crashing thread: Main Thread
[Show all 45 lines](#)

Timestamp > Advanced

- Allows Splunk to ignore timestamps found later in data
- May update the number of events extracted
- Warns if it cannot find a timestamp within the range

Using TIME_FORMAT

- Syntax: **TIME_FORMAT = <strftime-style format>**
- Examples:

Timestamp	TIME_FORMAT entry
2020-10-31	%Y-%m-%d
January 24, 2003	%B %d, %Y

- For more detail and other options, check:
 - **SPLUNK_HOME\etc\system\README\props.conf.spec**
 - docs.splunk.com/Documentation/Splunk/latest/Data/ConfigureTimestampRecognition
 - docs.splunk.com/Documentation/Splunk/latest/Data/Handleeventtimestamps

Splunk Web: Advanced Timestamp Extraction

Set Source Type

This page lets you see how the Splunk platform sees your data before indexing. If the events look correct and have the right timestamps, click "Next" to proceed. If not, use the options below to define proper event breaks and timestamps. If you cannot find an appropriate source type for your data, create a new one by clicking "Save As".

Source: /opt/log/crashlog/dreamcrusher.xml

View Event Summary

Source type: default

> Event Breaks

Timestamp

Determine how timestamps for the incoming data are defined.

Extraction

Time Zone

Timestamp format
A string in strftime() format that helps Splunk recognize timestamps. [Learn More](#)

Timestamp prefix
Timestamp is always prefaced by a regex pattern eg:
\d+abc123\d[2,4]

Lookahead
Timestamp never extends more than this number of characters into the event, or past the Regex if specified above.

List ▾ Format 20 Per Page ▾

1 2 3 4 5 6 7 8 ... Next >

	Time	Event
		timestamp = none
2	11/20/19 12:00:00.000 AM	<Interceptor> <AttackCoords>-80.33100097073213,25.10742916222947</AttackCoords> <Outcome>Interdiction</Outcome> <Infiltrators>23</Infiltrators> <Enforcer>Trenwood</Enforcer> <ActionDate>2019-11-20</ActionDate> <RecordNotes></RecordNotes> <NumEscaped>0</NumEscaped> <LaunchCoords>-80.23429525620114,24.08680387475695</LaunchCoords> <AttackVessel>Rustic</AttackVessel> </Interceptor> Collapse
3	11/29/19 12:00:00.000 AM	<Interceptor> <AttackCoords>-80.14622349209523,24.53605142362535</AttackCoords> <Outcome>Interdiction</Outcome> <Infiltrators>6</Infiltrators> <Enforcer>Cunningham</Enforcer> Show all 11 lines
4	11/14/19 12:00:00.000 AM	<Interceptor> <AttackCoords>-80.75496221688965,24.72483828554483</AttackCoords> <Outcome>Interdiction</Outcome>

Setting Time Zone Rules

- Use time zone offsets to ensure correct event time

props.conf

- Splunk applies time zones in this order:

1. A time zone indicator in the raw event data

- ▶ **-0800, GMT-8 or PST**

2. The value of a TZ attribute set in **props.conf**

- ▶ Checks the **host**, **source**, or **sourcetype** stanzas

- ▶ en.wikipedia.org/wiki/List_of_zoneinfo_timezones

3. If a forwarder is used, the forwarder-provided time zone is used

4. If all else fails, Splunk applies the time zone of the indexer's host server

```
[host::nyc*]
TZ = America/New_York

[source::mnt/cn_east/*]
TZ = Asia/Shanghai
```

Splunk Event Timestamp Processing

- 1 • Use **TIME_FORMAT** (from `props.conf`) to identify a timestamp in event
- 2 • If no **TIME_FORMAT** configured: Try to automatically identify timestamp from event
- 3 • If identify time+date, but no year: Determine a year
- 4 • If identify time, but no date: Try to find date in source name or file name
- 5 • If cannot identify a date: use file modification time
- 6 • Else no timestamp found:
 - If any timestamp from same source, use the most recent timestamp
 - If no timestamps, use the current system time when indexing the event

<http://docs.splunk.com/Documentation/Splunk/latest/Data/HowSplunkextractstimestamps>

Saving New Source Type

The diagram illustrates the workflow for saving a new source type. It starts with a configuration screen on the left, which includes a dropdown for 'Source type: _json', a 'Save As' button highlighted with a green box and arrow, and a table of settings. A green box and arrow point from the 'Save As' button to a 'Save Source Type' dialog on the right. The dialog shows fields for 'Name' (geojson), 'Description' (JavaScript Object Notation format. For more information, visit <http://json.org/>), 'Category' (Application), and 'App' (Search & Reporting). A green box and arrow point from the 'Save' button in the dialog to a text area below it. This text area contains the 'props.conf' configuration text for the new source type.

Source type: _json ▾

Save As

Timestamp

Advanced

Name	Value
CHARSET	UTF-8
INDEXED_EXTRACTI	json
KV_MODE	none
SHOULD_LINEMERG	false
category	Structured
description	JavaScript Object Notation for
disabled	false
pulldown_type	true
TIMESTAMP_FIELDS	time

New setting

Copy to clipboard

Apply settings

Save Source Type

Name: geojson

Description: JavaScript Object Notation format. For more information, visit <http://json.org/>

Category: Application

App: Search & Reporting

Save

Copy and paste this props.conf text:

```
[ _json ]
CHARSET=UTF-8
INDEXED_EXTRACTS=json
KV_MODE=none
SHOULD_LINEMERGE=false
category=Structured
description=JavaScript Object Notation format. For more
information, visit http://json.org/
disabled=false
pulldown_type=true
TIMESTAMP_FIELDS=time
```

When saved, the source type becomes a custom source type that can be re-used

- Copy and deploy sourcetype settings manually to your forwarders
- Alternately get settings from **props.conf** stanza for the new source type

Source Type Manager

Settings > Source types allows access to configured sourcetypes independent of the Add Data wizard

Source Types

New Source Type

Source types are used to assign configurations like timestamp recognition, event breaking, and field extractions to data indexed by Splunk. [Learn more](#)

12 Source Types

Show only popular

Category: Application ▾

App: All ▾

filter



20 per page ▾

Name ▾	Actions	Category ▾	App ▾
catalina Output produced by Apache Tomcat Catalina (System.out and System.err)	Edit Clone	Application	system
dc_mem_crash Dream Crusher server memory dump	Edit Clone Delete	Application	search
dcrusher_attacks Dream Crusher user interactions	Edit Clone Delete	Application	search
dreamcrusher.xml	Edit Clone Delete	Application	search
log4j Output produced by any Java 2 Enterprise Edition (J2EE) application server using log4j	Edit Clone	Application	system

Custom sourcetypes can be edited, deleted, and cloned

Question: Default `props.conf` Settings

What are the default settings in `props.conf`?

- A. `LINE_BREAKER=([\r\n])` and `SHOULD_LINEMERGE=true`
- B. `LINE_BREAKER=([\r\n])` and `SHOULD_LINEMERGE=false`
- C. `LINE_BREAKER=([\r\n]+)` and `SHOULD_LINEMERGE=true`
- D. `LINE_BREAKER=([\r\n]+)` and `SHOULD_LINEMERGE=false`

Answer: Default `props.conf` Settings

What are the default settings in `props.conf`?

- A. `LINE_BREAKER=([\r\n])` and `SHOULD_LINEMERGE=true`
- B. `LINE_BREAKER=([\r\n])` and `SHOULD_LINEMERGE=false`
- C. `LINE_BREAKER=([\r\n]+)` and `SHOULD_LINEMERGE=true`**
- D. `LINE_BREAKER=([\r\n]+)` and `SHOULD_LINEMERGE=false`

The default settings for `props.conf` include:

- **`LINE_BREAKER=([\r\n]+)`**: Any sequence of new lines or carriage returns
- **`SHOULD_LINEMERGE=true`**: Merges separate lines to make individual events

Question: Timestamp Performance

Which of the following features does not improve efficiency of timestamp extraction?

- A. TIME_FORMAT
- B. TIME_PREFIX
- C. TZ
- D. MAX_TIMESTAMP_LOOKAHEAD

Answer: Timestamp Performance

Which of the following features does not improve efficiency of timestamp extraction?

- A. TIME_FORMAT
- B. TIME_PREFIX
- C. TZ
- D. MAX_TIMESTAMP_LOOKAHEAD

Using **TZ** to set timezone rules ensures that the correct timezone is used, but does not improve timestamp extraction performance as does the other features mentioned.

Question: Event Timestamp Processing

If no timestamp is found in an event, various methods are used to approximate a timestamp for that event. Which of the following is not one of those methods?

- A. Using the current system time when indexing the event
- B. Using UNIX epoch: Jan 1, 1970
- C. Finding a date in source name or file name
- D. Using file modification time

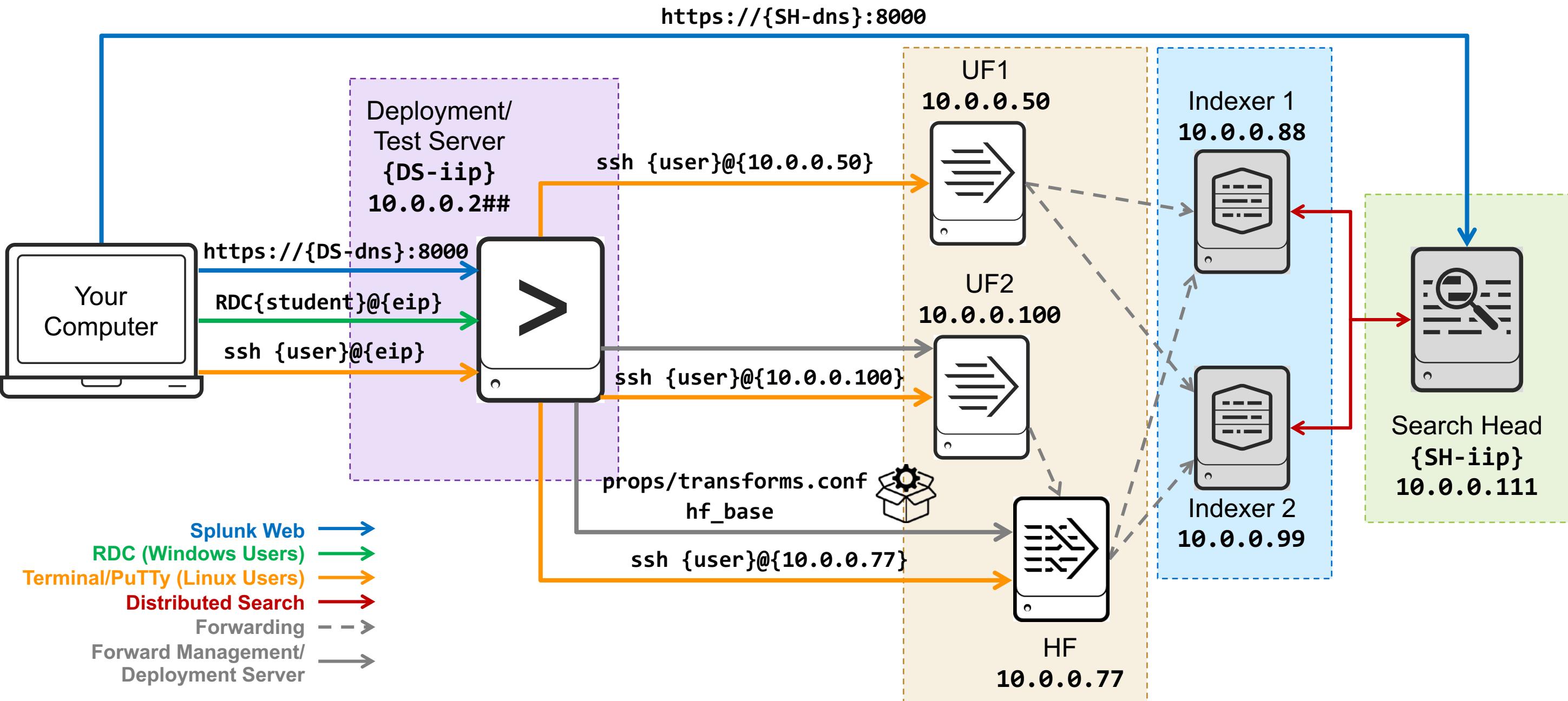
Answer: Event Timestamp Processing

If no timestamp is found in an event, various methods are used to approximate a timestamp for that event. Which of the following is not one of those methods?

- A. Using the current system time when indexing the event
- B. Using UNIX epoch: Jan 1, 1970**
- C. Finding a date in source name or file name
- D. Using file modification time

All of these methods may be used in Splunk event timestamp processing, except using the UNIX epoch time.

Module 12 Lab – Environment Diagram



Module 12 Lab

Time: 20-25 minutes

Description: Create a New Source Type

Tasks:

- Use preview to evaluate two custom file types:
 - A new log sample that contains multiple timestamps
 - A new log sample that contains multi-line events in XML format
- Apply a custom line breaking rule and custom timestamp rules and save as a new sourcetype

Module 13:

Manipulating Input Data

Module Objectives

- Explore Splunk transformation methods
- Create rulesets with Ingest Actions
- Mask data with Ingest Action rules
- Mask data with **SEDCMD** and **TRANSFORMS**
- Override sourcetype or host based upon event values

Modifying the Raw Input Data

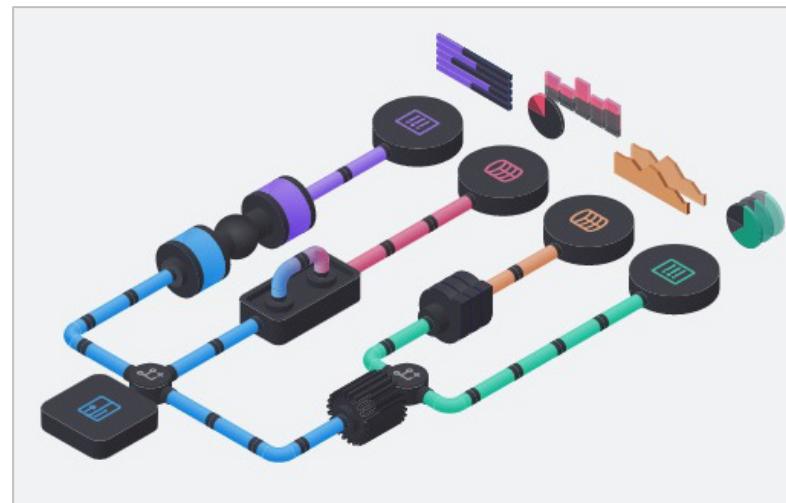
Sometimes necessary prior to indexing

- In cases of privacy concerns
 - **Healthcare**: Patient information
 - **Finance**: Credit card or account numbers
 - **Globalization**: Data transported across international borders
- According to business use cases
 - **Audit and security**: Route all events to the **web** index, except credit card transactions which are sent to the **credits** index

Should be performed with *extreme* care

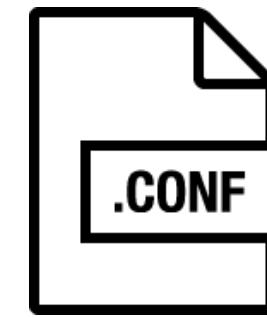
- Unlike all other modifications discussed, these changes modify the raw data (**_raw**) before it is indexed
- Indexed data will not be identical to the original data source

Splunk Transformation Methods



Ingest Action

- Uses Splunk Web
- Can mask, truncate, route or eliminate data
- Introduced in Splunk 9.0 (currently Linux only support)



Classic Configuration

- Uses **SEDCMD** to mask or truncate data
- Uses **TRANSFORMS** to mask, truncate, route or eliminate data

Using SEDCMD

- Per event simplified data modifications using UNIX "sed-like" syntax
 - Provides “search and replace” using regular expressions and substitutions
 - Supported on both Linux and Windows
- Example: Hide first 5 digits of account numbers in **vendor_sales.log**:

```
[22/Oct/2014:00:46:27] VendorID=9112 Code=B AcctID=4902636948  
[22/Oct/2014:00:48:40] VendorID=1004 Code=J AcctID=4236256056  
[22/Oct/2014:00:50:02] VendorID=5034 Code=H AcctID=8462999288
```

Replace with
AcctID=xxxxx99288

```
[source::.../vendor_sales.log]  
SEDCMD-1acct = s/AcctID=\d{5}(\d{5})/AcctID=xxxxx\1/g
```

props.conf

\1 Indicates the
capture group

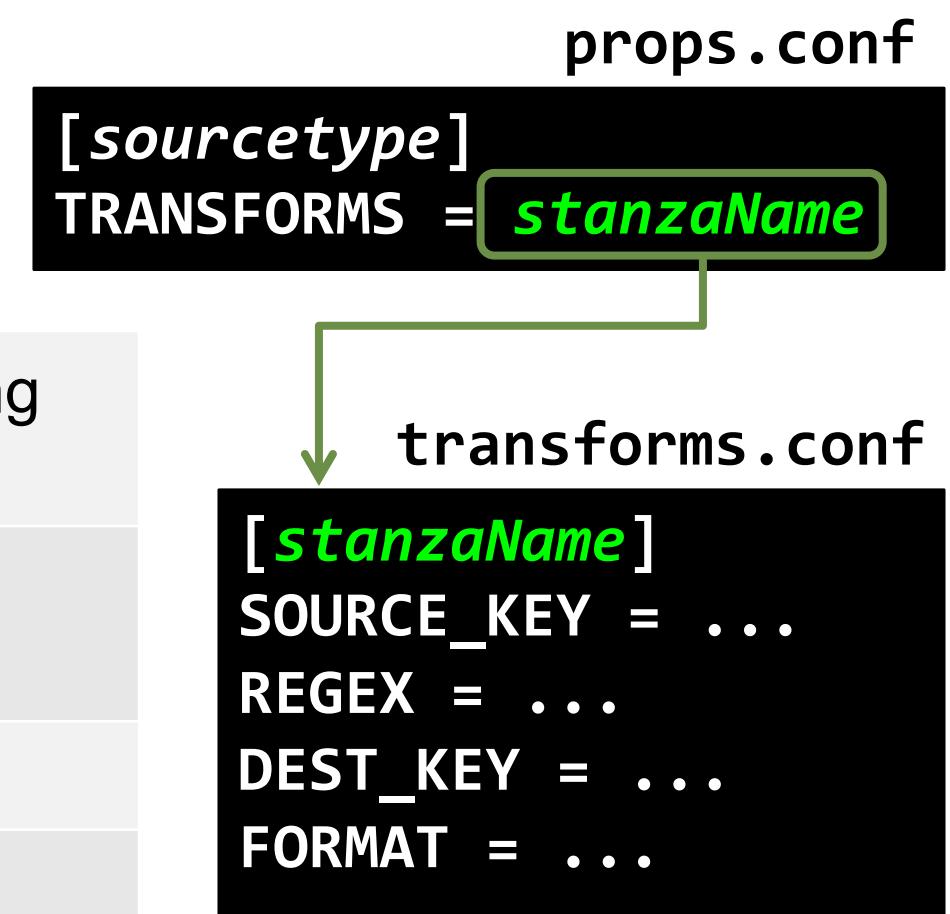
- Refer to: docs.splunk.com/Documentation/Splunk/latest/Data/Anonymizedata

Using TRANSFORMS

- Per event transformation based on REGEX pattern matches
- Invoked from **props.conf**
- Defined in **transforms.conf**
- Based on attributes:

SOURCE_KEY	Which field to use as source for pattern matching (default: <code>_raw</code> : unprocessed text of all events)
REGEX *	Events from the SOURCE_KEY that will be processed, with optional regex capture groups
DEST_KEY *	Where to write the processed data
FORMAT *	Controls how REGEX writes the DEST_KEY

* required



Masking Sensitive Data

```
[22/Apr/2014:00:46:27] VendorID=9112 CC_Num: 4217656647324534 Code=B  
[22/Apr/2014:00:48:40] Sent to checkout TransactionID=100763  
[22/Apr/2014:00:50:02] VendorID=5034 CC_Num: 6218651647508091 Code=H
```

props.conf

```
[source:....\\store\\purchases.log]  
TRANSFORMS-1ccnum = cc_num_anon
```

transforms.conf

```
[cc_num_anon]  
REGEX = (.*CC_Num:\s)\d{12}(\d{4}).*  
DEST_KEY = _raw  
FORMAT = $1xxxxxxxxxxxx$2
```

- For the purchases.log source, send to the cc_num_anon transformation processor.
- The label -1ccnum identifies this transform namespace and is used to determine sequence.

- When SOURCE_KEY is omitted, _raw is used.
- REGEX pattern finds two capture groups and rewrites the raw data feed with a new format.

```
[22/Apr/2014:00:46:27] VendorID=9112 CC_Num: xxxxxxxxxxxx4534 Code=B  
[22/Apr/2014:00:48:40] Sent to checkout TransactionID=100763  
[22/Apr/2014:00:50:02] VendorID=5034 CC_Num: xxxxxxxxxxxx8091 Code=H
```

Setting Per-Event Source Type

Should be your last option because it is more efficient to set the sourcetype during the inputs phase

```
[29/Apr/2017:07:08:32] VendorID=4119 Code=E AcctID=1808937180466558 Custom  
[29/Apr/2017:07:09:42] VendorID=5012 Code=N AcctID=7905045242265135  
[29/Apr/2017:07:11:10] VendorID=7015 Code=G AcctID=3283196485834211 Custom
```

props.conf

```
[source::udp:514]  
TRANSFORMS = custom_sourcetype
```

transforms.conf

```
[custom_sourcetype]  
SOURCE_KEY = _raw  
REGEX = Custom$  
DEST_KEY = MetaData:Sourcetype  
FORMAT = sourcetype::custom_log
```

- Check events in network input source
- If an event contains “Custom” at the end, assign the new sourcetype value **custom_log**
- When **MetaData:** key is used, its **FORMAT** value must be prefixed by:
 - **host::**
 - **source::**
 - **sourcetype::**

Setting Per-Event Host Name

```
[22/Apr/2014:00:46:27] sales accepted server:A01R2 SID=107570  
[22/Apr/2014:00:48:40] sales rejected server:B13R1 SID=102498  
[22/Apr/2014:00:50:02] sales accepted server:A05R1 SID=173560
```

props.conf

```
[sales_entries]  
TRANSFORMS-register = sales_host
```

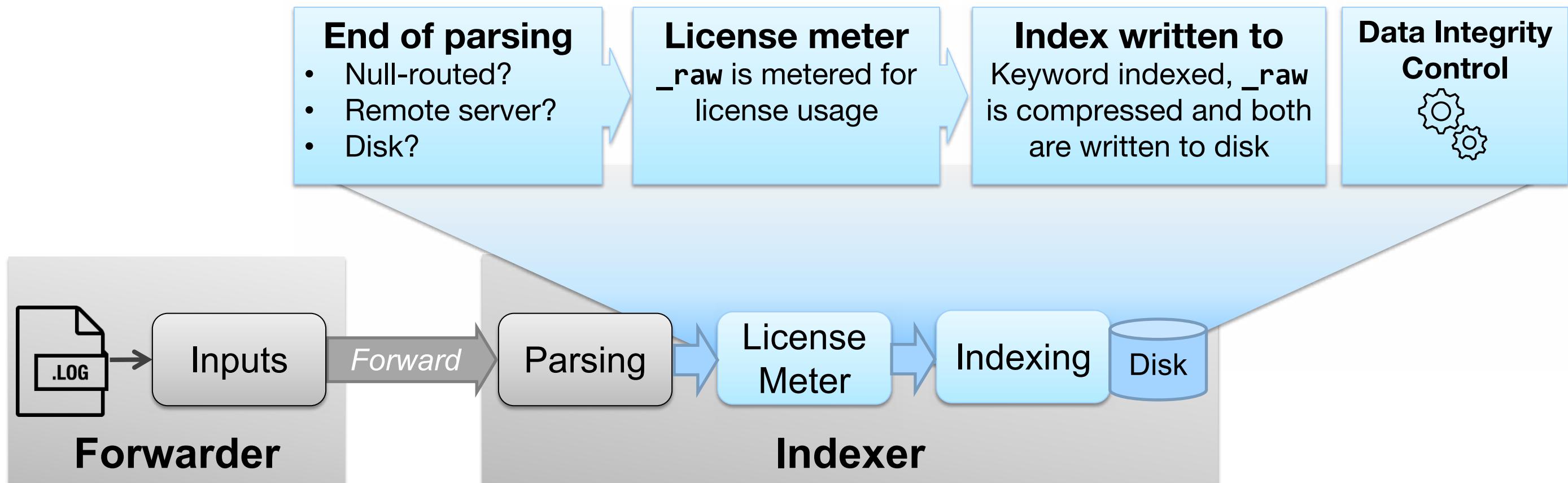
transforms.conf

```
[sales_host]  
SOURCE_KEY = _raw  
REGEX = server:(\w+)  
DEST_KEY = MetaData:Host  
FORMAT = host::$1
```

- Check each events in the **_raw** source
- If an event contains “**server:**”, capture the word and rewrite the value of the **MetaData:Host** key with the captured group
- When **MetaData:** key is used, its **FORMAT** value must be prefixed by:
 - **host::**
 - **source::**
 - **sourcetype::**

Indexing Phase Details

After the parsing phase, Splunk passes the fully processed events to the index processor

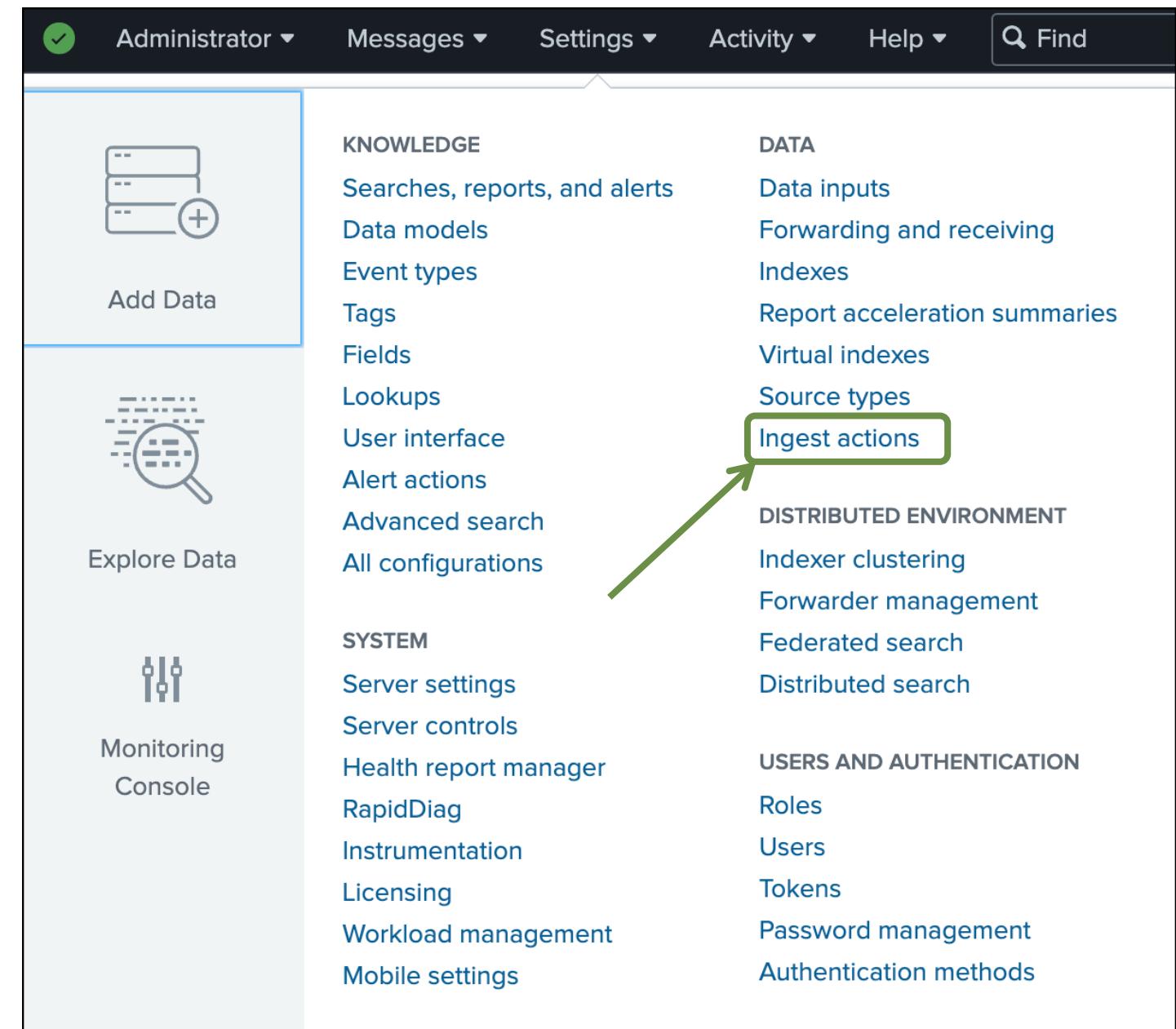


Persisted to Disk

- Indexed data is written to disk
 - Includes all modifications and extractions
 - Includes raw data (`_raw`) and metadata (source, sourcetype, host, timestamp, punct, etc.)
- Changes to **props.conf** or **transforms.conf**
 - Only applies to new data
 - Requires restarting the indexer, or re-loading by visiting:
<http://servername:splunkwebport/debug/refresh>
- Re-indexing is required to index old data with new settings

Using Ingest Actions

- Uses rulesets in Splunk Web
- Can mask, truncate, route or eliminate data
- Provides data previews
- Integrates automatically with deployment configurations
- Should be implemented on a Deployment Server dedicated to ingest action functions



Creating Rulesets

- Consist of one or more rules
 - Each rule can mask, truncate, route or eliminate data
- Require that the role has the following capabilities (**admin** role has these by default):
 - **list_ingest_ruleset**
 - **edit_ingest_ruleset**
- Can only have one ruleset per source type

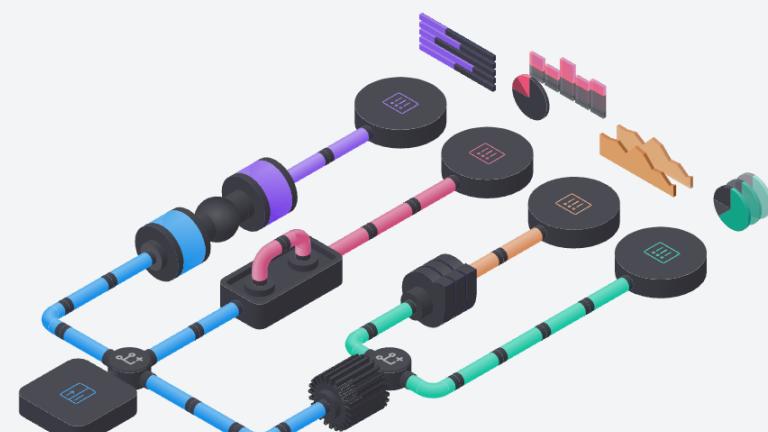
Ingest Actions

Create ingest-time rulesets to manipulate events based on their patterns. For more information, see [Splunk Docs](#).

You are logged into: splunk01

[Deploy ▾](#)

[Rulesets](#) [Destinations](#)



[New Ruleset](#)

Creating a New Ruleset (Index Data Preview)

Create New Ruleset

Access_Log_Ruleset

Rules for access_combined_wcookie source type

Event Stream: access_combined_wco 31KB

Preview using: Indexed Data (highlighted)

Sourcetype: access_combined_wcookie (highlighted)

Sample size: 100

Sample ratio: No event sampling

Time Range: Last 24 hours

Sample

+ Add Rule ▾

Data Preview for Event Stream

All Events (100) | Affected Events | Unaffected Events | *✓* | *✗*

i	Time	Event
>	8/25/2022 10:26:43.000 PM	91.214.92.22 - [26/Aug/2022:05:26:43] "GET /category.screen?categoryId=ARCADE&JSESSIONID=SD2SL8FF10ADFF4950 HTTP 1.1" 200 3469 "http://www.buttercupgames.com/cart.do?action=view&itemId=EST-26&productId=MB-AG-G07" "Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.2.28) Gecko/20120306 YFF3 Firefox/3.6.28 (.NET CLR 3.5.30729; .NET4.0C)" 229
>	8/25/2022 10:26:37.000 PM	host = splunk01 source = C:\opt\log\www2\access.log sourcetype = access_combined_wcookie
>	8/25/2022 10:26:03.000 PM	91.214.92.22 - [26/Aug/2022:05:26:37] "GET /category.screen?categoryId=SIMULATION&JSESSIONID=SD2SL8FF10ADFF4950 HTTP 1.1" 200 2148 "http://www.buttercupgames.com" "Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.2.28) Gecko/20120306 YFF3 Firefox/3.6.28 (.NET CLR 3.5.30729; .NET4.0C)" 760
>	8/25/2022 10:25:56.000 PM	host = splunk01 source = C:\opt\log\www2\access.log sourcetype = access_combined_wcookie
>	8/25/2022 10:25:45.000 PM	210.192.123.204 - [26/Aug/2022:05:26:03] "POST /cart.do?action=view&itemId=EST-27&JSESSIONID=SD9SL4FF3ADFF4964 HTTP 1.1" 200 2490 "http://www.buttercupgames.com/cart.do?action=addtocart&itemId=EST-27&productId=CU-PG-G06" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/534.57.10 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 959
>	8/25/2022 10:25:45.000 PM	host = splunk01 source = C:\opt\log\www2\access.log sourcetype = access_combined_wcookie
>	8/25/2022 10:25:45.000 PM	210.192.123.204 - [26/Aug/2022:05:25:56] "GET /product.screen?productId=CU-PG-G06&JSESSIONID=SD9SL4FF3ADFF4964 HTTP 1.1" 200 2394 "http://www.buttercupgames.com/cart.do?action=addtocart&productId=CU-PG-G06" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_4) AppleWebKit/534.57.10 (KHTML, like Gecko) Chrome/19.0.1084.46 Safari/536.5" 779
>	8/25/2022 10:25:45.000 PM	host = splunk01 source = C:\opt\log\www2\access.log sourcetype = access_combined_wcookie
>	8/25/2022 10:25:45.000 PM	210.192.123.204 - [26/Aug/2022:05:25:45] "GET /cart.do?action=view&itemId=EST-26&JSESSIONID=SD9SL4FF3ADFF4964 HTTP 1.1" 200 2776 "http://www.buttercupgames.com/product.screen?prod

Note i

Indexed Data is the default preview option when the Splunk instance has indexed data (and is not also a Deployment Server)

Creating a New Ruleset (Sample File Preview)

Create New Ruleset

Mask_Sales_Entries

Mask account codes in the sales_entries.log

Event Stream sales_entries 3.1KB

Data Preview for Event Stream

All Events (44) Affected Events Unaffected Events

i	Time	Event
>	6/21/2022 9:24:11.000 AM	Tue Jun 21 2022 16:24:11 Sent to checkout TransactionID=100763 host = EC2AMAZ-B95HUCT source = C:\Program Files\Splunk\var\run\splunk\dispatch\1661493039.131\sales_entries... sourcetype = sales_entries-sample-yvta312h1k
>	6/21/2022 9:24:11.000 AM	Tue Jun 21 2022 16:24:11 checkout response for TransactionID=100763 CustomerID=6i30kqk3 host = EC2AMAZ-B95HUCT source = C:\Program Files\Splunk\var\run\splunk\dispatch\1661493039.131\sales_entries... sourcetype = sales_entries-sample-yvta312h1k
>	6/21/2022 9:24:11.000 AM	Tue Jun 21 2022 16:24:11 ecomm engine response TransactionID=100763 CustomerID=6i30kqk3 accepted host = EC2AMAZ-B95HUCT source = C:\Program Files\Splunk\var\run\splunk\dispatch\1661493039.131\sales_entries... sourcetype = sales_entries-sample-yvta312h1k
>	6/21/2022 9:24:13.000 AM	host = EC2AMAZ-B95HUCT source = C:\Program Files\Splunk\var\run\splunk\dispatch\1661493039.131\sales_entries... sourcetype = sales_entries-sample-yvta312h1k
>	6/21/2022 9:25:29.000 AM	Tue Jun 21 2022 16:25:29 Sent to checkout TransactionID=100764 host = EC2AMAZ-B95HUCT source = C:\Program Files\Splunk\var\run\splunk\dispatch\1661493039.131\sales_entries... sourcetype = sales_entries-sample-yvta312h1k

Preview using

Indexed Data Sample File

Sourcetype

sales_entries

Drop your file anywhere or [browse..](#)

Max file upload size of 500 MB

Only the first 1,000 events can be used to preview

or Paste from Clipboard

+ Add Rule ▾

**Using a file with sample entries:
sales_entries_samples.log**

Note

Sample File is the only preview option available for Deployment Servers.

Masking with Regular Expression

The screenshot shows the Splunk masking interface. On the left, under the MASK section, a rule titled "Mask with Regular Expression" is selected. A green arrow points from this section to a callout box containing two bullet points:

- Use regular expression with capture groups
- Splunk uses PCRE regex

Below the MASK section, the FILTER and ROUTE sections are visible. In the FILTER section, there is a "Mask with Regex" entry. The ROUTE section contains a "Route to Destination" entry.

In the center, a "Data Preview for Mask" table shows event details. The table has columns for i, Time, and Event. The preview shows four events, with the fourth event's AcctCode field highlighted with a red box and a green border, indicating it is the target of the masking rule.

i	Time	Event
>	6/21/2022 9:24:11.000 AM	Tue Jun 21 2022 16:24:11 Sent to checkout TransactionID=100763 host = EC2AMAZ-B95HUCT source = C:\Program Files\Splunk\var\run\splunk\dispatch\1661493039.131\sales_entries_... sourcetype = sales_entries-sample-yvta312h1k
>	6/21/2022 9:24:11.000 AM	Tue Jun 21 2022 16:24:11 checkout response for TransactionID=100763 CustomerID=6i30kqk3 host = EC2AMAZ-B95HUCT source = C:\Program Files\Splunk\var\run\splunk\dispatch\1661493039.131\sales_entries_... sourcetype = sales_entries-sample-yvta312h1k
>	6/21/2022 9:24:11.000 AM	Tue Jun 21 2022 16:24:11 ecomm engine response TransactionID=100763 CustomerID=6i30kqk3 host = EC2AMAZ-B95HUCT source = C:\Program Files\Splunk\var\run\splunk\dispatch\1661493039.131\sales_entries_... sourcetype = sales_entries-sample-yvta312h1k
>	6/21/2022 9:24:12.000 AM	Tue Jun 21 2022 16:24:12 TransactionID=100763 AcctCode=8333-4577XXXX host = EC2AMAZ-B95HUCT source = C:\Program Files\Splunk\var\run\splunk\dispatch\1661493039.131\sales_entries_... sourcetype = sales_entries-sample-yvta312h1k

Previewing and Saving the Ruleset

Edit Ruleset

Mask_Sales_Entries

Mask account codes in the sales_entries.log

Event Stream sales_entries

Data Preview for Event Stream

All Events Affected Events Unaffected Events

Cancel Preview Config Save

Preview Mask_Sales_Entries

You are seeing a preview of Ruleset configuration. Copy/paste the stanzas below into your props.conf and transforms.conf files to deploy manually in your environment.

Caution: Manually editing Ruleset configurations may become incompatible with the user interface and not display in the user interface.

props.conf

transforms.conf

Preview Base Search:

This is the SPL used to generate the sample dataset.

SID:

1661493963.155

[Open in Search ↗](#)

Post-process Search:

This SPL is applied to the sample dataset to simulate and preview the effect of the currently selected rule.

```
eval INPUT._raw=coalesce($_raw$,"__NULL__")
| eval _raw=replace(_raw,"(AcctCode=\d{4})-\d{4}","\\1-XXXX")
```

[Open in Search ↗](#)

Standalone Rule Deployment

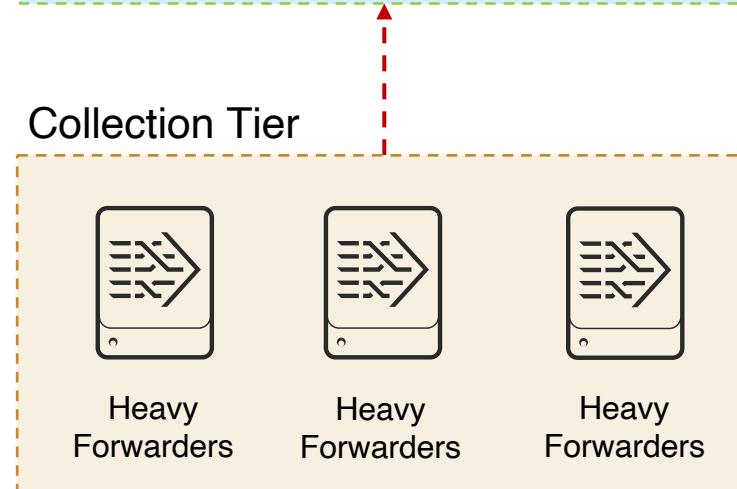
Standalone:

`SPLUNK_HOME/etc/apps/
splunk_ ingest_actions/local`



Heavy Forwarder:

`SPLUNK_HOME/etc/apps/
splunk_ ingest_actions/local`



- Manually distribute rulesets by copying config files from **splunk_ ingest_actions** app to appropriate instances (other indexers or HFs)

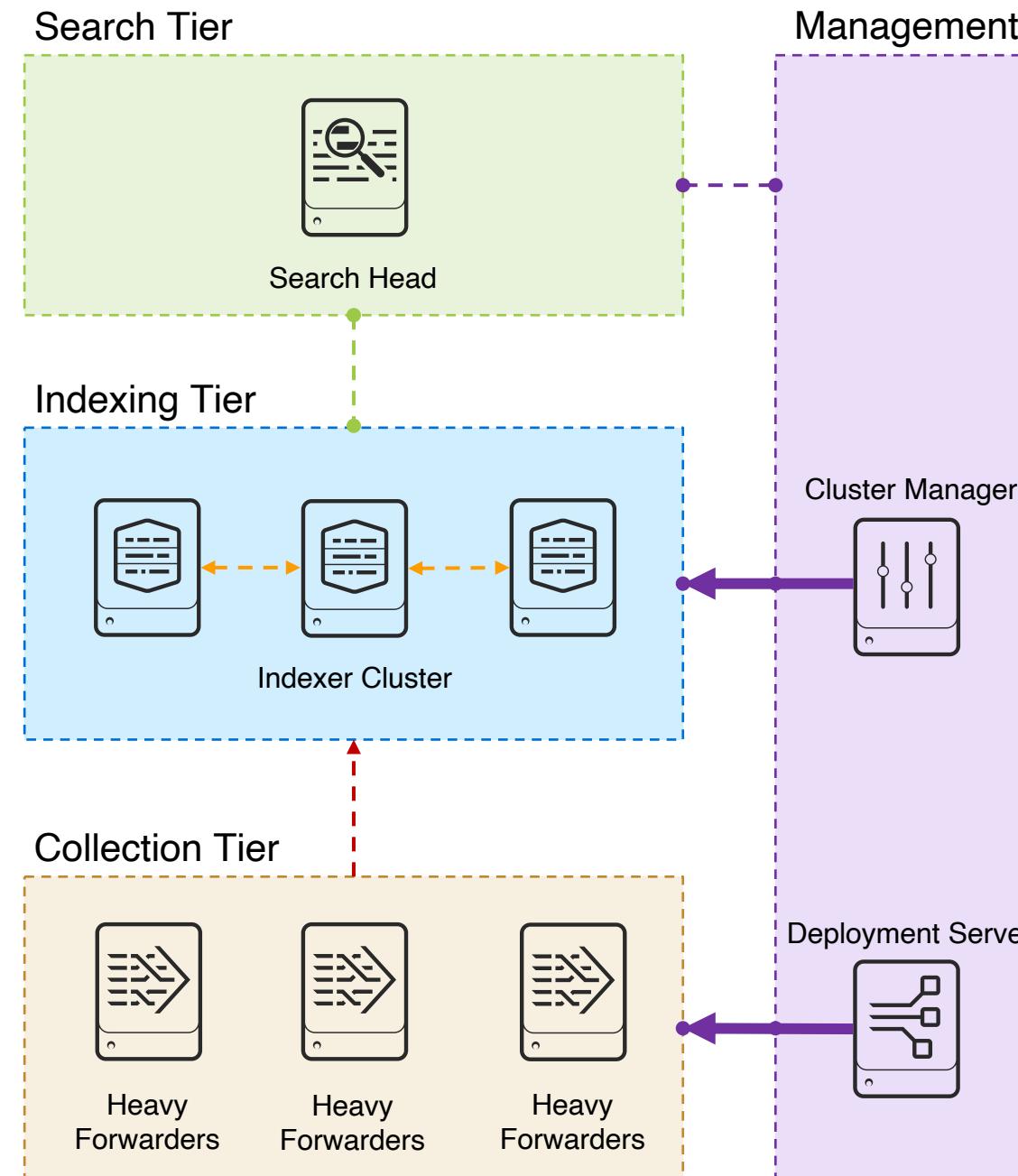
Distributed Rule Deployment

Indexers:

`SPLUNK_HOME/etc/apps/
splunk_ingest_actions/local`

Heavy Forwarder:

`SPLUNK_HOME/etc/apps/
splunk_ingest_actions/local`



- CM automatically distributes to Indexers during cluster bundle push
- Only single CM supported
- DS automatically distributes to HF during app distribution

Cluster Manager:

`SPLUNK_HOME/etc/manager-apps/
splunk_ingest_actions/local`

Deployment Server:

`SPLUNK_HOME/etc/deployment-apps/
splunk_ingest_actions/local`

Order of Ingest Action and Classic Rules

- Ingest Action rulesets are processed after existing transformation rules in the following order:
 1. Heavy Forwarder transforms
 2. Heavy Forwarder Ingest Action ruleset transforms *OR*
 3. Indexer transforms
 4. Indexer Ingest Action ruleset transforms
- Ingest Action rulesets on HF and Indexers will accept previously parsed data

Question: Methods For Masking Data

Which of the following is not a method that Splunk allows for masking data?

- A. Ingest Actions
- B. Using SEDCMD
- C. Using TRANSFORMS
- D. Using REST API

Answer: Methods For Masking Data

Which of the following is not a method that Splunk allows for masking data?

- A. Ingest Actions
- B. Using SEDCMD
- C. Using TRANSFORMS
- D. Using REST API**

Answer here

Question: Supported IA Rules

Which of the following is not supported by Ingest Action rulesets?

- A. Masking data
- B. Compressing data
- C. Routing data
- D. Eliminating data

Answer: Supported IA Rules

Which of the following is not supported by Ingest Action rulesets?

- A. Masking data
- B. Compressing data**
- C. Routing data
- D. Eliminating data

Ingest Action rulesets can contain rules that mask, truncate, route or eliminate data.

Question: About **SEDCMD**

Which of the following is not true about **SEDCMD**?

- A. Uses `props.conf`
- B. Uses `transforms.conf`
- C. Used to mask or truncate raw data
- D. Cannot eliminate unwanted events

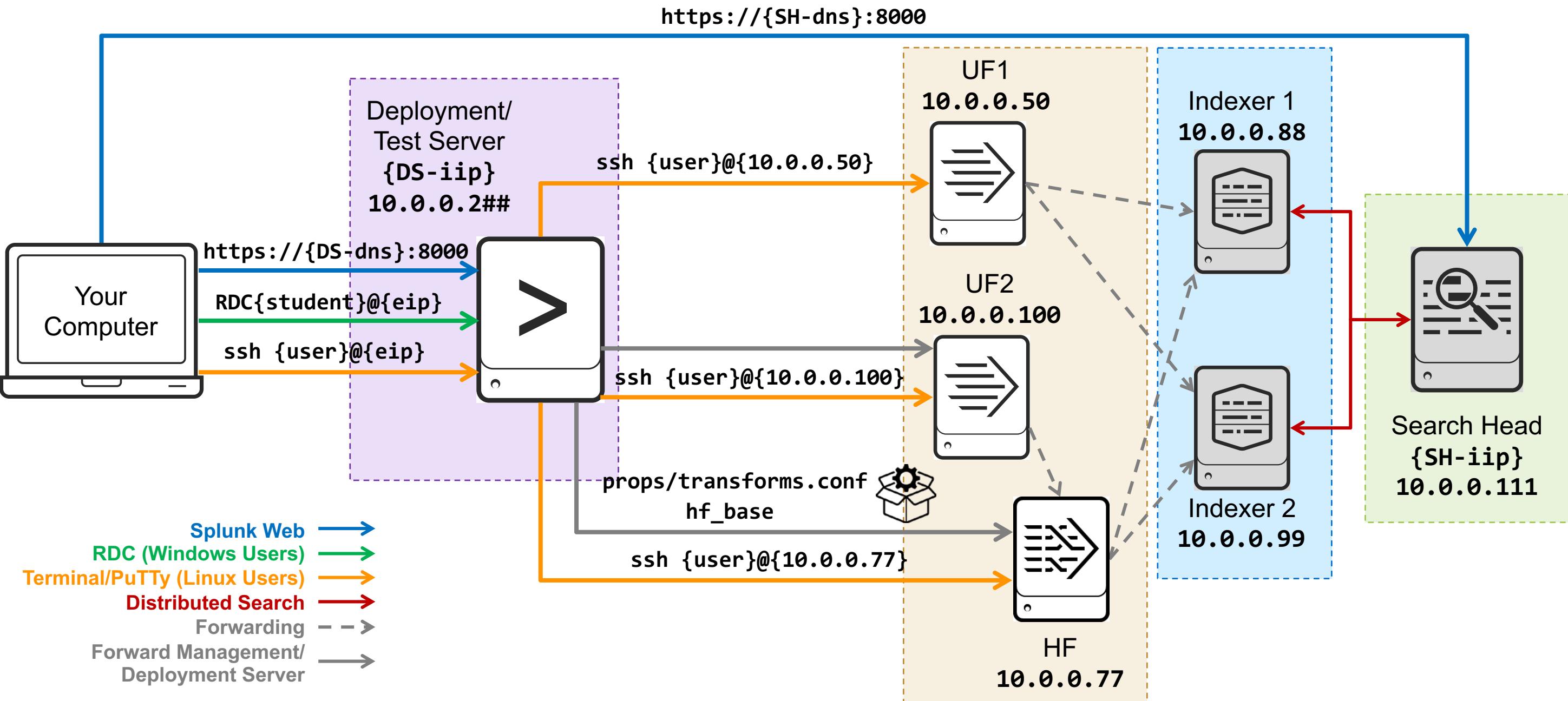
Answer: About SEDCMD

Which of the following is not true about **SEDCMD**?

- A. Uses `props.conf`
- B. Uses `transforms.conf`**
- C. Used to mask or truncate raw data
- D. Cannot eliminate unwanted events

To mask, truncate, or eliminate unwanted events you must use **TRANSFORMS**, which uses **props.conf** and **transforms.conf**.

Module 13 Lab – Environment Diagram



Module 13 Lab

Time: 30

Description: Manipulating Input Data

Tasks:

- Create an Ingest Action Ruleset on the DS to mask sensitive entries in **sales_entries.log**
- Verify input data from UF2 is properly masked by the HF
- Use **props.conf** and **transforms.conf** to mask sensitive entries in **vendor_sales.log**
- Copy **props.conf** and **transforms.conf** to the **hf_base** app
- Verify input data from UF2 is properly masked by the HF

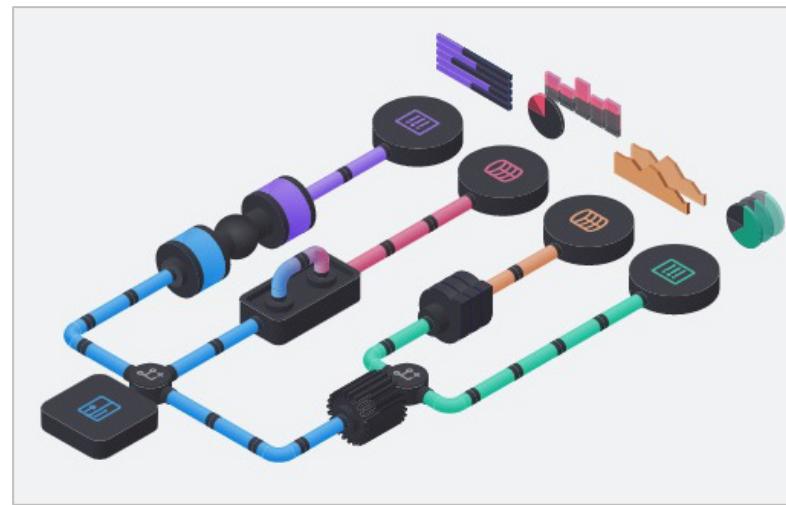
Module 14:

Routing Input Data

Module Objectives

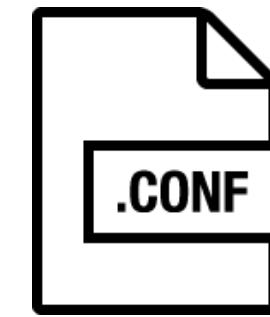
- Filter data with Ingest Action rules
- Route data with Ingest Action rules
- Route data with **TRANSFORMS**

Splunk Transformation Methods



Ingest Action

- Uses Splunk Web
- Can mask, truncate, route or eliminate data
- Introduced in Splunk 9.0



Classic Configuration

- Uses **SEDCMD** to mask or truncate data
- Uses **TRANSFORMS** to mask, truncate, route or eliminate data

Per-Event Index Routing

Again, if possible, specify the index for your inputs during the input phase (**inputs.conf**)

props.conf

```
[mysrctype]
TRANSFORMS-itops = route_errs_warns
```

transforms.conf

```
[route_errs_warns]
REGEX = (Error|Warning)
DEST_KEY = _MetaData:Index
FORMAT = itops
```

If **Error** or **Warning** is found in the incoming **_raw**, change its **index** field value to **itops**

Filtering Unwanted Events

- Route specific unwanted events to the **null queue**
 - Events discarded at this point do NOT count against your daily license quota

props.conf

```
[WinEventLog:System]
TRANSFORMS = null_queue_filter
```

transforms.conf

```
[null_queue_filter]
REGEX = (?i)^EventCode=(592|593)
DEST_KEY = queue
FORMAT = nullQueue
```

- The **(?i)** in the **REGEX** means “ignore case.”
- Events with an **eventcode** of **592** or **593** should not be indexed
- Route to **queue** and use **nullQueue** format to discard events

Routing Events to Groups using HF

Route specific events to different groups using the HF
(another use case for HF)

`props.conf`

```
[default]
TRANSFORMS-routing=errorRouting
```

```
[syslog]
TRANSFORMS-routing=syslogRouting
```

`transforms.conf`

```
[errorRouting]
REGEX = error
DEST_KEY=_TCP_ROUTING
```

```
FORMAT = errorGroup
```

```
[syslogRouting]
REGEX = .
DEST_KEY=_TCP_ROUTING
FORMAT=syslogGroup
```

`outputs.conf`

```
[tcpout]
defaultGroup=everythingElseGroup
```

```
[tcpout:errorGroup]
server=10.1.1.200:9999
```

```
[tcpout:syslogGroup]
server=10.1.1.197:9996,10.1.1.198:9997
```

```
[tcpout:everythingElseGroup]
server=10.1.1.250:9998
```

Filtering Input Data with Regular Expressions

The screenshot illustrates the process of filtering input data using regular expressions in Splunk. On the left, a sidebar shows options for MASK (Mask with Regular Expression), FILTER (Filter using Regular Expression, highlighted with a green box and an arrow pointing to the main window), and ROUTE (Route to Destination). The main window displays a 'Create New Ruleset' dialog for 'Badge_Access_US'. It includes a description 'Badge access in US cities', a data preview table showing two events (one from London and one from Boston), and a configuration section for a 'Filter using Regex' rule. The rule details are: Source Field: _raw, Drop Events Matching Regular Expression: London, and an 'Apply' button. A yellow callout box on the right lists the benefits of filtering:

- View affected events in red or by using filters
- View the reduction in event ingest data
- Filtered (eliminated) data does not count against daily license quota

Create New Ruleset
Badge_Access_US
Badge access in US cities

Data Preview for Filter

All Events (134)	Affected Events (71)	Unaffected Events (63)
6/22/2022 4:56:00.000 AM	Jun 22 2022 11:56:00 Address=1.1.1.R2 Address_Description=London Device=Proximity Reader Event_Description=Access Granted: Door Used rfid=145297537706 host = EC2AMAZ-B95HUCT source = C:\Program Files\Splunk\var\run\splunk\dispatch\1661494496.159\history_acces... sourcetype = badge_access-sample-d5pglyjlhy	6/22/2022 4:56:18.000 AM
		Jun 22 2022 11:56:18 Address=1.1.1.R2 Address_Description=Boston Device=Proximity Reader Event_Description=Access Granted: Door Used rfid=374765319282 host = EC2AMAZ-B95HUCT

Event Stream badge_access 19KB

Filter using Regex /London/ ↓ 53% | 9.4KB

Filter using regular expression

Source Field: _raw

Drop Events Matching Regular Expression: London

Learn more ↗

Apply

+ Add Rule ▾

Previewing and Saving the Ruleset

Create New Ruleset

Badge_Access_US

Badge access in US cities

> Event Stream badge_access 19KB Data Preview for Filter All Events (134) Affected Events (71) Unaffected Events (63) </> □

Preview Badge_Access_US

You are seeing a preview of Ruleset configuration. Copy/paste the stanzas below into your props.conf and transforms.conf files to deploy manually in your environment.

Caution: Manually editing Ruleset configurations may become incompatible with the user interface and not display in the user interface.

props.conf [Copy](#)

```
[badge_access]
RULESET-Badge_Access_US = _rule:Badge_Access_US:filter:rege
RULESET_DESC-Badge_Access_US = Badge access in US cities
```

transforms.conf [Copy](#)

```
[ _rule:Badge_Access_US:filter:regex:f0p7ce1a]
INGEST_EVAL = queue=if(match(_raw, "London"), "nullQueue",
STOP_PROCESSING_IF = queue == "nullQueue"
```

Preview Base Search:
This is the SPL used to generate the sample dataset.

SID:
1661494496.159

[Open in Search](#)

Post-process Search:
This SPL is applied to the sample dataset to simulate and preview the effect of the currently selected rule.

```
eval queue=if(match(_raw, "London"), "nullQueue", null())
```

[Open in Search](#)

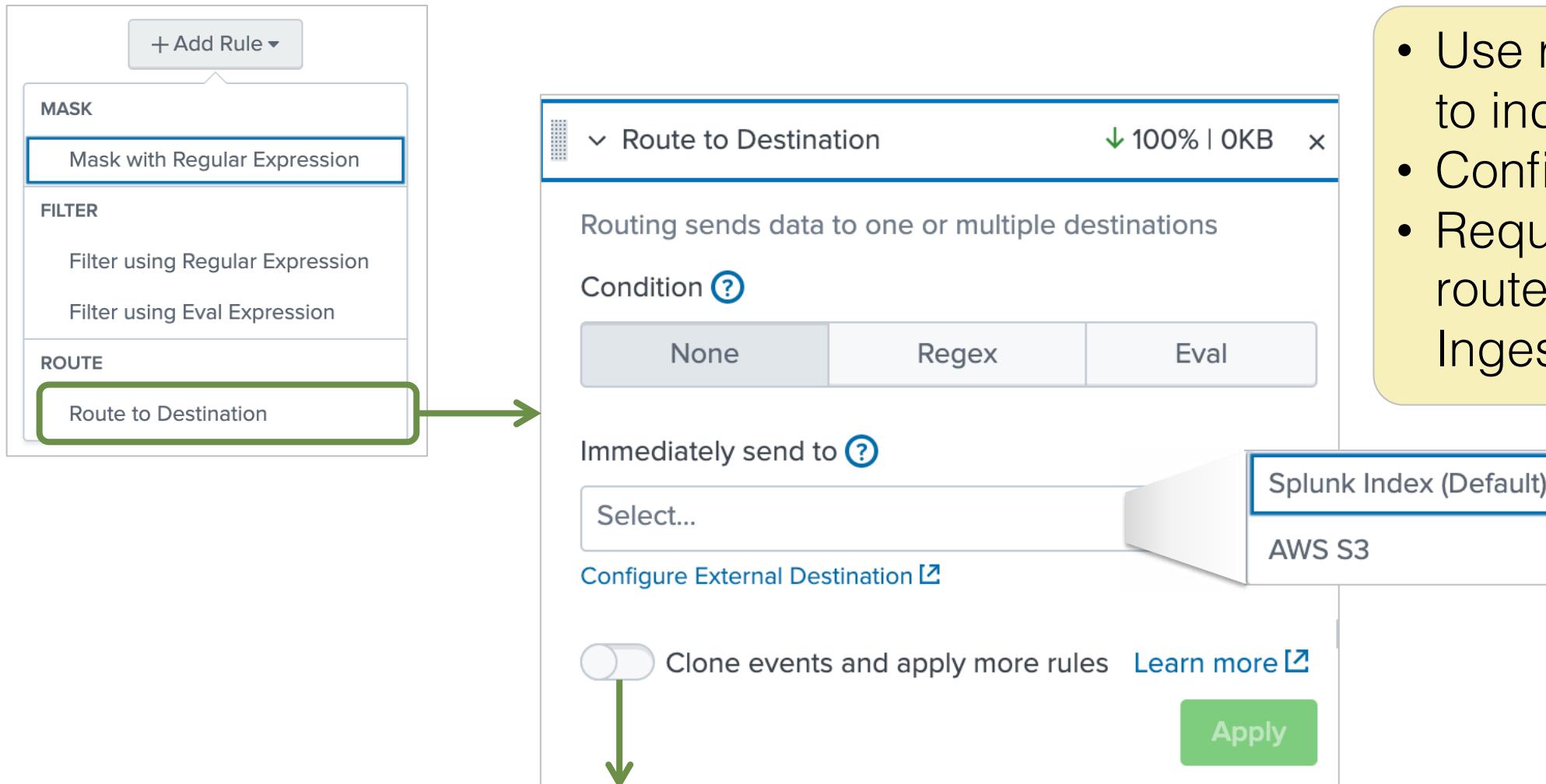
Filtering Input Data with Eval Expressions

The screenshot shows the Splunk Rules Editor interface. On the left, a sidebar lists rule types: MASK, FILTER, and ROUTE. Under MASK, 'Mask with Regular Expression' is selected. Under FILTER, 'Filter using Regular Expression' and 'Filter using Eval Expression' are listed; 'Filter using Eval Expression' is highlighted with a green border and has a green arrow pointing to it from the sidebar. Under ROUTE, 'Route to Destination' is listed. The main panel shows an event stream named 'badge_access'. A 'Filter using Regex' step is shown above a 'Filter using Eval' step. The 'Filter using Eval' step has a dropdown menu open, displaying the expression 'len(_raw) > 155'. A yellow callout box contains two bullet points: '• Use eval expressions when warranted' and '• Use multiple rules for the ruleset as needed'. At the bottom right of the main panel is a green 'Apply' button.

- Use eval expressions when warranted
- Use multiple rules for the ruleset as needed

docs.splunk.com/Documentation/Splunk/latest/SearchReference/CommonEvalFunctions

Routing Data to Destinations



- Use route to destination to send to index, AWS S3, or both
- Configure AWS S3 on the HF
- Requires Splunk 9.0.1 and later to route to multiple indexes using Ingest Actions

- Applies currently defined rules, and routes the stream to the destination
- Applies any additionally defined rules against the event stream and routes that subset to the destination defined in the next Route to Destination rule

Question: Eliminating Data with Ingest Actions

Which Ingest Action rule is used to eliminate data?

- A. Mask with Regular Expression
- B. Filter using Regular Expression
- C. Routing to Destination: Index
- D. Routing to Destination: Null

Answer: Eliminating Data with Ingest Actions

Which Ingest Action rule is used to eliminate data?

- A. Mask with Regular Expression
- B. Filter using Regular Expression**
- C. Routing to Destination: Index
- D. Routing to Destination: Null

The Filter using Regular Expression and Filter using Eval Expression rules are used to eliminate unwanted events.

Question: TRANSFORMS and FORMAT

Given this **TRANSFORMS** configuration, what does **errorGroup** define?

- A. Format for discarding unwanted events
- B. Server group to route events to
- C. Settings for changing event metadata
- D. Group events for masking using regex

`props.conf`

```
[default]
TRANSFORMS-routing=errorRouting
```

`transforms.conf`

```
[errorRouting]
REGEX = error
DEST_KEY=_TCP_ROUTING
FORMAT = errorGroup
```

Answer: TRANSFORMS and FORMAT

Given this **TRANSFORMS** configuration, what does **errorGroup** define?

- A. Format for discarding unwanted events
- B. Server group to route events to**
- C. Settings for changing event metadata
- D. Group events for masking using regex

Using **DEST_KEY=_TCP_ROUTING** routes events to server group **errorGroup**, as configured in **outputs.conf**.

props.conf

```
[default]
TRANSFORMS-routing=errorRouting
```

transforms.conf

```
[errorRouting]
REGEX = error
DEST_KEY=_TCP_ROUTING
FORMAT = errorGroup
```

outputs.conf

```
[tcpout:errorGroup]
server=10.1.1.200:9999
```

Module 14 Lab

Time: 30 min

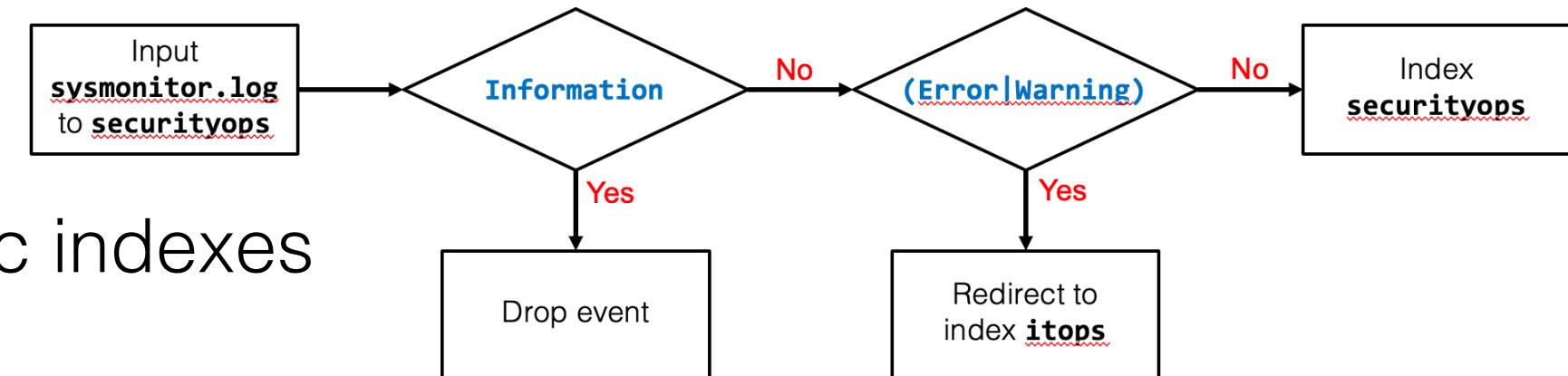
Description: Routing Input Data

Tasks:

- Use Ingest Actions to drop unwanted events



- Use **props.conf** and **transforms.conf** to:
 - Redirect events to specific indexes
 - Drop unwanted events

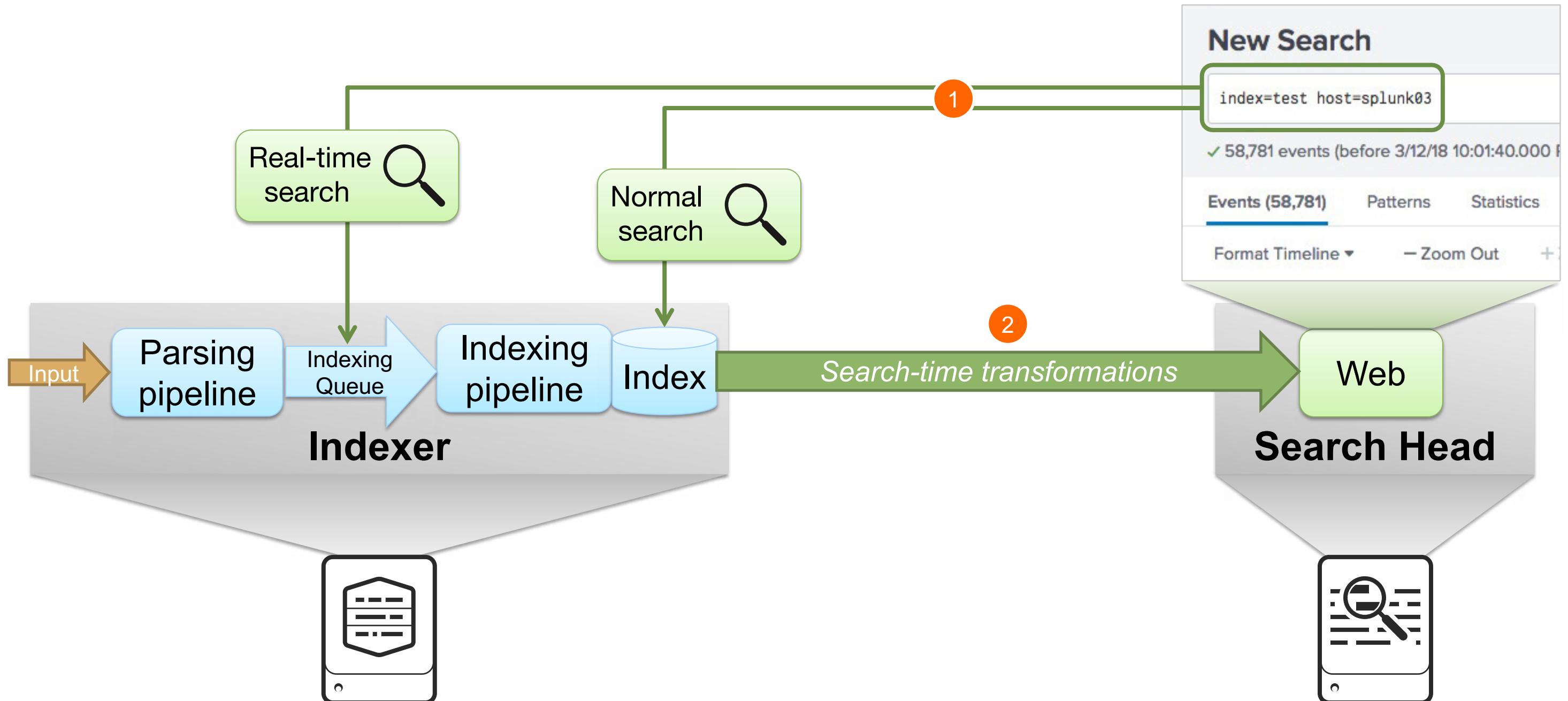


Module 15: Supporting Knowledge Objects

Module Objectives

- Define default and custom search time field extractions
- Identify the pros and cons of indexed time field extractions
- Configure indexed field extractions
- Describe default search time extractions
- Manage orphaned knowledge objects

Search Phase: The Big Picture

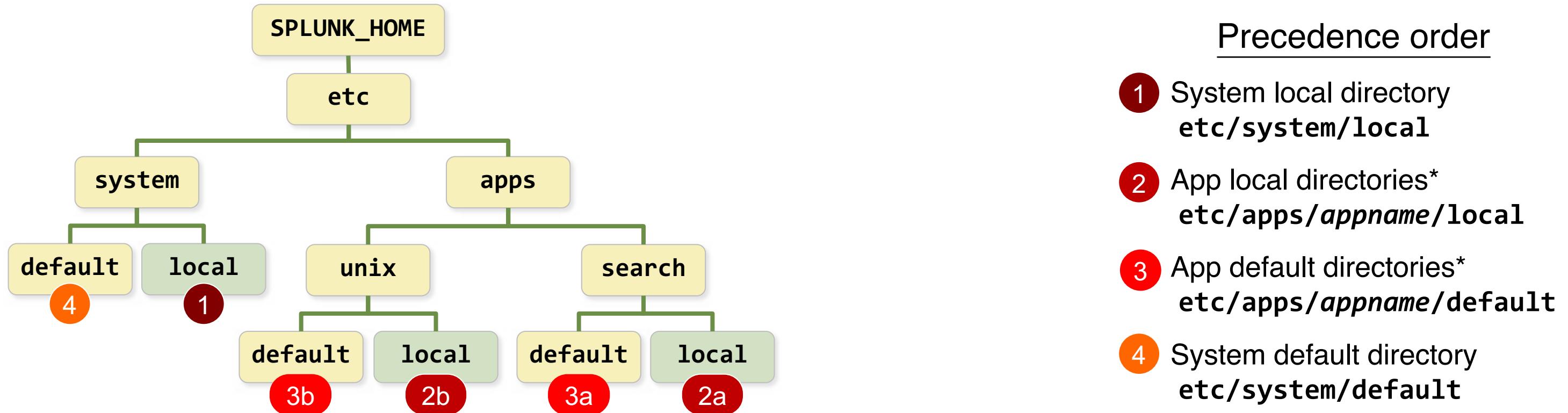


File Context and Index-time versus Search-time

	Global Context	App/User Context
<i>Used during:</i>	Index-time	Search-time
<i>Used by:</i>	<ul style="list-style-type: none">• User-independent tasks• Background tasks• Input, parsing, indexing	<ul style="list-style-type: none">• User-related activity• Searching• Search-time processing
<i>Example use-case:</i>	A network input to collect syslog data	Mary's private report in the Search app
<i>Example files:</i>	inputs.conf outputs.conf props.conf	macros.conf savedsearches.conf props.conf

docs.splunk.com/Documentation/Splunk/latest/Admin/Wheretofindtheconfigurationfiles

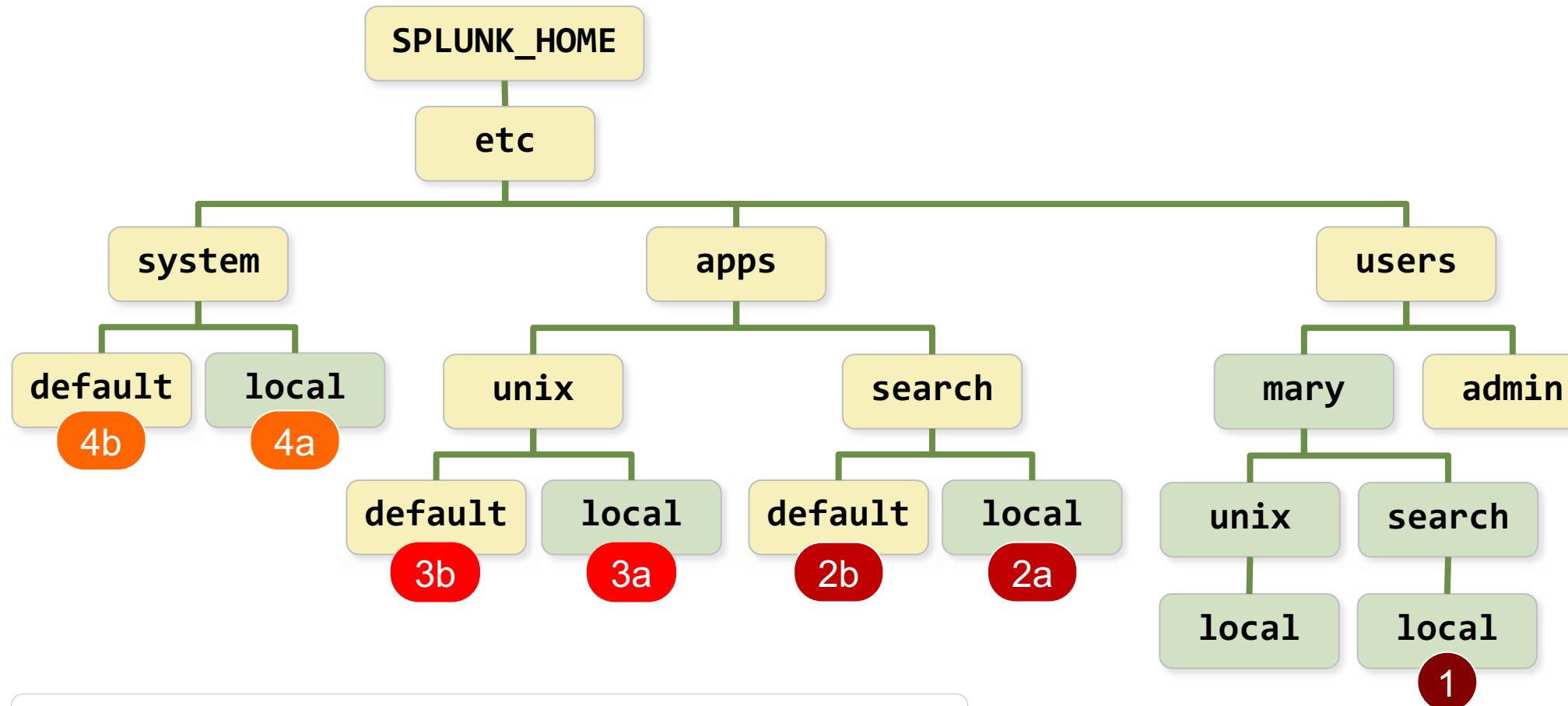
Review: Index-Time Precedence (Global Context)



Note

* When determining priority of app directories in **global** context (for steps 2 and 3), Splunk uses *lexicographical* order. (Files in apps directory "A" have higher priority than files in apps directory "B".)

Search-Time Precedence (App/User Context)



Note

* If objects from the app are exported globally with **.meta** file setting, evaluate all other app directories using *reverse lexicographical* order. (Files in apps directory "B" have higher priority than directory "A".)

Precedence order

- 1 Current user directory for app
`etc/users/user/appname/local`
- 2 App directory - running app
`etc/apps/appname/local`
`etc/apps/appname/default`
- 3 App directories - all other apps*
`etc/apps/appname/local`
`etc/apps/appname/default`
- 4 System directories
`etc/system/local`
`etc/system/default`

Indexed Field Extraction

- Fields are generally extracted at search-time
- During index-time, event data is stored in the index on disk
 - Default fields are extracted and added automatically
 - Custom fields are added based on customizations (by the administrator)
- Certain use cases result in indexed fields
 - Inputs phase (usually on the forwarder) for structured inputs
 - Parsing phase (usually on the indexer) for fields that may be negatively impacting search performance
- Add custom indexed fields only if necessary
 - Can negatively impact indexing performance and search times
 - Increases the size of the searchable index

Pros/Cons of Indexed Field Extractions

PROs	CONs
<ul style="list-style-type: none">• Provision the extraction during the input or parsing phase• Can configure on the universal forwarder• Auto-formatting• Can drop useless headers and comments	<ul style="list-style-type: none">• Increased storage size (2-5x the original size consumed on the indexer)• Static field names: additional step required for late-binding use cases• Possible performance implications• Less flexible: changes to fields require a re-index of the dataset, or only apply to new data

- Recommendations:
 - For frequently re-configured delimited sources, use indexed extractions (example: **IIS**)
 - For static CSV, use **REPORT** and **DELIMS**, or other search-time extractions
 - Use a dedicated index

Configuring Indexed Field Extractions

Define additional attributes in **props.conf**, **transforms.conf**, and fields in **fields.conf**

File	Splunk instance	Example
props.conf	Indexer, Heavy Forwarder	<pre>[testlog] TRANSFORMS-netscreen = netscreen-error</pre>
transforms.conf	Indexer, Heavy Forwarder	<pre>[netscreen-error] REGEX = device_id=\[\w+\](?<error_code>[^:]++) FORMAT = error_code::"\$1" WRITE_META = true</pre>
fields.conf	Search Head	<pre>[error_code] INDEXED=true</pre>

Structured Data Field Extraction Example

- Indexed extractions are input phase **props.conf** settings
 - In this scenario, the settings belong on forwarder
 - Check **props.conf.spec** for more options

```
[my_structured_data]
INDEXED_EXTRACTION = w3c
HEADER_FIELD_LINE_NUMBER = 4
TIMESTAMP_FIELDS = date, time
```

```
#Software: Microsoft Internet Information Services 7.5
#Version: 1.0
#Date: 2015-06-08 00:00:00
#Fields: date time cs-method cs-uri-stem cs-uri-query c-ip cookie referer cs-host sc
2015-01-08 00:00:00 POST AutoComplete.asmx/GetCompletionList - 10.175.16.79
cApproved=1;+fParticipant=000000695607440|urn:System-Services:GatewayTokenService_n
format:persistent|http://www.acme.com/2015/06/attributes/credentialidentifier; &nest
fc2df5;+style=normal https://search.acme.com/Account/Account.aspx?redirect=https://d
200 1113 0
...
```

Source type: iis ▾ Save As

> Event Breaks

> Timestamp

Advanced

Name	Value
CHARSET	UTF-8
INDEXED_EXTRACTI <ins>TION</ins>	w3c
MAX_TIMESTAMP_L <ins>E</ins>	32
SHOULD_LINEMERG	false
category	Web
description	W3C Extended log format pro
disabled	false
pulldown_type	true

Previewing Structured Data

Add Data

Select Source Set Source Type Input Settings Review Done

< Back Next >

Set Source Type

This page lets you see how Splunk sees your data before indexing. If the events look correct and have the right timestamps, click "Next" to proceed. If not, use the options below to define proper event breaks and timestamps. If you cannot find an appropriate source type for your data, create a new one by clicking "Save As".

Source: **Traffic_Violations.csv**

View Event Summary

Source type: csv ▼ Save As

Timestamp

Extraction: Auto, Current time, Advanced...

Time zone: Auto

Timestamp format: A string in strftime() format that helps Splunk recognize timestamps. [Learn More](#)

Timestamp fields:

Table ▾ Format 20 Per Page ▾ ◀ Prev 1 2 3 4 5 6 7 8 ... Next >

	_time	Accident	Agency	Alcohol	Arrest Type	Article	Belts	Charge	Color	Commercial License
1	9/24/13 5:11:00.000 PM	No	MCP	No	A - Marked Patrol	Transportation Article	No	13-401(h)	BLACK	No
2	8/29/17 10:19:00.000 AM	No	MCP	No	A - Marked Patrol	Transportation Article	No	21-201(a1)	GREEN	No
3	12/1/14 12:52:00.000 PM	No	MCP	No	A - Marked Patrol	Transportation Article	No	21-403(b)	SILVER	No

Splunk automatically identifies structured data and parses the event boundaries and field names

- Produces an **indexed extraction stanza**
- If you see a timestamp warning, indicate where to find a timestamp by specifying a field name

Indexed Field Extractions – Caveat

- Splunk software does not parse structured data that has been forwarded to an indexer
 - If you have configured **props.conf** on the targeted forwarder with **INDEXED_EXTRACTIONS** and its associated attributes, the forwarded data skips the following queues on the indexer:
 - ▶ Parsing
 - ▶ Aggregation
 - ▶ Typing

[http://docs.splunk.com/Documentation/Splunk/latest/Forwarding/Routeandfilterdata
#Caveats for routing and filtering structured data](http://docs.splunk.com/Documentation/Splunk/latest/Forwarding/Routeandfilterdata#Caveats_for_routing_and_filtering_structured_data)

Default Search Time Field Extractions

- Provided by Splunk for common source types
- Can be discovered by Splunk from your search results
 - Automatically detects key/value pairs (e.g. **a=1**)
- Can be added with add-ons and apps

Splunk App for Unix and Linux	Has many search time fields for standard UNIX logs, such as secure.log , messages.log , and so on
Splunk App for Windows	Has many defaults for Windows data
For other data	Look for an app on splunkbase.splunk.com specifically designed for that type of data

Custom Search Time Field Extractions

SPL

- Use **rex** (or similar) commands in the search language
- Requires knowledge of regular expressions (REGEX)
- All roles can use this command

Field Extractor

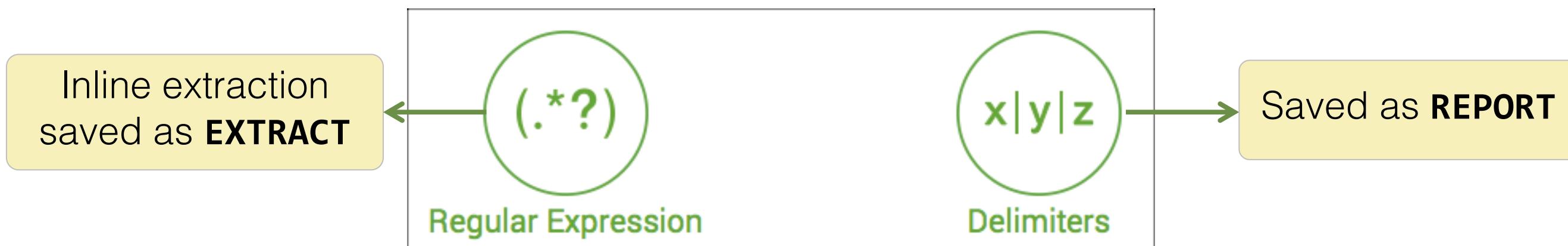
- Found in Splunk Web
- Handles REGEX-based and delimiter-based extractions
- Knowledge of regular expressions helpful, but not required

Configuration files

- Provides additional advanced extraction options
- Knowledge of REGEX required
- Available only to admins

Field Extractions and `props.conf`

- Field extraction happens during index-time (indexed fields) and/or search-time (extracted fields)
- Search-time extractions can be inline or a field transform
- Use extraction directives
 - **EXTRACT** (inline extraction)
 - Defined in `props.conf` as single field extraction
 - **REPORT** (field transform)
 - Defined in `transforms.conf`
 - Invoked from `props.conf`



REPORT Extractions in props.conf

- **REPORT** references a transform defined separately in **transforms.conf**
- In **transforms.conf**, you can
 - Define field extractions using delimiters
 - Apply other advanced extraction techniques
- For full details on **REPORT**, see:

docs.splunk.com/Documentation/Splunk/latest/Knowledge/Createandmaintainsearch-timefieldextractionsthroughconfigurationfiles

Using EXTRACT and REPORT in props.conf

- Applies to this sourcetype
- The REGEX pattern defines extracted field

Arbitrary namespace you assign to this extraction.
Useful for ordering multiple transactions

props.conf

```
[tradelog]
EXTRACT-1type = type:\s(?<acct_type>\S+)
```

Extracted field name

```
[sysmonitor]
REPORT-sysmon = sysmon-headers
KV_MODE = none
```

Process this stanza in transforms.conf

transforms.conf

```
[sysmon-headers]
DELIMS = ","
FIELDS = Time,EventCode,EventType,Type,ComputerName,LogName,RecordNumber
```

Lookups

- A Splunk data enrichment knowledge object
 - Uses stanzas defined in **transforms.conf** and **props.conf**
 - Used *only* during search time
- Four types:

Lookup type	Description
File-based	Uses a CSV file stored in the lookups directory
KV Store	Requires collections.conf that defines fields
External	Uses a python script or an executable in the bin directory
Geospatial	Uses a kmz saved in the lookups directory to support the choropleth visualization

Add new

Lookups » Lookup definitions » Add new

Destination app: search

Name *:

Type: File-based
 External
 KV Store
 Geospatial
geo_ddw_countries.csv

Lookup file *:

Create and manage lookup table files.

Configure time-based lookup
 Advanced options

Other Search Time Knowledge Objects

- KOs are stored in configuration files:
 - **macros.conf**, **tags.conf**, **eventtypes.conf**, **savedsearches.conf**, etc.
 - See docs and ***.spec** files in **SPLUNK_HOME/etc/system/README**
- Create or modify KOs using:
 - Splunk Web (automatically updates **.conf** files)
 - Editing **.conf** files manually (requires admin rights)
 - Use **btool** to verify changes
 - Splunk Web: Advanced edit (supports some system settings)

Search name	RSS feed	Scheduled time	Display view	Owner	App	Alerts	Sharing	Status	Actions
quake_L24h	None	None	emaxwell	search	0	Private Permissions	Enabled Disable	Run Advanced edit	Clone Move Delete
quake_L24H	None	None	admin	search	0	Private Permissions	Enabled Disable	Run Advanced edit	Clone Move Delete
Top five sourcetypes	None	None	No owner	search	0	App Permissions	Enabled Disable	Run Advanced edit Clone	

Orphaned Knowledge Objects (KOs)

What are orphaned knowledge objects?

- KOs without a valid owner
- Occurs when a Splunk account is deactivated and the KOs associated with that account remain in the system

Issues with orphaned knowledge objects

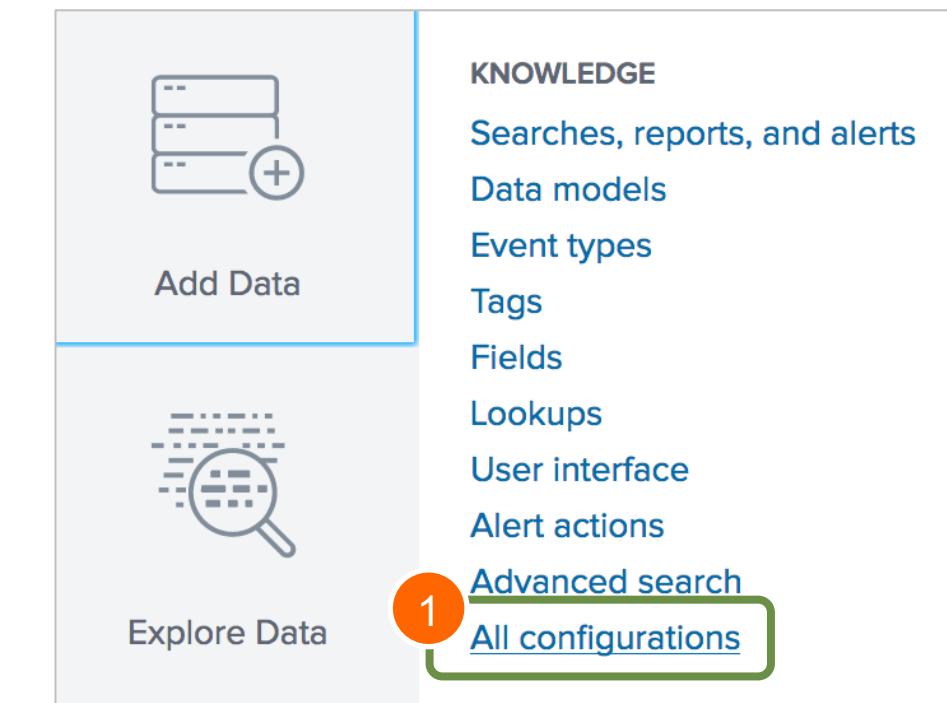
- Can cause performance problems and security concerns
- Searches that refer to an orphaned lookup may not work
- Search scheduler cannot run a report on behalf of a nonexistent owner

Locating Orphaned Knowledge Objects

- Splunk runs a default search on a daily schedule to detect orphaned scheduled reports
- Report on orphaned KO using any of these methods:
 - Click **Messages**, then click the message link to access the alerts dashboard
 - Run the search from **Search > Dashboards > Orphaned Scheduled Searches, Reports, Alerts**
 - Run the MC Health Check search to detect orphaned knowledge objects

Reassigning Knowledge Objects

- Requires **admin** role capability
- Possible for both orphaned and owned KOs
- Performed in Splunk Web with:
 1. Select **Settings > All configurations**
 2. Click **Reassign Knowledge Objects**



A screenshot of the 'All configurations' page in Splunk Web. The page title is 'All configurations' and it shows 'Showing 1-25 of 263 items'. There are several filter options at the top: App (Instrumentation (splu...)), Owner (Any), Visible in the App (Visible in the App), and a search bar with a 'filter' placeholder and a magnifying glass icon. Below the filters is a pagination section with '25 per page' and a set of numbered buttons from 1 to 10. A green rounded rectangle highlights the 'Reassign Knowledge Objects' button, which is located in the top right corner of the page area. A red circle containing the number 2 is placed over this highlighted button.

Reassigning Knowledge Objects (cont.)

Reassign Knowledge Objects

Select knowledge objects and reassign them to another user. [Learn more](#)

263 Knowledge Objects All Orphaned Object type: All ▾ All Objects ▾ App: Instrumentation (splunk_instrumentation) ▾ Filter by Owner ▾ filter 10 per page ▾

Edit Selected Knowledge Object (0) ▾

Name	Actions	Object type
ActiveDirectory : EXTRACT-GUID	Reassign	props-extract
ActiveDirectory : EXTRACT-SID	Reassign	props-extract
ActiveDirectory : REPORT-MESSAGE	Reassign	props-extract
PerformanceMonitor : REPORT-MESSAGE	Reassign	props-extract

Note You can also reassign multiple knowledge objects by selecting the check boxes next to the objects, then selecting Edit Selected Knowledge Objects > Reassign.

• Use the filter options at the top to locate the objects you want to reassign
• The Orphaned button displays all shared, orphaned objects

1. Click Reassign
2. Select a new owner from the New Owner drop-down menu
3. Click Save

Reassign Entity

⚠️ Knowledge object ownership changes can have side effects such as giving saved searches access to previously inaccessible data or making previously available knowledge objects unavailable. Review your knowledge objects before you reassign them. [Learn more](#)

Name ActiveDirectory : EXTRACT-GUID
Type props-extract
Owner nobody
New Owner Select an owner ▾ 2

Lookup an owner 3

Administrator (admin)
SH_alf (alf)
SH_beta (beta)
(emaxwell)
SH_nic (nic)
Nobody

Question: Contents of `props/transforms.conf`

The `props.conf` and `transforms.conf` files do not store which of the following?

- A. Field Extractions
- B. Line Break configurations
- C. Lookups
- D. Saved Searches

Answer: Contents of `props/transforms.conf`

The `props.conf` and `transforms.conf` files do not store which of the following?

- A. Field Extractions
- B. Line Break configurations
- C. Lookups
- D. Saved Searches

`props.conf` and `transforms.conf` store Splunk's processing properties and configure data transformations. Saved searches are stored in `savedsearches.conf`.

Question: Lookup Types

Which of the following is not one of the four lookup types used during search time?

- A. File-based
- B. KV Store
- C. Internal
- D. Geospatial

Answer: Lookup Types

Which of the following is not one of the four lookup types used during search time?

- A. File-based
- B. KV Store
- C. Internal
- D. Geospatial

Lookup types include: 1. File-based (uses a **CSV** file), 2. KV Store (requires **collections.conf**), 3. External (uses a python script or executable), and 4. Geospatial (uses a **kmz** to support choropleth visualizations).

Question: Reassigning KOs

Which users can reassign knowledge objects?

- A. Users with the `user` role
- B. Users with the `power` role
- C. Users with the `splunk-system-role` role
- D. Users with the `admin` role

Answer: Reassigning KOs

Which users can reassign knowledge objects?

- A. Users with the `user` role
- B. Users with the `power` role
- C. Users with the `splunk-system-role` role
- D. Users with the `admin` role

Only users with the `admin` role capability can reassign knowledge objects. This is possible for both orphaned and owned KOs.

Module 15 Lab

Time: 5-10 minutes

Description: Supporting Knowledge Objects

Tasks:

- Create a knowledge object (report)
- Search for orphaned knowledge objects
- Assign the report to the user, **emaxwell**

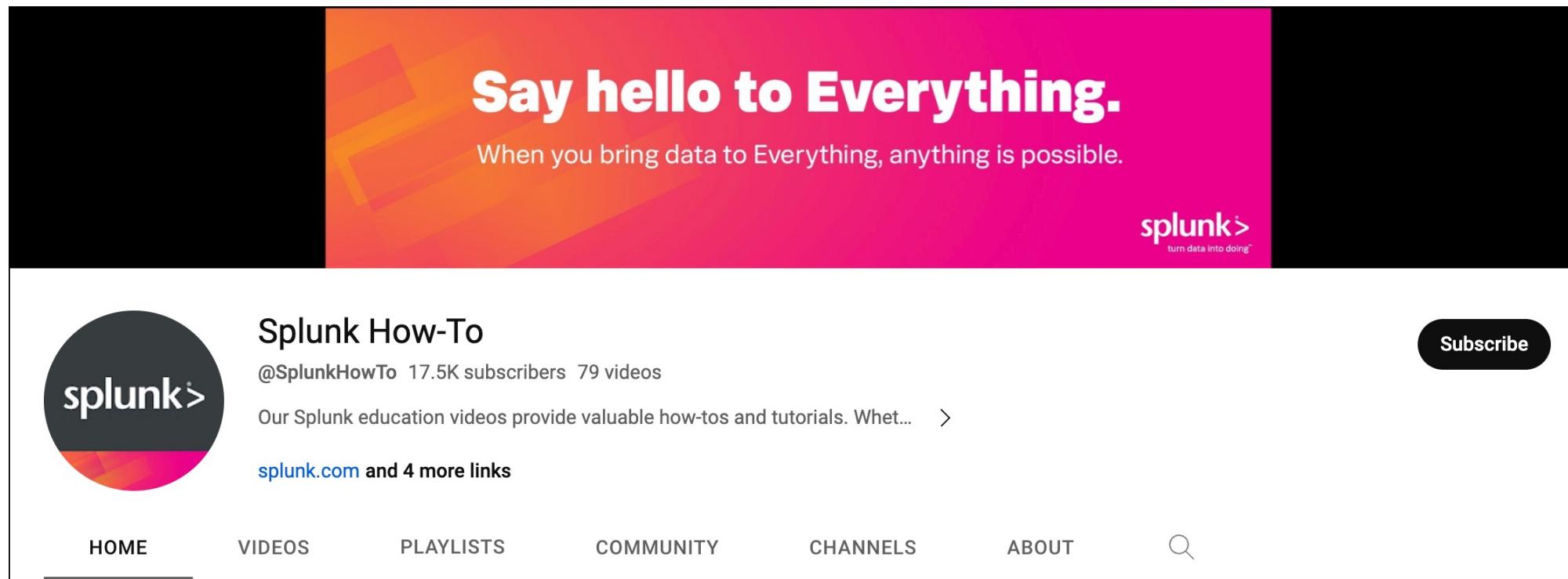
Wrap-up Slides

Community

- Splunk Community Portal
community.splunk.com
 - Answers
 - Discussions
 - Splunk Trust
 - User Groups
 - Ideas
- Splunk Blogs –
splunk.com/blog/
- Splunk Apps –
splunkbase.com
- Splunk Dev Google Group
groups.google.com/forum/#!forum/splunkdev
- Splunk Docs on Twitter –
twitter.com/splunkdocs
- Splunk Dev on Twitter –
twitter.com/splunkdev
- Splunk Live! –
splunklive.splunk.com
- .conf – conf.splunk.com

Splunk How-To Channel

- Check out the Splunk Education How-To channel on YouTube:
splk.it/How-To
- Free, short videos on a variety of Splunk topics



Support Programs

- Web
 - Documentation: dev.splunk.com and docs.splunk.com
- Splunk Lantern
 - Guidance from Splunk experts
 - lantern.splunk.com
- Global Support
 - Support for critical issues, a dedicated resource to manage your account – 24 x 7 x 365
 - Web: splunk.com/index.php/submit_issue
- Enterprise, Cloud, ITSI, Security Support
 - Web: splunk.com/en_us/about-splunk/contact-us.html#tabs/customersupport
 - Phone: (855) SPLUNK-S or (855) 775-8657

Support ^

Support Portal

Submit a case ticket

Splunk Answers

Ask Splunk experts questions

Contact Us

Contact our customer support

Product Security Updates

Keep your data secure

System Status

Splunk Mobile

- Free app available to all Splunk Cloud and Splunk Enterprise customers
- Analyze data and receive actionable alerts on-the-go with mobile-friendly dashboards
- iOS and Android
- See the [Product Brief](#)
- Download for iOS splk.it/ios and Android splk.it/android



Thank You

splunk>