



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

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Course Prerequisites

- Required:
 - Splunk Fundamentals 1
 - Splunk Fundamentals 2
- Strongly Recommended:
 - 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 field extractions

Course Outline

Module 1: Introducing Splunk Data Administration

Module 2: Getting Data In - Staging

Module 3: Forwarder Configuration

Module 4: Heavy Forwarders & Forwarder Management

Module 5: Monitor Inputs

Module 6: Network and Scripted Inputs

Module 7: Windows and Agentless Inputs

Module 8: Fine-tuning Inputs

Module 9: Parsing Phase and Data Preview

Module 10: Manipulating Raw Data

Module 11: Supporting Knowledge Objects

Data Administrator vs System Administrator

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

Splunk System Administrator

System management

- Install, configure, and manage Splunk components
- Install and manage Splunk apps
- Manage Splunk licensing
- Manage Splunk indexes
- Manage Splunk users and authentication
- Manage Splunk configuration files
- Monitor MC and respond to system health alerts

Module 1: Introducing Splunk Data Administration

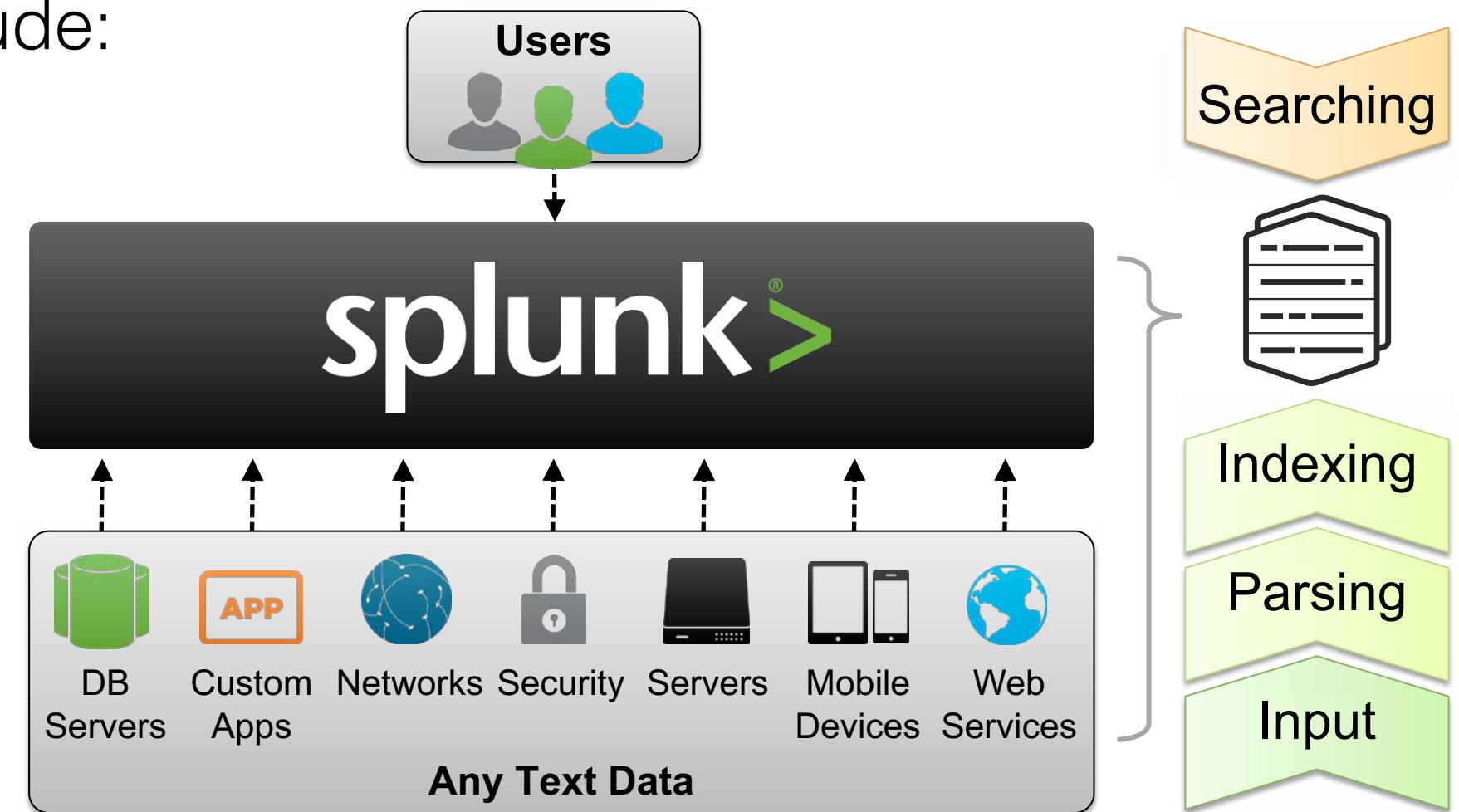
Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Module 1 Objectives

- Provide an overview of Splunk
- Describe the four phases of the distributed model
- Identify Splunk configuration files and directories
- Describe index-time and search-time precedence
- Use **btool** to retrieve Splunk configuration information

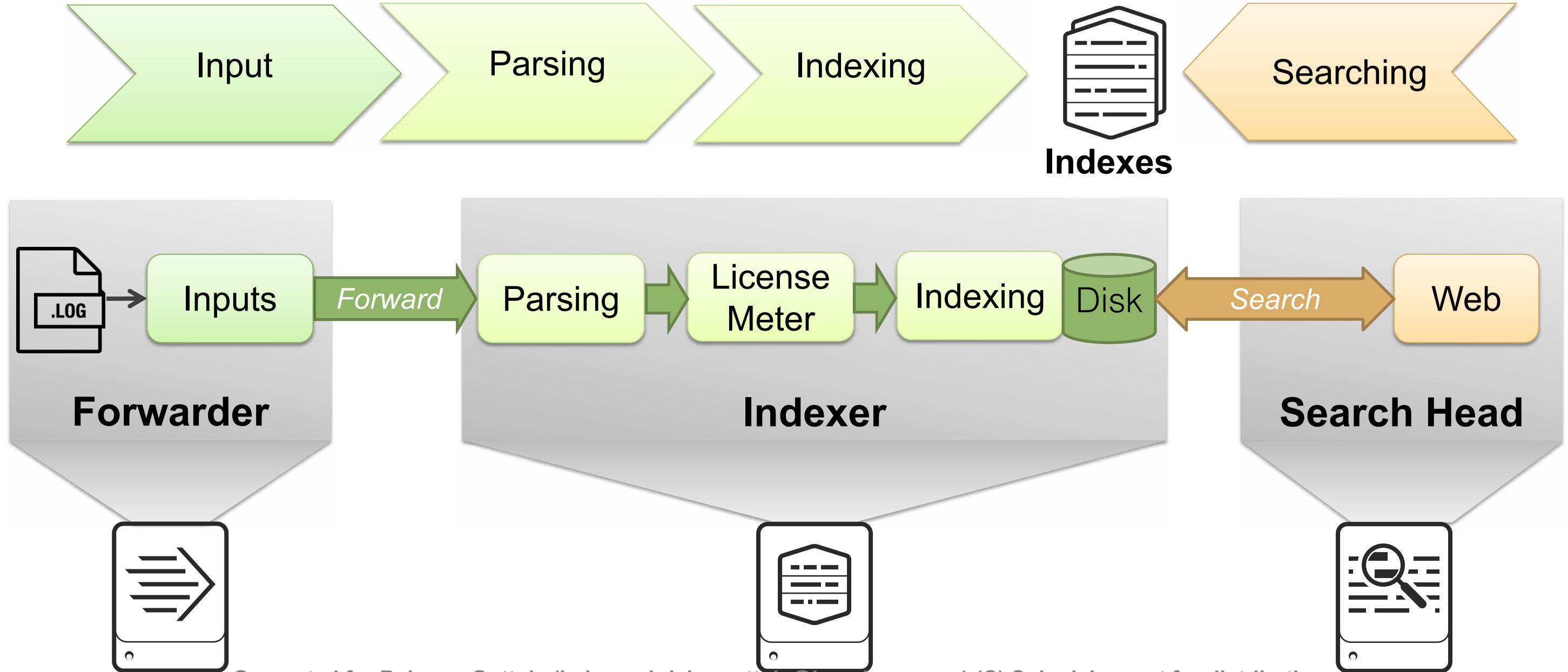
Splunk Overview

- 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



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

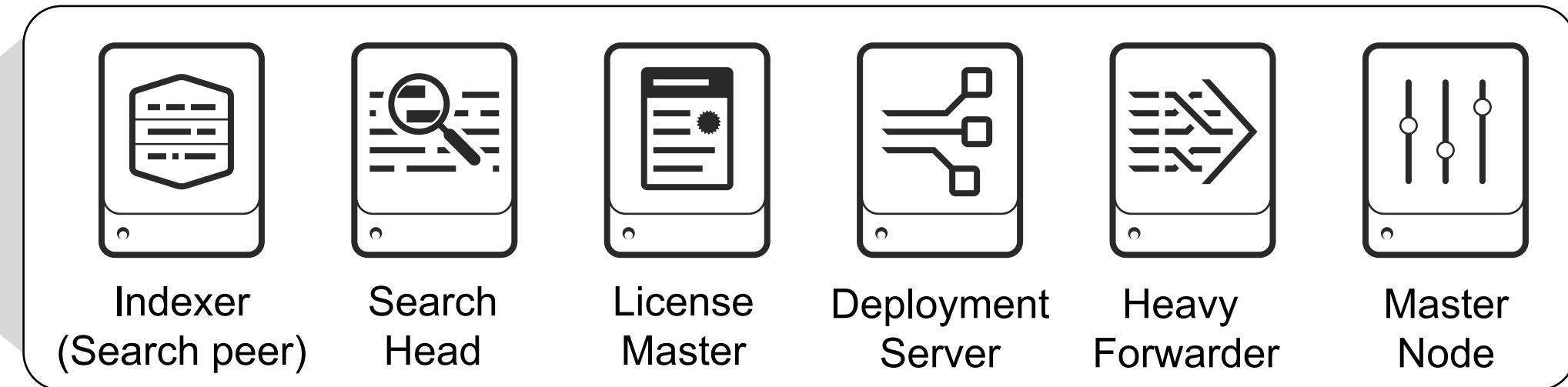
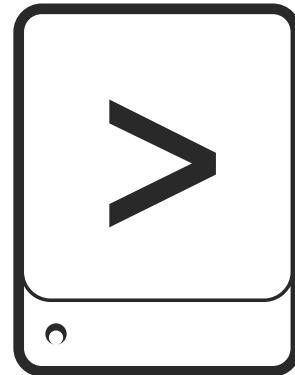
The Four Phases of the Distributed Model



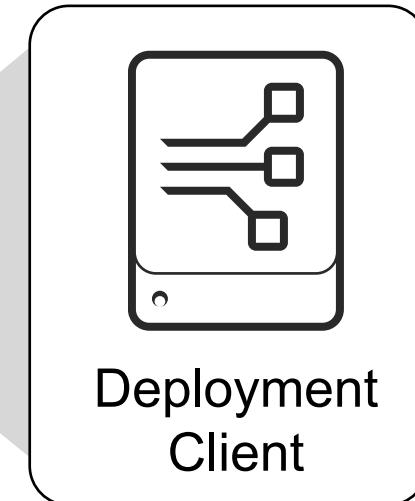
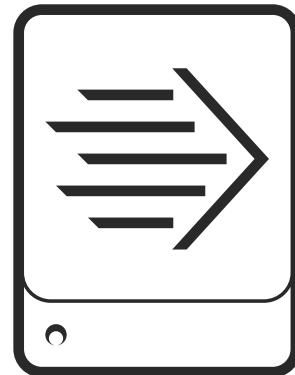
Generated for Balaram Guttula (balaram.krishn.guttula@pmchase.com) (C) Splunk Inc, not for distribution

Software in Splunk Enterprise packages

**Splunk
Enterprise
package**



**Universal
Forwarder
package**



Note

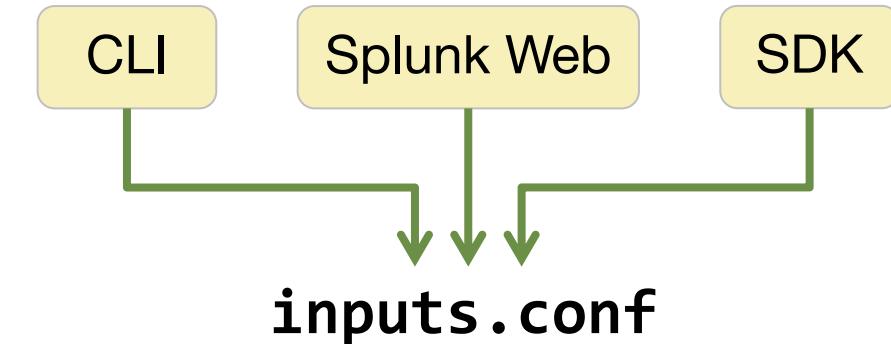
The System Administrator is responsible for installing and configuring Splunk components.

Splunk Configuration Files



Configuration Files (.conf)

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



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

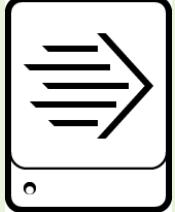
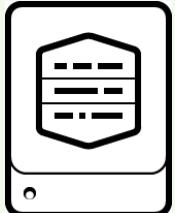
```
[monitor:///var/log/httpd]  
sourcetype = access_common  
ignoreOlderThan = 7d  
index = web
```

Note

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

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

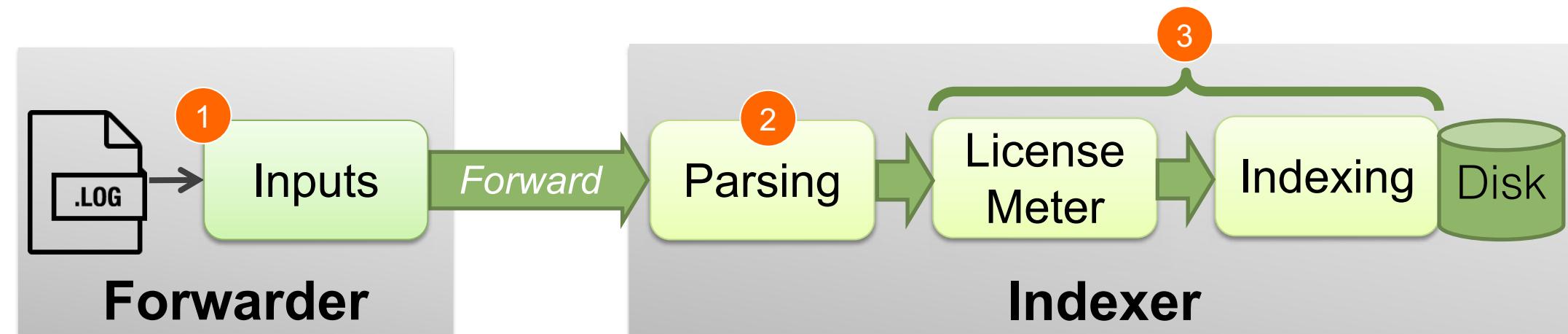
Commonly Used Splunk Configuration Files

Component	inputs.conf	props.conf	outputs.conf
Universal Forwarder 	What data to collect (production logs)	Limited parsing (such as character encoding, refine MetaData, event breaks*)	Where to forward the data (generally to Indexers)
Indexer 	What data to collect; which ports to listen to	Refine MetaData at event level, event breaks, Time Extraction, TZ, data transformation	Not generally needed (generally Indexer does not forward data)
Search Head 	What data to collect (internal Splunk logs)	Field Extractions (search- time), lookups, and so on	Where to forward the data (such as the internal Splunk logs to an indexer)

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

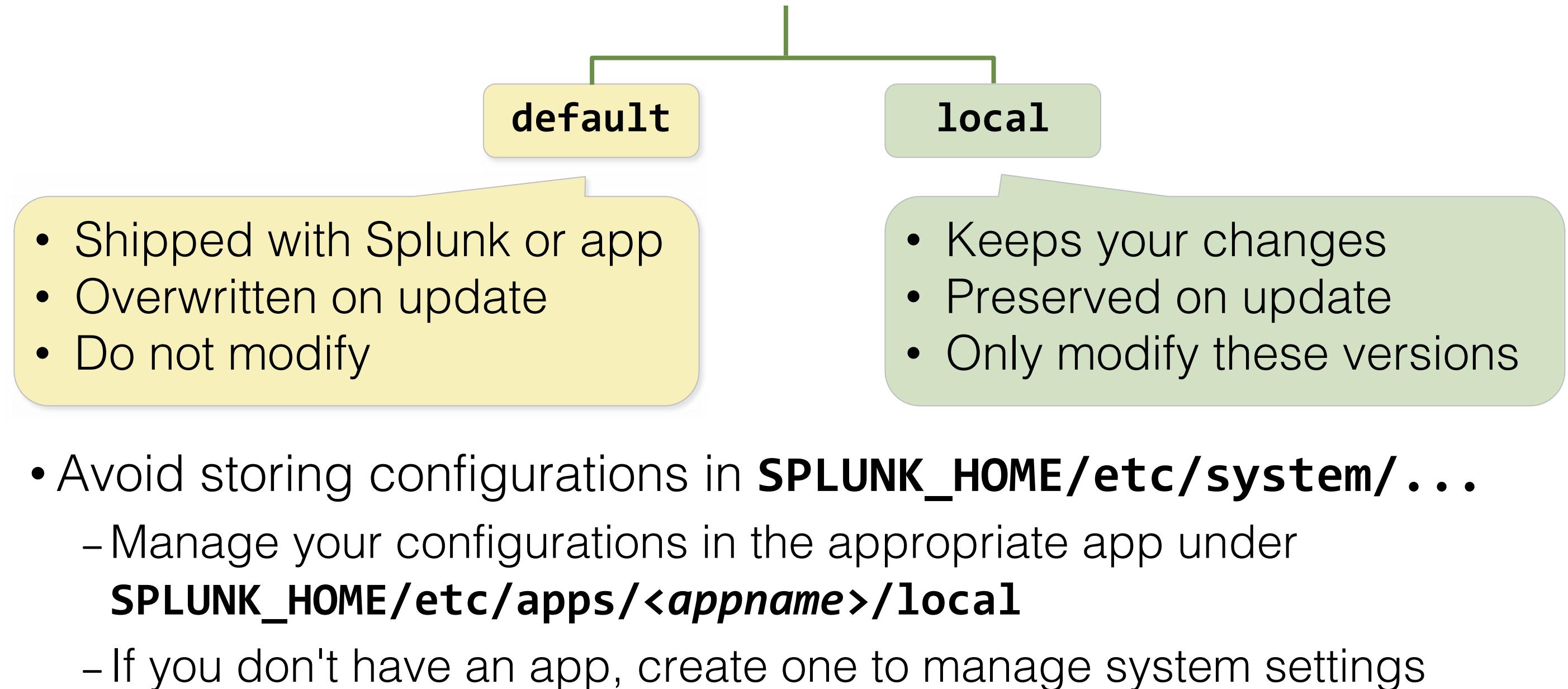
Index-Time Process

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



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

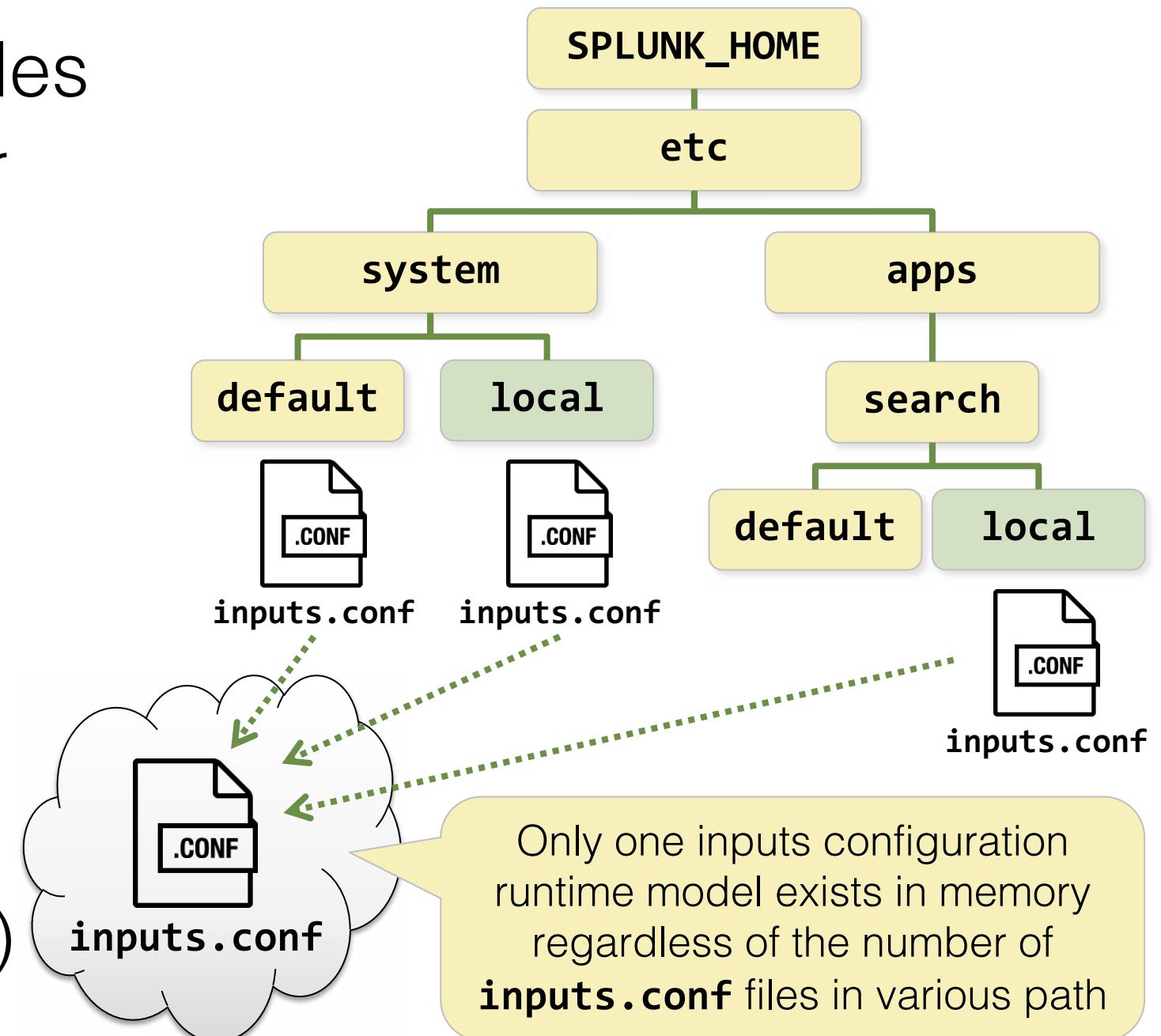
Default vs. Local Configuration



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

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)



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

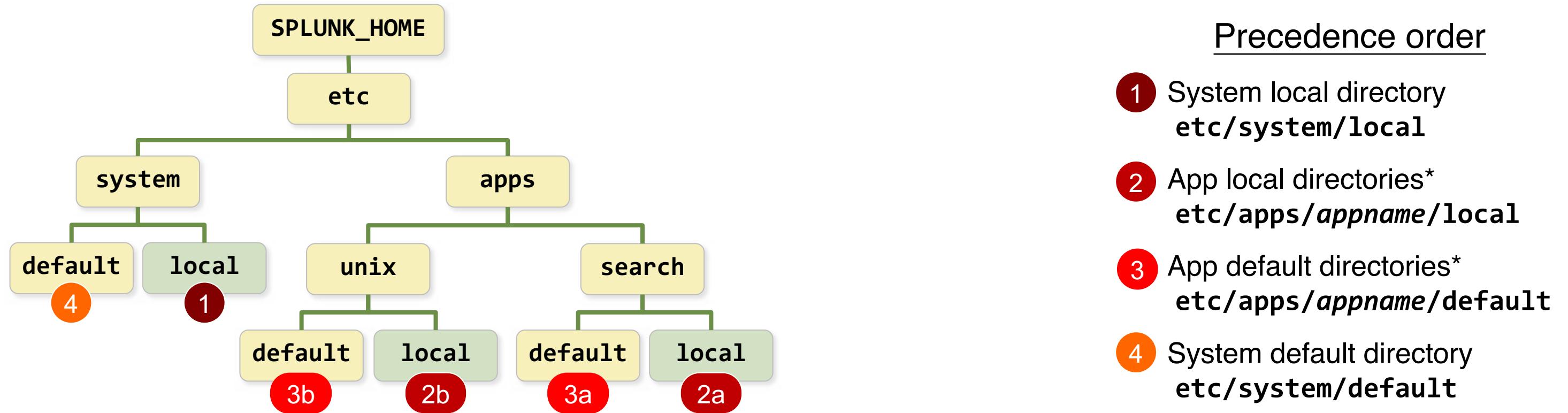
File Context and Index-time vs Search-time

	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

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

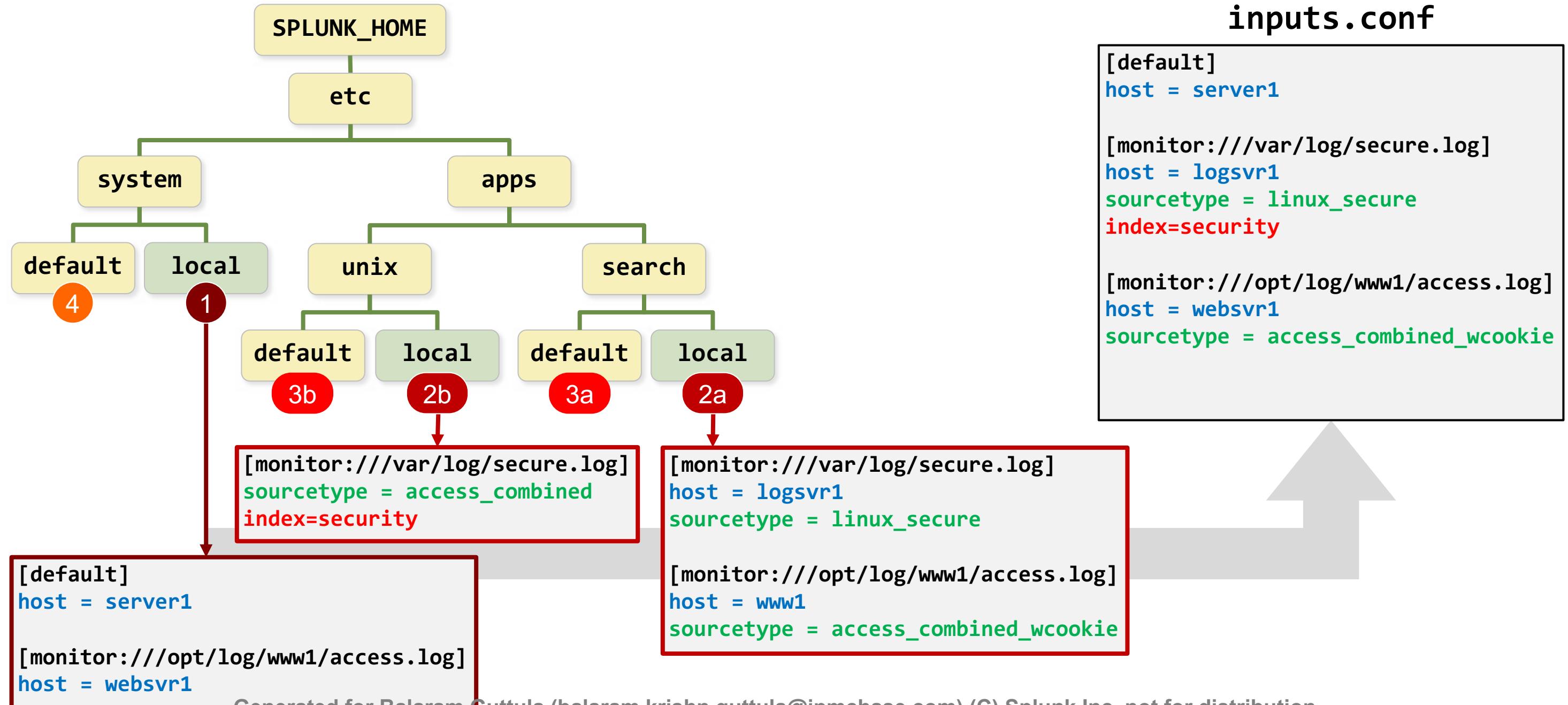
Index-Time Precedence (Global Context)



Note

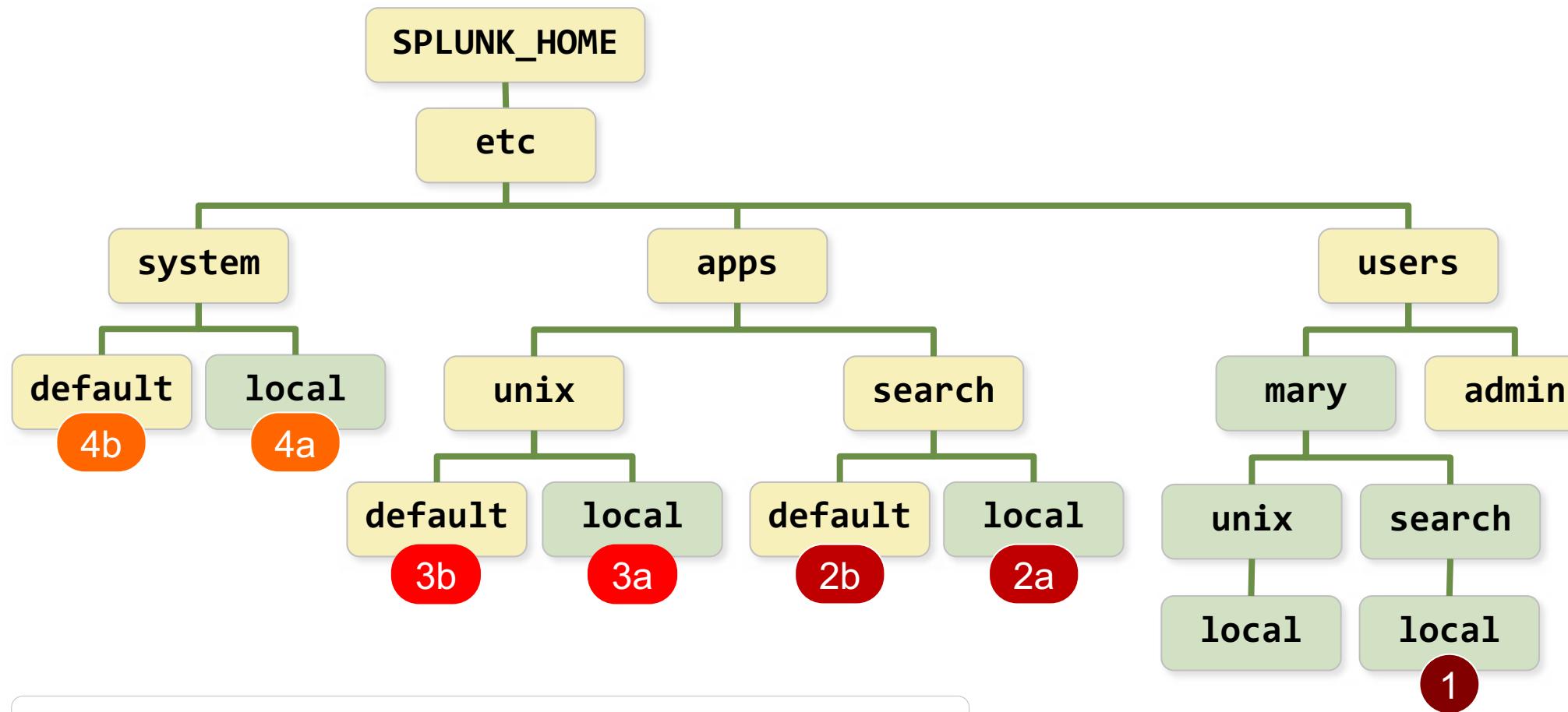
* When determining priority of app directories in **global** context (for steps 2 and 3), Splunk uses *lexicographical*/order. (Files in apps directory "A" have higher priority than files in apps directory "B".)

Example of Index-Time Precedence with `inputs.conf`



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Search-Time Precedence (App/User Context)



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`

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".)

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Validating the Splunk configuration

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**

Validating the on-disk configuration

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

Configuration Validation with **btool**

- **splunk btool <conf-name> list [options]**
 - Shows on-disk configuration for requested file
 - Run **splunk btool check** each time Splunk starts
 - Useful for checking the configuration scope and permission rules
 - Use **--debug** to display the exact .conf file location
 - Add **--user= <user> --app=<app>** to see the user/app context layering

- Examples:

```
splunk help btool
```

```
splunk btool check
```

```
splunk btool inputs list
```

```
splunk btool inputs list monitor:///var/log
```

```
splunk btool inputs list monitor:///var/log --debug
```

docs.splunk.com/Documentation/Splunk/latest/Troubleshooting/Usebtooltotroubleshootconfigurations

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Example using btool

Scenario: What are the **/var/log/secure.log** input configurations and where are they specified?

```
> splunk btool inputs list monitor:///var/log/secure.log --debug
```

etc/apps/search/local/inputs.conf	[monitor:///var/log/secure.log]
etc/apps/search/local/inputs.conf	host = logsvr1
etc/apps/unix/local/inputs.conf	index = security
etc/apps/search/local/inputs.conf	sourcetype = linux_secure

etc/apps/unix/local/inputs.conf

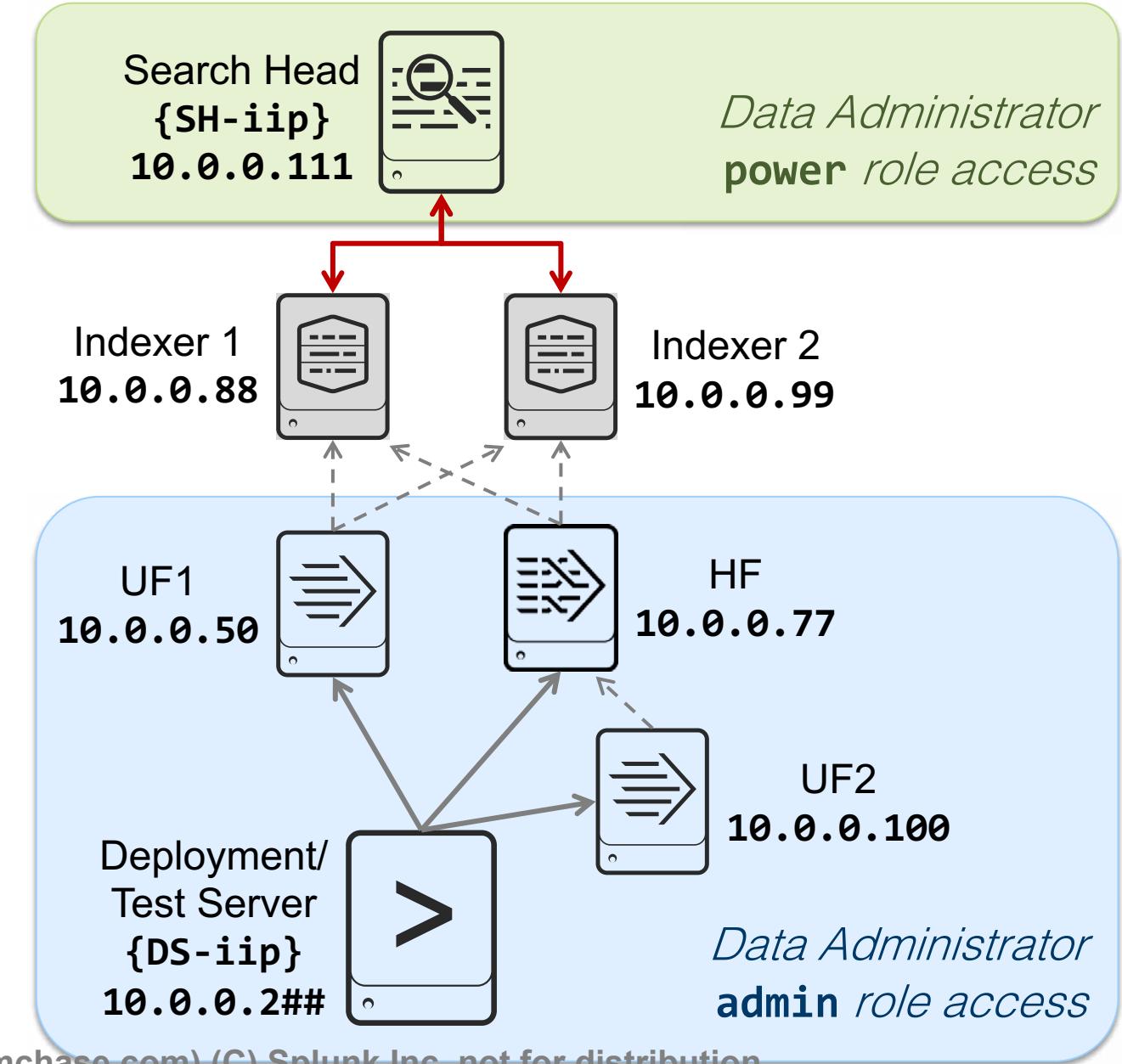
```
[monitor:///var/log/secure.log]
sourcetype = access_combined
index = security
```

etc/apps/search/local/inputs.conf

```
[monitor:///var/log/secure.log]
host = logsvr1
sourcetype = linux_secure
```

Data Administrator Access Scenario

Splunk instance	Access
Search Head (search / verify data configs)	power role
Indexers	No access
Deployment/Test Server	
Forwarders (data sources and inputs)	admin role



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Module 1 Knowledge Check

- Which installer will the System Admin use to install the heavy forwarder?
- Which configuration file tells a Splunk instance to ingest data?
- True or False. The best place to add a parsing configuration on an indexer would be **SPLUNK_HOME/etc/system/local directory** as it has the highest precedence.

Module 1 Knowledge Check – Answers

- Which installer will the System Admin use to install the heavy forwarder?

Splunk Enterprise

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

inputs.conf

- True or False. The best place to add a parsing configuration on an indexer would be the **SPLUNK_HOME/etc/system/local** directory, as it has the highest precedence.

False. Best practice is to put the configuration in an *app's local directory (**SPLUNK_HOME/etc/apps/<appname>/local**).*

Module 1 Lab Exercise – Discover Lab Environment

Time: 15 minutes

Tasks:

- Log into search head and test/deployment server
- Discover Splunk Enterprise lab environment
- Use CLI to connect to Splunk components

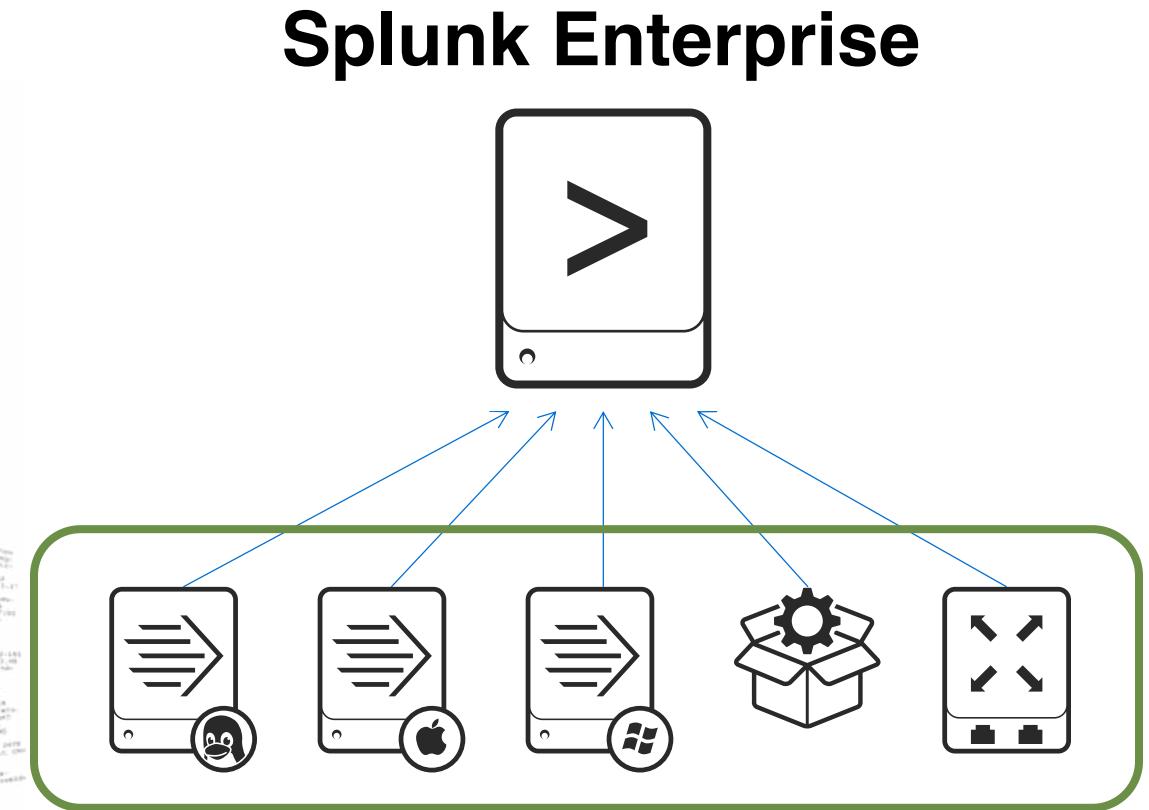
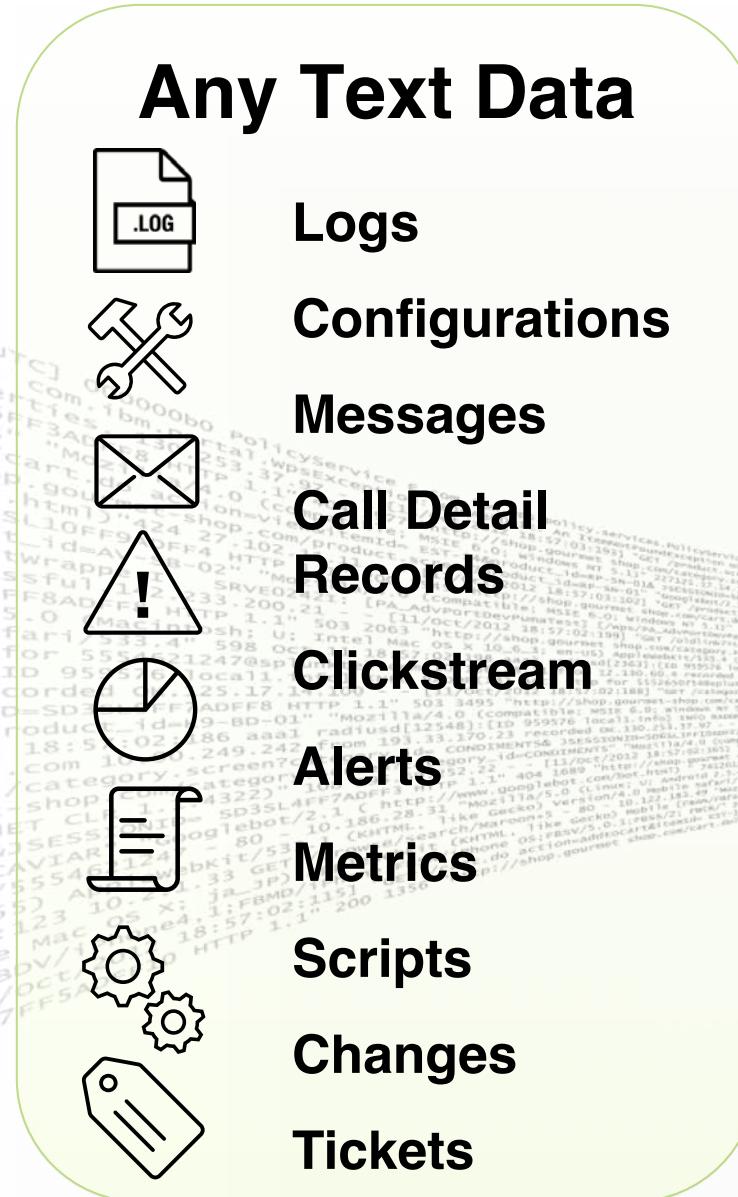
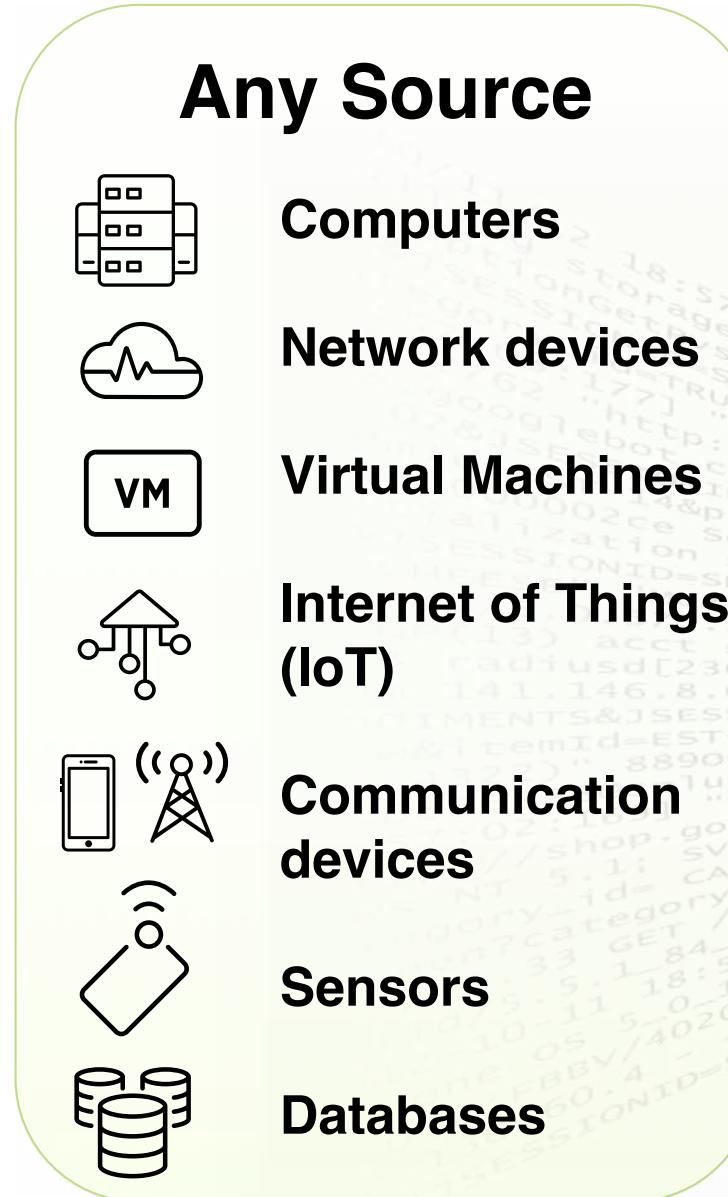
Module 2: Getting Data In – Staging

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Module Objectives

- Describe data input types and default metadata settings
- Describe differences between the input and parsing phase
- Configure initial input testing with Splunk Web

Forwarding Data to Splunk



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Data Input Types

- Splunk supports many types of data input
 - **Files and directories:** monitoring text files and/or directory structures containing text files
 - **Network data:** listening on a port for network data
 - **Script output:** executing a script and using the output from the script as the input
 - **Windows logs:** monitoring Windows event logs, Active Directory, etc.
 - **HTTP:** using the HTTP Event Collector
 - And more...
- You can add data inputs with:
 - Apps and add-ons from Splunkbase
 - Splunk Web
 - CLI
 - Directly editing **inputs.conf**

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Default Metadata Settings

- When Splunk indexes a data source, metadata values are assigned
 - Metadata is applied to the entire source
 - Defaults are used if alternates are not specified
 - Overriding values can be performed at input time or later

Metadata	Default
source	Path of input file, network hostname:port, or script name
host	Splunk hostname of the inputting instance (usually a forwarder)
sourcetype	Uses the source filename if Splunk cannot automatically determine
index	Defaults to the main index

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

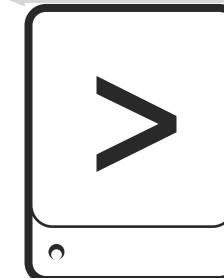
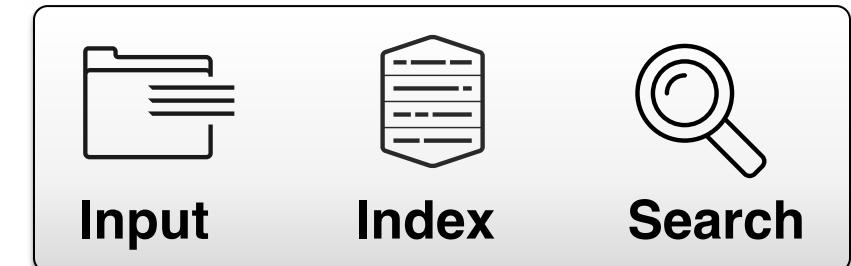
Input Phase vs. Parsing Phase

Input phase	Parsing phase
<ul style="list-style-type: none">• Most efficient, but less flexible• Acquires data from source• Sets initial metadata fields: source, sourcetype, host, index, etc.• Converts character encoding• Operates on the entire data stream• Most configuration done in inputs.conf on forwarder<ul style="list-style-type: none">- Some configuration is in props.conf	<ul style="list-style-type: none">• Less efficient, but finer control• Breaks data into events with timestamps• Applies event-level transformations• Fine-tunes metadata settings from inputs phase• Operates on individual events• Most configuration done in props.conf on indexer<ul style="list-style-type: none">- Also: transforms.conf

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Testing Indexes with Input Staging

- Production data typically resides
 - On a forwarder (remote system)
 - Not on the indexer
- Test input data
 - Use Splunk Web > Add Data
 - Sample a log file on a test server
 - Check to see if **sourcetype** and other settings are applied correctly
 - Delete the test data, change your test configuration, and repeat as necessary



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. At the top, there is a navigation bar with the following items: Administrator, Messages, Settings (with a red circle containing the number 1), Activity, Help, and Find. Below the navigation bar is a main menu with several categories: KNOWLEDGE (Searches, reports, and alerts; Data models; Event types; Tags; Fields; Lookups; User interface), DATA (Forwarding and receiving; Indexes; Report acceleration summaries; Virtual indexes; Source types), and a section for LOCAL inputs (Type: Files & Directories, Index a local file or monitor an entire directory; Type: HTTP Event Collector, 0 inputs). A green arrow points from the 'Add Data' icon on the Home page to the 'Data inputs' link in the main menu. Another green arrow points from the 'Data inputs' link to the '+ Add new' button in the LOCAL inputs section. A third green arrow points from the '+ Add new' button to the 'Add new' link in the LOCAL inputs table.

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

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

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

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

4 data sources in total

Get data into Splunk

Or get data in with the following methods

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

Splunk Enterprise universal forwarders

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](#)

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Add Data Menu (cont.)



Upload

files from my computer

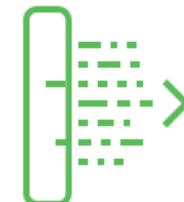
Local log files
Local structured files (e.g. CSV)
[Tutorial for adding data](#) ↗



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

- Allows uploading local files that only get indexed once
- Useful for testing or data that is created once and never gets updated
- Does not update **inputs.conf**

Monitor

- Provides one-time or continuous monitoring of files, directories, http events, network ports, or data gathering scripts located on Splunk Enterprise instances
- Useful for testing inputs

Forward

- Remote machines gather and forward data to indexers over a receiving port
- Main source of input in production environments
- Requires installation of forwarders

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Select Source

Add Data

Select Source Set Source Type Input Settings Review Done

Next >

Select the **Files & Directories** option to configure a monitor input

1

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.

Configure this instance to monitor files and directories for data. To monitor all objects in a directory, select the directory. Then, Splunk monitors all objects within the directory. Then, it indexes data sources in the directory. To a...
Specify the source with absolute path to a file or directory, or use the **Browse** button
2

File or Directory ? /opt/log/www1/access.log
On Windows: c:\apache\apache.error.log or \\hostname\apache\apache.error.log. On Unix: /var/log or /mnt/www01/var/log.

Continuously Monitor Index Once
3

Whitelist ?
Blacklist ?
For ongoing monitoring
For one-time indexing, or testing; Does not create a stanza in **inputs.conf**

Generated for Balaram Guttula (balaram.krishn.guttula@pmchase.com) (C) Splunk Inc, not for distribution

Select Source From Windows Instance

- On Windows Splunk instances, there are additional Windows-specific source options
- To monitor a shared network drive, enter the path manually:

*nix:	<host>/<path>
Windows:	\\<host>\<path>< b=""></host>\<path><>

- Make sure Splunk has read access to the mounted drive

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.

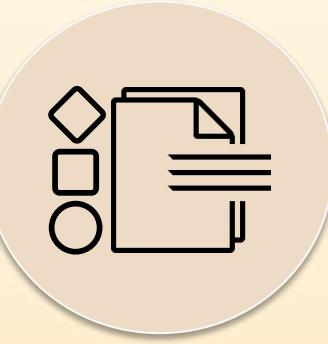
Registry monitoring

Have the Splunk platform index the local Windows Registry, and monitor it for changes.

Active Directory monitoring

Monitor Active Directory objects and their relationships.

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

Note 

Sourcetypes can be created or defined by the administrator, or by installing apps, which often define custom source types for their inputs.

Source Type	Description
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

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Set Source Type (Data Preview Interface)

Set Source Type

This page lets you see how the Splunk platform sees your data before indexing it. You can use this interface to define proper event breaks and source types.

Automatically determined for major data types

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

View Event Summary

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

Event

Time

1 6/20/19 10:38:36.000 PM

2 6/20/19 10:38:41.000 PM

3 6/20/19 10:38:45.000 PM

4 6/20/19 10:38:48.000 PM

5 6/20/19 10:38:50.000 PM

6 6/20/19 10:38:52.000 PM

7 6/20/19 10:38:54.000 PM

8 6/20/19 10:38:56.000 PM

9 6/20/19 10:38:58.000 PM

10 6/20/19 10:39:00.000 PM

11 6/20/19 10:39:02.000 PM

12 6/20/19 10:39:04.000 PM

13 6/20/19 10:39:06.000 PM

14 6/20/19 10:39:08.000 PM

15 6/20/19 10:39:10.000 PM

16 6/20/19 10:39:12.000 PM

17 6/20/19 10:39:14.000 PM

18 6/20/19 10:39:16.000 PM

19 6/20/19 10:39:18.000 PM

20 6/20/19 10:39:20.000 PM

21 6/20/19 10:39:22.000 PM

22 6/20/19 10:39:24.000 PM

23 6/20/19 10:39:26.000 PM

24 6/20/19 10:39:28.000 PM

25 6/20/19 10:39:30.000 PM

26 6/20/19 10:39:32.000 PM

27 6/20/19 10:39:34.000 PM

28 6/20/19 10:39:36.000 PM

29 6/20/19 10:39:38.000 PM

30 6/20/19 10:39:40.000 PM

31 6/20/19 10:39:42.000 PM

32 6/20/19 10:39:44.000 PM

33 6/20/19 10:39:46.000 PM

34 6/20/19 10:39:48.000 PM

35 6/20/19 10:39:50.000 PM

36 6/20/19 10:39:52.000 PM

37 6/20/19 10:39:54.000 PM

38 6/20/19 10:39:56.000 PM

39 6/20/19 10:39:58.000 PM

40 6/20/19 10:40:00.000 PM

41 6/20/19 10:40:02.000 PM

42 6/20/19 10:40:04.000 PM

43 6/20/19 10:40:06.000 PM

44 6/20/19 10:40:08.000 PM

45 6/20/19 10:40:10.000 PM

46 6/20/19 10:40:12.000 PM

47 6/20/19 10:40:14.000 PM

48 6/20/19 10:40:16.000 PM

49 6/20/19 10:40:18.000 PM

50 6/20/19 10:40:20.000 PM

51 6/20/19 10:40:22.000 PM

52 6/20/19 10:40:24.000 PM

53 6/20/19 10:40:26.000 PM

54 6/20/19 10:40:28.000 PM

55 6/20/19 10:40:30.000 PM

56 6/20/19 10:40:32.000 PM

57 6/20/19 10:40:34.000 PM

58 6/20/19 10:40:36.000 PM

59 6/20/19 10:40:38.000 PM

60 6/20/19 10:40:40.000 PM

61 6/20/19 10:40:42.000 PM

62 6/20/19 10:40:44.000 PM

63 6/20/19 10:40:46.000 PM

64 6/20/19 10:40:48.000 PM

65 6/20/19 10:40:50.000 PM

66 6/20/19 10:40:52.000 PM

67 6/20/19 10:40:54.000 PM

68 6/20/19 10:40:56.000 PM

69 6/20/19 10:40:58.000 PM

70 6/20/19 10:41:00.000 PM

71 6/20/19 10:41:02.000 PM

72 6/20/19 10:41:04.000 PM

73 6/20/19 10:41:06.000 PM

74 6/20/19 10:41:08.000 PM

75 6/20/19 10:41:10.000 PM

76 6/20/19 10:41:12.000 PM

77 6/20/19 10:41:14.000 PM

78 6/20/19 10:41:16.000 PM

79 6/20/19 10:41:18.000 PM

80 6/20/19 10:41:20.000 PM

81 6/20/19 10:41:22.000 PM

82 6/20/19 10:41:24.000 PM

83 6/20/19 10:41:26.000 PM

84 6/20/19 10:41:28.000 PM

85 6/20/19 10:41:30.000 PM

86 6/20/19 10:41:32.000 PM

87 6/20/19 10:41:34.000 PM

88 6/20/19 10:41:36.000 PM

89 6/20/19 10:41:38.000 PM

90 6/20/19 10:41:40.000 PM

91 6/20/19 10:41:42.000 PM

92 6/20/19 10:41:44.000 PM

93 6/20/19 10:41:46.000 PM

94 6/20/19 10:41:48.000 PM

95 6/20/19 10:41:50.000 PM

96 6/20/19 10:41:52.000 PM

97 6/20/19 10:41:54.000 PM

98 6/20/19 10:41:56.000 PM

99 6/20/19 10:41:58.000 PM

100 6/20/19 10:42:00.000 PM

101 6/20/19 10:42:02.000 PM

102 6/20/19 10:42:04.000 PM

103 6/20/19 10:42:06.000 PM

104 6/20/19 10:42:08.000 PM

105 6/20/19 10:42:10.000 PM

106 6/20/19 10:42:12.000 PM

107 6/20/19 10:42:14.000 PM

108 6/20/19 10:42:16.000 PM

109 6/20/19 10:42:18.000 PM

110 6/20/19 10:42:20.000 PM

111 6/20/19 10:42:22.000 PM

112 6/20/19 10:42:24.000 PM

113 6/20/19 10:42:26.000 PM

114 6/20/19 10:42:28.000 PM

115 6/20/19 10:42:30.000 PM

116 6/20/19 10:42:32.000 PM

117 6/20/19 10:42:34.000 PM

118 6/20/19 10:42:36.000 PM

119 6/20/19 10:42:38.000 PM

120 6/20/19 10:42:40.000 PM

121 6/20/19 10:42:42.000 PM

122 6/20/19 10:42:44.000 PM

123 6/20/19 10:42:46.000 PM

124 6/20/19 10:42:48.000 PM

125 6/20/19 10:42:50.000 PM

126 6/20/19 10:42:52.000 PM

127 6/20/19 10:42:54.000 PM

128 6/20/19 10:42:56.000 PM

129 6/20/19 10:42:58.000 PM

130 6/20/19 10:43:00.000 PM

131 6/20/19 10:43:02.000 PM

132 6/20/19 10:43:04.000 PM

133 6/20/19 10:43:06.000 PM

134 6/20/19 10:43:08.000 PM

135 6/20/19 10:43:10.000 PM

136 6/20/19 10:43:12.000 PM

137 6/20/19 10:43:14.000 PM

138 6/20/19 10:43:16.000 PM

139 6/20/19 10:43:18.000 PM

140 6/20/19 10:43:20.000 PM

141 6/20/19 10:43:22.000 PM

142 6/20/19 10:43:24.000 PM

143 6/20/19 10:43:26.000 PM

144 6/20/19 10:43:28.000 PM

145 6/20/19 10:43:30.000 PM

146 6/20/19 10:43:32.000 PM

147 6/20/19 10:43:34.000 PM

148 6/20/19 10:43:36.000 PM

149 6/20/19 10:43:38.000 PM

150 6/20/19 10:43:40.000 PM

151 6/20/19 10:43:42.000 PM

152 6/20/19 10:43:44.000 PM

153 6/20/19 10:43:46.000 PM

154 6/20/19 10:43:48.000 PM

155 6/20/19 10:43:50.000 PM

156 6/20/19 10:43:52.000 PM

157 6/20/19 10:43:54.000 PM

158 6/20/19 10:43:56.000 PM

159 6/20/19 10:43:58.000 PM

160 6/20/19 10:44:00.000 PM

161 6/20/19 10:44:02.000 PM

162 6/20/19 10:44:04.000 PM

163 6/20/19 10:44:06.000 PM

164 6/20/19 10:44:08.000 PM

165 6/20/19 10:44:10.000 PM

166 6/20/19 10:44:12.000 PM

167 6/20/19 10:44:14.000 PM

168 6/20/19 10:44:16.000 PM

169 6/20/19 10:44:18.000 PM

170 6/20/19 10:44:20.000 PM

171 6/20/19 10:44:22.000 PM

172 6/20/19 10:44:24.000 PM

173 6/20/19 10:44:26.000 PM

174 6/20/19 10:44:28.000 PM

175 6/20/19 10:44:30.000 PM

176 6/20/19 10:44:32.000 PM

177 6/20/19 10:44:34.000 PM

178 6/20/19 10:44:36.000 PM

179 6/20/19 10:44:38.000 PM

180 6/20/19 10:44:40.000 PM

181 6/20/19 10:44:42.000 PM

182 6/20/19 10:44:44.000 PM

183 6/20/19 10:44:46.000 PM

184 6/20/19 10:44:48.000 PM

185 6/20/19 10:44:50.000 PM

186 6/20/19 10:44:52.000 PM

187 6/20/19 10:44:54.000 PM

188 6/20/19 10:44:56.000 PM

189 6/20/19 10:44:58.000 PM

190 6/20/19 10:45:00.000 PM

191 6/20/19 10:45:02.000 PM

192 6/20/19 10:45:04.000 PM

193 6/20/19 10:45:06.000 PM

194 6/20/19 10:45:08.000 PM

195 6/20/19 10:45:10.000 PM

196 6/20/19 10:45:12.000 PM

197 6/20/19 10:45:14.000 PM

198 6/20/19 10:45:16.000 PM

199 6/20/19 10:45:18.000 PM

200 6/20/19 10:45:20.000 PM

201 6/20/19 10:45:22.000 PM

202 6/20/19 10:45:24.000 PM

203 6/20/19 10:45:26.000 PM

204 6/20/19 10:45:28.000 PM

205 6/20/19 10:45:30.000 PM

206 6/20/19 10:45:32.000 PM

207 6/20/19 10:45:34.000 PM

208 6/20/19 10:45:36.000 PM

209 6/20/19 10:45:38.000 PM

210 6/20/19 10:45:40.000 PM

211 6/20/19 10:45:42.000 PM

212 6/20/19 10:45:44.000 PM

213 6/20/19 10:45:46.000 PM

214 6/20/19 10:45:48.000 PM

215 6/20/19 10:45:50.000 PM

216 6/20/19 10:45:52.000 PM

217 6/20/19 10:45:54.000 PM

218 6/20/19 10:45:56.000 PM

219 6/20/19 10:45:58.000 PM

220 6/20/19 10:46:00.000 PM

221 6/20/19 10:46:02.000 PM

222 6/20/19 10:46:04.000 PM

223 6/20/19 10:46:06.000 PM

224 6/20/19 10:46:08.000 PM

225 6/20/19 10:46:10.000 PM

226 6/20/19 10:46:12.000 PM

227 6/20/19 10:46:14.000 PM

228 6/20/19 10:46:16.000 PM

229 6/20/19 10:46:18.000 PM

230 6/20/19 10:46:20.000 PM

231 6/20/19 10:46:22.000 PM

232 6/20/19 10:46:24.000 PM

233 6/20/19 10:46:26.000 PM

234 6/20/19 10:46:28.000 PM

235 6/20/19 10:46:30.000 PM

236 6/20/19 10:46:32.000 PM

237 6/20/19 10:46:34.000 PM

238 6/20/19 10:46:36.000 PM

239 6/20/19 10:46:38.000 PM

240 6/20/19 10:46:40.000 PM

241 6/20/19 10:46:42.000 PM

242 6/20/19 10:46:44.000 PM

243 6/20/19 10:46:46.000 PM

244 6/20/19 10:46:48.000 PM

245 6/20/19 10:46:50.000 PM

246 6/20/19 10:46:52.000 PM

247 6/20/19 10:46:54.000 PM

248 6/20/19 10:46:56.000 PM

249 6/20/19 10:46:58.000 PM

250 6/20/19 10:47:00.000 PM

251 6/20/19 10:47:02.000 PM

252 6/20/19 10:47:04.000 PM

253 6/20/19 10:47:06.000 PM

254 6/20/19 10:47:08.000 PM

255 6/20/19 10:47:10.000 PM

256 6/20/19 10:47:12.000 PM

257 6/20/19 10:47:14.000 PM

258 6/20/19 10:47:16.000 PM

259 6/20/19 10:47:18.000 PM

260 6/20/19 10:47:20.000 PM

261 6/20/19 10:47:22.000 PM

262 6/20/19 10:47:24.000 PM

263 6/20/19 10:47:26.000 PM

264 6/20/19 10:47:28.000 PM

265 6/20/19 10:47:30.000 PM

266 6/20/19 10:47:32.000 PM

267 6/20/19 10:47:34.000 PM

268 6/20/19 10:47:36.000 PM

269 6/20/19 10:47:38.000 PM

270 6/20/19 10:47:40.000 PM

271 6/20/19 10:47:42.000 PM

272 6/20/19 10:47:44.000 PM

273 6/20/19 10:47:46.000 PM

274 6/20/19 10:47:48.000 PM

275 6/20/19 10:47:50.000 PM

276 6/20/19 10:47:52.000 PM

277 6/20/19 10:47:54.000 PM

278 6/20/19 10:47:56.000 PM

279 6/20/19 10:47:58.000 PM

280 6/20/19 10:48:00.000 PM

281 6/20/19 10:48:02.000 PM

282 6/20/19 10:48:04.000 PM

283 6/20/19 10:48:06.000 PM

284 6/20/19 10:48:08.000 PM

285 6/20/19 10:48:10.000 PM

286 6/20/19 10:48:12.000 PM

287 6/20/19 10:48:14.000 PM

288 6/20/19 10:48:16.000 PM

289 6/20/19 10:48:18.000 PM

290 6/20/19 10:48:20.000 PM

291 6/20/19 10:48:22.000 PM

292 6/20/19 10

Pretrained Source Types

- Built-in source types shipped with Splunk
- Can be added to 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>

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Input Settings

The screenshot shows the 'Add Data' wizard at the 'Input Settings' step. The progress bar indicates the following steps: 'Select Source' (green), 'Set Source Type' (green), 'Input Settings' (gray), 'Review' (white), and 'Done' (white). The 'Input Settings' section contains three main configuration sections: 'App context', 'Host', and 'Index'.

- App context:** A dropdown menu is set to 'Search & Reporting (search)'. A callout bubble points to this field with the text: '• App Context determines where input configuration is saved
• For Search & Reporting (search): **SPLUNK_HOME/etc/apps/search/local**'.
- Host:** The 'Host field value' is set to 'splunk01'. A callout bubble points to this field with the text: 'By default, the default host name in General settings is used'. Another callout bubble points to the 'Constant value' radio button with the text: 'Select index where input will be stored'.
- Index:** The 'Index' dropdown is set to 'itops'. A callout bubble points to this field with the text: 'Select index where input will be stored'.

Generated for Balaram Guttula (balaram.krishn.guttula@pmchase.com) (C) Splunk Inc, not for distribution

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' on the left, a progress bar with five steps ('Select Source', 'Set Source Type', 'Input Settings', 'Review', 'Done') where the first four are green and the last one is grey, and buttons for '< Back' and 'Submit >' on the right. The main content area is titled 'Review' and displays the following configuration details:

Input Type	File Monitor
Source Path	/opt/log/www1/access.log
Continuously Monitor	Yes
Source Type	access_combined
App Context	search
Host	splunk01
Index	itops

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

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:

The screenshot shows the 'Add Data' wizard in Splunk. The progress bar at the top indicates the following steps: 'Select Source' (green dot), 'Set Source Type' (green dot), 'Input Settings' (green dot), 'Review' (green dot with a checkmark), and 'Done' (gray dot). Below the progress bar, a success message says 'File input has been created successfully.' It also provides 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.'

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

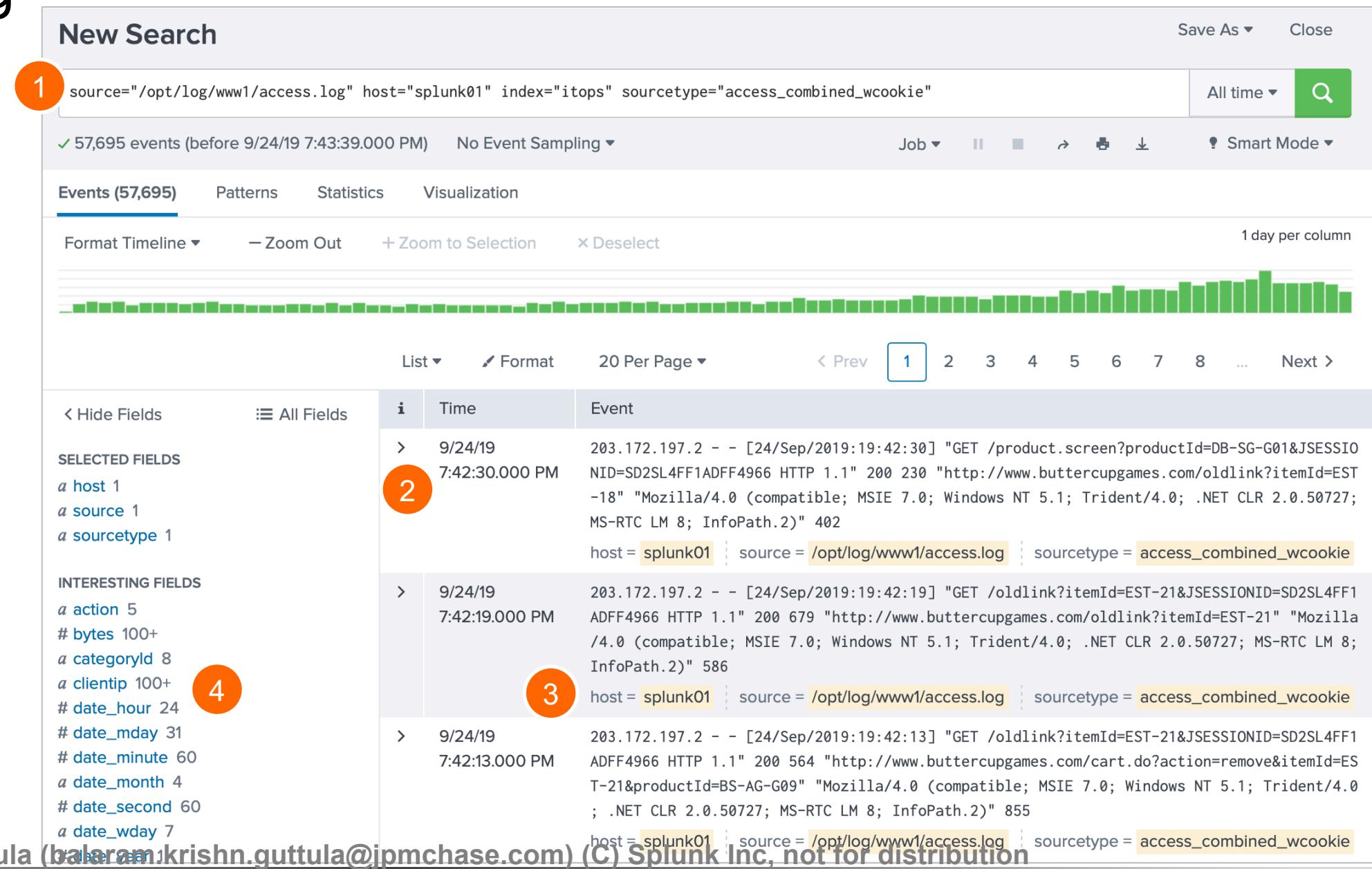
Note



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

Verify your Input

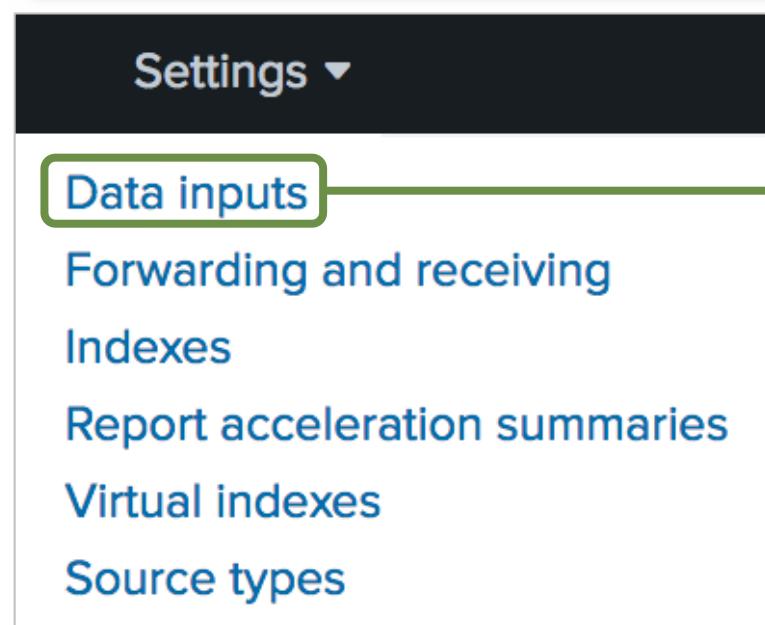
1. Click Start Searching or search for **index=<test_idx>**
2. Verify the event timestamps
3. Confirm the host, source, and sourcetype field values
4. Check the auto-extracted field names



Generated for Balaram Guttula (balaram.krishn.guttula@jmpchase.com) (C) Splunk Inc, not for distribution

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
Scripts Run custom scripts to collect or generate more data.	5	+ Add new

Forwarded inputs Inputs handled by remote instances but configured from this deployment server

Type	Inputs	Actions
Windows Event Logs Collect event logs from forwarders.	0	+ Add new

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Viewing Configured Inputs: Files & Directories

Files & directories

Data inputs » Files & directories

Showing 1-10 of 10 items

filter 

25 per page ▾

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	14	introspection_generator_addon	Enabled Disable	
\$SPLUNK_HOME/var/log/splunk	Constant Value	Automatic	_internal	33	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		_telemetry	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	

Launched Add Data wizard

New Local File & Directory

Indexing destination

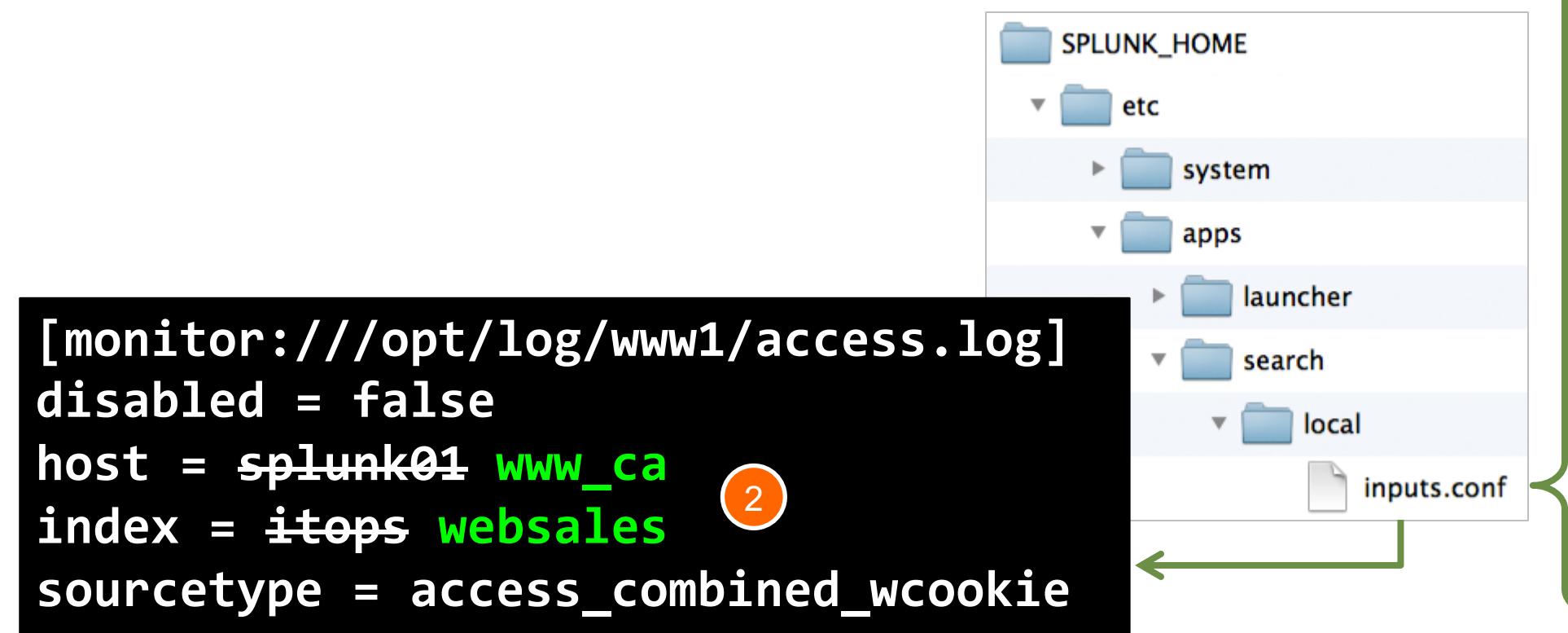
Location of configuration (app context)

Click to edit existing input settings

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Viewing and Updating `inputs.conf`

1. Adding monitored data into Splunk updates the target app's `local/inputs.conf` file
2. Optionally edit `inputs.conf` manually



You can tell Splunk to continuously collect data from a file or direct

More settings

Host

Tell Splunk how to set the value of the host field in your events fro

Set host constant value

Specify method for getting h

Host field value: splunk01

Source type

Tell Splunk what kind of data this is so you can group it with other

can specify what you want if Splunk gets it wrong.

Set the source type: Manual

When this is set to automatic sourcetypes placeholder nar

Source type *: access_combined_wcookie

Index

Set the destination index for this source.

Index: itops

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Module 2 Knowledge Check

- When you configure the inputs using **Settings > Add Data**, under what directory is the **inputs.conf** created?
- True or False. You cannot change the sourcetype when you go through the **Settings > Add Data** wizard.
- True or False. Splunk will not create an **inputs.conf** file when you use the Upload option in **Settings > Add Data**.

Module 2 Knowledge Check – Answers

- ❑ When you configure the inputs using **Settings > Add Data**, under what directory is the **inputs.conf** created?

It depends on the **App Context** setting on the **Input Setting** stage. Best practice is to put the configuration file in the local directory of your app. If you have clustering enabled, then the **SPLUNK_HOME/etc/system/local** may not be the highest in the precedence order. More details are available in the Cluster Administration course.

- ❑ True or False. You cannot change the sourcetype when you go through the **Settings > Add Data** wizard.

False. You can change the source type from the dropdown. In fact, you can even create a new source type. We will learn how to do this in Module 9.

- ❑ Splunk will not create an **inputs.conf** file when you use the **Upload** option in **Settings > Add Data**.

True. Upload is a one-time process, so Splunk does not create an **inputs.conf**.

Module 2 Lab Exercise – Add a Local Data Input

Time: 20 minutes

Tasks:

- Create all local indexes required on the deployment/test server
- Index a file on the deployment server
- Verify the indexed events with their metadata values
- View the stanza in the saved **inputs.conf** file

Module 3: Forwarder Configuration

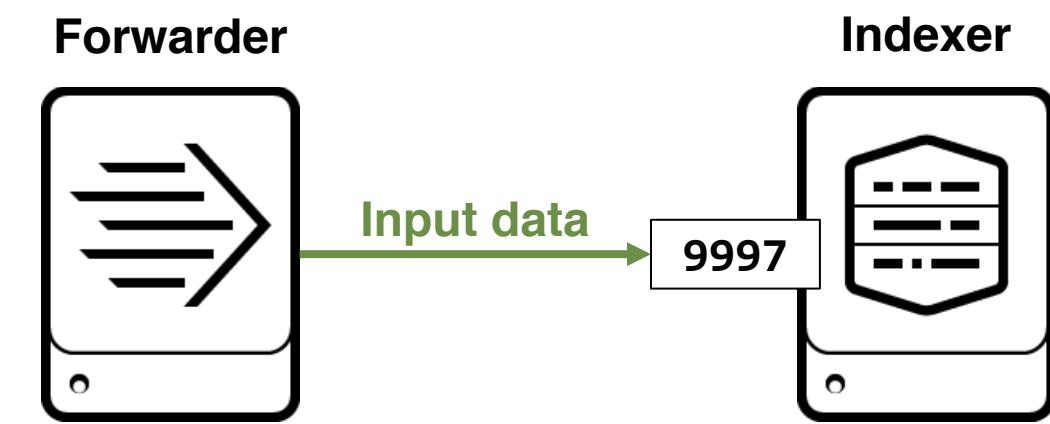
Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Module 3 Objectives

- Understand the role of production indexers and forwarders
- Understand the functionality of Universal Forwarders
- Configure forwarders
- Identify additional forwarder options

Forwarders and Indexers

- In production environments:
 - Indexers run on dedicated servers
 - Most input data is on remote servers
- Install **forwarders** on remote servers to
 - Gather the data
 - Send it over the network to indexers
- Configure indexers to listen on a receiving port for the forwarded data

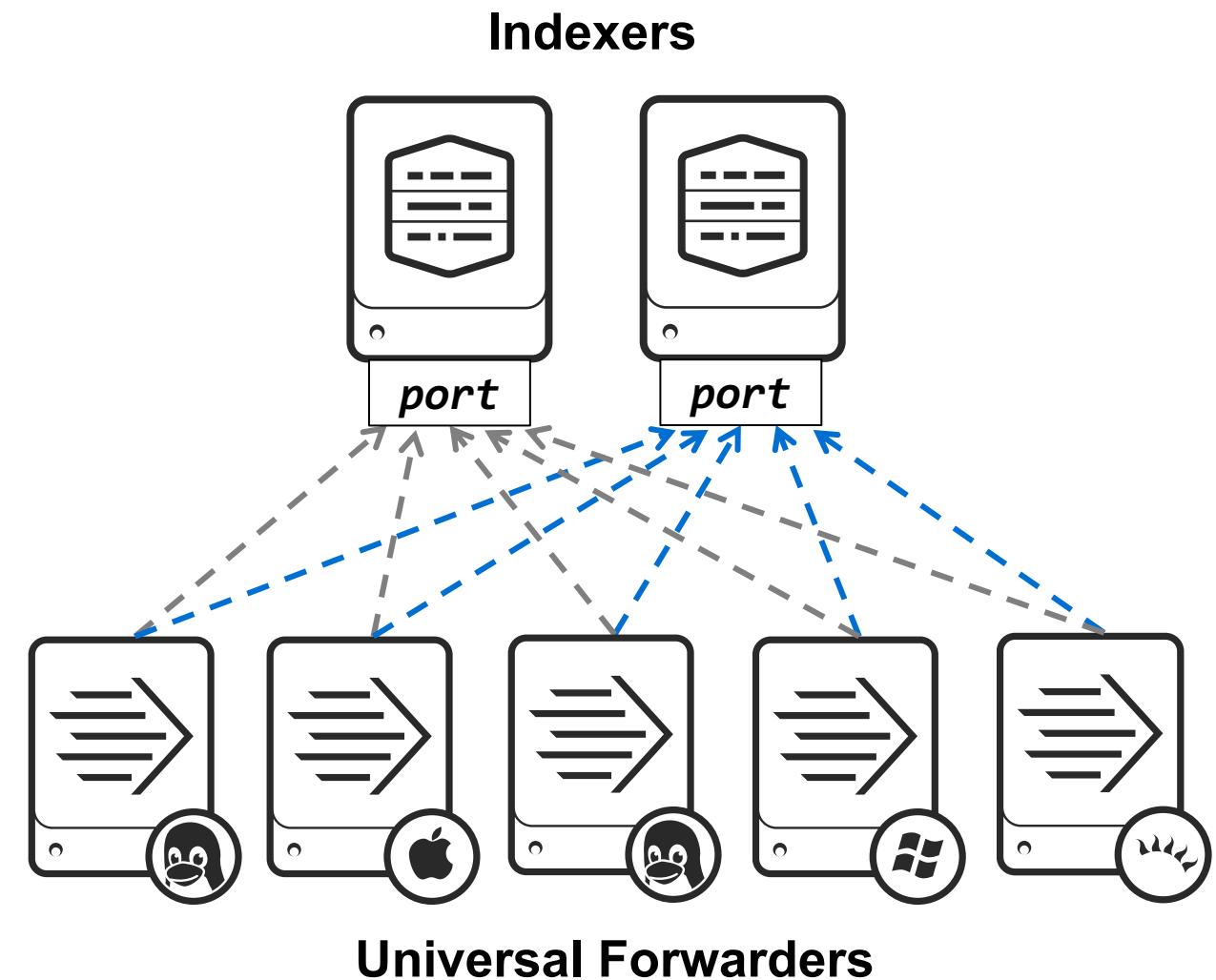


Understanding Universal Forwarders



Universal Forwarders (UF)

- Gathers data from a host, sending it over the network to receiving ports on indexers
- 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)



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Installing Splunk Universal Forwarder

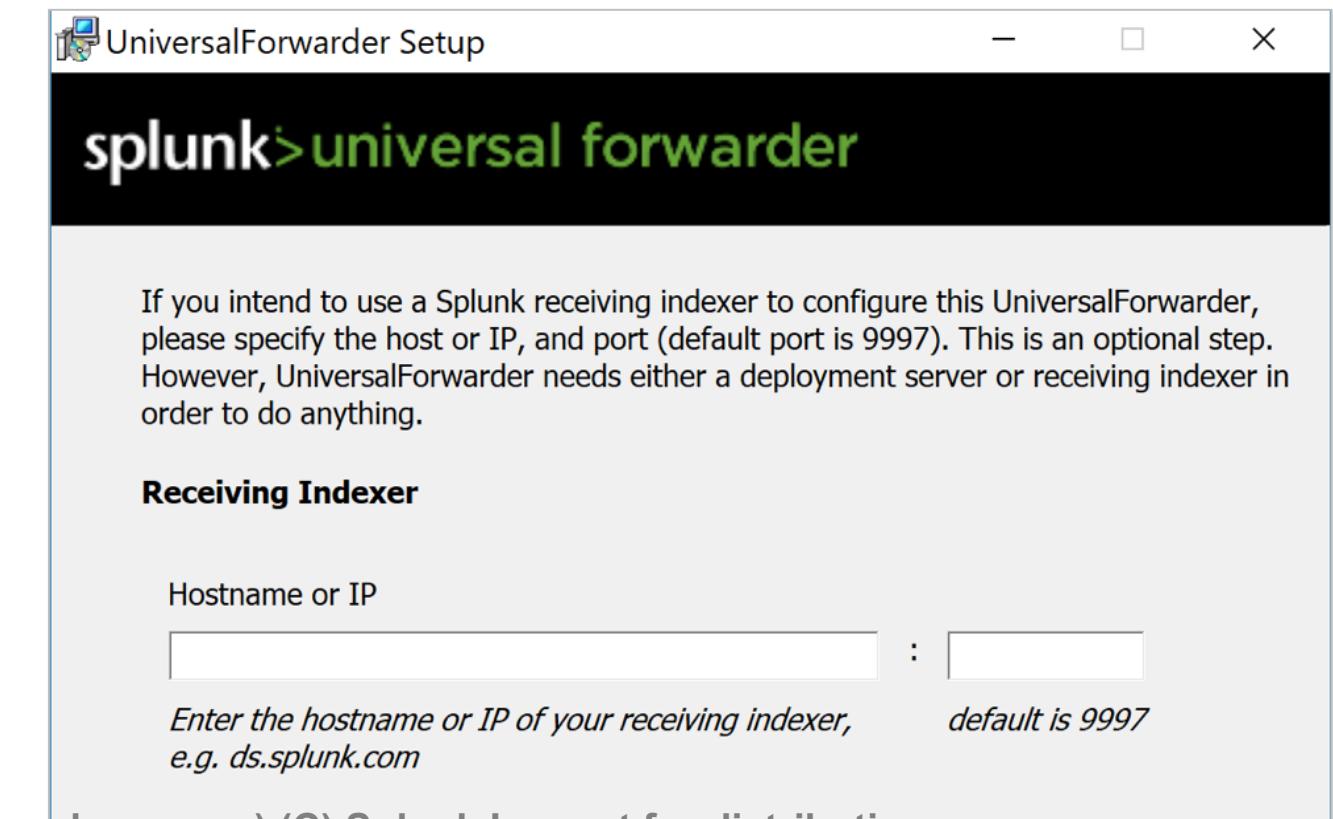
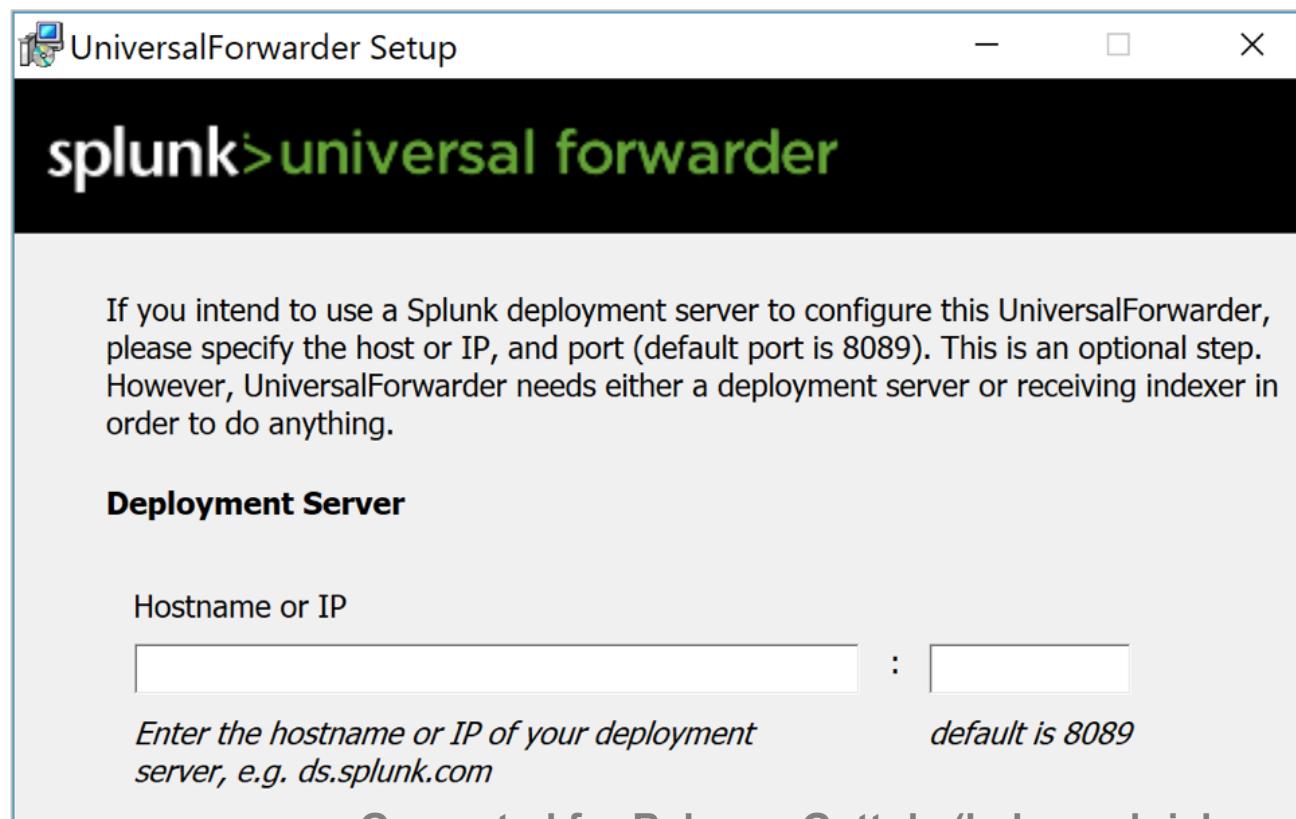
	*NIX	Windows
<i>Download</i>	www.splunk.com/en_us/download/universal-forwarder.html	
<i>Install</i>	<ul style="list-style-type: none">• Un-compress .tgz, .rpm, or .deb file in the path Splunk will run from• Default SPLUNK_HOME is: /opt/splunkforwarder	<ul style="list-style-type: none">• Execute .msi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

- Same **splunk** command-line interface in **SPLUNK_HOME/bin**
 - Same commands for start/stop, restart, etc.
 - An admin account and password are required
- When installing large numbers of forwarders, use an automated method

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Using the Interactive Windows Installer

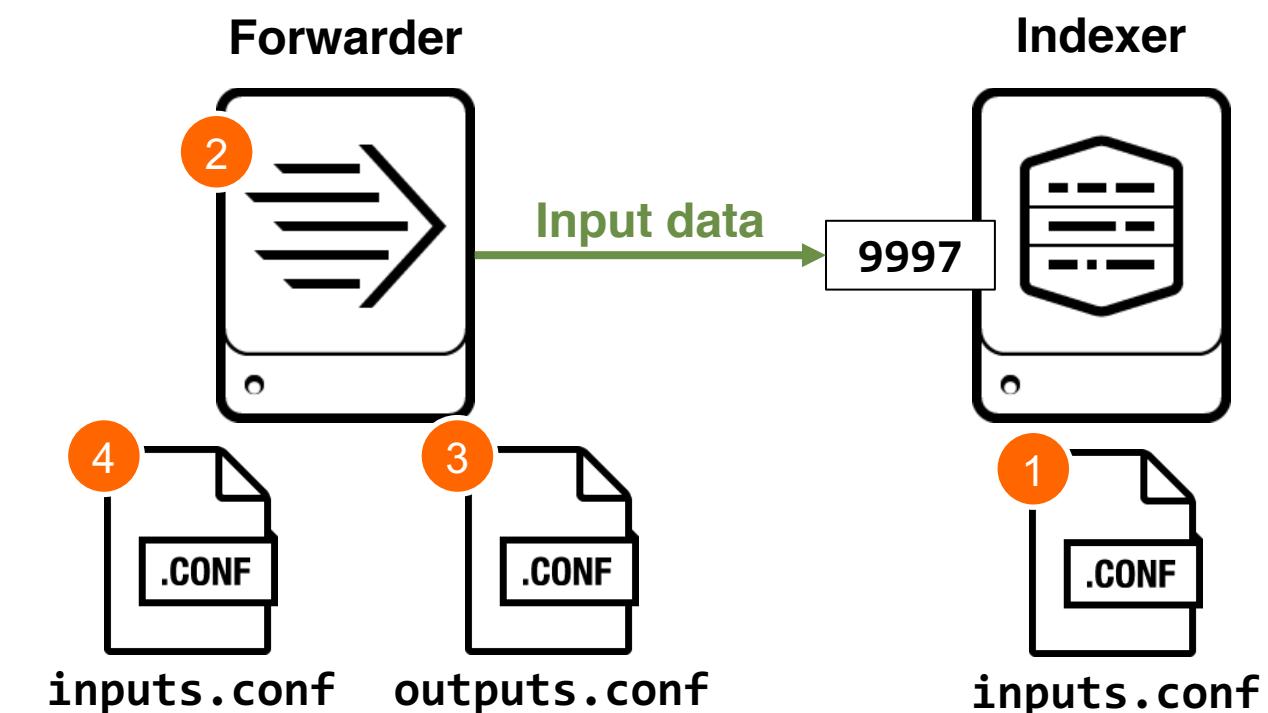
- Most forwarder settings can be configured using the installer wizard
 - Can run as a local or domain user without local administrator privileges
- CLI installation is available for scripted installations
docs.splunk.com/Documentation/Forwarder/latest/Forwarder/InstallWindowsuniversalforwarderfromthecommandline



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Universal Forwarder Configuration Steps

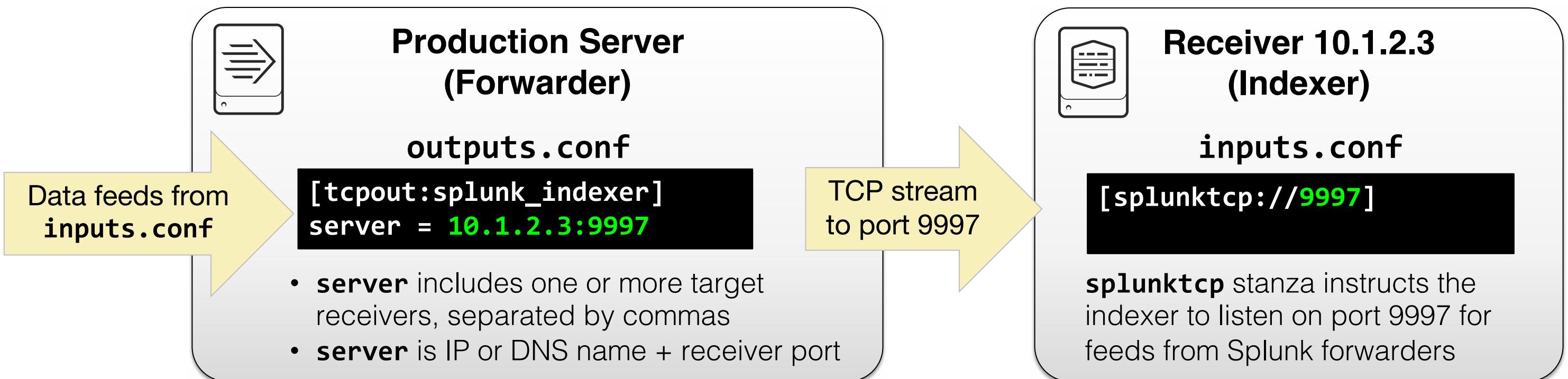
1. Set up a receiving port on each indexer
 - Task only needs to be performed once
2. Download and install Universal Forwarder
3. Set up forwarding on each forwarder, by either:
 - Editing **outputs.conf** manually
 - Using Deployment Server
4. Add inputs on forwarders, by either:
 - Editing **inputs.conf** manually
 - Using Deployment Server
 - Running Splunk commands (CLI)



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Forwarder outputs.conf File

- Points the forwarder to the receivers
- Can specify additional options for load balancing, SSL, compression, alternate indexers, and indexer acknowledgement



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Defining Target Indexers on the Forwarder

- Execute on the forwarder for each destination indexer:

splunk add forward-server <indexer:receiving_port>

- For example, **splunk add forward-server 10.1.2.3:9997** configures the **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

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Configuration and Connection Validation

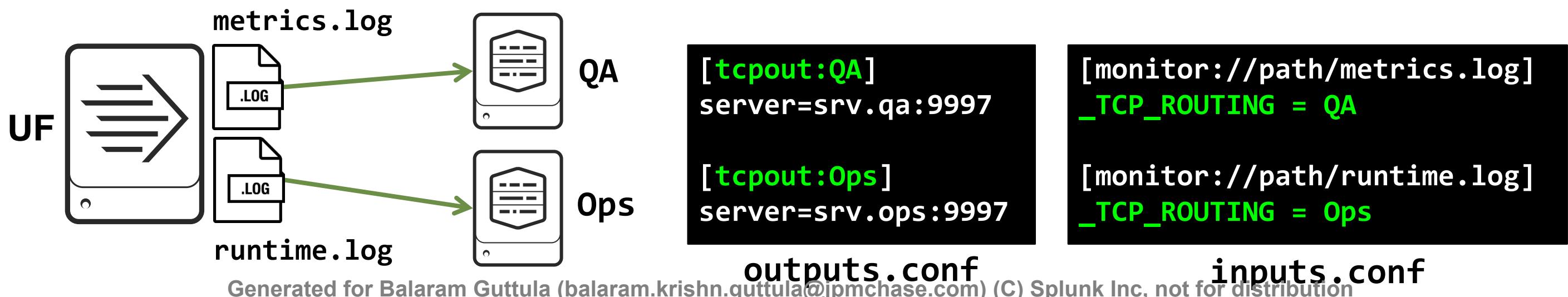
- After running **splunk add forward-server**:
 - Forwarder should be communicating with the indexer
 - Forwarder Splunk logs are automatically sent to indexer's **_internal** index
- To check the configuration:
 - On indexer, run: **splunk display listen**
 - On forwarder, run: **splunk list forward-server**
- To verify successful connection:
 - Search: **index=_internal host=<forwarder_hostname>**
- To remove the target indexer setting:
 - On forwarder, run: **splunk remove forward-server <indexer:port>**

Troubleshooting Forwarder Connection

- Is the forwarder sending data to the indexer?
 - Check **SPLUNK_HOME/var/log/splunk/splunkd.log** on forwarder:
tail -f var/log/splunk/splunkd.log | egrep 'TcpOutputProc|TcpOutputFd'
- Does the indexer receive any data on the listening port?
 - Search on indexer:
**index=_internal sourcetype=splunkd component=TcpInputConfig OR
(host=<uf> component=StatusMgr)**
 - To get the hostname **<uf>**, run on the forwarder:
splunk show default-hostname
- Check the configuration files

Selectively Forwarding Data to Indexers

- Universal forwarder can route based on sources
 - Define multiple tcpout stanzas in **outputs.conf**
 - Specify **_TCP_ROUTING** identifying the **tcpout** stanza names in each source in **inputs.conf**
- Example:
 - QA team wants **metrics.log** sent to the QA team's indexer and Ops team wants **runtime.log** sent to the operations indexer



Generated for Balaram Guttula (balaram.krishn.guttula@pmchase.com) (C) Splunk Inc, not for distribution

Additional Forwarding Options



Compressing the feed



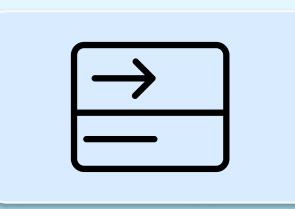
Securing the feed



Automatic load balancing to multiple indexers



Indexer acknowledgement to forwarder

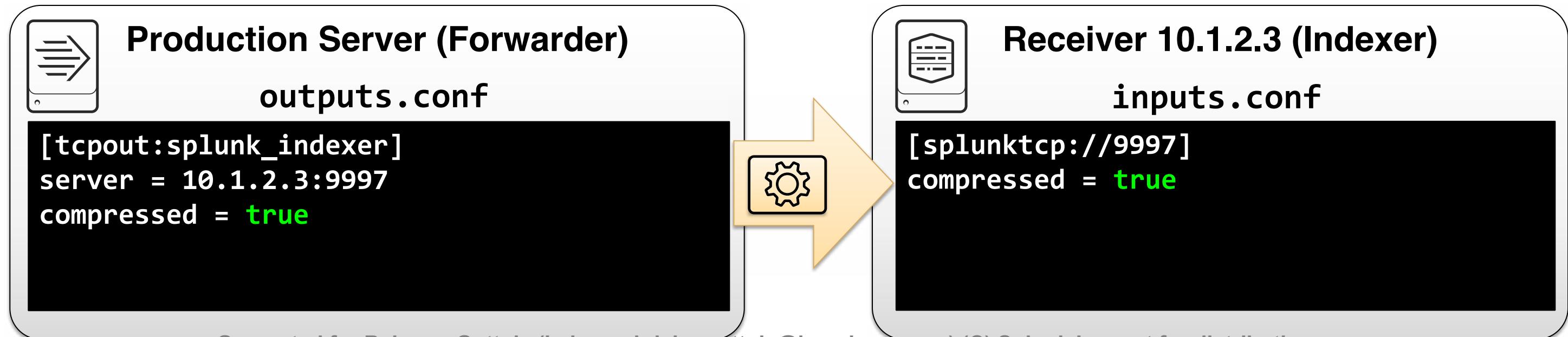


Forwarder queue size

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Compressing the Feed

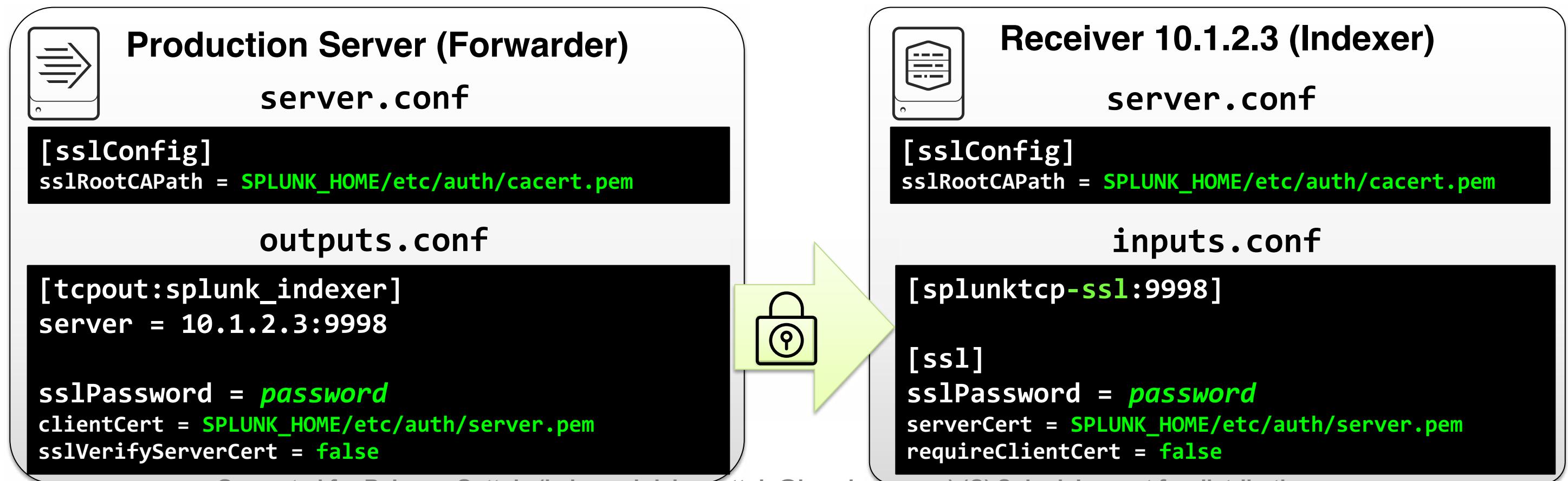
- Reduces network utilization
- Increases CPU utilization slightly
- Compression can be set at either the forwarder or the indexer
 - Compress select feeds by setting compression on the forwarder
 - Compress all feeds by setting compression on the indexer



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Securing the Feed with SSL

- Encrypts the feed
- Automatically compresses the feed
- Increases CPU utilization



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

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/latest/Security/Aboutsecuringdatafromforwarders

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

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

wiki.splunk.com/Community:Splunk2Splunk_SSL_DefaultCerts

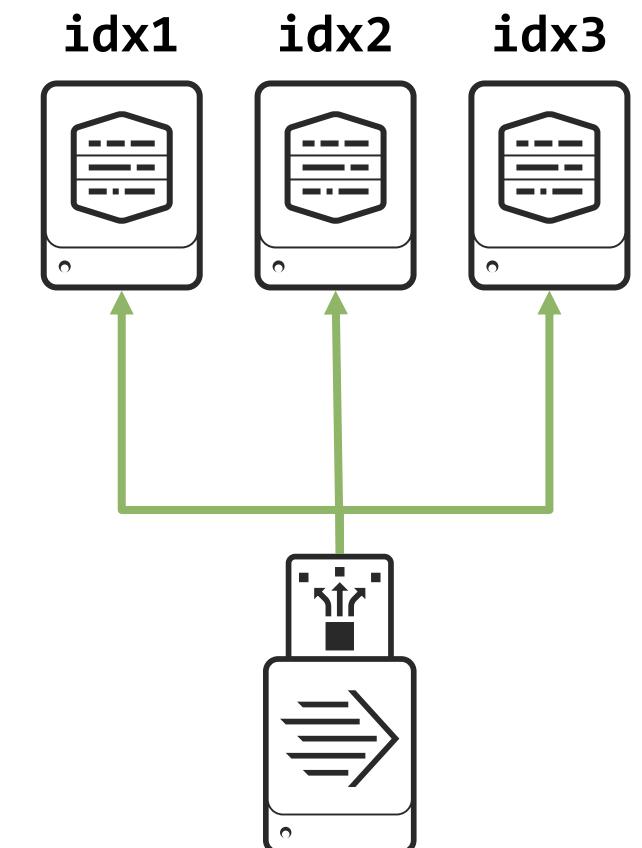
wiki.splunk.com/Community:Splunk2Splunk_SSL_SelfSignedCert_NewRootCA

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

Defining Event Boundary on UF

- Event boundaries
 - Detecting when one event ends and another starts
 - Normally determined on the indexer (where data is parsed)
- 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 the event breaker settings on UF per sourcetype in **props.conf**

- Single line event

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

- Multi-line event

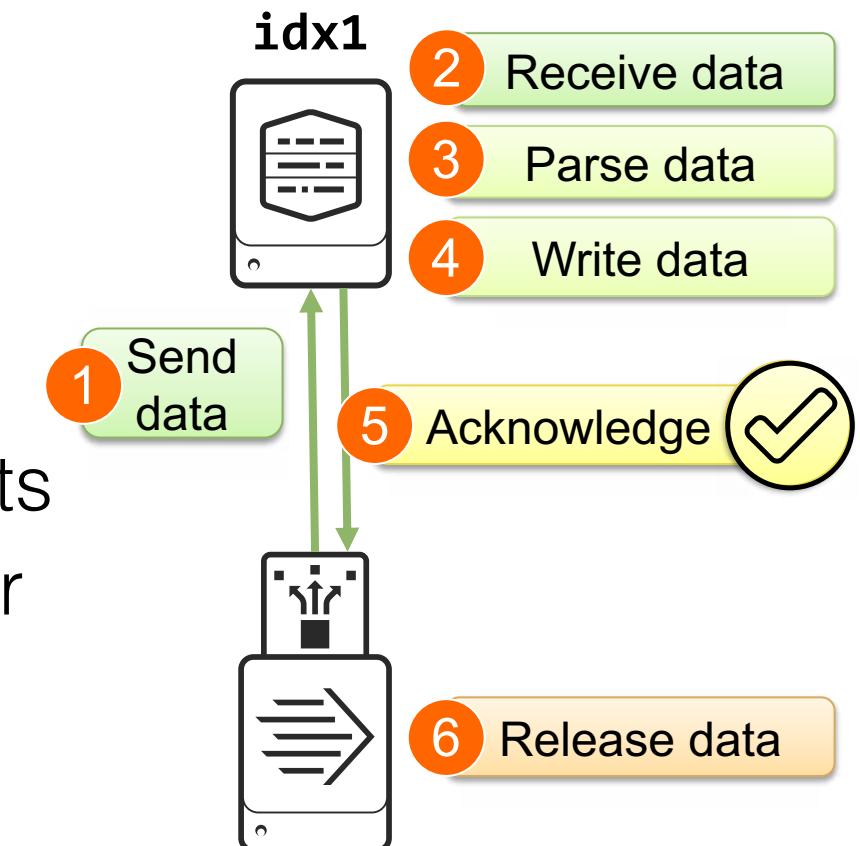
```
[my_log4j]
EVENT_BREAKER_ENABLE = true
EVENT_BREAKER = ([\r\n]+)\d\d\d\d-\d\d-\d\d
```

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

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Indexer Acknowledgement

- Disabled by default (**useACK=false**)
- Enabled with **useACK=true** in **outputs.conf**
- Guards against loss of forwarded data
 1. **Forwarder:** Send data to indexer
 2. **Indexer:** Receives the block of data
 3. **Indexer:** Parses the data
 4. **Indexer:** Writes the data to the file system as events
 5. **Indexer:** Sends acknowledgement to the forwarder
 6. **Forwarder:** Releases data from memory
 - ▶ If no acknowledgement is received from indexer, forwarder instead resends the data

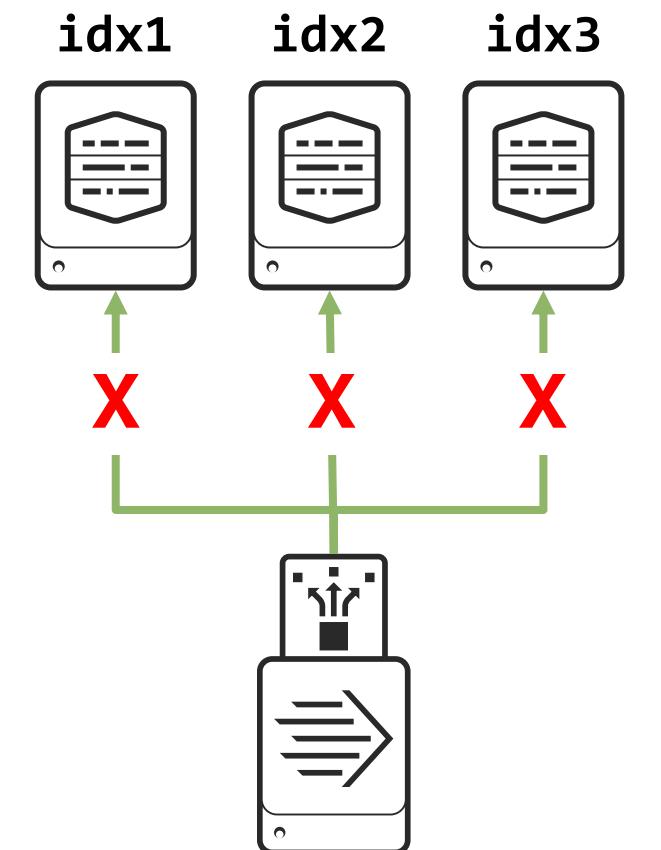


docs.splunk.com/Documentation/Splunk/latest/Forwarding/Protectagainstlossofin-flightdata

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

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

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Forwarding Resources

- Overview of forwarders

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

- Forwarder deployment overview

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

- Overview of enterprise installation

- Link at the bottom of the web page has example install packages and Windows install

wiki.splunk.com/Deploying_Splunk_Light_Forwarders

Useful Commands

Command	Operation
From the Forwarder	
splunk add forward-server	Configures the forwarder to connect the receiving indexer
splunk list forward-server	Displays the current receiving indexer configuration
splunk remove forward-server	Removes the receiving indexer 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

Module 3 Knowledge Check

- If the forwarder is set to send its data to 2 indexers at 30 second intervals, does it switch exactly at the 30th second?
- True or False. Turning SSL on between the forwarder and the receiver automatically compresses the feed.
- What configuration file on the forwarder defines where data is to be forwarded to?

Module 3 Knowledge Check - Answers

- If the forwarder is set to send its data to 2 indexers at 30 second intervals, does it switch exactly at the 30th second?

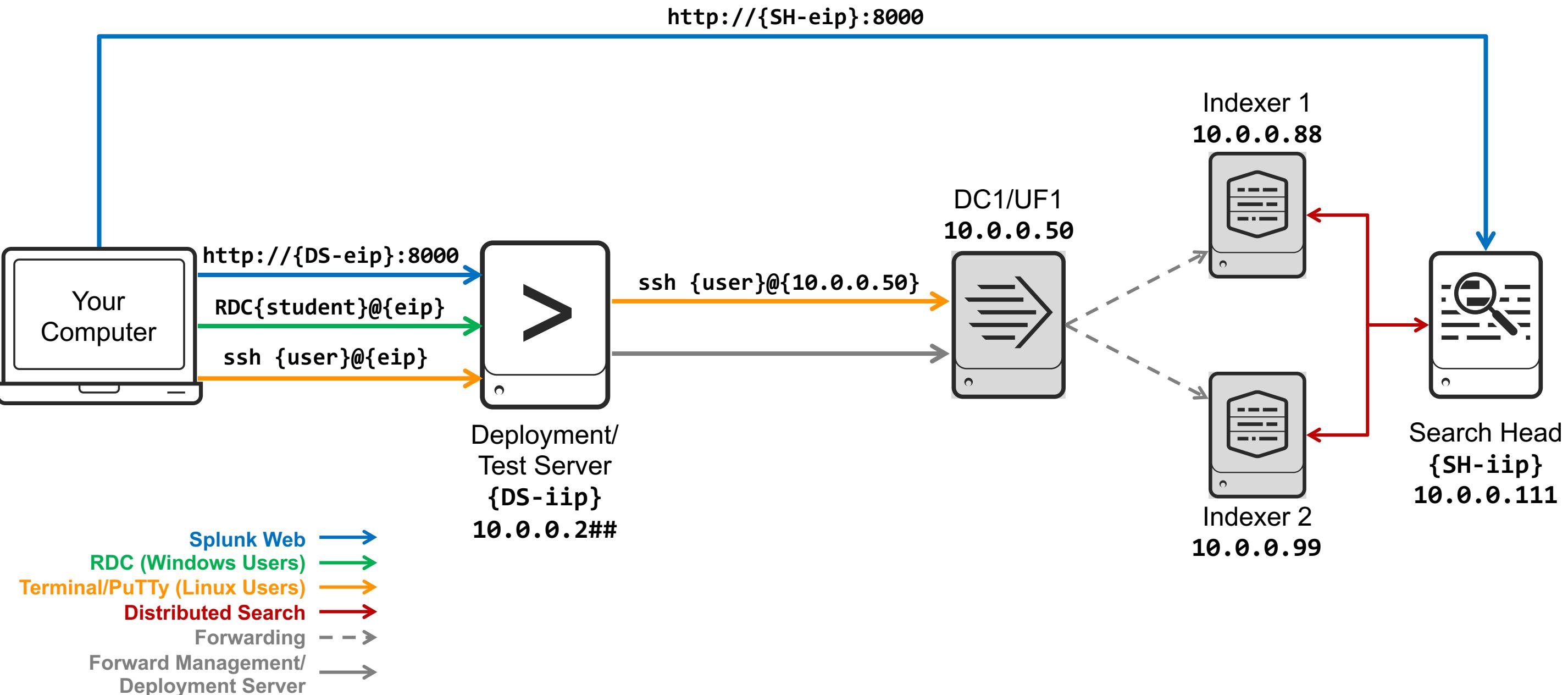
Not always. The forwarder does not want to send half an event to indexer1 and the other half to indexer2. To avoid this situation, for example, if the forwarder is tailing a file, then it waits for an EOF or a pause in IO activity before it switches.
- True or False. Turning SSL on between the forwarder and the receiver automatically compresses the feed.

True

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

outputs.conf

Module 3 Lab Exercise – Environment Diagram



Module 3 Lab Exercise – Setting Up Forwarders

Time: 20 – 25 minutes

Tasks:

- Configure forwarder to send data to the Indexer 1 (**10.0.0.88**) and Indexer 2 (**10.0.0.99**)
- Confirm forwarder connection from your search head

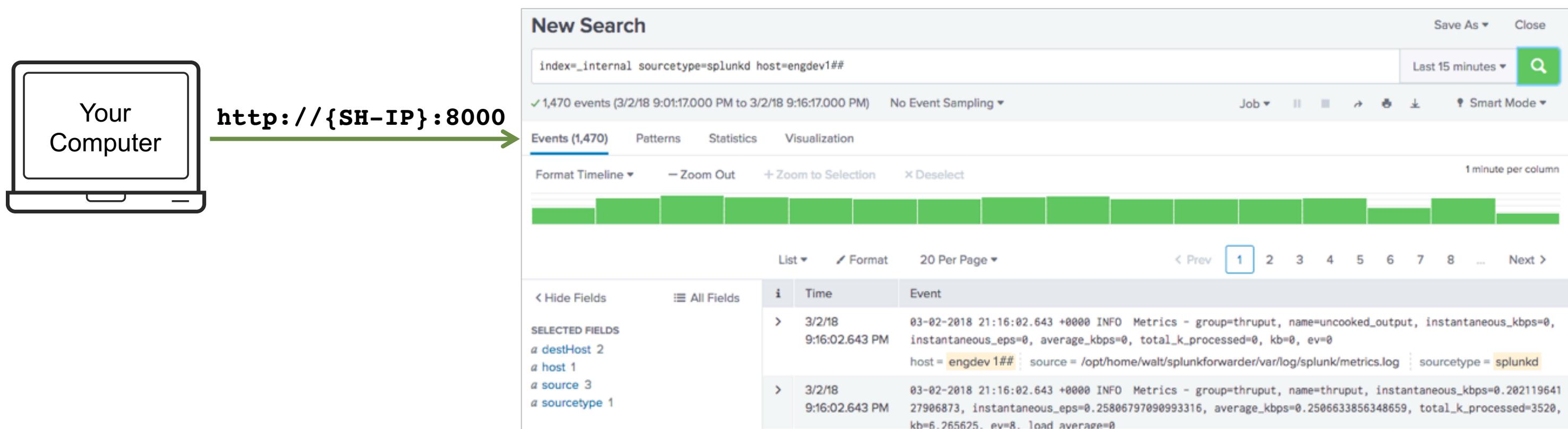
Lab notes:

- You have a login on a remote Linux host that is your forwarder
- This lab exercise only establishes the connection between your UF and Indexer

Module 3 Lab Exercise – Setting up Forwarders (cont.)

Verification: Run a search to get forwarded internal logs from UF#1

index=_internal sourcetype=splunkd host=engdev1##



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

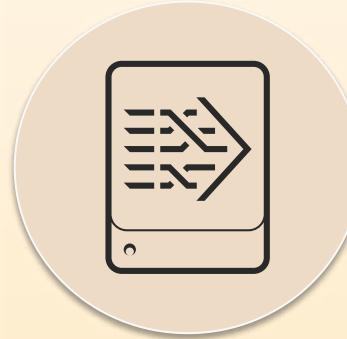
Module 4: Heavy Forwarders & Forwarder Management

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Module Objectives

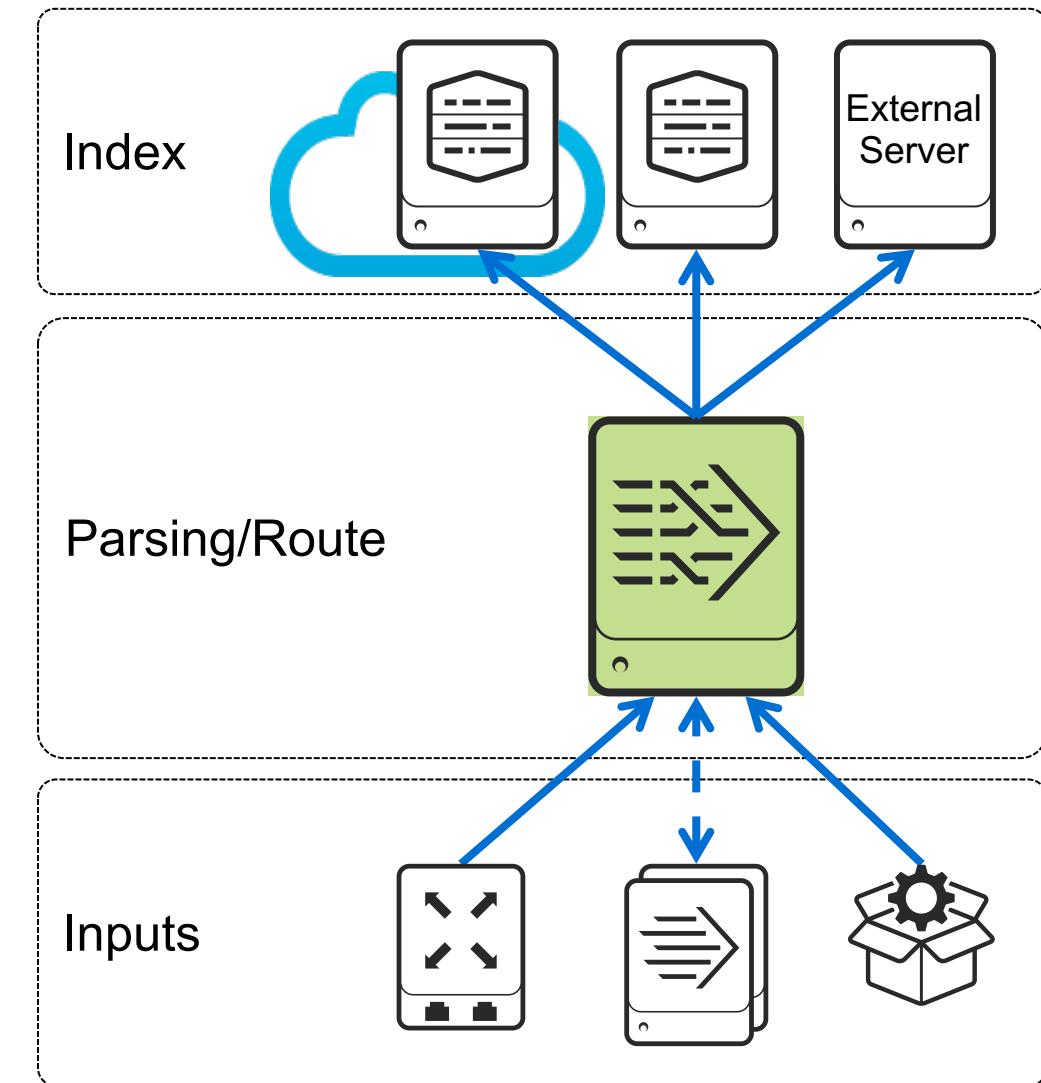
- Introduction to the Heavy Forwarder (HF)
 - Describe what the heavy forwarder is and use cases
 - Perform heavy forwarder configuration
 - Deploy an app to the heavy forwarder
- Using Splunk Forwarder Management
 - Describe Splunk Deployment Server (DS)
 - Manage forwarders using deployment apps
 - Configure deployment clients and client groups
 - Monitor forwarder management activities

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
- Cannot perform distributed searches



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

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

- Able to do all UF tasks, and more
- Required by some apps, add-ons, or input types (such as HEC, DBconnect)
- Provides Splunk Web, if needed
- Supports complex, event-level routing to separate indexers or indexer clusters
- Can anonymize or mask data before forwarding to an indexer
- Predictable version of Python

HF Configuration – Inputs

- HF can receive data from other Splunk instances and UFs
- Setup can be performed using either:
 - CLI: **splunk enable listen <port>**
 - Deploying an **inputs.conf** file from the deployment server with the following stanza:

```
[splunktcp://<port>]
```

Note 

During the lab exercises, you will manually set up the HF port using CLI.

HF Configuration – Outputs

- To configure the HF to forward the data to the indexers use either:
 - CLI to set up forwarding manually:
splunk add forward-server:<indexer:&ilistening_port>
 - Deploy an **outputs.conf** file from the DS with the following entry:

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

[tcpout-server://<indexer:&ilistening_port>]

[tcpout:default-autolb-group]
disabled = false
server = <indexer:&ilistening_port>
```

Note 

In the lab exercises, you will configure forwarding using the deployment server (DS).

HF Configuration – Client

- To configure the HF as a deployment client to the Deployment Server by hostname or IP:

```
splunk set deploy-poll <deploymentServer:port>
```

- This creates a **deploymentclient.conf** file with the following stanza and setting:

```
[target-broker:<deploymentServer>]  
targetUri = 10.0.0.2##:8089
```

HF Configuration – Optimizations

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

outputs.conf

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

- Disable Splunk Web on the HF:

web.conf

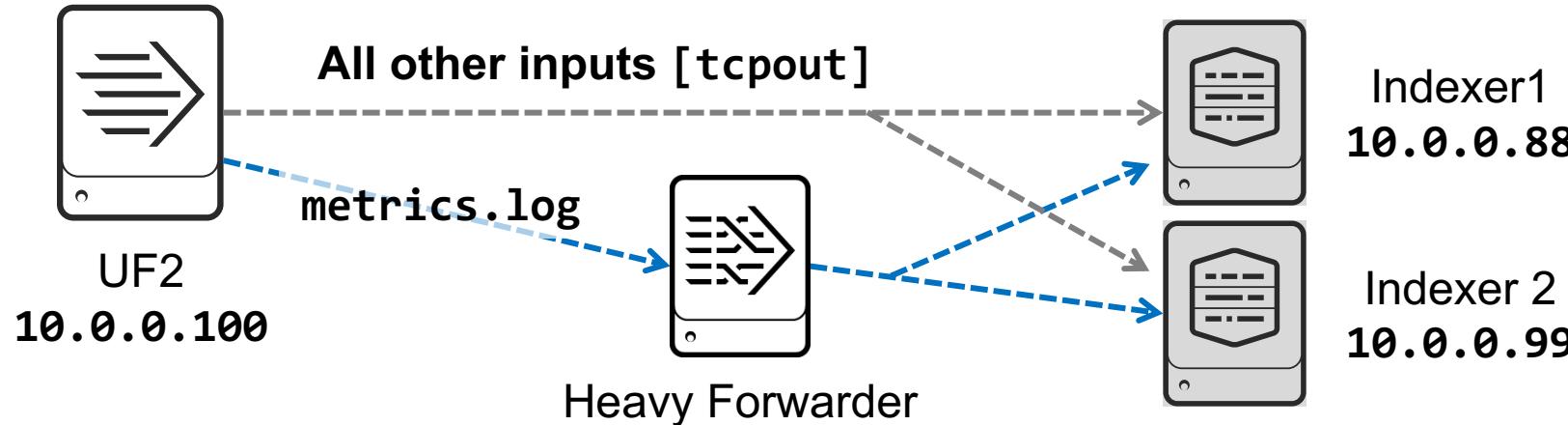
```
[settings]
startwebserver = 0
```

Selectively Routing Data

- Selectively send data from the UF directly to the indexers
- Route UF data through a HF to parse before sending to indexers

Note

It is best practice to send your data to the indexers. Because of the limitations of the lab environment for this class, you will route all data from UF2 through the HF to the indexers.



inputs.conf on UF2

```
[monitor://path/metrics.log]
_TCP_ROUTING = HF_TheParser

[monitor://path/runtime.log]
```

outputs.conf on UF2

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

[tcpout:default-autolb-group]
disabled=false
server = 10.0.0.88:9997,10.0.0.99:9997

[tcpout:HF_TheParser]
server=10.0.0.77:99XX
```

Module 4 HF Knowledge Check

- True or False. The HF has a GUI.
- True or False. The UF and the HF can be used to mask data before transmitting to indexers.
- True or False. The listening port has to be 8089.

Module 4 HF Knowledge Check – Answers

- True or False. The HF has a GUI.

True.

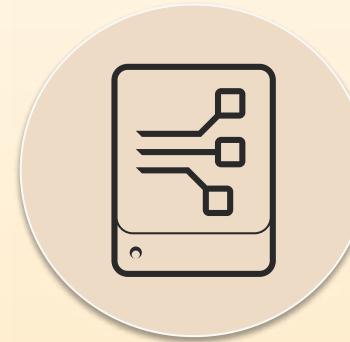
- True or False. The UF and the HF can be used to mask data before transmitting to indexers.

False. Only the HF, specifically a Splunk Enterprise instance, can perform data masking.

- True or False. The default listening port is 8089.

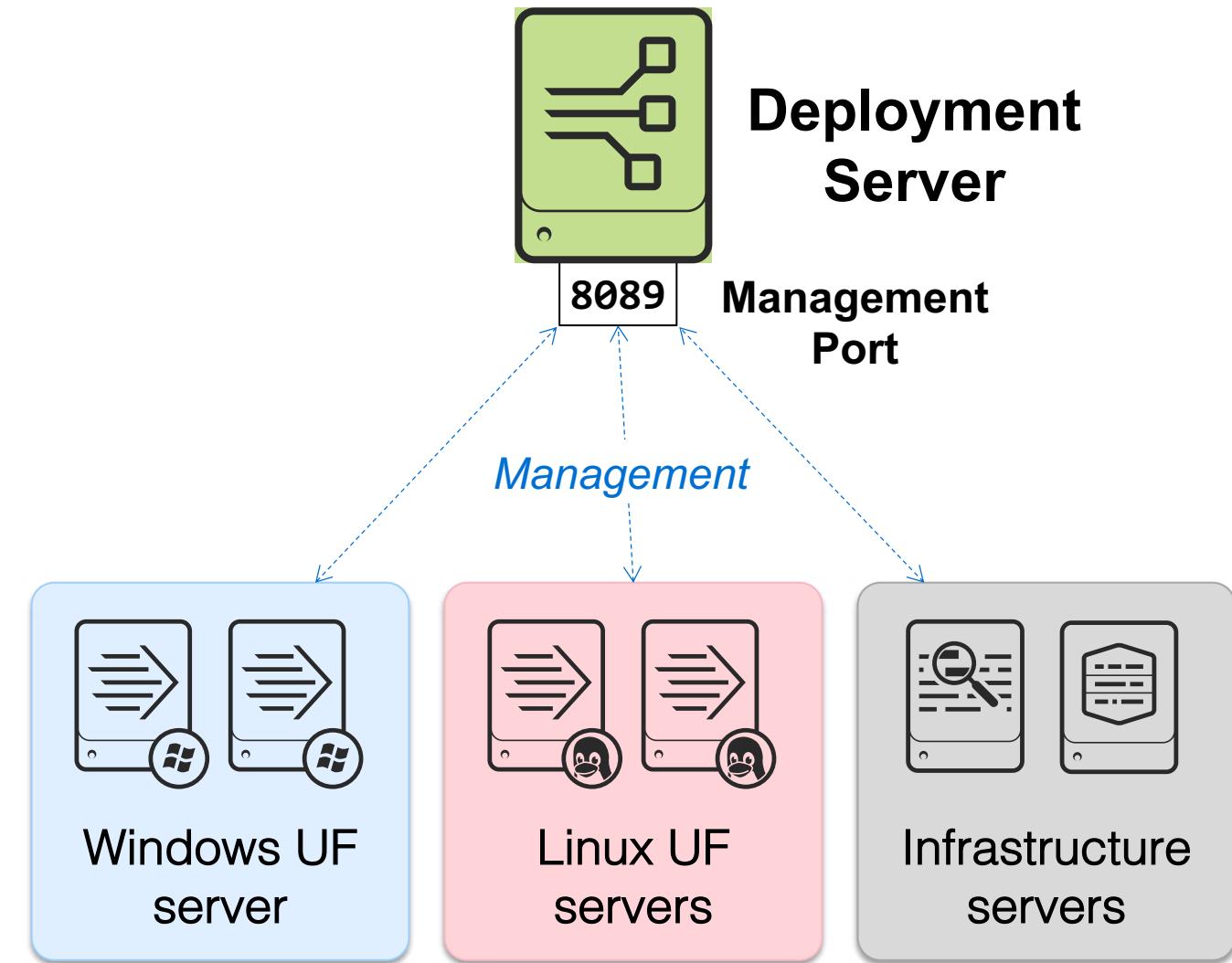
False. 8089 is the default management port. The listening port can be any port.

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



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Deployment Server Components

Deployment Apps

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

Server Class

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

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

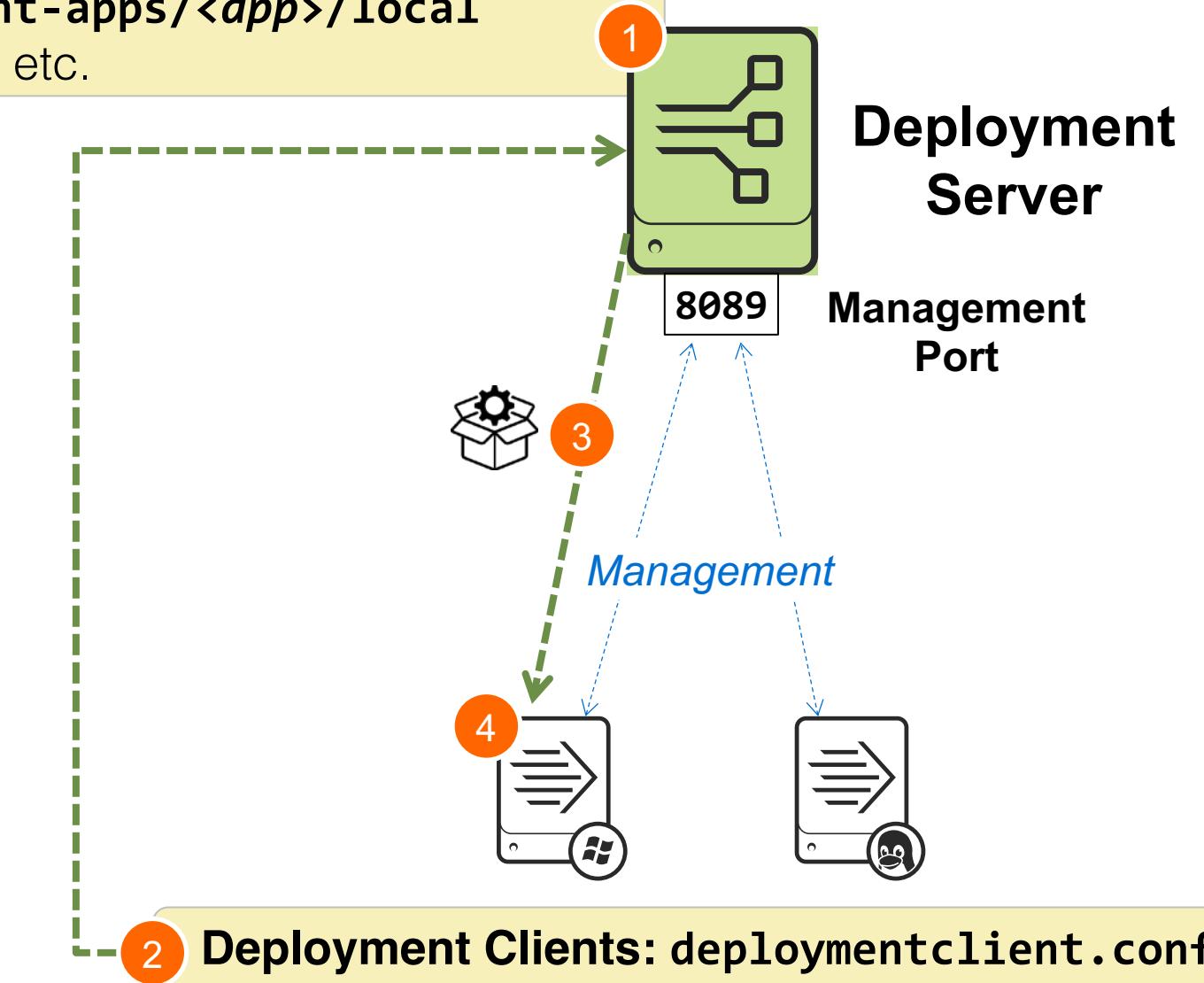
Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Deployment Server Configuration

Configuration on DS:

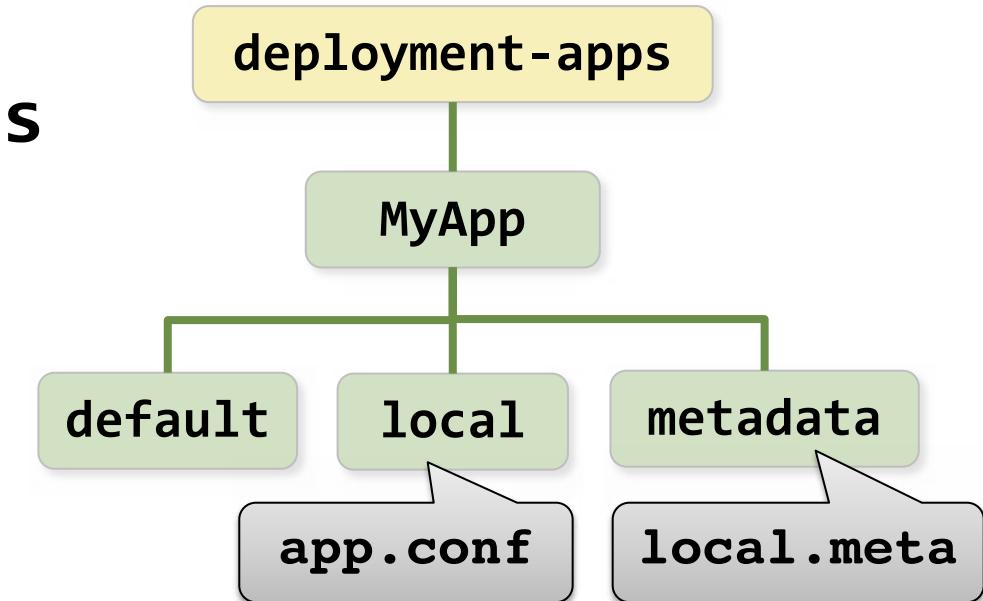
Map clients to apps: **SPLUNK_HOME/etc/apps/<app>/local/serverclass.conf**
App repository: **SPLUNK_HOME/etc/deployment-apps/<app>/local**
Apps/configs to deploy: **outputs.conf, inputs.conf, etc.**

1. Configure DS, server class configuration, and app packages
2. Configure instances as deployment clients; phones home to deployment server
3. Client downloads subscribed apps as directed by DS
4. Client uses configuration; for example sending data to indexers configured in **outputs.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

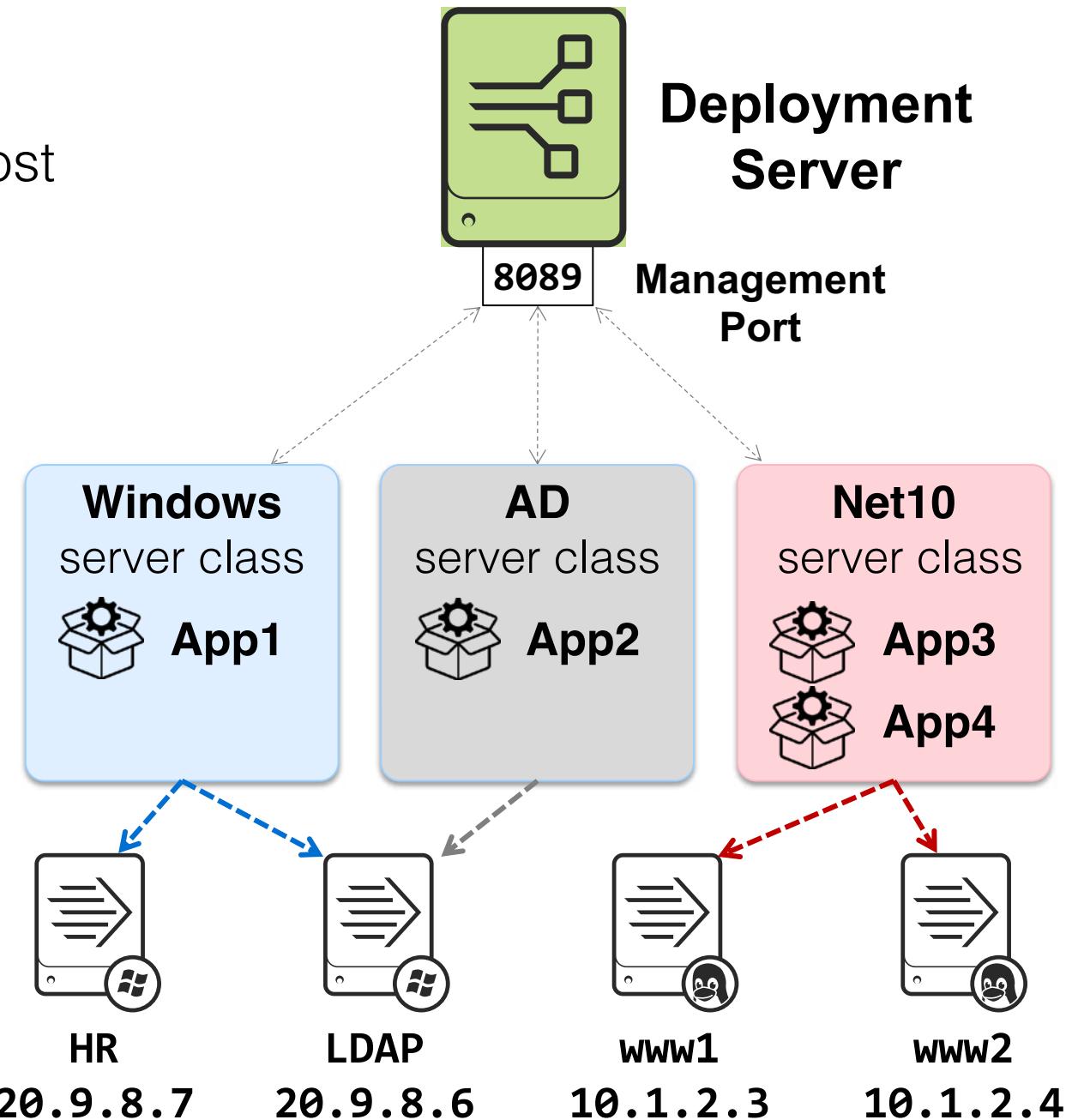
- Can be downloaded from Splunkbase
- Installed on a Splunk instance:
 - Using the Deployment Server
 - Using CLI on the instance
 - Manually by installing the app
- Deploy to **SPLUNK_HOME/etc/apps**
- Comes with documentation for details about settings for **inputs.conf**, and so on

The screenshot shows the Splunkbase App Search Results page. At the top, there's a search bar with placeholder text "Search App by keyword, technology...". To the right are "My Account" and "Support & Services" buttons. Below the header, the title "App Search Results" is displayed. On the left, a sidebar contains filters for "PRODUCTS & SOLUTIONS" (with categories like DevOps, Security, Fraud & Compliance, IT Operations, Utilities, Business Analytics, IoT & Industrial Data), "TECHNOLOGIES", "APP TYPE", "APP CONTENTS", "SPLUNK VERSION" (set to 7.3), "CIM VERSION", and "SPLUNK BUILT & OTHER". Above the results, there are three selected filters: "Product & Solutions: Splunk Enterprise", "Category: Utilities", and "Category: IT Operations". The results section shows 1-20 of 793 results, ordered by "Newest". Each result card includes the app icon, name, install count, and a small orange checkmark. The cards shown are: Cisco ACI Add-on for Splunk (373 installs), Avi Networks Add-on for Splunk (31 installs), HTTP Event Push (11 installs), Sunburst Viz (199 installs), Semicircle Donut Chart Viz (206 installs), Day Night Map Viz (96 installs), Clock Viz (109 installs), Circlepack Viz (134 installs), NetFlow LOGIC (partially visible), and two other green circular icons.

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

What's a Server Class?

- Maps groups of clients to deployment apps
 - Clients can be grouped based on client name, host name, IP address, DNS name, or machine types
- Example:
 - Windows server class
 - Assigned to Windows systems
 - Installs App1
 - AD server class
 - Assigned to Active Directory servers
 - Installs App2
 - Net10 server class
 - Assigned to hosts on **10.1.2.*** subnet
 - Installs App3 and App4



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Enabling Forwarder Management

1. On deployment server: Add one or more apps in
SPLUNK_HOME/etc/deployment-apps
2. In Forwarder Management UI: Create one or more server classes
3. On forwarders: Setup the deployment client
 - Run **splunk set deploy-poll <DS:port>** using the **splunkd** port on the deployment server (default **8089**)
 - Run **splunk restart**
4. On deployment server: Verify deployment clients and deployment status
5. On forwarders: Verify **SPLUNK_HOME/etc/apps** folder for deployed apps

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

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](#)

Select the Server Classes tab

1

2

Forwarder Management

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

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

New Server Class

Name Cancel Save

Enter a name for the new server class

3

2 New Server Class

The screenshot shows the Splunk Forwarder Management interface. In the top left, there's a summary section with metrics for clients, deployment errors, and total downloads. Below it, a message says 'No server classes' with a link to 'create one'. A callout 'Select the Server Classes tab' points to the 'Server Classes (0)' link, which is highlighted with a green box and numbered 1. A green arrow labeled 2 points from the 'create one' link to the 'Server Classes (1)' link in the main navigation bar of the lower section. Another callout 'Enter a name for the new server class' points to the 'Name' input field in the 'New Server Class' dialog, which is also highlighted with a green box and numbered 3. The 'Save' button in the dialog is also highlighted with a green box and numbered 2.

Generated for Balaram Guttula (balaram.krishn.guttula@pmchase.com) (C) Splunk Inc, not for distribution

Selecting Apps for the Server Class

The screenshot illustrates the process of selecting apps for a server class named "uf_base".

Step 1: On the main "Server Class: uf_base" page, a green "Add Apps" button is highlighted with a red circle containing the number 1.

Step 2: A modal window titled "Edit Apps" shows the "uf_base" app selected in the "Unselected Apps" list. The "uf_base" item is highlighted with a red circle containing the number 2. A yellow callout bubble with the text "Select app to move it to Selected Apps" points to the "uf_base" item.

Step 3: The "Selected Apps" list contains the "uf_base" app. The "Save" button is highlighted with a red circle containing the number 3.

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Post Deployment Behavior Setting

Server Class: uf_base

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

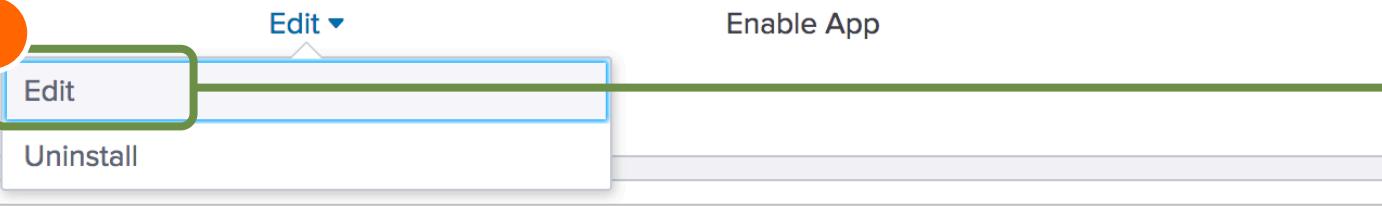
[Back to Forwarder Management](#)

Apps Edit

Deployed Successfully ▾ filter

1 Apps 10 Per Page ▾

Name	Actions	After Installation	Clients
uf_base	Edit ▾ 1 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

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

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

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

Include (whitelist)

ip-10*

Can be client name, host name.
Examples: 185.2.3.* , fwdr-*
Learn more ↗

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

- Supports wildcards
- **Exclude** takes precedence over **Include**

- In addition to **include/exclude**, you can further filter based on machine types
- The list is based on the clients that have connected to this deployment server

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

splunk® turn data into doing™

102

Splunk Enterprise Data Administration
Copyright © 2020 Splunk, Inc. All rights reserved | 29 May 2020

Configuring Deployment Clients

- Configure your forwarders to be deployment clients

- Run **splunk set deploy-poll <DS:port>**

- DS** = deployment server hostname or IP
- port** = splunkd port
- Creates **deploymentclient.conf** in **SPLUNK_HOME/etc/system/local**

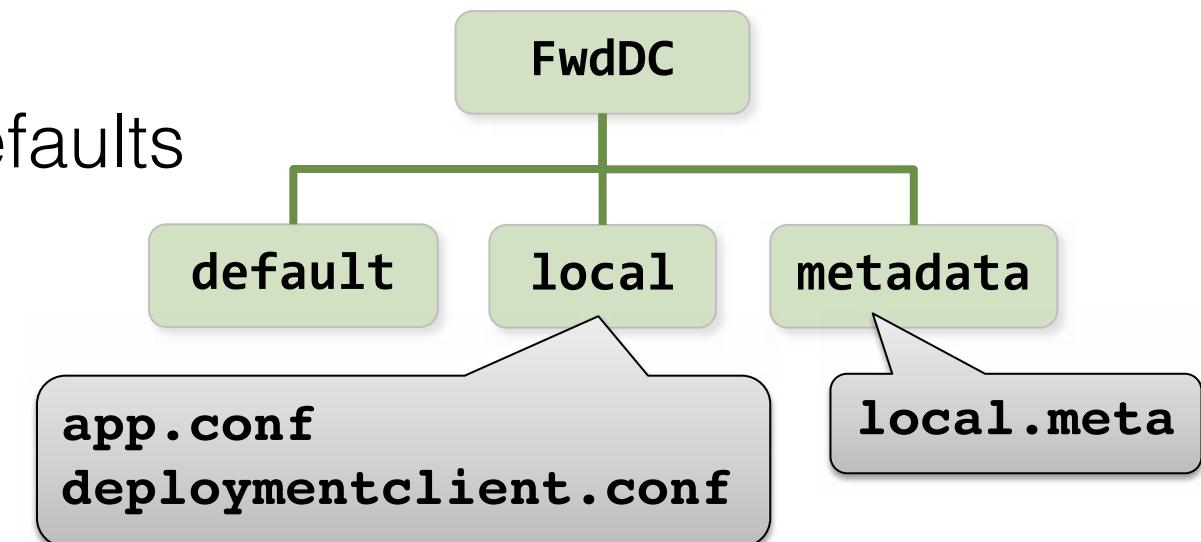
(Alternately create **deploymentclient.conf** manually)

- Create an app and place the **.conf** file in the **local** directory
- Restart the deployment clients: **splunk restart**

- Edit **[deployment-client]** stanza to override defaults

- Can be part of initial deployment app
- Contains **phoneHomeIntervalInSecs** setting
(default is 60 seconds)

```
deploymentclient.conf
[target-broker:deploymentServer]
targetUri = splunk_server:8089
...
[deployment-client]
clientName = webserver_1
phoneHomeIntervalInSecs = 600
```

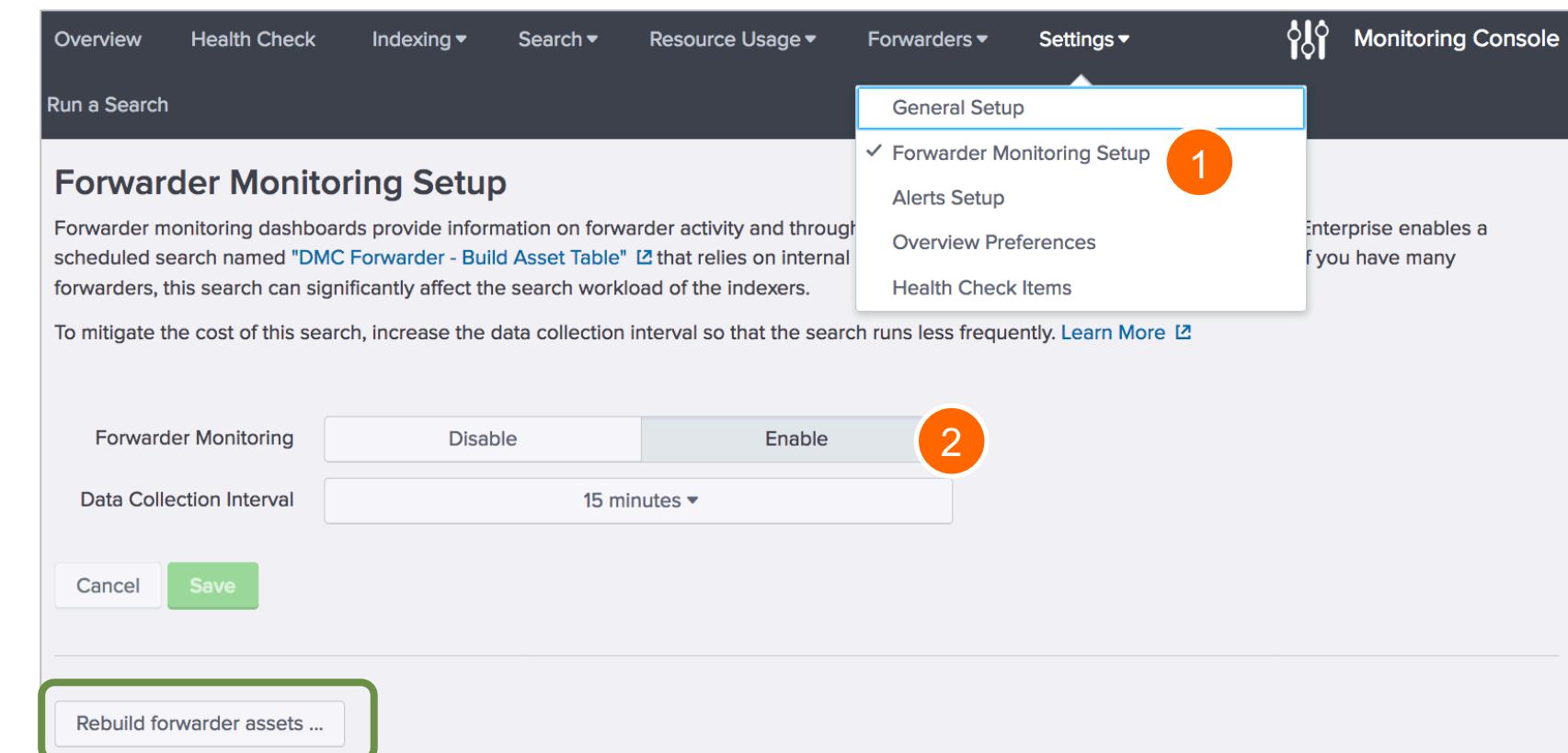


Configuring Deployment Clients (cont.)

- On the deployment client (usually a forwarder)
 - Confirm that the expected app directories and contents have arrived in **SPLUNK_HOME/etc/apps**
- If the app is changed on the Deployment Server, then the forwarder will load the updated app after its next phone-home
 - To change the app settings using Forwarder Management, use the app's Edit menu associated with the server class
 - To change inputs for an app, go to **Settings > Data Inputs > Forwarded Inputs**
- If the post-deployment behavior option is set, the forwarder is restarted
- On the deployment client
 - Use **splunk show deploy-poll** to check the deployment server settings
 - Use **splunk list forward-server** to check the indexer destination settings

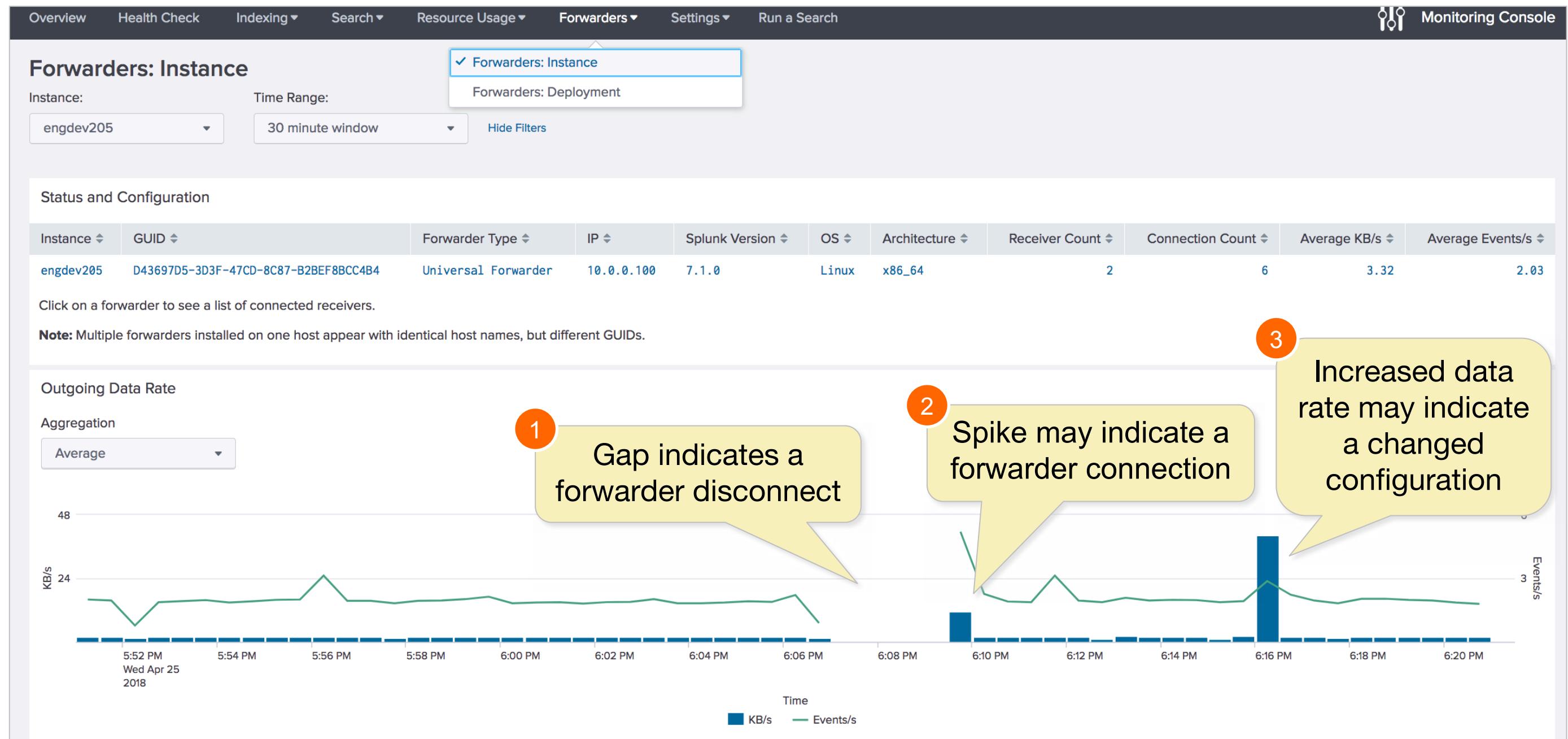
Forwarder Monitoring with Monitoring Console

- Provides valuable information on forwarder activity and throughput
- Enabled with:
 1. Monitoring Console: Settings > Forwarder Monitoring Setup
 2. Forwarder Monitoring: Enable
- 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



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

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 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
splunk list forward-server	Displays the current forward server configuration

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Module 4 Knowledge Check

- On the DS, what is the difference between the apps sitting in the **SPLUNK_HOME/etc/apps** folder versus the **SPLUNK_HOME/etc/deployment-apps** ?
- When an app is deployed from the DS to the client, where will you find that app on the client by default?
- True or False. Deployment clients poll the DS on port 9997.

Module 4 Knowledge Check – Answers

- On the DS, what is the difference between the apps sitting in the **SPLUNK_HOME/etc/apps** folder versus the **SPLUNK_HOME/etc/deployment-apps**?

The apps in the **.../etc/apps** folder are for the Deployment Server and the apps in the **.../etc/deployment-apps** are apps for deployment to a client.

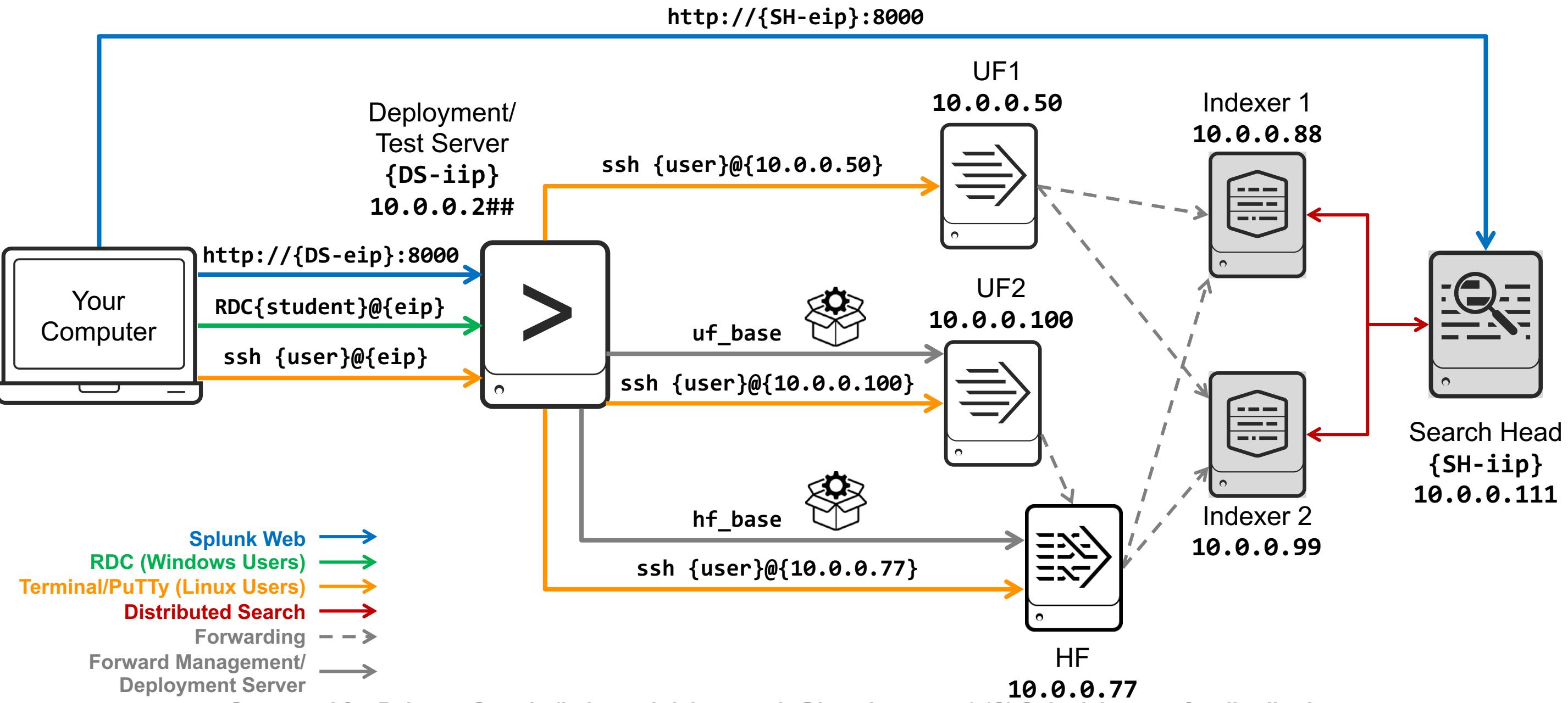
- When an app is deployed from the Deployment Server to the client, where will you find that app on the client by default?

Apps by default are deployed from the DS to the client in the **SPLUNK_HOME/etc/apps** folder.

- True or False. Clients poll the DS on port 9997.

False. Clients poll the DS on it's management port (**8089** by default.)

Module 4 Lab Exercise – Environment Diagram



Module 4 Lab Exercise – Forwarder Management

Time: 25 – 30 minutes

Tasks:

- Enable Forwarder Management by copying the **uf_base** and **hf_base** apps into the **SPLUNK_HOME/etc/deployment-apps** folder on the deployment/test server
- Configure UF2 as a deployment client to the deployment/test server
- Enable listening port on HF for data being transmitted from UF2
- Configure the HF as a deployment client to the deployment/test server
- Create two server classes to manage UF2 and the HF from the deployment/test server
- Confirm deployment of **hf_base** app on HF and **uf_base** app on UF2

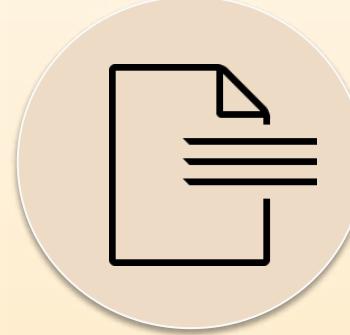
Module 5: Monitor Inputs

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Module Objectives

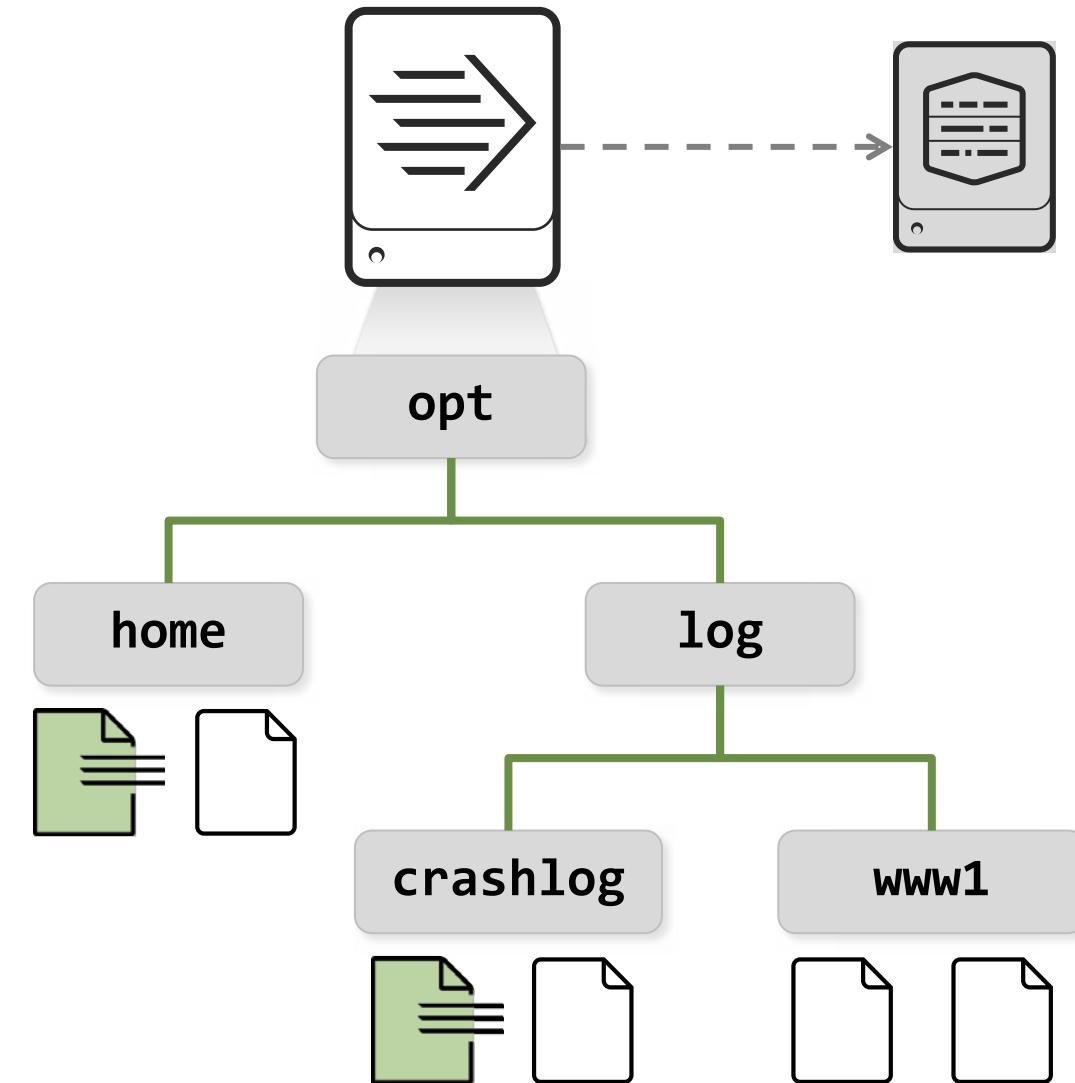
- Create file and directory monitor inputs
- Use optional settings for monitor inputs
- Deploy a remote monitor input

Monitoring Input Files



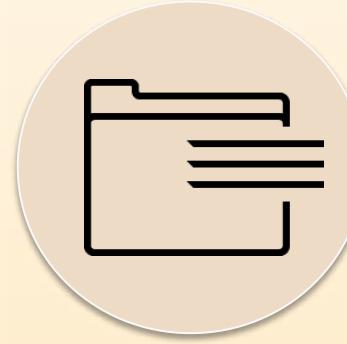
Monitoring Files

- 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**



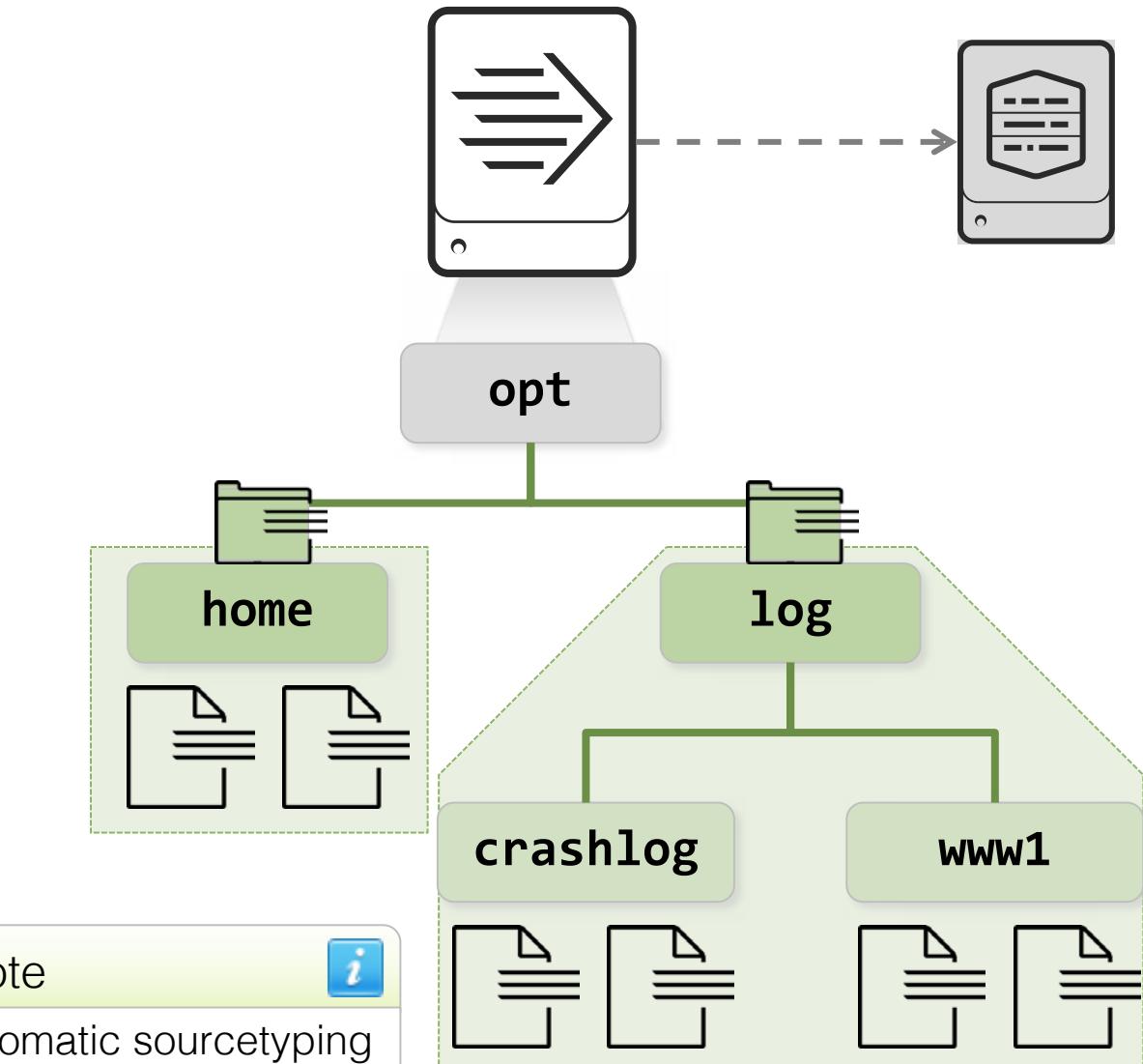
Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Monitoring Input Directories



Monitoring Directories

- Defines a directory tree as data source
- Recursively traverses directory for 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



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

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
 - The `sourcetype`, `host`, `index`, etc. are optional
 - Default `host` is defined in **SPLUNK_HOME
etc/system/local/inputs.conf**
 - Default `source` is the fully-qualified file name
 - Default `sourcetype` is `automatic`
- 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://C:\logs\system.log]
[monitor://C:\logs\]
[monitor:///var/log/]
```

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 since 60 days ago with **ignoreOlderThan = 60d**)
- Completely ignores files with modification time outside the time window (even if the file is updated later)

Whitelist and Blacklist

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

Example: Using Whitelist to Include Files

- Files/directories that match the regular expression are indexed
- The syntax for blacklists 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

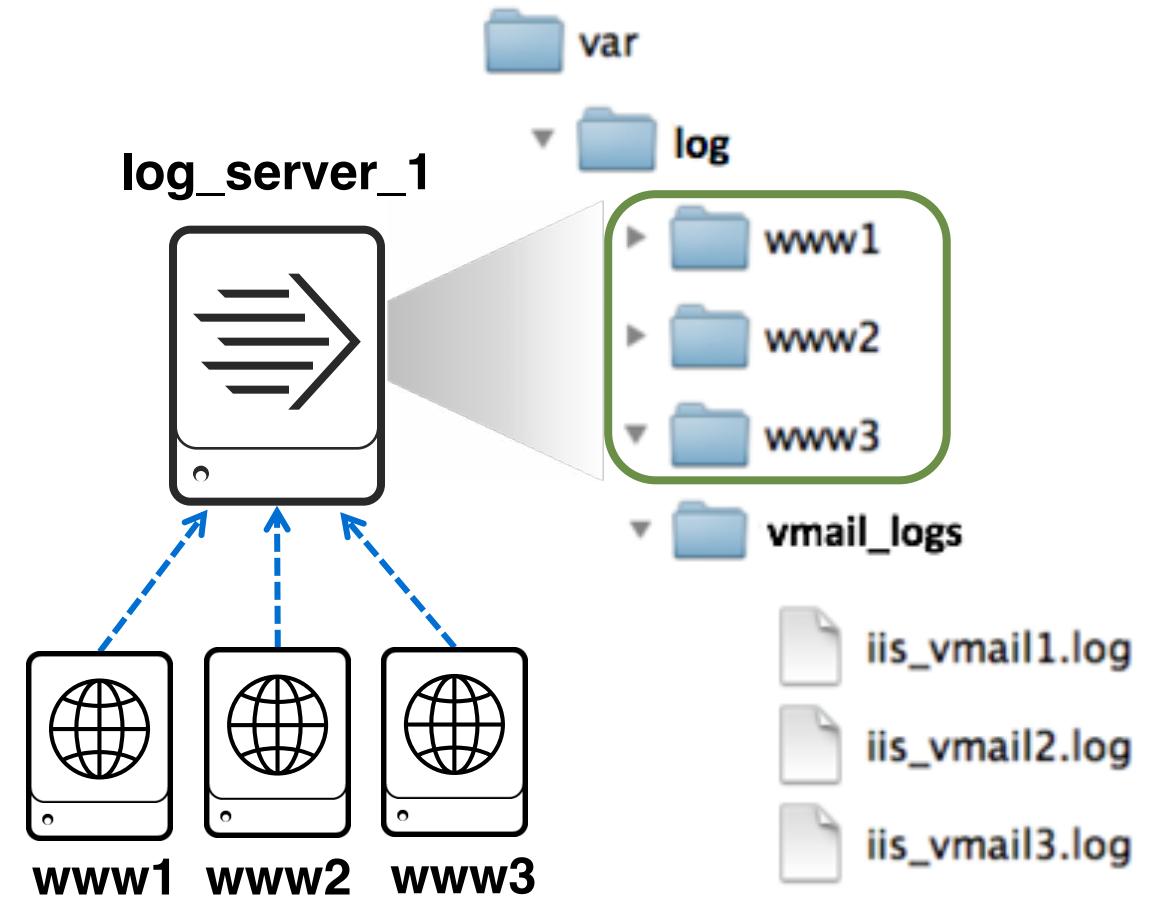
```
[monitor:///var/log/www1/]
whitelist = /query\.log$|/my\.log$
```

✓ /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, currently at the 'Input Settings' step. On the left, a file tree shows a 'var' folder containing a 'log' folder with three subfolders: 'www1', 'www2', and 'www3'. 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', 'Regular expression on path', and 'Segment in path', with 'Segment in path' being selected (indicated by a red border). Below the input field is a large black box containing the configuration command:

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

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

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

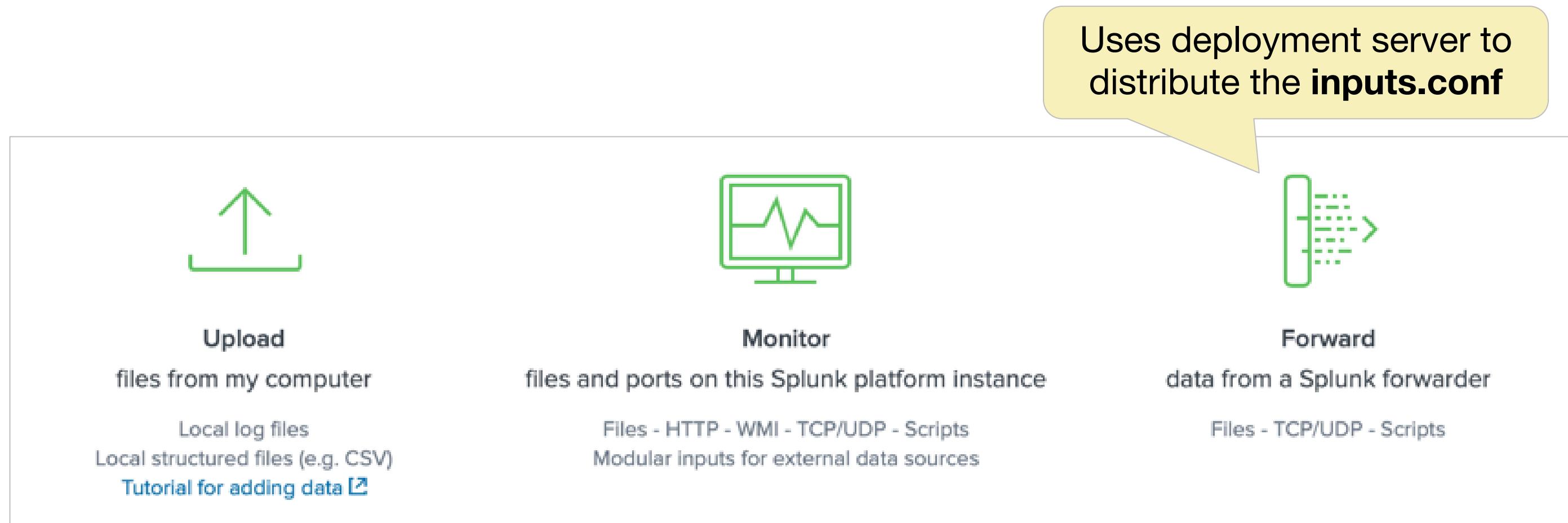
The screenshot shows the 'Add Data' wizard in Splunk, specifically the 'Input Settings' step. On the left, a file tree shows a directory structure: 'var' -> 'log' -> 'vmail_logs'. Inside 'vmail_logs', there are three files: 'iis_vmail1.log', 'iis_vmail2.log', and 'iis_vmail3.log'. A green arrow points from the 'vmail_logs' folder to the 'Regular expression?' input field. The 'Regular expression?' field contains the value `\w+(vmail.+)\.log$`. To the right of this field is a red box highlighting the radio button for 'Regular expression on path'. Below this, there are other options: 'Constant value' (radio button not selected) and 'Segment in path' (radio button not selected). A green arrow points from the 'Regular expression?' field down to a black box at the bottom. The black box contains the configuration command:

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

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Creating a Remote Data Input

After deployment clients are working, optionally create deployment apps for configuring inputs on the clients



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Creating a Remote Data Input (cont.)

Add Data

Select Forwarders Select Source Input Settings Review Done

Next >

Select Forwarders

Create or select a server class for data inputs. Use this page only in a single-instance Splunk environment.

To enable forwarding of data from deployment clients to this instance, set the output con

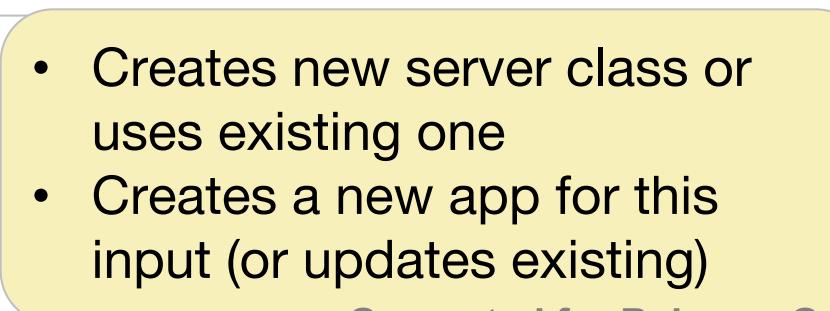
Select Server Class	New	Existing
---------------------	-----	----------

Available host(s)

LINUX ip-10-0-0-100

add all »

New Server Class Name eng_webservers



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

Add Data

Select Forwarders Select Source Input Settings Review Done

Next >

Select Forwarders

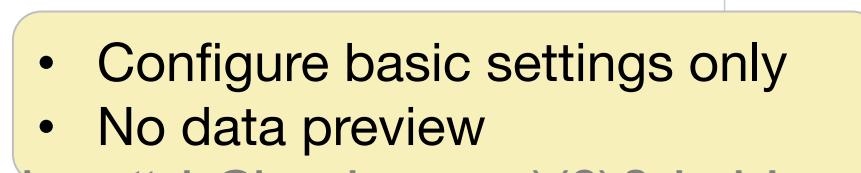
Configure selected Splunk Universal Forwarders to monitor both existing an file or directory. If you choose to monitor a directory, you can only assign a s the data within that directory. If a directory contains different log files from va sources, configure individual file monitor inputs for each type of log file (you opportunity to set individual source types this way). If the specified directory subdirectories, Splunk recursively examines them for new files. [Learn More](#)

File or Directory ? /opt/log/www2

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

Whitelist ? optional

Blacklist ? optional



• Configure basic settings only
• No data preview

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Editing Remote Data Input

The screenshot shows the Splunk interface for managing remote data inputs. On the left, a sidebar lists 'Forwarded inputs' and 'Files & Directories'. A green arrow points from the 'Forwarded inputs' link to the 'Type' section of the main configuration panel. A red circle with the number '1' highlights the 'Files & Directories' link in the sidebar. A red circle with the number '2' highlights the second item in the list of files/directories. A red circle with the number '3' highlights the 'More settings' link in the top right of the configuration panel.

Forwarded inputs ←

Type

Windows Event Logs
Collect event logs from forwarders.

Files & Directories
Monitor files or directories on forwarders.

Files & directories
Data inputs » Files & directories
Showing 1-2 of 2 items
filter

Source path	Host	Source type
/opt/log	None	Automatic
2 /opt/www2/access.log	None	Automatic

3 You can tell Splunk to continuously collect data from a file or directory (keep indexing data as it comes in), or index a static file and then stop.

More settings

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

Set host constant value
Specify method for getting host field for events coming from this source.

Host field value

Source type
Tell Splunk what kind of data this is so you can group it with other data of the same type when you search. Splunk does this automatically, but you can specify what you want if Splunk gets it wrong.

Set the source type Automatic
When this is set to automatic, Splunk classifies and assigns the sourcetype automatically, and gives unknown sourcetypes placeholder names.

Index
Set the destination index for this source.

Index test

Advanced options

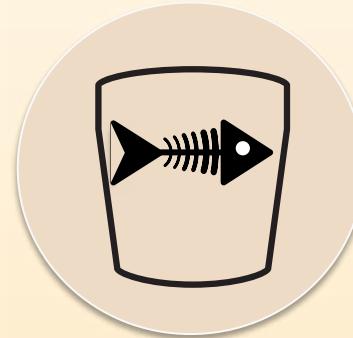
Whitelist
Specify a regex that files from this source must match to be monitored by Splunk.

Blacklist
Specify a regex that files from this source must NOT match to be monitored by Splunk.

Cancel **Save**

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

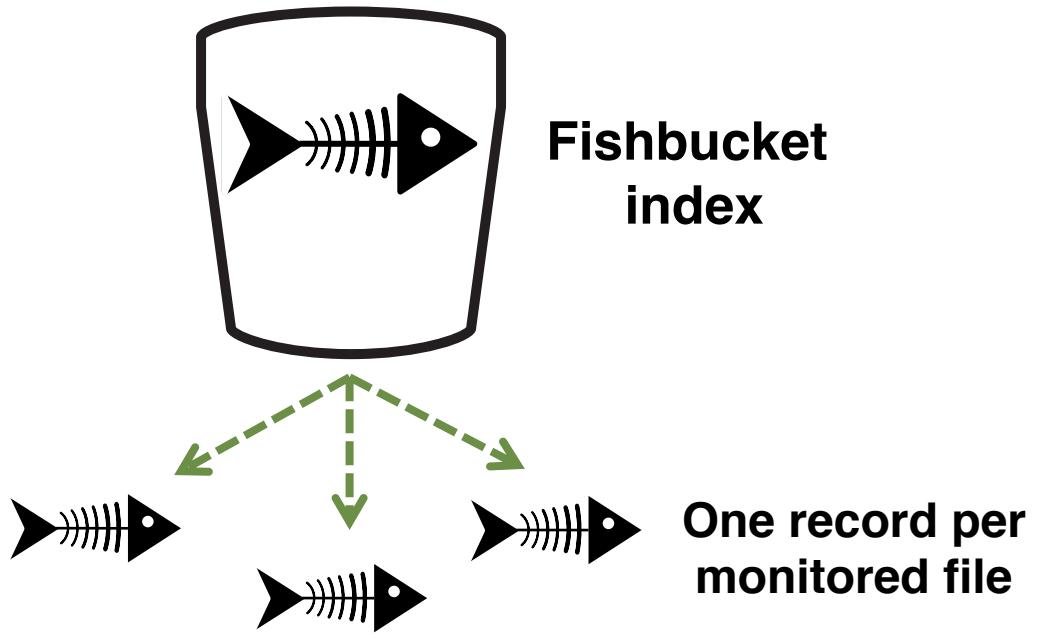
What is the Fishbucket?



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**

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution



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
 - 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. This results in more license usage.

Module 5 Knowledge Check

- True or False. You can use the wildcards ... and * in the whitelist and blacklist.
- True or False. The **host_regex** setting in **inputs.conf** can extract the host from the filename only.
- After a file monitor is set up and is running, if you change the host value, will the new host value be reflected for already ingested data?
- In our environment, we have a UF, an Indexer and a SH. Which instance contains the **_thefishbucket**?

Module 5 Knowledge Check – Answers

- True or False. You can use the wildcards, ... and * in the whitelist and blacklist.
False. The wildcards, ... and * are meant for the stanzas.
- True or False. The **host_regex** setting in **inputs.conf** can extract the host from the filename only.
False. It can extract the host from the path of the file.
- After a file monitor is set up and is running, if you change the host value, will the new host value be reflected for already ingested data?
No. All changes apply to the new data only. To reflect changes for your old data, you need to delete and re-ingest the old data.
- In our environment, we have a UF, an Indexer and a SH. Which instance contains the **_thefishbucket**?
Each instance will have its own local **_fishbucket.**

Module 5 Lab Exercise – File Monitors

Time: 20 – 25 minutes

Tasks:

- To test-collect remote data from UF#2, add a remote directory monitor input to the **test** index
- Modify the **inputs.conf** file using the following caveats and re-deploy
 - ▶ Send the source logs to the **sales** index
 - ▶ Override the **default-host** name value
 - ▶ To monitor only the **www.*** sub-directories, use **whitelist**
 - Exclude the indexing of the **secure.log** files, use **blacklist**

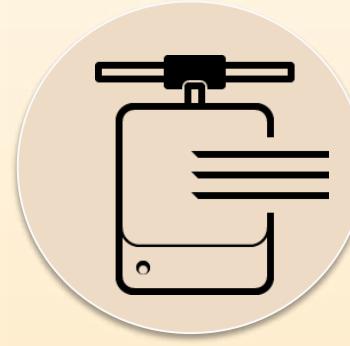
Module 6: Network and Scripted Inputs

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Module Objectives

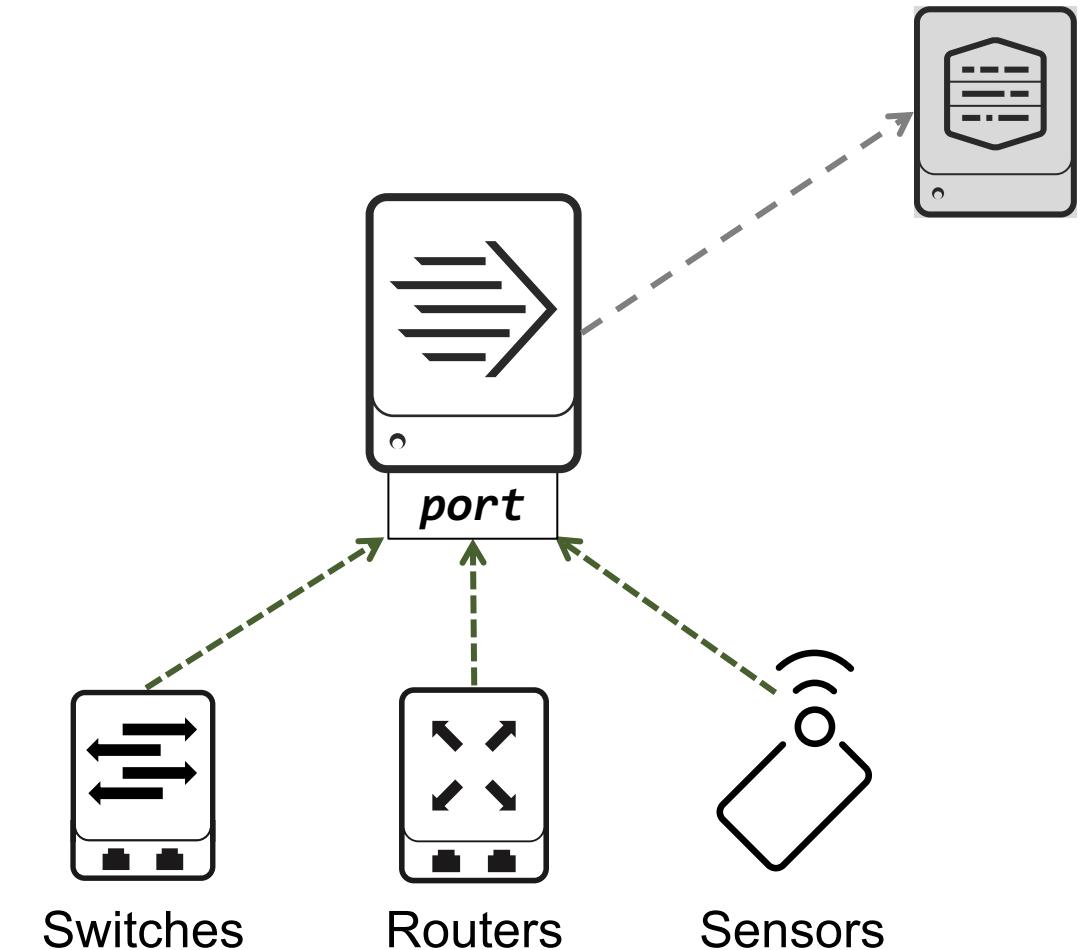
- Create network (TCP and UDP) inputs
- Describe optional settings for network inputs
- Create a basic scripted input

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 Splunk to listen on a network port.

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

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

TCP UDP

Port ? Example: 514

Source name override ? host:port

Only accept connection from ? 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

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

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 explicitly setting 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', with 'Custom' selected), a text input field containing 'dnsserver', and a 'More settings' checkbox which is checked.

Source
Source name override <input type="text" value="dcrusher9001"/>
If set, overrides the default source v
Source type
Set sourcetype for all events from this source.
Set sourcetype <input type="button" value="Manual"/>
Source type <input type="text" value="dnslog"/>
If this field is left blank, the default v
Host
Set host <input type="radio"/> IP <input type="radio"/> DNS <input checked="" type="radio"/> Custom
<input type="text" value="dnsserver"/>
Index
Set the destination index for this source.
Index <input type="text" value="test"/>

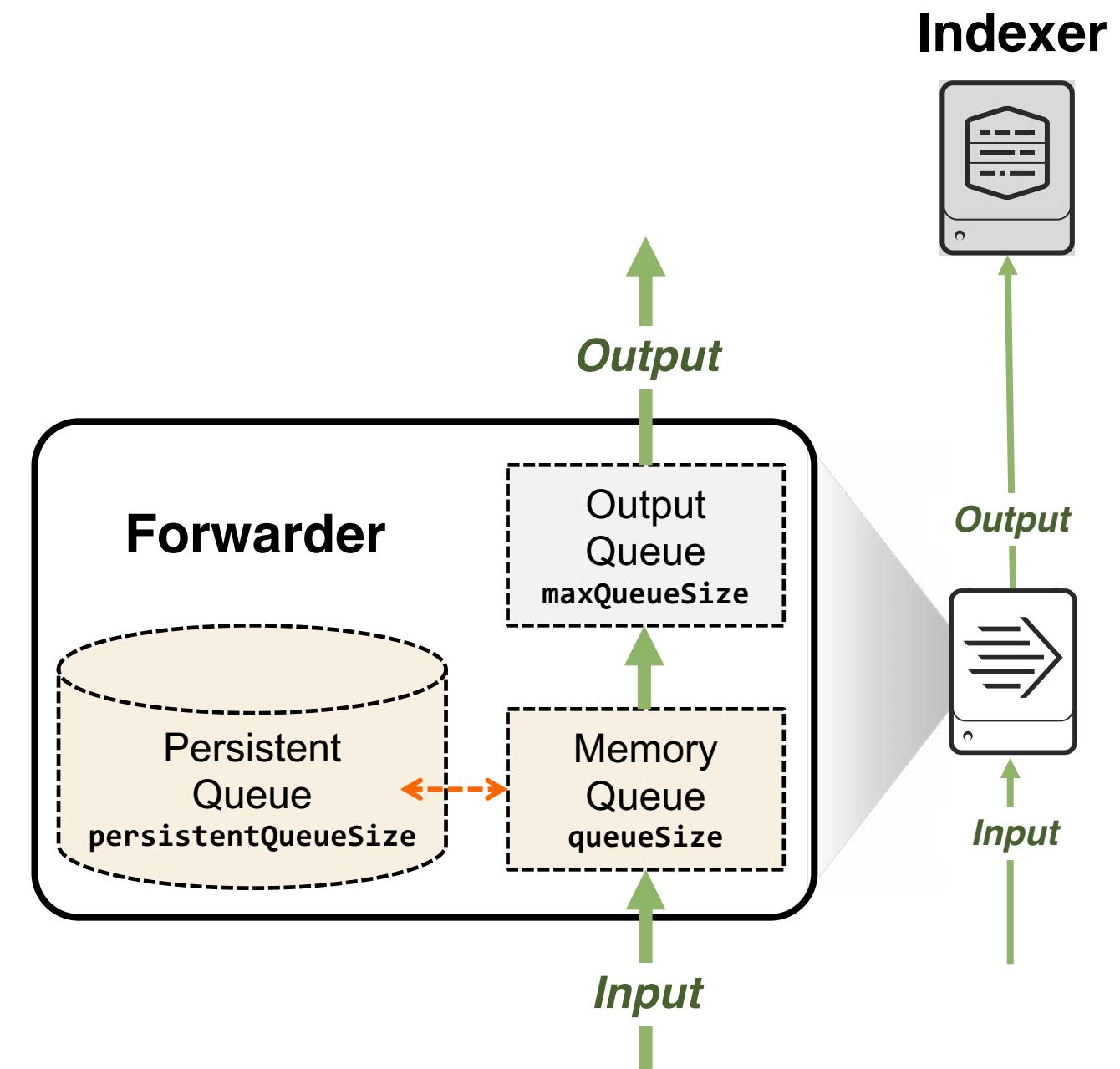
Network Input: Sender Filtering Options

- 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

- Queues provide input flow control
 - Applies to TCP, UDP, scripted input
 - Controls network data bursts, slow resources, or slow forwarding
 - If indexers can't be reached: Data is stored in the output queue
 - If the output queue is full: Data is stored in the memory queue
 - If the memory queue is full: Data is stored in the persistent queue on disk
- Persistent queue preserves across restarts
 - Not a solution for input failure



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Network Input: Setting Queue Attributes

- Memory queue
 - Set with **queueSize** attribute (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** attribute (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/syslog1/syslog.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

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

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 scripted input using the 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>
```

Use **passAuth** to run the script as the specified OS user – Splunk passes the auth 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 Splunk 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
```

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

Script Name

Command ?

Interval Input ?

Interval ? In Seconds

Cron Schedule

Source name override ?

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Editing Scripted Inputs

The image shows two screenshots of the Splunk web interface for managing data inputs.

Left Screenshot (List View):

- Header:** Script
- Breadcrumbs:** Data inputs » Script
- Text:** Showing 1-1 of 1 item
- Search Bar:** filter
- Table Headers:** Command, Interval, Last run, Status
- Data Row:** \$SPLUNK_HOME/etc/apps/_server_app_devserver_vmstat/bin/myvmstat.sh (highlighted with a green box)

Right Screenshot (Edit View):

- Header:** \$SPLUNK_HOME/etc/apps/_server_app_devserver_vmstat/bin/myvmstat.sh
- Breadcrumbs:** Data inputs » Script » \$SPLUNK_HOME/etc/apps/_server_app_devserver_vmstat/bin/myvmstat.sh
- Buttons:** New Remote Script
- Source Section:**
 - 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 Section:**
 - 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 Section:**
 - Host field value:** (Input field)
- Index Section:**
 - Set the destination index for this source:** (Input field)
 - Index:** (Dropdown menu)
 - default
 - history
 - itops
 - main (selected)
 - summary
 - test

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

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

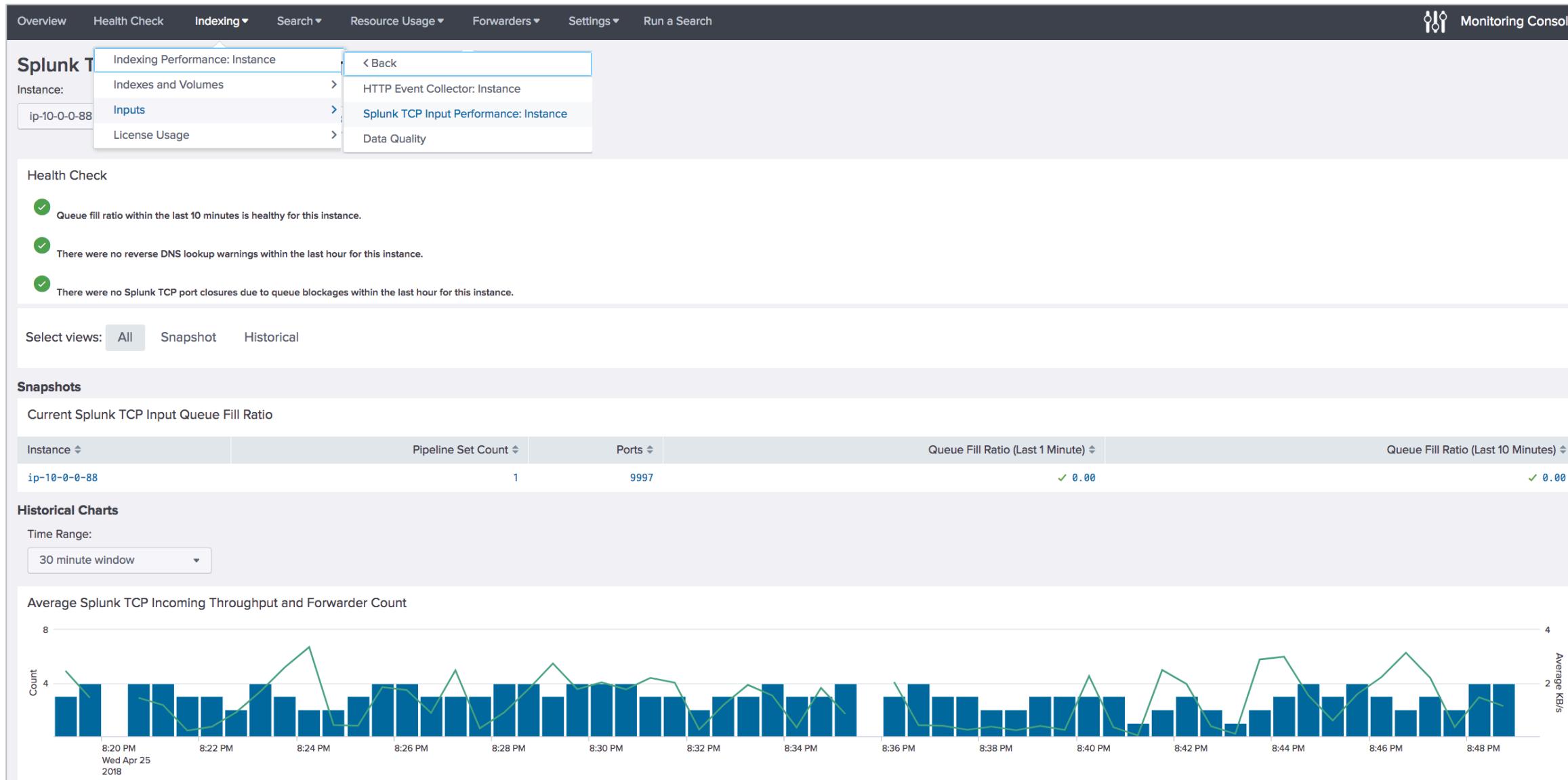
- Allows the use of Splunk's simple configuration of monitoring files
- Takes advantage of the file system and Splunk's robust file monitoring capabilities
- Can easily recover even when forwarder goes down
- Configured with a scripted log file:
 1. Schedule the script to run using an external scheduler (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

Monitoring with MC: Splunk TCP Inputs

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



Generated for Balaram Guttula (balaram.krishn.guttula@pmchase.com) (C) Splunk Inc, not for distribution

Module 6 Knowledge Check

- True or False. Persistent Queue and Memory Queue can be applied to Network as well as Scripted inputs.
- Why is it a Best Practice to send data to a syslog collector that writes into a directory structure and then have a UF/HF ingest the data from the directory structure?
- True or False. An interval setting for scripted inputs can be specified in cron syntax.
- Is it possible to use the host value and not the DNS name or IP address for a TCP input? How?

Module 6 Knowledge Check – Answers

- True or False. Persistent Queue and Memory Queue can be applied to Network as well as Scripted inputs.

True.

- Why is it a Best Practice to send data to a syslog collector that writes into a directory structure and then have a UF/HF ingest the data from the directory structure?

If the UF has to be restarted, the **_fishbucket** will prevent data loss.

- True or False. An interval setting for scripted inputs can be specified in cron syntax.

True. You can specify the interval in either number of seconds or cron syntax.

- Is it possible to use the host value and not the DNS name or IP address for a TCP input? How?

Yes, it is possible. Under the stanza in **inputs.conf** set the **connection_host** to none and specify the host value.

Module 6 Lab Exercise – Network Input

Time: 15 – 20 minutes

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
- Deploy the app to your forwarder

Lab Notes:

- Your instructor will run a script to send TCP data ports on the forwarder
- Use your assigned port to listen for the TCP data
- Deploy a remote scripted input

Module 7: Windows and Agentless Inputs

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Module Objectives

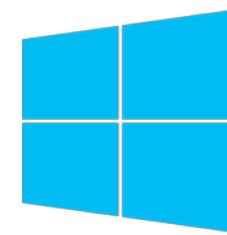
- Identify Windows specific **inputs.conf** stanzas and attributes
- Understand and configure Splunk HTTP Event Collector (HEC) agentless input
- Monitor HEC using MC (Monitoring Console)

Windows-Specific Inputs

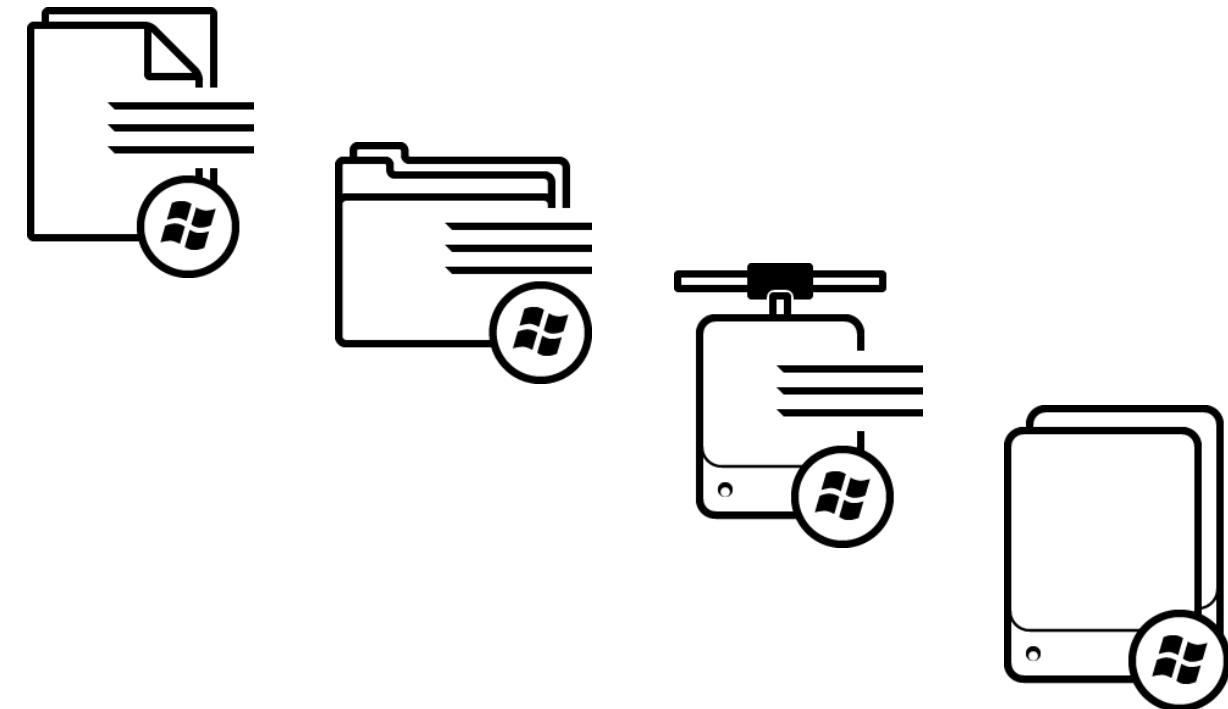


Windows-Specific Inputs

- 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



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Windows-Specific Input Types

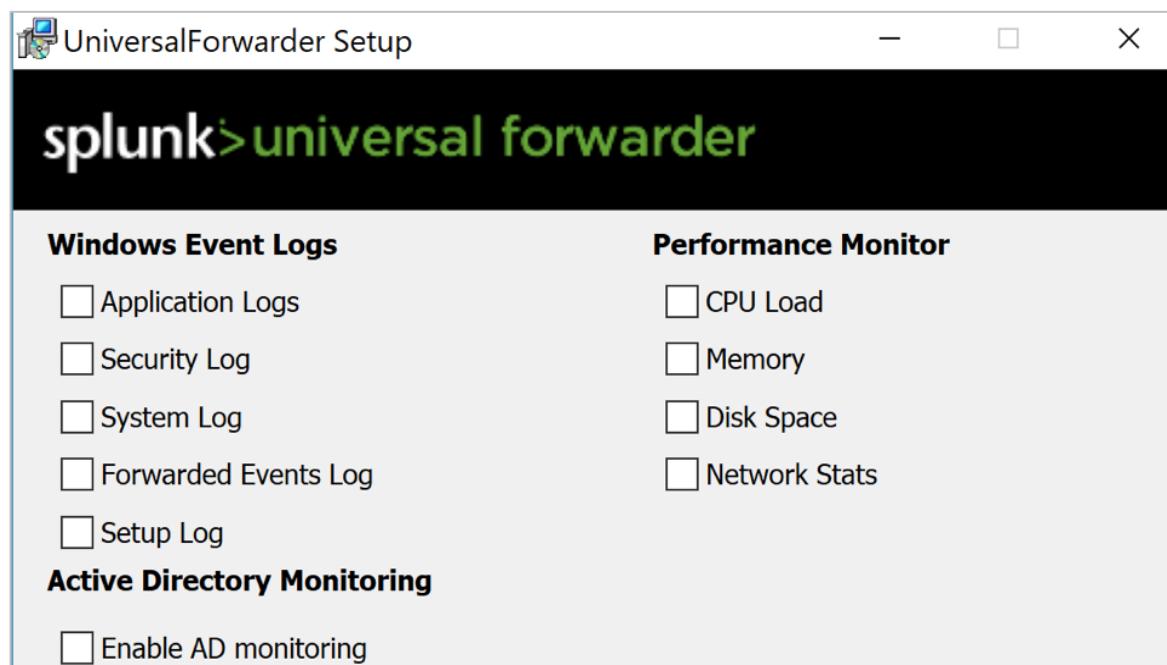
Input Type	Description
Event Log*	Consumes data from the Windows OS logs
Performance*	Consumes performance monitor data
Active Directory	Monitors changes in an Active Directory server
Registry	Monitors changes in a Windows registry
Host	Collects data about a Windows server
Network	Monitors network activity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

* Supports both local and remote (WMI) data collection

Local Windows Inputs Syntax

- Configure inputs during the Windows Forwarder installation
- Optionally, configure manually
 - See **inputs.conf.spec** and **inputs.conf.example** for details on setting up each Windows input type

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



Note i

While you can configure Windows inputs manually, Splunk recommends that you prepare the stanza using Splunk Web UI because it is easy to mistype the values for event log channels.

Windows Inputs: Using the Manager UI

Add Data  [Select Source](#) [Input Settings](#) [Review](#) [Done](#) [< Back](#) [Next >](#)

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.

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

Configure this instance to monitor local Windows Event Log channels where installed applications, services, and system processes send data. This monitor runs once for every Event Log input that you define. [Learn More](#)

Select Event Logs Available item(s) [add all »](#) Selected items

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

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

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Windows Input Filtering Options

- Filter out non-essential events using whitelist and blacklist
 - Provides ability to keep specific events and filter out low-value events
 - Configure up to 10 whitelist and 10 blacklist per stanza
 - Set whitelist and blacklist based on event field names and regex:
 - **whitelist = <List> | key=regex [key=regex]**
 - **blacklist = <List> | key=regex [key=regex]**
 - In case of a conflict, the blacklist prevails

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

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Windows Remote Inputs With WMI

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

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Configuring WMI Inputs

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

```
[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

The screenshot shows the 'Add Data' configuration page for a PowerShell input. On the left, a sidebar lists various monitoring options like TCP / UDP, Remote Performance Monitoring, Registry monitoring, Active Directory monitoring, Local Windows host monitoring, Local Windows network monitoring, Local Windows print monitoring, Scripts, and Powershell v3 Modular Input. The main panel is titled 'PowerShell v1 or v3' and contains fields for 'name' (set to 'RunningProcesses'), 'Command or Script Path' (empty), 'Cron Schedule' (set to '/10 * * * *'), 'More settings' (checkbox checked), 'Source type' (set to 'Automatic'), 'Host' (set to 'splunk01'), and 'Index' (set to 'default'). A large black callout box at the bottom right contains the PowerShell command template: [powershell://<name>] script = <command> schedule = [<number>]<>|<cron>].

PowerShell v1 or v3

Command or a script file

Blank field executes once only

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

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Windows Inputs Resources

- About Windows data

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

- General information about event log

<https://docs.microsoft.com/en-us/windows/desktop/wes/windows-event-log>

- Performance Counters Portal

<https://docs.microsoft.com/en-us/windows/desktop/PerfCtrs/performance-counters-portal>

- Performance Counters Reference

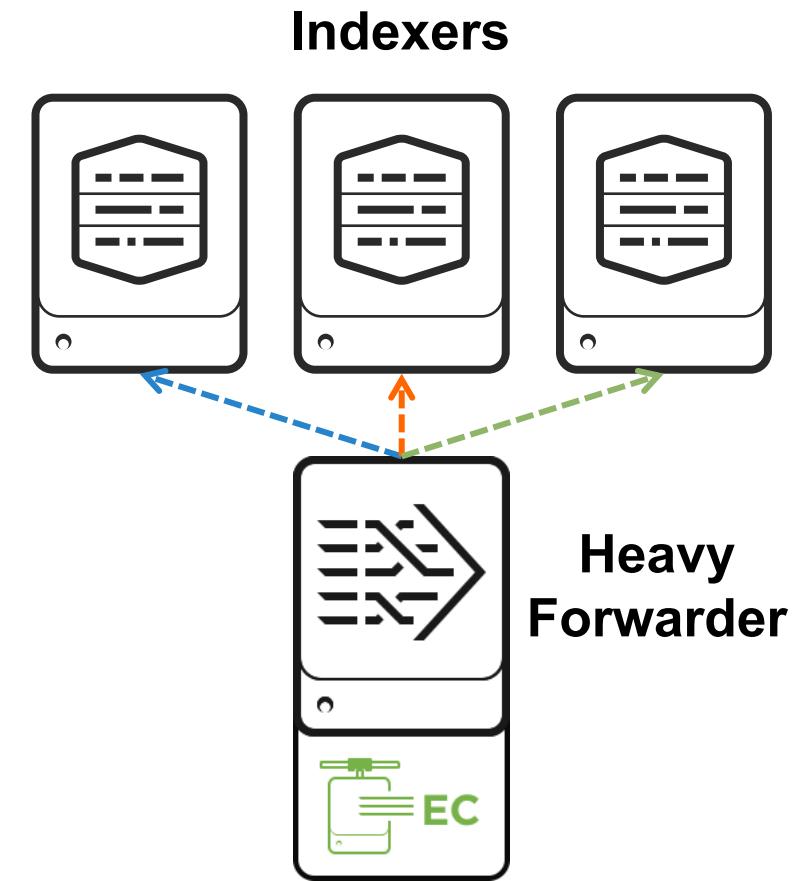
<https://docs.microsoft.com/en-us/windows/desktop/PerfCtrs/performance-counters-reference>

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

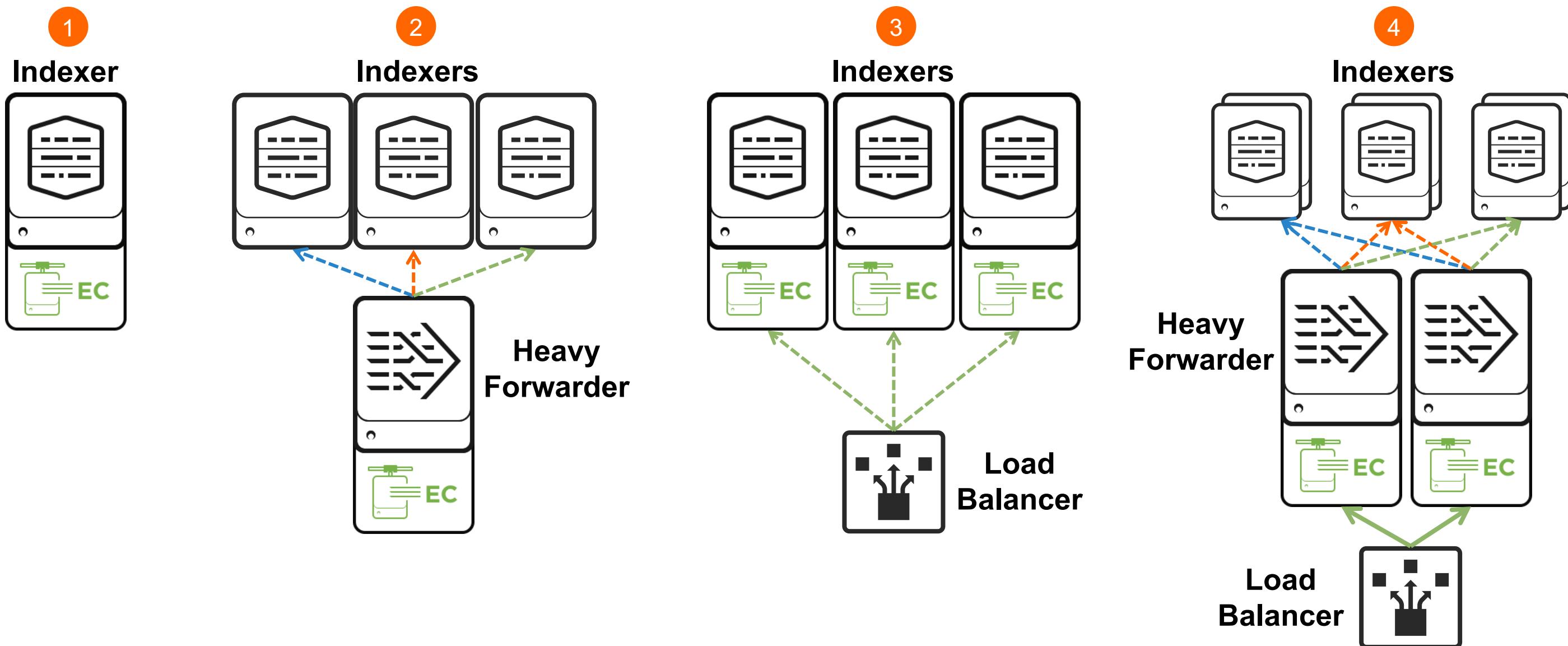


Event collector enabled
to receive HTTP events

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Distributed HEC Deployment Options

HEC can scale by taking advantage of Splunk distributed deployment



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

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 toolbar has buttons for 'Global Settings' (disabled) and 'New Token' (highlighted with a red circle labeled '1'). The main area shows a table with one token named 'iot_sensors' (highlighted with a green box labeled '2'). The table columns are 'Name', 'Actions', 'Token Value', 'Source Type', 'Index', and 'Status'. The token value 'af58d9a4-4df6-4fda-a209-1c3988e1ceaf' is also highlighted with a green box labeled '2'. The status is listed as 'Disabled'.

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

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

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 (which support 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"}
}'
```

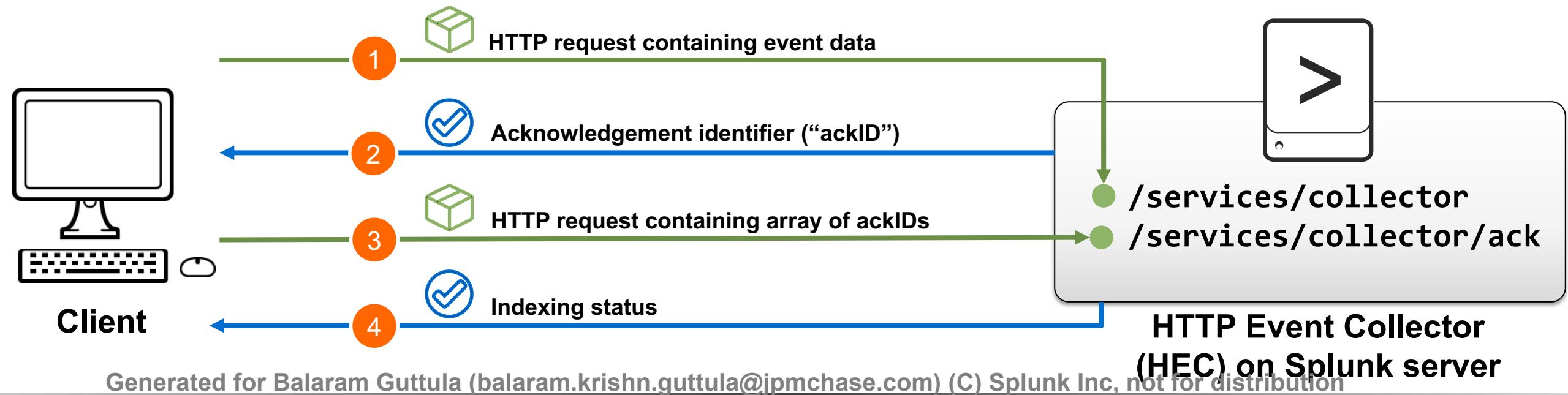
Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

HTTP Event Collector Options

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

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



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

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
- Once an **ACK** has been received, it is released from memory
- Client polling functionality is not built into Splunk and requires custom programming

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

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

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

Generated for Balaram Guttula (balaram.krishn.guttula@pmchase.com) (C) Splunk Inc, not for distribution

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

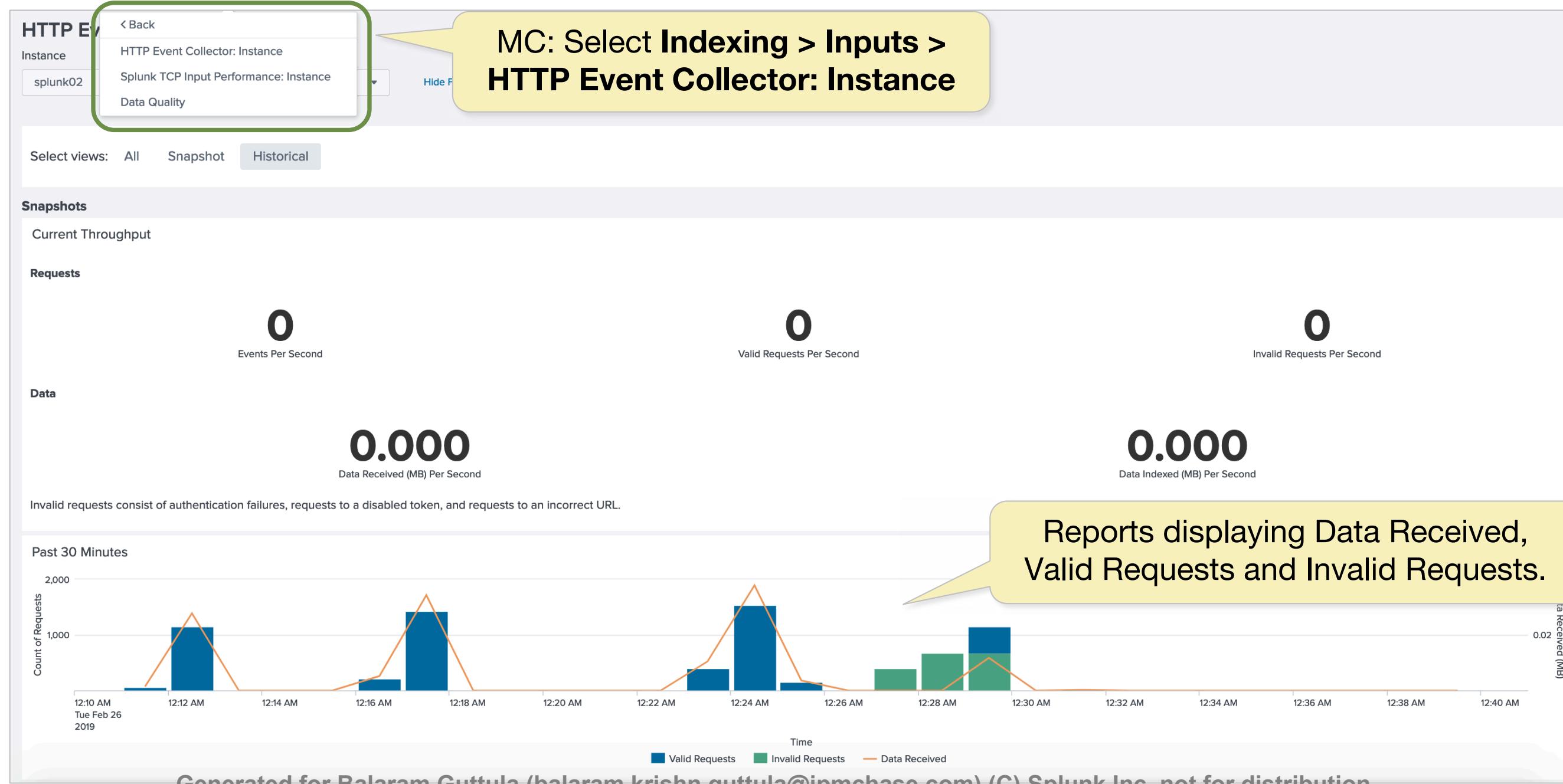
- Example:
 - Splunk admins want to limit who can access the HEC endpoints
- Solution:
 - Manually add the dedicated server settings in **inputs.conf**
- Configuration Notes:
 - Available attributes under the **[http]** stanza
 - Configure a specific SSL cert for HEC and client certs
 - Enable cross-origin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

inputs.conf

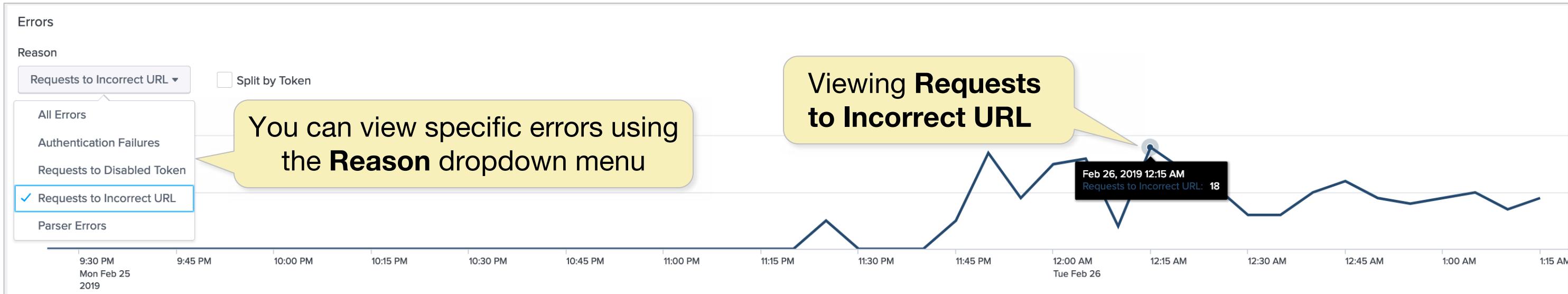
```
[http]
enableSSL = 1
crossOriginSharingPolicy = *.splunk.com
acceptFrom = "!45.42.151/24, !57.73.224/19, *"
```

Generated for Balaram Guttula (balaram.krishn.guttula@pmchase.com) (C) Splunk Inc, not for distribution

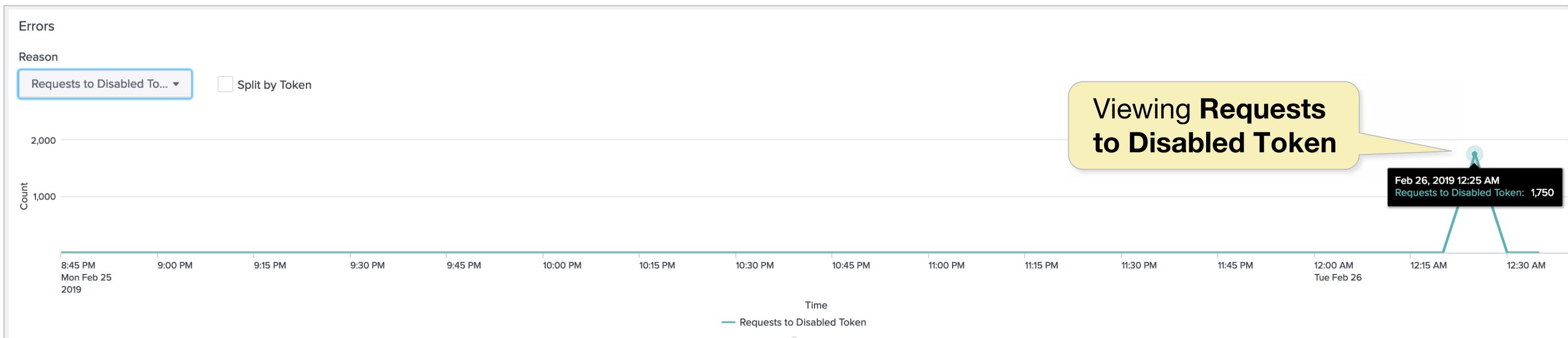
Monitoring HEC with MC



Monitoring HEC with MC – Viewing Errors



Viewing Requests to Incorrect URL



Viewing Requests to Disabled Token

Generated for Balaram Guttula (balaram.krishn.guttula@jmpchase.com) (C) Splunk Inc, not for distribution

HTTP Event Collector (HEC) documentation

- Refer to:
 - Introduction to Splunk HTTP Event Collector
dev.splunk.com/view/event-collector/SP-CAAAE6M
 - Blogs: Tips & Tricks on HTTP Event Collector
blogs.splunk.com/2015/10/06/http-event-collector-your-direct-event-pipe-to-splunk-6-3

Module 7 Knowledge Check

- True or False. You can set up a windows input using a UF on the windows server and send the data to an Indexer running on Linux.
- True or False. You can collect Active Directory data from a Windows Server remotely using **wmi.conf**.
- True or False. Event Collector can be set up on a UF.
- True or False. Data can be sent in json or any raw data format to the event collector.

Module 7 Knowledge Check – Answers

- True or False. You can set up a windows input using a UF on the windows server and send the data to an Indexer running on Linux.
True.
- True or False. You can collect Active Directory data from a Windows Server remotely using **wmi.conf**.
False. Only event logs and performance monitoring logs can be collected using wmi.conf.
- True or False. Event Collector can be set up on a UF.
False. Event collector can be set up on an Indexer or HF.
- True or False. Data can be sent in json or any raw data format to the event collector.
True.

Module 7 Lab Exercise – HTTP Event Collector

Time: 10 – 15 minutes

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 8: Fine-tuning Inputs

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

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

Testing New Inputs

- Use Data Preview
 - Evaluate new data sources without actually inputting them
- Create a test index
 - Send test inputs to this index
 - Delete index and start again as desired
 - Does not require **splunkd** restart
- Create a Splunk test environment
 - Use a laptop, virtual machine, or spare server
 - Send test inputs for new sources and evaluate, prior to production
 - Should be running same version of Splunk as in production

Things to Get Right at Index Time

Input phase

- Host
- Source type
- Source
- Index

Parsing phase

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

* Optional

What if I Don't Get It Right?

- On a testing / development system
 - This is what test/dev Splunk setups are for!
 - Clean or delete/recreate the test index, change your configurations, and try again
 - May need to reset the fishbucket
- On a production server
 - Leave the 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

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]  
attribute = value
```

example

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

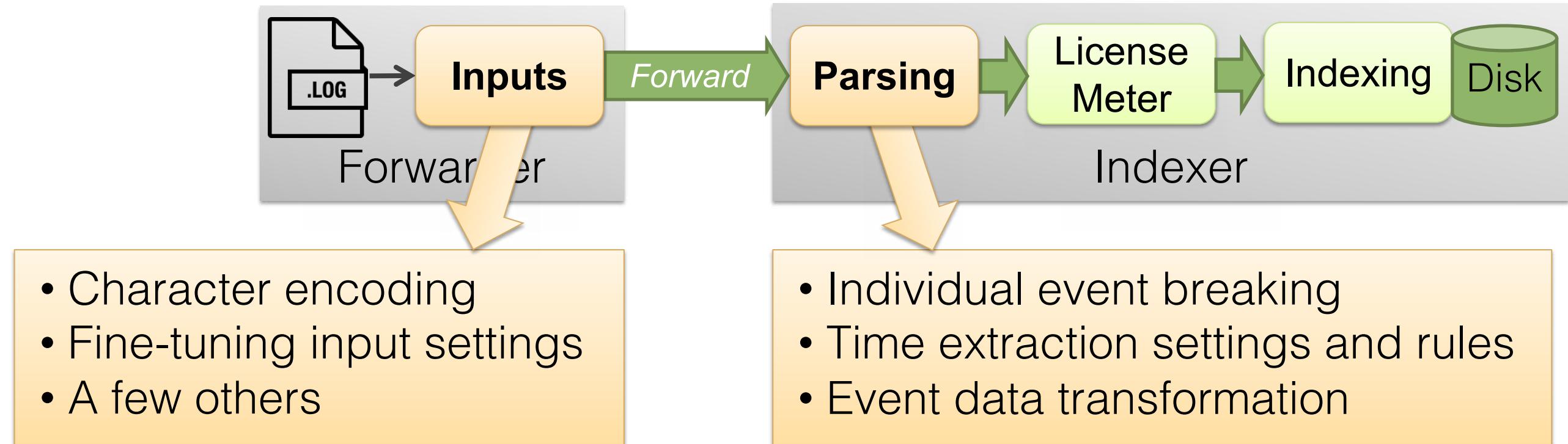
```
[host::nyc*]  
TZ = US/Eastern
```

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

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

Input and Parsing Phases and **props.conf**

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



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

[wiki.splunk.com/Where do I configure my Splunk settings](https://wiki.splunk.com/Where_do_I_configure_my_Splunk_settings)

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**

Module 8 Knowledge Check

- ❑ In the **props.conf** example below, what is **sendmail**?

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

- ❑ Examine the **props.conf** example below. Is this an acceptable format for the stanzas?

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

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

Module 8 Knowledge Check – Answers

- ❑ In the **props.conf** example below, what is **sendmail**?

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

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

- ❑ Examine the **props.conf** example below. Is this an acceptable format for the stanzas?

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

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

No. You cannot use a wildcard with source types in **props.conf**.

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Module 8 Lab Exercise – Fine-Tuning Inputs

Time: 10 – 15 minutes

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 9: Parsing Phase and Data Preview

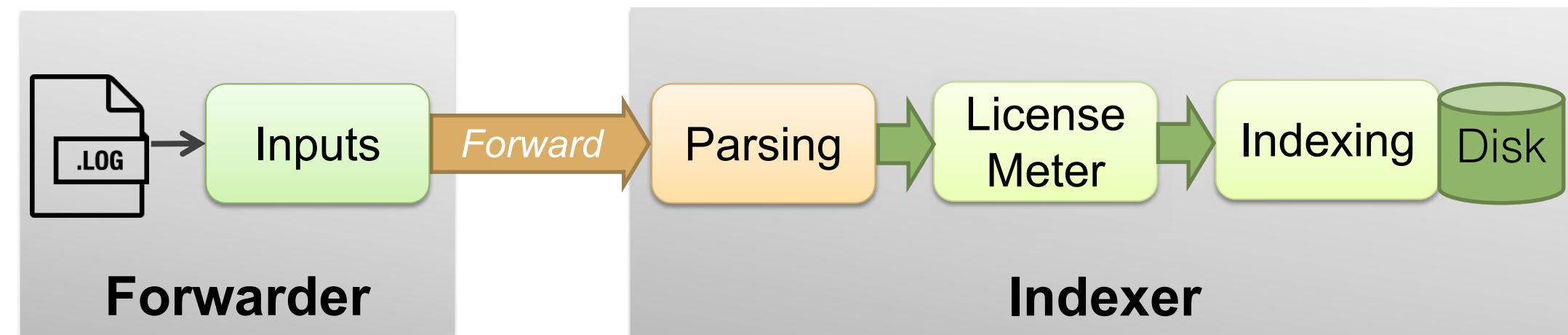
Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

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

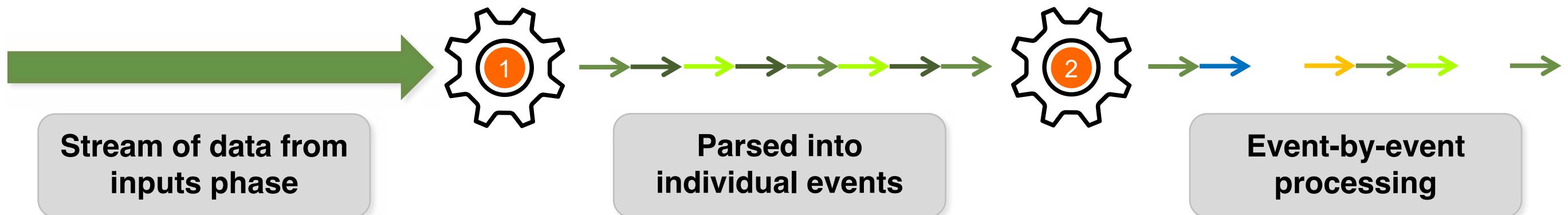
- Occurs as data arrives at the indexer (or heavy forwarder)
- Breaks up input data stream into discrete **events**, each with a timestamp and time zone
- Creates, modifies, and redirects events
 - Applies additional transformation steps to modify the metadata fields or modify raw data



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

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

- Performed in two steps:
 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]+)**

2. Line merging (optional)
 - Takes separate lines and merges them 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**)
 - Disabled if set to **false**, which improves performance

docs.splunk.com/Documentation/Splunk/latest/Data/Configureeventlinebreaking
Generated for Balaram Guttula (balaram.krishn.guttula@pmchase.com) (C) Splunk Inc, not for distribution

Event Boundary Examples

Monitored input: Single line input with 3 events

```
[19/Sep/2019:18:22:32] VendorID=7033 Code=E AcctID=4390644811207834 ↵
[19/Sep/2019:18:22:48] VendorID=1239 Code=K AcctID=5822351159954740 ↵
[19/Sep/2019:18:22:59] 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" = "2014-08-21 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
```

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Date/timestamp Extraction

- Correct date/timestamp extraction is essential
- 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
- Splunk works well with standard date/time format and well-known data types
- Custom timestamp extraction is specified in **props.conf**

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
 - ▶ Example data with "date-like" code at the start of the line

1989/12/31 16:00:00 ed May 23 15:40:21 2015 ERROR UserManager - Exception thrown

Start looking here for date/timestamp

props.conf

[*my_custom_source_or_sourcetype*]

TIME_PREFIX = \d{4}/\d{2}/\d{2} \d{2}:\d{2}:\d{2} \w+\s

Using **MAX_TIMESTAMP_LOOKAHEAD**

- Syntax: **MAX_TIMESTAMP_LOOKAHEAD = <integer>**
- Specifies how many characters to look beyond the start of the line for a timestamp
 - Works in conjunction with **TIME_PREFIX**
 - If set, it starts counting from the point the **TIME_PREFIX** indicates Splunk should start looking for the date/timestamp
 - Improves efficiency of timestamp extraction
 - The complete timestamp string must be present within the specified range

Using **TIME_FORMAT**

- Syntax: **TIME_FORMAT** = <*strftime-style format*>
- Specifies the format of the timestamp using a **strftime()** expression
 - For example, **2015-12-31** would be **%Y-%m-%d**
- 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

Setting Time Zones – Splunk's 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
 3. If a forwarder is used, the forwarder-provided time zone is used
 - ▶ en.wikipedia.org/wiki/List_of_zoneinfo_timezones
 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>

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

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 your sandbox environment or test index to perfect your settings before taking a new data input into the production environment
- Use Data Preview configuration settings to create new source types

Previewing Unstructured Data

The screenshot shows the Splunk interface for previewing unstructured data. On the left, a detailed log entry is displayed with various fields highlighted by green boxes and numbered 1 and 2. A yellow callout box points to the event preview area on the right.

Log Entry (Number 1):

```
[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: [0xF0097000009A300]
RBP: [0x0000000000002000]
RSP: [0x004B000000000D00]
RAX: [0x00042000010D0000]
RBX: [0x3005000000100000]
RCX: [0xE0E00000C010000]
RDX: [0x0000000A00000C00]
EFL: [0x00000000000020000]

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.buttercupgames.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 Preview (Number 2):

Splunk makes its best attempt to identify event boundaries and timestamps; however, if you are more familiar with the data, provide more info

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 5898. Crashing thread: Main Thread Show all 25 lines
2 8/24/18 1:07:11.000 AM	Linux /usr13.eng.buttercupgames.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

Generated for Balaram Guttula (balaram.krishn.guttula@pmchase.com) (C) Splunk Inc, not for distribution

Previewing Unstructured Data (cont.)

Add Data < Back Next >

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/crash-2019-03-06-00_46_26.log [View Event Summary](#)

Source type: Select Source Type [Save As](#)

> Event Breaks
✓ Timestamp

Determine how timestamps for the incoming data are defined.

Extraction Auto Curr... Adva... 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

Time Event

1 3/6/19 [167154] 2019-03-06 00:46:26
12:46:26.000 AM Received fatal signal 6 (Aborted).
Cause:
Signal sent by PID 6241 running under UID 5898.
Crashing thread: Main Thread
[Show all 45 lines](#)

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Previewing Unstructured Data (cont.)

Another example – previewing **XML** file as unstructured data

The screenshot shows the 'Add Data' wizard with five steps: 'Select Source' (green dot), 'Set Source Type' (green dot), 'Input Settings' (grey dot), 'Review' (grey dot), and 'Done' (grey dot). The current step is 'Set Source Type'. The page title is 'Set Source Type'. A sub-section title 'Event Preview' is present. A tooltip with the following text is overlaid on the interface:

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

Event Preview

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

Source type: default ▾

Save As

List ▾

1

2

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.

3/6/18 8:16:05.000 PM

3/6/18 8:16:05.000 PM

Sebastiano Jiménez,

View Event Summary

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

Generated for Balaram Guttula (balaram.krishn.guttula@pmchase.com) (C) Splunk Inc, not for distribution

Previewing Unstructured Data (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/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 regex that determines how the raw text stream is broken into initial events, before line breaking takes place.
Sets SHOULD_LINEMERGE = false and BREAKER to the user-provided regular expression.
Results to `([\r\n]+)`, meaning data is broken into an event for each line, delimited by any number of carriage return or newline characters.
The regex must contain a capturing group -- a pair of parentheses which defines an identified subcomponent 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

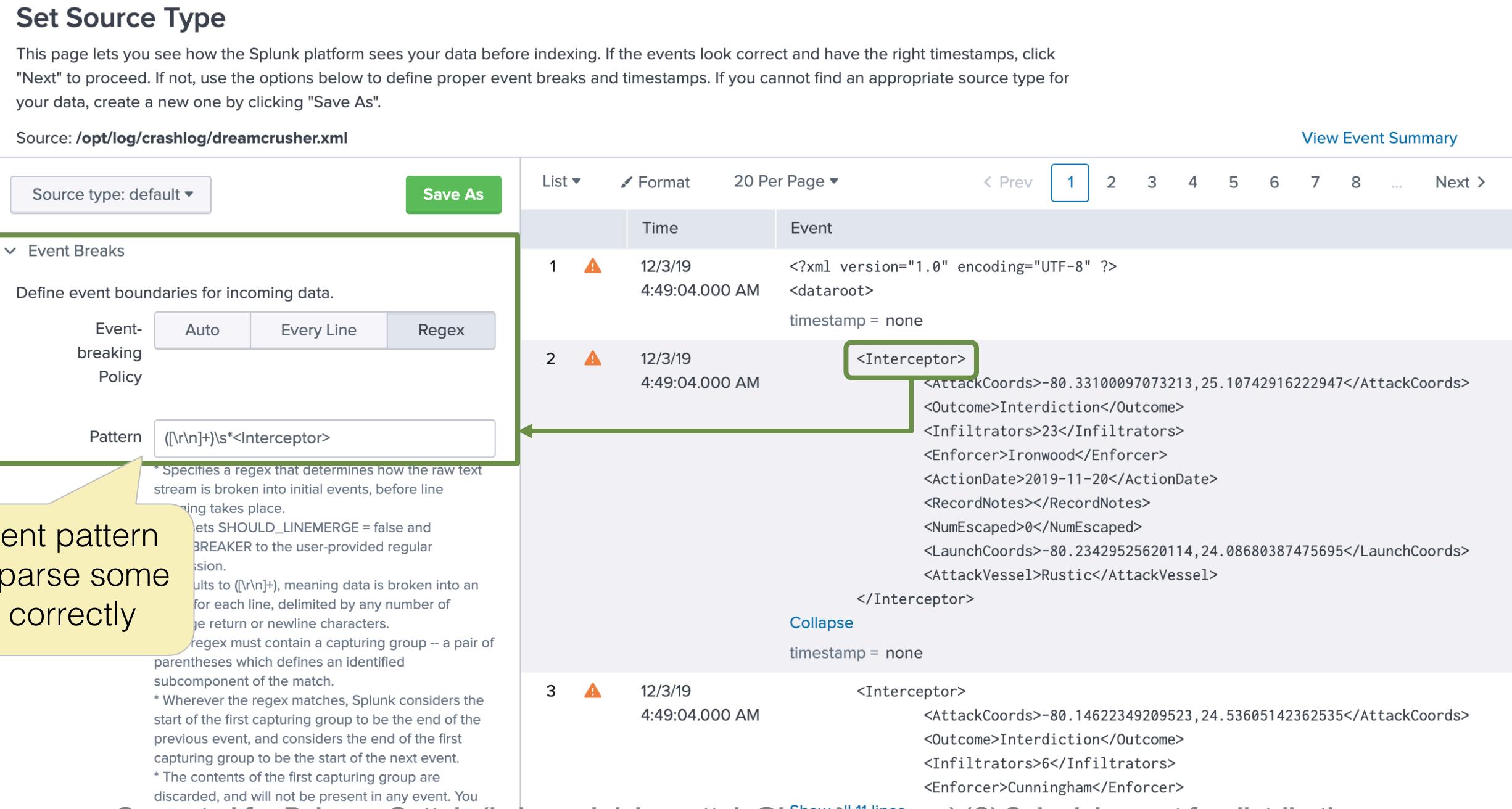
List ▾ Format 20 Per Page ▾

< 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>-80.14622349209523,24.53605142362535</AttackCoords> <Outcome>Interdiction</Outcome> <Infiltrators>6</Infiltrators> <Enforcer>Cunningham</Enforcer>

Show all 11 lines

Enter event pattern prefix to parse some events correctly



Generated for Balaram Guttula (balaram.krishn.guttula@pmchase.com) (C) Splunk Inc, not for distribution

splunk® turn data into doing™

210

Splunk Enterprise Data Administration
Copyright © 2020 Splunk, Inc. All rights reserved | 29 May 2020

Previewing Unstructured Data (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/crashlog/dreamcrusher.xml

Source type: default List ▾

> 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.

Click **Timestamp > Advanced** to access time zone, timestamp prefix, and timestamp definitions needed to extract the correct time from the data

#	Date	Time	Event Data
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>
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>
4	11/14/19	12:00:00.000 AM	<Interceptor> <AttackCoords>-80.75496221688965,24.72483828554483</AttackCoords> <Outcome>Interdiction</Outcome>

[Collapse](#) [Show all 11 lines](#)

Generated for Balaram Guttula (balaram.krishn.guttula@jmpchase.com) (C) Splunk Inc, not for distribution

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 ▾ (arrow)

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: Comma-separated list of field names

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

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

splunk® turn data into doing™

212

Splunk Enterprise Data Administration
Copyright © 2020 Splunk, Inc. All rights reserved | 29 May 2020

Saving New Source Type

The screenshot shows the configuration of a new source type in Splunk. On the left, the 'Advanced' settings for '_json' are displayed, including fields like CHARSET (UTF-8), INDEXED_EXTRACTI (json), KV_MODE (none), SHOULD_LINEMERG (false), category (Structured), description (JavaScript Object Notation for), disabled (false), pulldown_type (true), and TIMESTAMP_FIELDS (time). A green box highlights the 'Save As' button. An arrow points from this button to the 'Save Source Type' dialog on the right.

Save Source Type Dialog:

- Name: geojson
- Description: JavaScript Object Notation format. For more information, visit <http://json.org/>
- Category: Application
- App: Search & Reporting

A green box highlights the 'Save' button, which is also targeted by an arrow from the 'Save As' button. Below the dialog, a text area contains the props.conf text for the new source type:

```
[_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

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Source Type Manager

Click **Settings > Source types** to view and access the configured source types 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 source types you create
can be edited, deleted, and cloned

Generated for Balaram Guttula (balaram.krishn.guttula@jmpchase.com) (C) Splunk Inc, not for distribution

Module 9 Knowledge Check

- True or False. Time extraction can be done using **props.conf** on the UF and the HF.
- True or False. Event boundaries can be defined using **props.conf** at the UF.
- True or False. When extracting a timestamp, if the parser finds the indexer's OS time, it will use that as the first preference.

Module 9 Knowledge Check – Answers

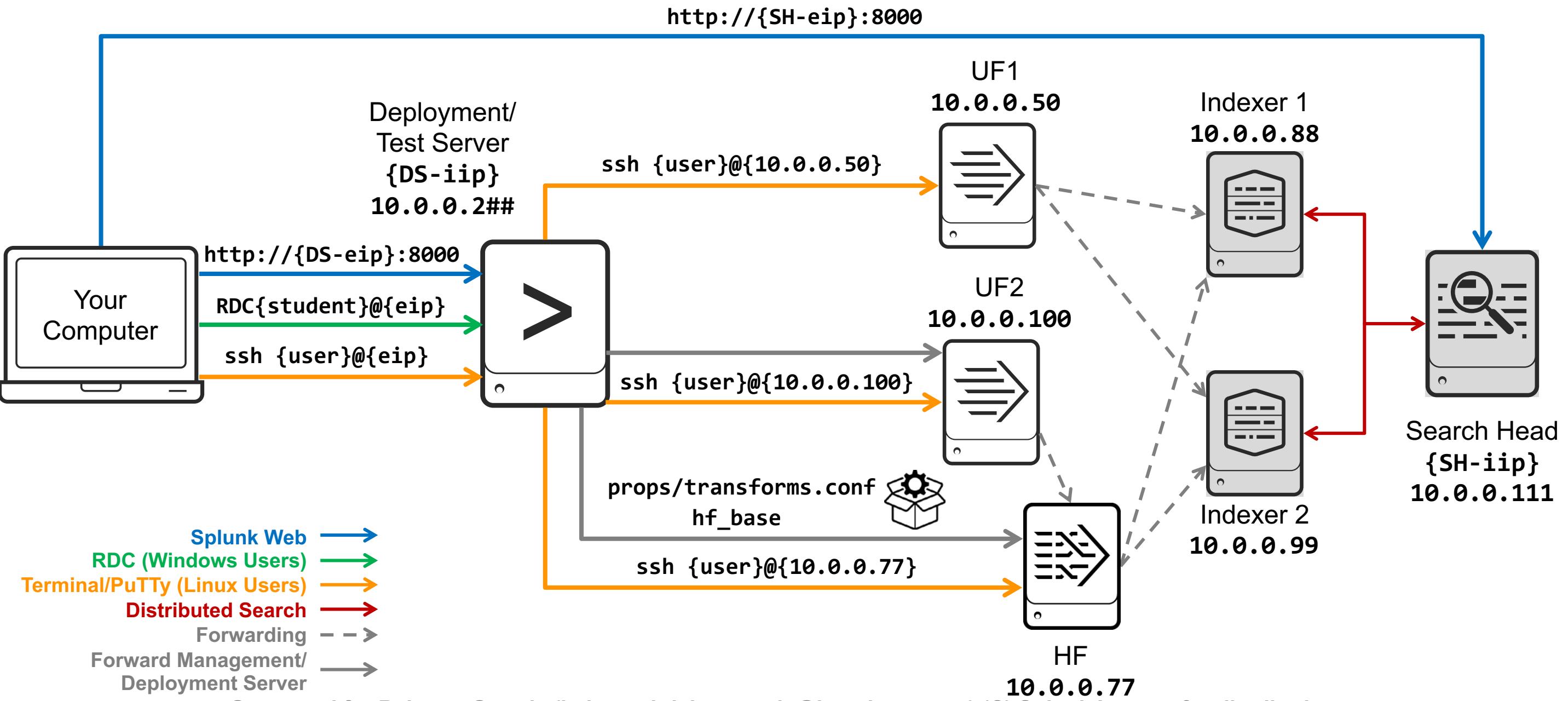
- True or False. Time extraction can be done using **props.conf** on the UF and the HF.

False. You will learn how to specify Time Extraction if the file contains a header line. But if it does not contain a header line, then time has to be extracted on the HF/ Indexer.
- True or False. Event boundaries can be defined using **props.conf** at the UF.

True. You may want to define event boundaries for certain event types at the UF level. Remember the more you do at the UF level, the more resources you will need.
- True or False. When extracting a timestamp, if the parser finds the indexer's OS time, it will use that as the first preference.

False. When all else fails, the Indexer's OS time is used as the *last* preference.

Module 9 Lab Exercise – Environment Diagram



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Module 9 Lab Exercise – Create a New Source Type

Time: 20 – 25 minutes

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 10:

Manipulating Raw Data

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Module Objectives

- Explain how data transformations are defined and invoked
- Use transformations with **props.conf** and **transforms.conf** to:
 - Mask or delete raw data as it is being indexed
 - Override sourcetype or host based upon event values
 - Route events to specific indexes based on event content
 - Prevent unwanted events from being indexed
- Use **SEDCMD** to modify raw data

Modifying the Raw 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: Routing all events to the **web** index, except credit card transactions which are sent to the **credits** index

Should be performed with 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

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Splunk Transformation Methods

- When possible, define meta field values during the input phase
 - Most efficient to use **inputs.conf**
- Splunk provides two methods of raw data transformations:

SEDCMD

- Uses only **props.conf**
- Only used to mask or truncate raw data

TRANSFORMS

- Uses **props.conf** and **transforms.conf**
- More flexible
- Transforms matching events based on source, source type, or host

Using SEDCMD

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

```
[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
```

vendor_sales.log

Replace with **AcctID=xxxxx99288**

```
[source::.../vendor_sales.log]
```

```
SEDCMD-1acct = s/AcctID=\d{5}(\d{5})/AcctID=xxxxx\1/g
```

props.conf

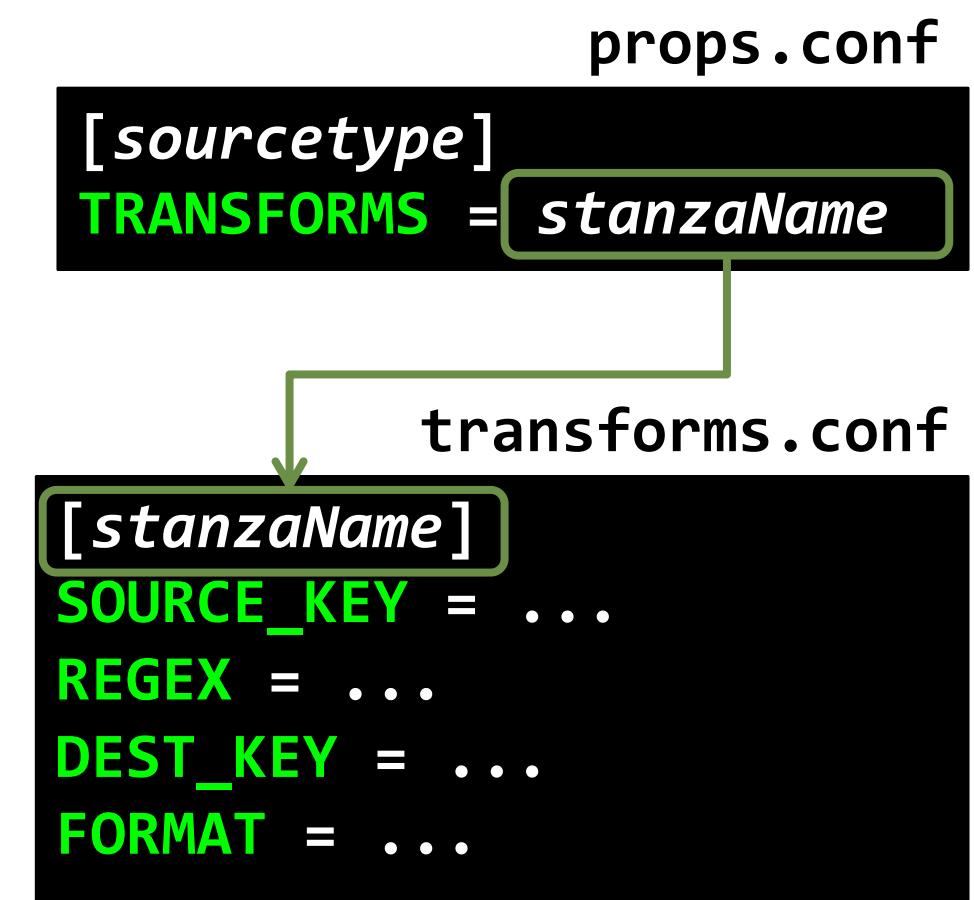
- For more examples, see:

Indicates the capture group

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

Using TRANSFORMS

- Per event transformation is based on REGEX pattern matches
- Define the transformation in **transforms.conf**
- Invoke the transformation from **props.conf**
- Transformation is based on the following attributes:
 - **SOURCE_KEY** indicates which data stream to use as the source for pattern matching (default: `_raw`)
 - **REGEX** identifies the events from the **SOURCE_KEY** that will be processed (required)
 - Optionally specifies regex capture groups
 - **DEST_KEY** indicates where to write the processed data (required)
 - **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 = $1xxxxxxxxxxxxx$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.
- This 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: xxxxxxxxxxxxxxxx4534 Code=B  
[22/Apr/2014:00:48:40] Sent to checkout TransactionID=100763  
[22/Apr/2014:00:50:02] VendorID=5034 CC_Num: xxxxxxxxxxxxxxxx8091 Code=H
```

Generated for Balaram Guttula (balaram.krishn.guttula@jmpchase.com) (C) Splunk Inc, not for distribution

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::**

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Per-Event Index Routing

Again, if at all 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

- You can 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

You can 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
```

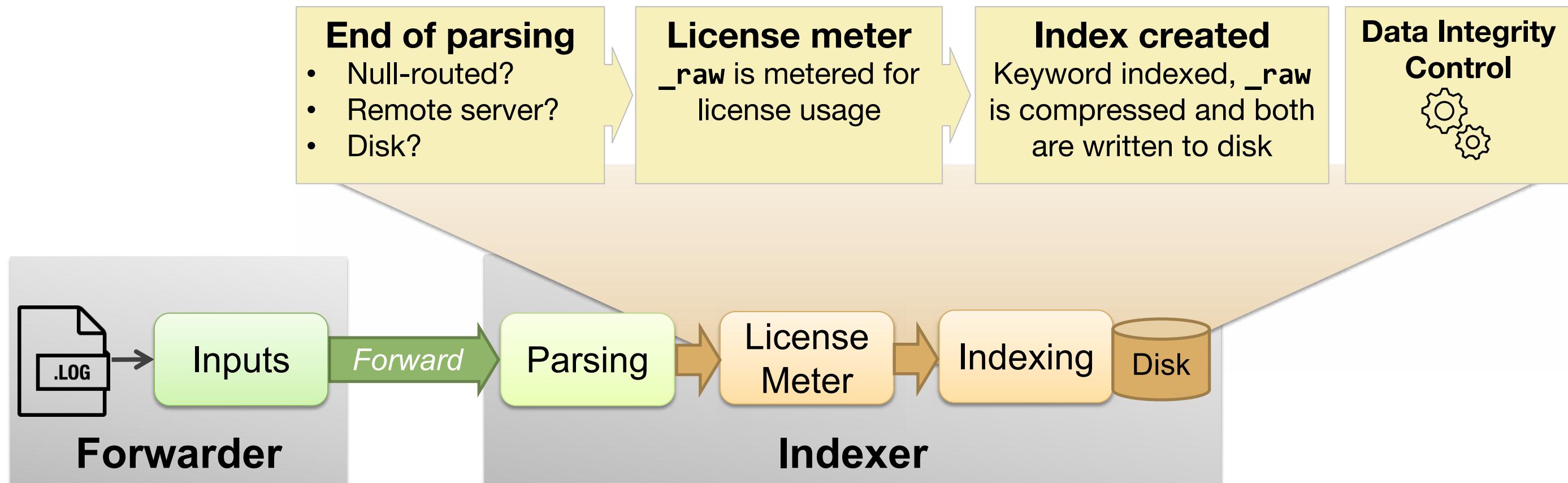
```
[tcpout:syslogGroup]
server=10.1.1.197:9996,10.1.1.198:9997
```

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

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Indexing Phase Details

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



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

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

Module 10 Knowledge Check

- True or False. **sedcmd** can be used to eliminate unwanted events.
- True or False. When using **transforms.conf**, the **SOURCE_KEY** is set to **_raw** by default.
- In the **props.conf** file example below, what is **itops**?

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

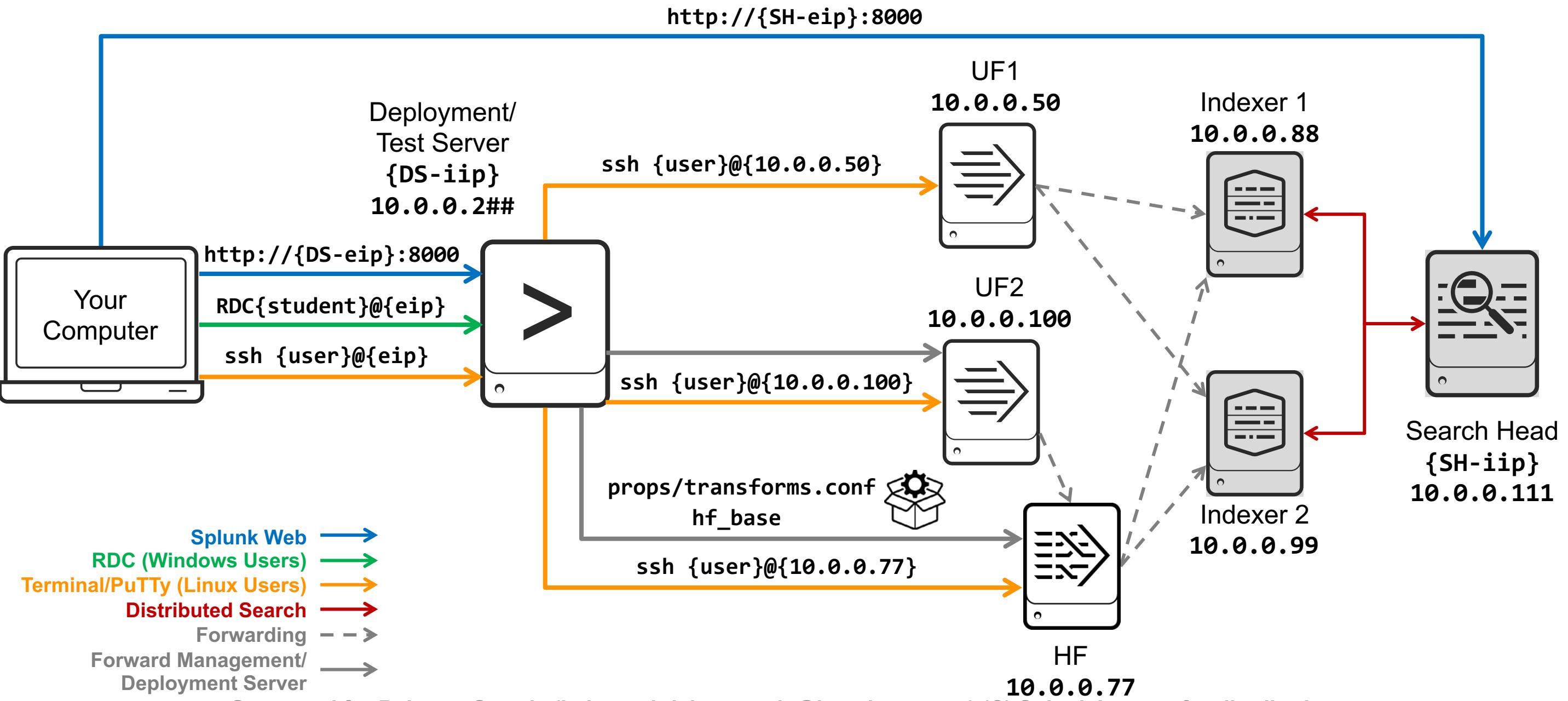
Module 10 Knowledge Check – Answers

- ❑ True or False. **sedcmd** can be used to eliminate unwanted events.
False. You have to use **transforms.conf**. **sedcmd** can only be used to mask or truncate data.
- ❑ True or False. When using **transforms.conf**, the **SOURCE_KEY** is set to **_raw** by default.
True. If you do not specify the **SOURCE_KEY** in **transforms.conf**, it defaults to **_raw**.
- ❑ In the **props.conf** file example below, what is **itops**?

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

Itops is the namespace and is used to determine the sequence.

Module 10 Lab Exercise – Environment Diagram



Module 10 Lab Exercise – Manipulating Data

Time: 10 - 15 minutes (20 - 25 minutes with optional lab)

Tasks:

- Use **transforms.conf** to:
 - Mask sensitive data
 - Redirect events to specific indexes
 - Drop unwanted events
- (Optional lab) Use **props.conf** to:
 - Drop in sequence the filtering and redirecting events

Module 11:

Supporting Knowledge

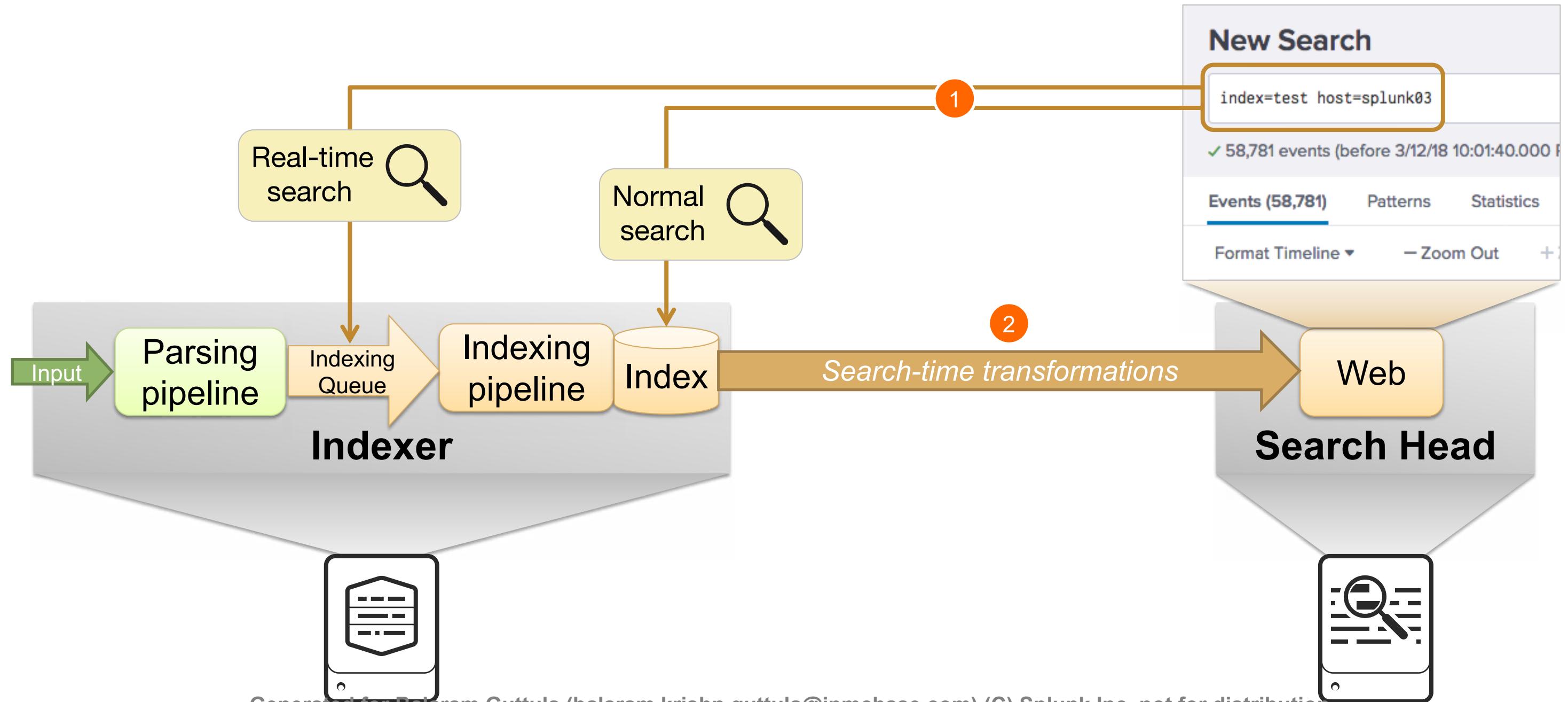
Objects

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Module Objectives

- Define default and custom search time field extractions
- Discuss 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



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Index-Time Field Extraction

- 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 data admin)
- Fields are generally extracted at search-time
- Certain use cases perform field extraction at index-time
 - On the forwarder for structured inputs
 - On the indexer for fields that may be negatively impacting search performance
- Add index-time custom fields only if necessary
 - Can negatively impact indexing performance and search times
 - Increases the size of the searchable index

Pros/Cons of Index Time Field Extractions

PROs	CONs
<ul style="list-style-type: none">Provision the extraction during the input or parsing phaseCan configure on the universal forwarderAuto-formattingCan 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 casesPossible performance implicationsLess 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

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_EXTRATIONS = 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	w3c
MAX_TIMESTAMP_L	32
SHOULD_LINEMERG	false
category	Web
description	W3C Extended log format pro
disabled	false
pulldown_type	true

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

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>

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

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
 - *nix app: has many search time fields for standard UNIX logs
 - ▶ For example, **secure.log**, **messages.log**, etc.
 - Windows app: has many defaults for Windows data
 - For other data: look for an app on <http://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

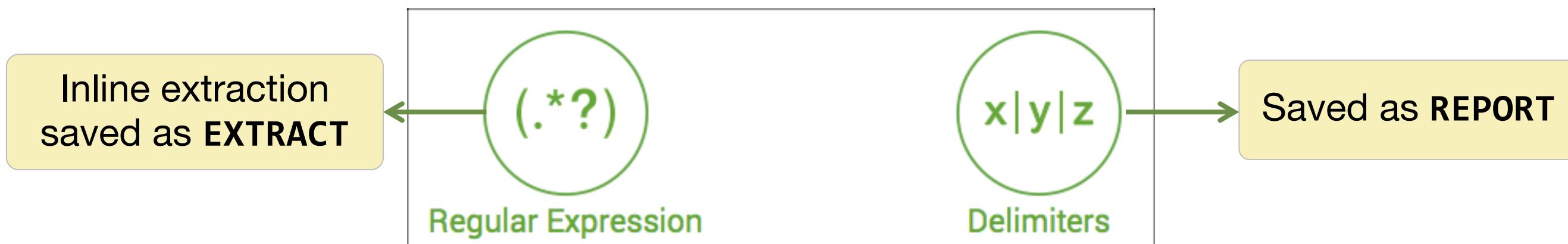
Configuration files

- Provides additional advanced extraction options
- Knowledge of REGEX required
- Available only to admins

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

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**



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

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_ip_countries.csv

Lookup file *:

Create and manage lookup table files.

Configure time-based lookup
 Advanced options

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

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

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

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

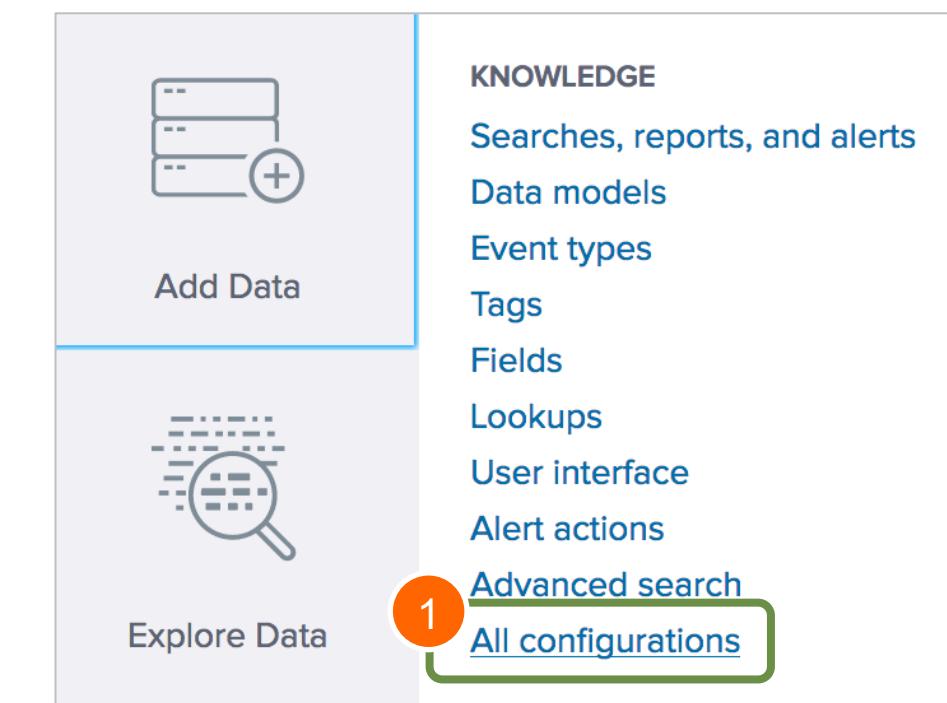
Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

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'. It shows 1-25 of 263 items. Filter options include App (Instrumentation (splu...)), Owner (Any), Visible in the App, filter, and 25 per page. A green box highlights the 'Reassign Knowledge Objects' button, which is numbered 2.

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

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 Search

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 checkboxes next to the objects and selecting Edit Selected Knowledge Objects > Reassign.

1. Click Reassign
2. Select a new owner from the New Owner dropdown 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 Search

Administrator (admin)
SH_alf (alf)
SH_beta (beta)
(emaxwell)
SH_nic (nic) 3

- Use the filter options at the top to locate the objects you want to reassign
- The **Orphaned** button displays all shared, orphaned objects

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) © Splunk Inc, not for distribution

Module 11 Knowledge Check

- True or False. **props.conf** and **transforms.conf** are used to store Field Extractions, Lookups, Saved Searches and macros.
- True or False. Any user belonging to any user role can reassign any KO.
- True or False. When you select the REGEX option in the Field Extractor in the GUI, it uses **props.conf** and **transforms.conf** in the background.

Module 11 Knowledge Check – Answers

- ❑ True or False. **props.conf** and **transforms.conf** are used to store Field Extractions, Lookups, Saved Searches and macros.

False. They are used only for Field Extractions and Lookups.

- ❑ True or False. Any user belonging to any user role has the ability to reassign any KO.

False. Only users belonging to the **admin** role can assign any KO.

- ❑ True or False. When you are using Splunk Web and select the REGEX option in the Field Extractor, it uses **props.conf** and **transforms.conf** in the background.

False. It only uses **props.conf**. Delimiter based extractions entries in **props.conf** and **transforms.conf** are manually created.

Module 11 Lab Exercise – KO Administration

Time: 5 – 10 minutes

Tasks:

- Create a knowledge object (report)
- Search for orphaned knowledge objects
- Assign the report to the user, **emaxwell**

Course Wrap-up

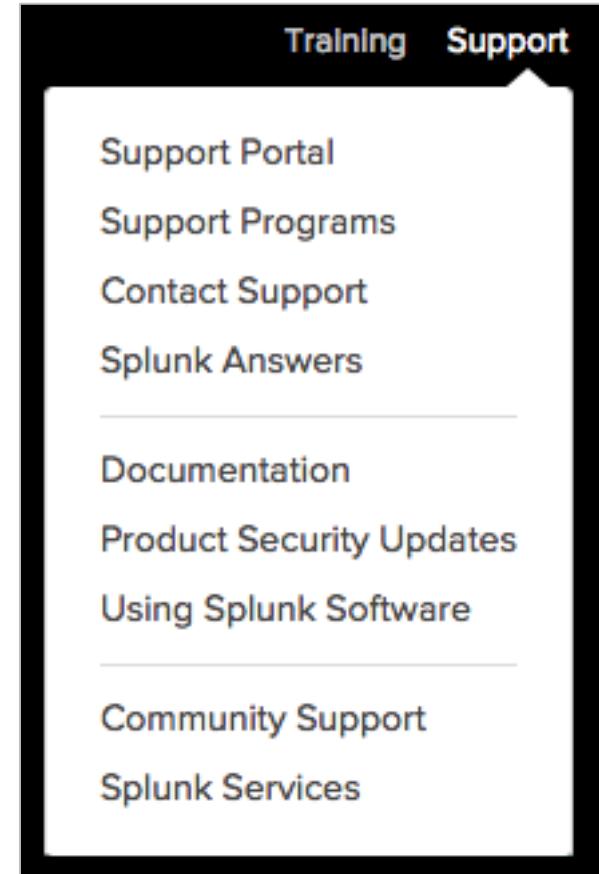
Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Community

- Splunk Community Portal
splunk.com/en_us/community.html
- Splunk Answers
answers.splunk.com
- Splunkbase
splunkbase.splunk.com/
- Splunk Blogs
splunk.com/blog/
- Splunk Live!
<http://splunklive.splunk.com/>
- Splunk .conf
conf.splunk.com
- Splunk Wiki
wiki.splunk.com
- Slack User Groups
splk.it/slack
- Splunk Dev Google Group
groups.google.com/forum/#!forum/splunkdev
- Splunk Docs on Twitter
twitter.com/splunkdocs
- Splunk Dev on Twitter
twitter.com/splunkdev
- IRC Channel
#splunk on the EFNet IRC server

Support Programs

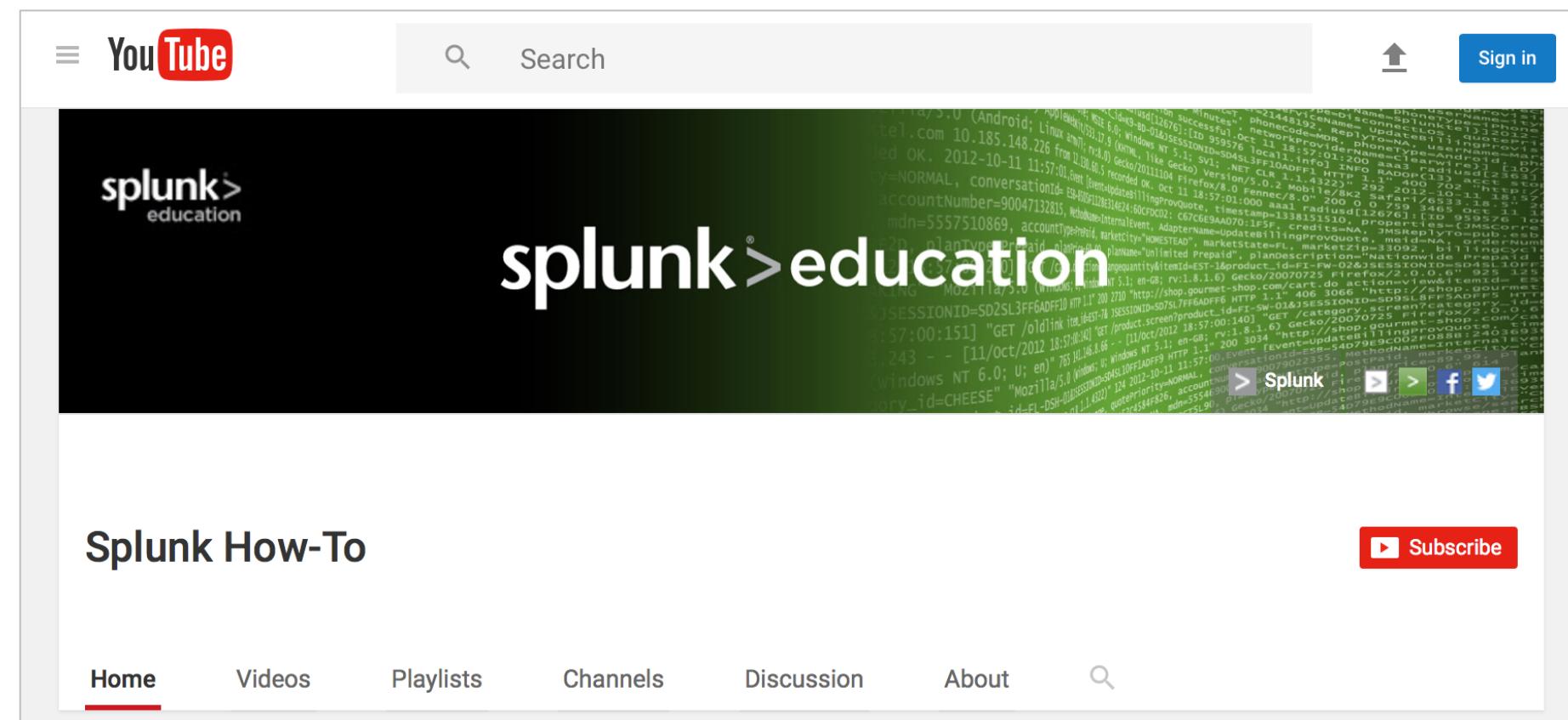
- Web
 - Documentation: docs.splunk.com and dev.splunk.com
 - Wiki: wiki.splunk.com
- Global Support
 - Support for critical issues, a dedicated resource to manage your account – 24 x 7 x 365
 - Web: splunk.com/index.php/submit_issue
 - Phone: (855) SPLUNK-S or (855) 775-8657
- Enterprise Support
 - Access customer support by phone and manage your cases online 24 x 7 (depending on support contract)



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

YouTube: The Splunk How-To Channel

- In addition to the roster of Splunk Education training courses, check out our How-To channel: <http://www.youtube.com/c/SplunkHowTo>
- Useful, short videos on a variety of Splunk topics



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Splunk Certification

- Splunk Certification program

https://www.splunk.com/en_us/training/faq-training.html

- Program information

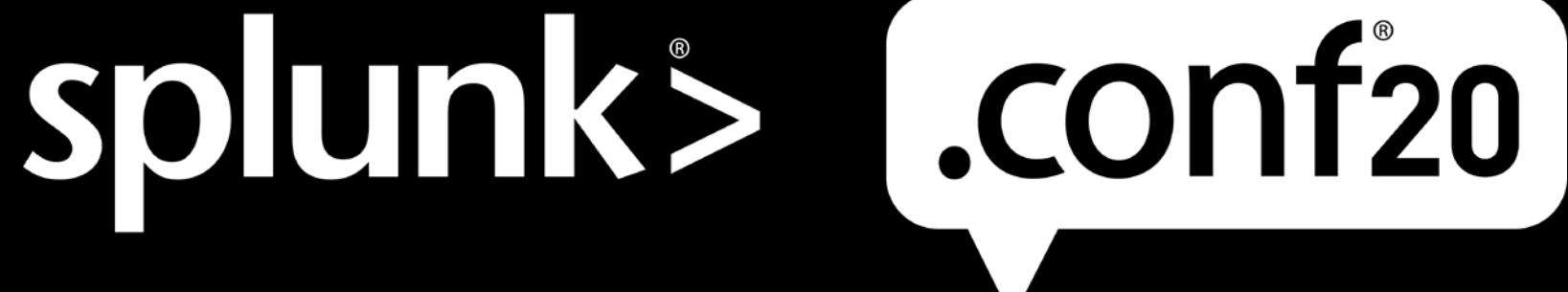
<https://www.splunk.com/pdfs/training/Splunk-Certification-Candidate-Handbook.pdf>

- Exam registration

<https://www.splunk.com/pdfs/training/Exam-Registration-Tutorial.pdf>

- If you have further questions, send an email to:

certification@splunk.com



October 20-21, 2020

Join us for two days of innovation featuring dozens of educational sessions and numerous opportunities to do amazing things with data.

“ Splunk makes our imagination the only limit to unlocking and understanding our data. ”

- IT Specialist, US Public Sector

Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution

Learn more at
conf.splunk.com



Thank You



Generated for Balaram Guttula (balaram.krishn.guttula@jpmchase.com) (C) Splunk Inc, not for distribution