Splunk4Admins

Advanced Data Routing with Forwarders and Ingest Actions

splunk>
a CISCO company

Forward-looking statements

This presentation may contain forward-looking statements that are subject to the safe harbors created under the Securities Act of 1933, as amended, and the Securities Exchange Act of 1934, as amended. All statements other than statements of historical facts are statements that could be deemed forward-looking statements. These statements are based on current expectations, estimates, forecasts, and projections about the industries in which we operate and the beliefs and assumptions of our management based on the information currently available to us. Words such as “expects,” “anticipates,” “targets,” “goals,” “projects,” “intends,” “plans,” “believes,” “momentum,” “seeks,” “estimates,” “continues,” “endeavors,” “strives,” “may,” variations of such words, and similar expressions are intended to identify such forward-looking statements. In addition, any statements that refer to (1) our goals, commitments, and programs; (2) our business plans, initiatives, and objectives; and (3) our assumptions and expectations, including our expectations regarding our financial performance, products, technology, strategy, customers, markets, acquisitions and investments are forward-looking statements. These forward-looking statements are not guarantees of future performance and involve significant risks, uncertainties and other factors that may cause our actual results, performance or achievements to be materially different from results, performance or achievements expressed or implied by the forward-looking statements contained in this presentation. Readers are cautioned that these forward-looking statements are only predictions and are subject to risks, uncertainties, and assumptions that are difficult to predict, including those identified in the “Risk Factors” section of Cisco’s most recent report on Form 10-Q filed on February 20, 2024 and its most recent report on Form 10-K filed on September 7, 2023, as well as the “Risk Factors” section of Splunk’s most recent report on Form 10-Q filed with the SEC on November 28, 2023. The forward-looking statements made in this presentation are made as of the time and date of this presentation. If reviewed after the initial presentation, even if made available by Cisco or Splunk, on Cisco or Splunk’s website or otherwise, it may not contain current or accurate information. Cisco and Splunk undertake no obligation to revise or update any forward-looking statements for any reason, except as required by law.

In addition, any information about new products, features, functionality or our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not be incorporated into any contract or other commitment or be relied upon in making a purchasing decision. We undertake no commitment, promise or obligation either to develop the features or functionalities described, in beta or in preview (used interchangeably), or to include any such feature or functionality in a future release. The development, release, and timing of any features or functionality described for our products remains at our sole discretion.

splunk>
a CISCO company

Please introduce yourself!

- Name
- Company/organisation
- Role
- Are you currently using Splunk?
- What are you interested in using Splunk for?

REGISTRATION 5 MINS

Enroll in Today’s Workshop

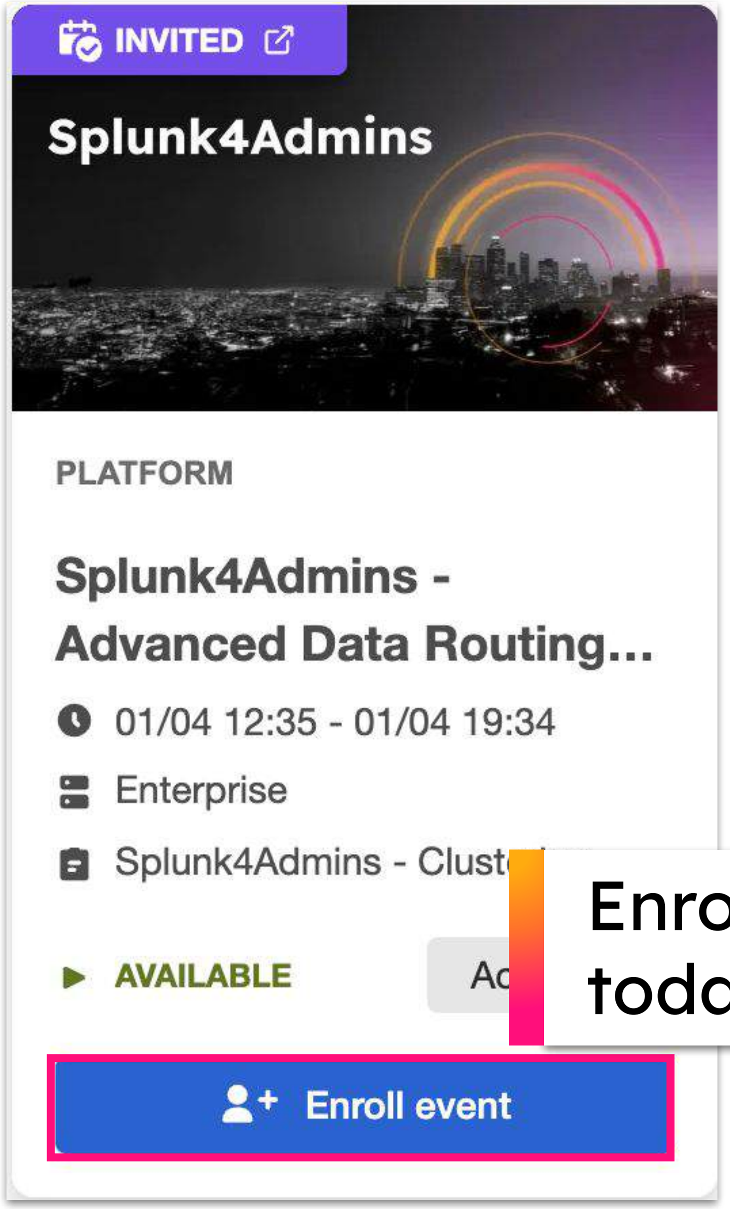
Tasks

1. Get a splunk.com account if you don’t have one yet: https://splk.it/SignUp
2. Enroll in the Splunk Show workshop event: https://show.splunk.com/event/<eventID>
3. Download the hands-on lab guide: https://splk.it/S4A-ADR-Lab-Guide

   Contains step-by-step instructions for all of today’s exercises!

4. Download a copy of today’s slide deck: https://splk.it/S4A-ADR-Attendee

Goal



PLATFORM

Splunk4Admins - Advanced Data Routing...

01/04 12:35 - 01/04 19:34

Enterprise

Splunk4Admins - Clust

AVAILABLE

Enroll in today’s event

Overview

Advanced Data Routing with Forwarders and Ingest Actions

The goal of this workshop is to help attendees understand some of the more advanced data routing functions available in Splunk as well as review general Ingest Actions configurations and features.

General Expectations

- Duration 60 to 90 Minutes
- General understanding of Splunk forwarding
- Experience with Splunk configuration files
- Knowledge of Splunk Data Routing

Audience

Audience	Recommendation
Those who are interested in how to observe and detect issues within Splunk Enterprise Search	
Splunk Admins	
Power User Certified	Required
Splunk Admin enabled \ Splunk Admin Certified	Preferred

Workshop Agenda

- Introduction
- **Review of Splunk Indexing**
  - MELT (Metrics, Events, Logs, Traces)
  - Getting Data In Review
  - Overview of GDI Topology
  - HEC
  - Splunk Forwarders (UF, HF)
  - Splunk OTEL Collector
- **Data Processing with Forwarders**
  - What is a Sourcetype
  - GDI efficiency and props.conf
  - Advanced Forwarder Controls
  - Cloned Forwarding
  - Example Forwarding Topology
- **Data Processing with Ingest Actions**
  - Overview of Ingest Actions
  - Ingest Action Examples
  - Benefits of Ingest Actions
  - Ingest Actions RULESET dataflow
  - Deploy a Ruleset
  - IA Pipeline Operations and Metrics

Introduction

The purpose of this workshop is to review Splunk pipeline processing and introduce data forwarding and advanced data routing features.

Topics covered in this Workshop are:

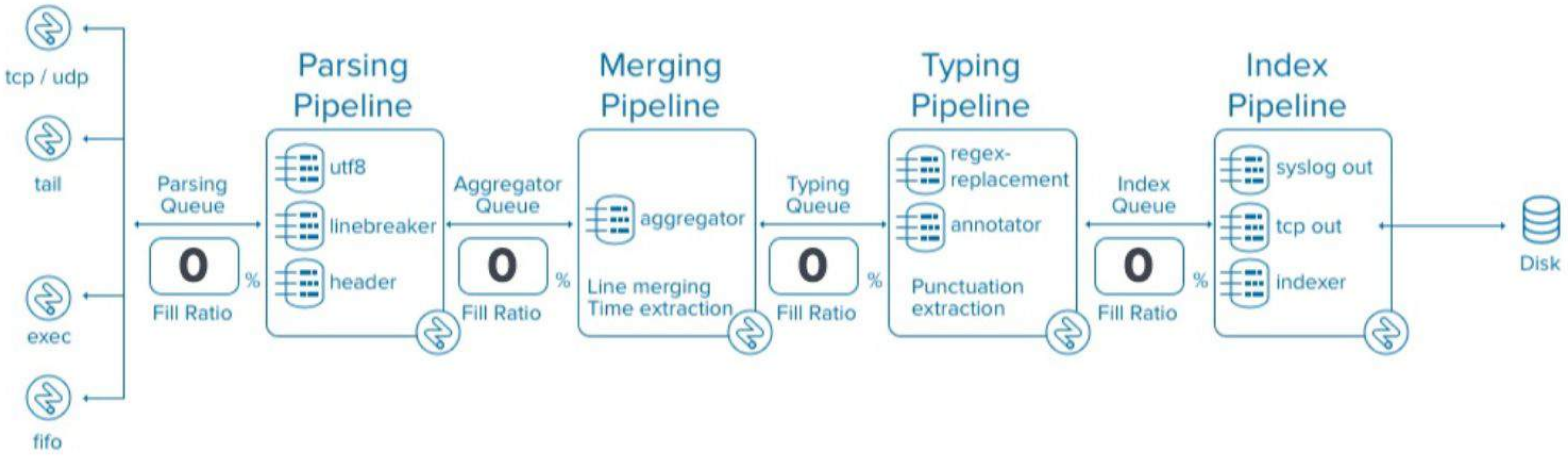
- Review of Splunk Indexing
- Data Pipelines
- Data routing and Advanced Data routing
- Features and functions of Splunk Ingest Actions

Workshop Agenda

- **Introduction**
- **Review of Splunk Indexing**
  - MELT (Metrics, Events, Logs, Traces)
  - Getting Data In Review
  - Overview of GDI Topology
  - HEC
  - Splunk Forwarders (UF, HF)
  - Splunk OTEL Collector
- **Data Processing with Forwarders**
  - What is a Sourcetype
  - GDI efficiency and props.conf
  - Advanced Forwarder Controls
  - Cloned Forwarding
  - Example Forwarding Topology
- **Data Processing with Ingest Actions**
  - Overview of Ingest Actions
  - Ingest Action Examples
  - Benefits of Ingest Actions
  - Ingest Actions RULESET dataflow
  - Deploy a Ruleset
  - IA Pipeline Operations and Metrics

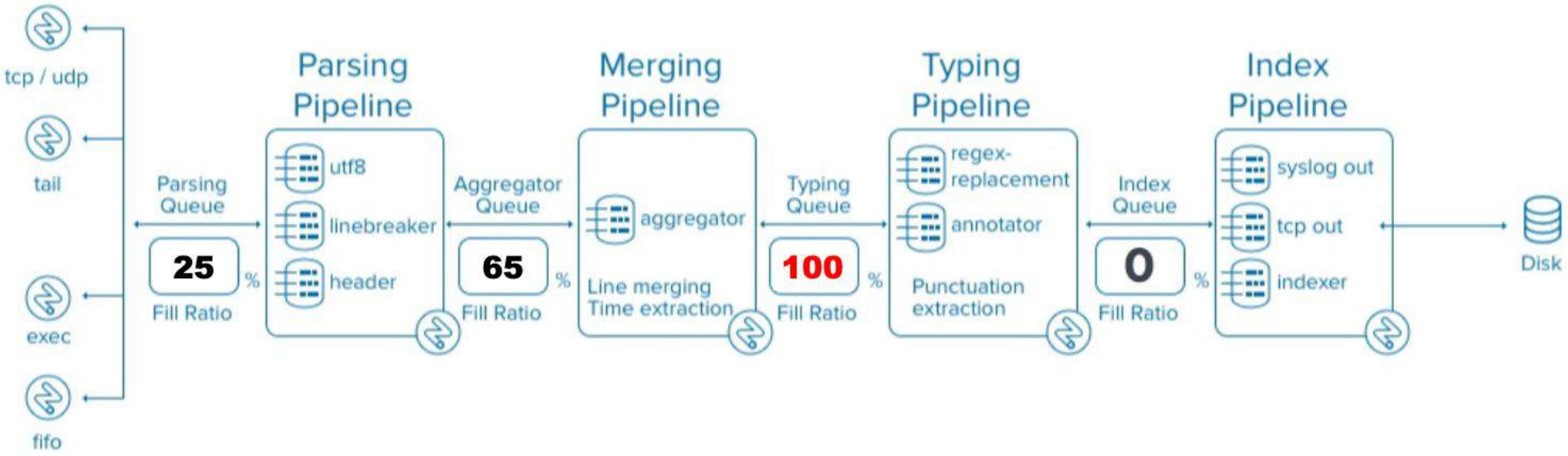
Review of Splunk Indexing

Splunk Event Pipeline



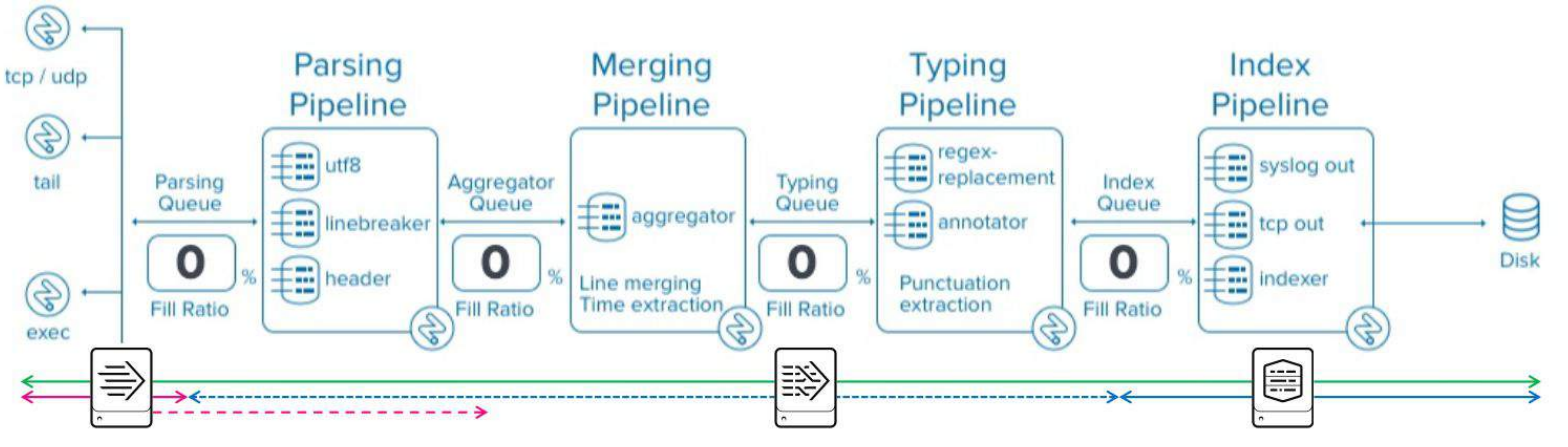
Review of Splunk Indexing (2/3)

Example: Data advancing through the Splunk Pipeline



Review of Splunk Indexing (3/3)

- Data is treated as individual events after the **Merging Pipeline**
- A pipeline-set can span across the multiple Splunk instances
- With extra resources, Splunk can run multiple pipeline-sets in parallel



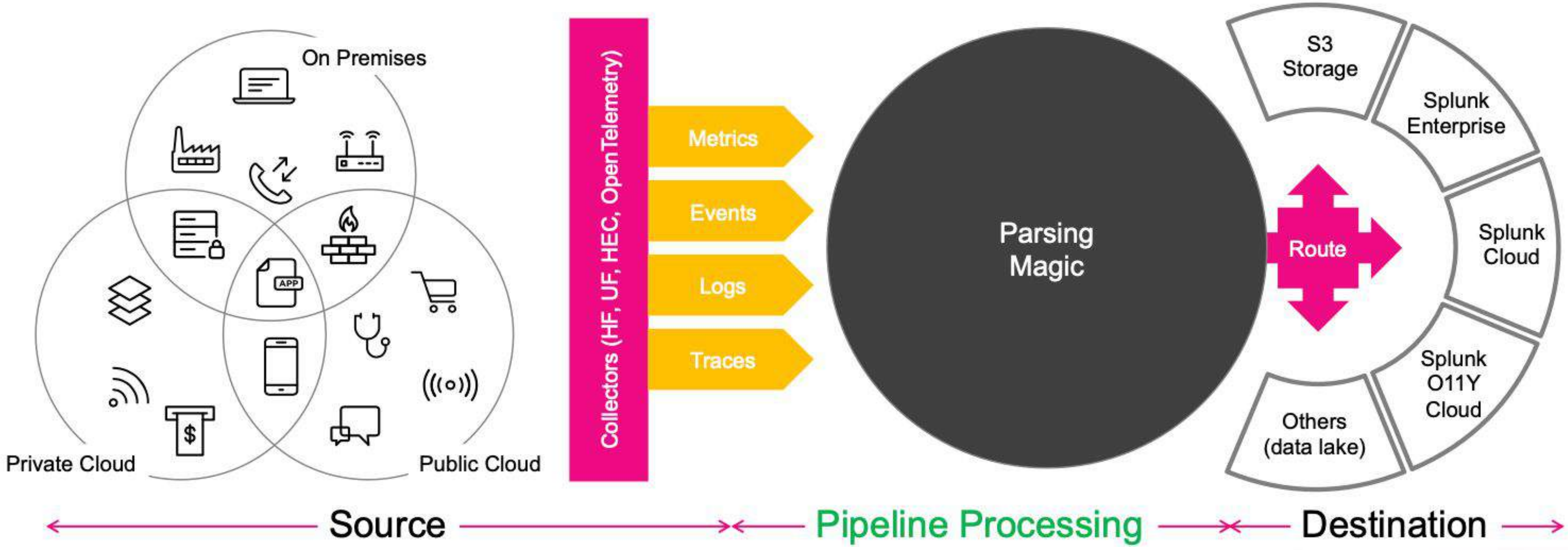
What is MELT?

- Metrics – a series of point-in-time measurements of something
- Events – structured logs or metrics that describe a behavior
- Logs – timestamped records that are not metrics nor traces
- Traces – a representation of the interactions across multiple services

https://www.splunk.com/en_us/blog/learn/melt-metrics-events-logs-traces.html

Overview of Splunk GDI

- GDI topology diagram



Splunk HTTP Event Collector (HEC)

- Uses simple HTTP endpoints to push data into Splunk pipelines from anywhere without a collector agent
- Highly performant and able to scale-out
- Splunk Enterprise/Cloud, HF, Edge Processor are HEC receivers

https://docs.splunk.com/Documentation/Splunk/latest/Data/UsetheHTTPEventCollector

Splunk Universal Forwarders (UF)

- Reliable and versatile remote data collector
- Uses minimum system footprint but provides rich set of features
- tcpout utilizes persistent queues
- The scripted and network inputs can utilize persistent queues
- Forwards unparsed data, except in certain cases
- Limited event-parsing capabilities

https://docs.splunk.com/Documentation/Forwarder/9.4.0/Forwarder/Abouttheuniversalforwarder

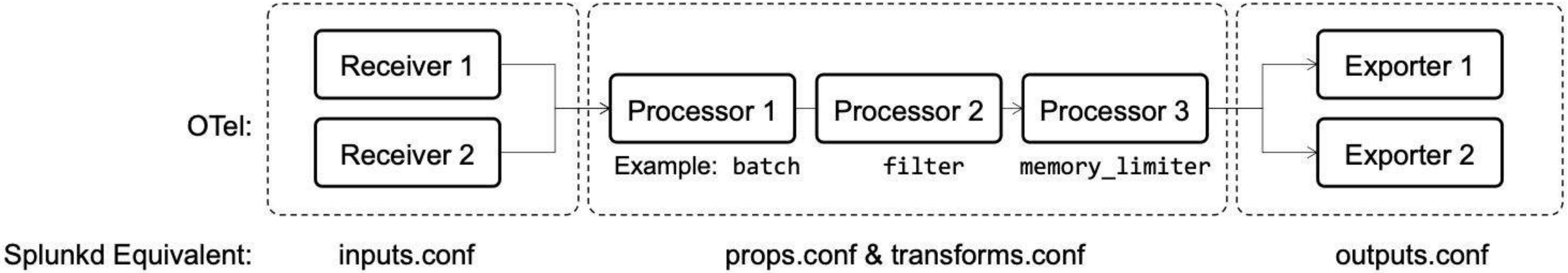
Splunk Heavy Forwarder (HF)

- Splunk Enterprise instance that retains the capabilities of an indexer
- Receives data via HTTP Event Collector (HEC) and from UFs (S2S)
- Fully-parses data before routing data

https://docs.splunk.com/Documentation/Splunk/9.4.1/Forwarding/Typesofforwarders

Splunk Open Telemetry Collector (OTel Agent)

- Uses an open-source pipeline framework to collect data
  - Vendor agnostic collector runs on each monitored host



OTel:

Receiver 1

Receiver 2

Processor 1

Processor 2

Processor 3

Example: `batch` `filter` `memory_limiter`

Exporter 1

Exporter 2

Splunkd Equivalent: inputs.conf props.conf & transforms.conf outputs.conf

https://docs.splunk.com/observability/en/gdi/opentelemetry/opentelemetry.html

https://docs.splunk.com/observability/en/gdi/opentelemetry/collector-with-the-uf.html#collector-with-the-uf

Lab 1

Workshop Agenda

- **Introduction**
- **Review of Splunk Indexing**
  - MELT (Metrics, Events, Logs, Traces)
  - Getting Data In Review
  - Overview of GDI Topology
  - HEC
  - Splunk Forwarders (UF, HF)
  - Splunk OTEL Collector
- **Data Processing with Forwarders**
  - What is a Sourcetype
  - GDI efficiency and props.conf
  - Advanced Forwarder Controls
  - Cloned Forwarding
  - Example Forwarding Topology
- **Data Processing with Ingest Actions**
  - Overview of Ingest Actions
  - Ingest Action Examples
  - Benefits of Ingest Actions
  - Ingest Actions RULESET dataflow
  - Deploy a Ruleset
  - IA Pipeline Operations and Metrics

Data Processing with Forwarders

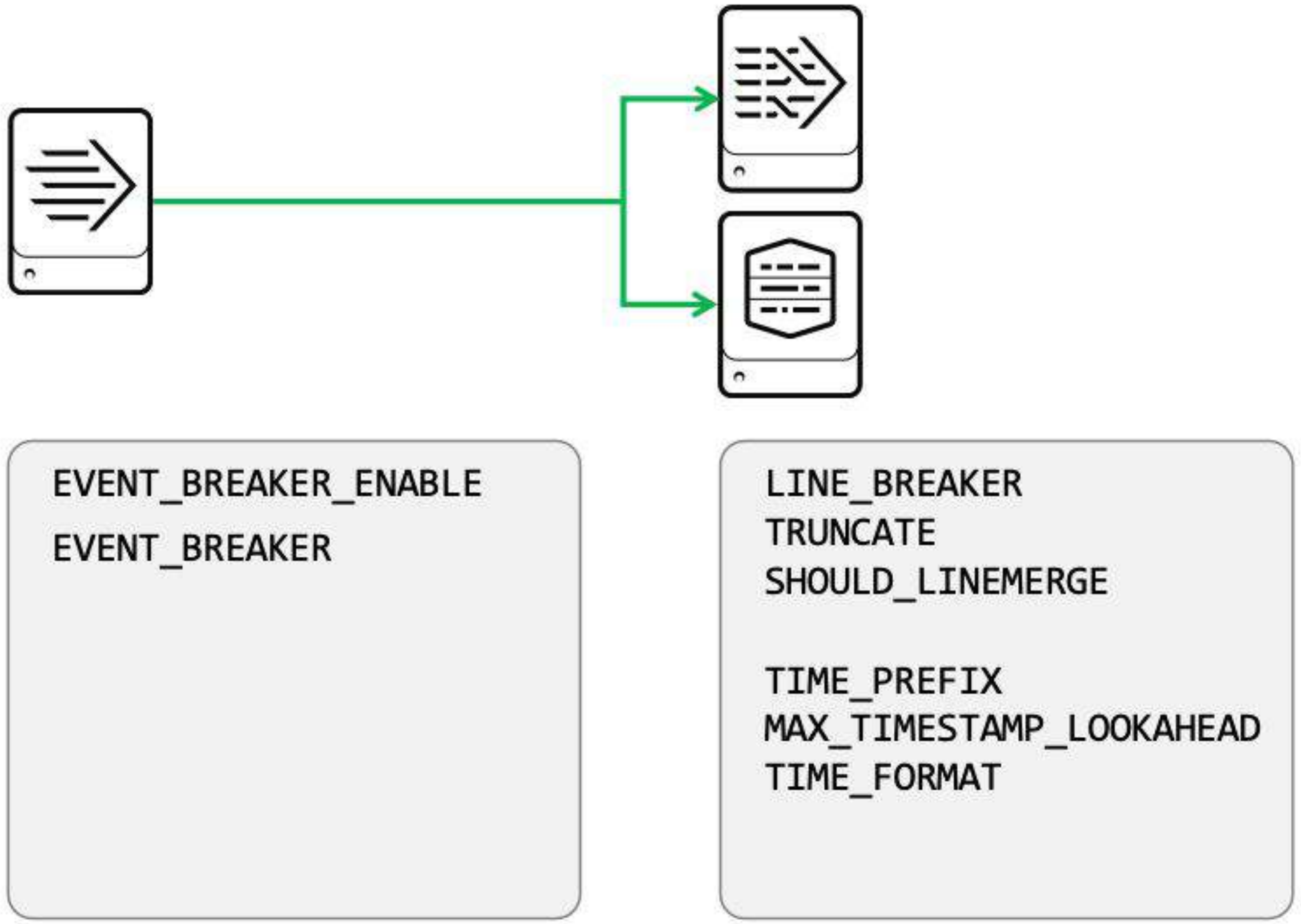
What is a Sourcetype in Splunk?

- Primary unit of partitioning data streams into similar event categories
    - Pretrained source types are included in Splunk
    - Splunk learns the data patterns of unknown source types
    - Splunk auto-assigns a sourcetype when you do not assign a value
- To change source type, you can:
    - Define a custom source type and override the automatic assignment
    - Assign a new value per-event basis
    - Rename source types (search time)

https://docs.splunk.com/Documentation/Splunk/9.4.1/Data/Whysourcetypesmatter

GDI Efficiency in props.conf

- To maximize GDI efficiency, always set these props.conf settings per source type
- The event breaker settings are applied on the UF’s
- The rest is applied on the first instance of Splunk Enterprise that executes the merging pipeline



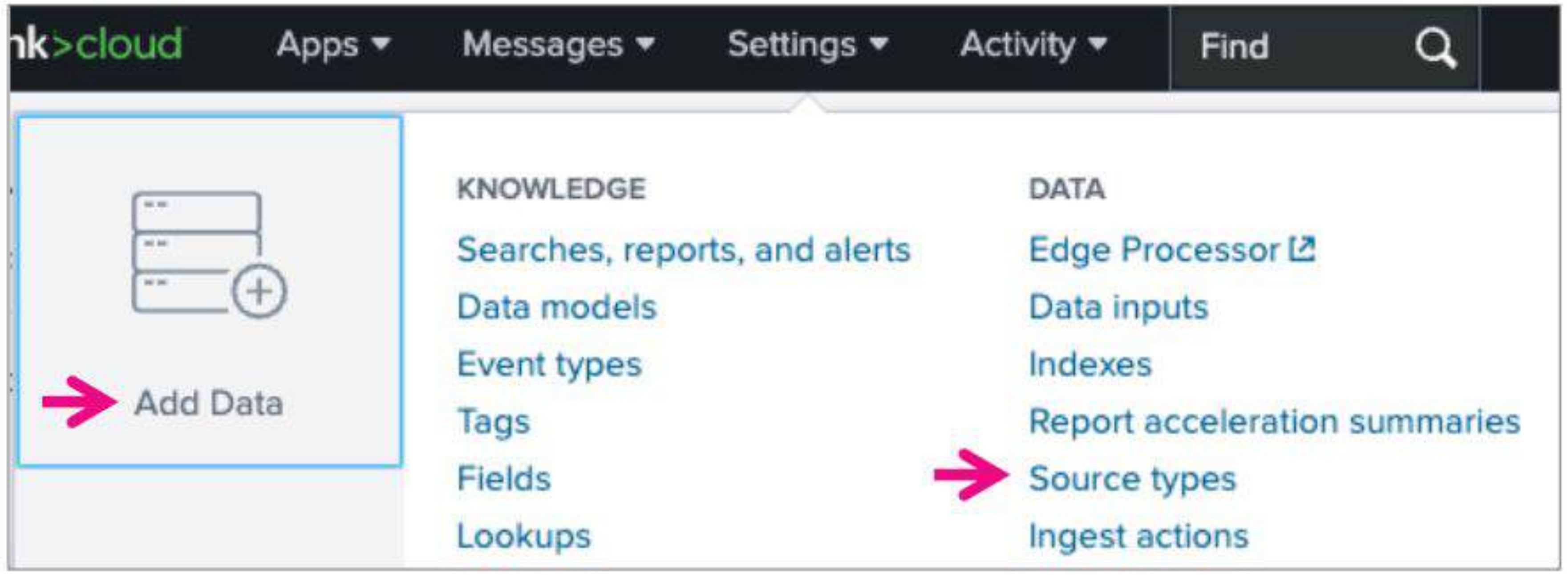
```
EVENT_BREAKER_ENABLE
EVENT_BREAKER
```

At the source UF

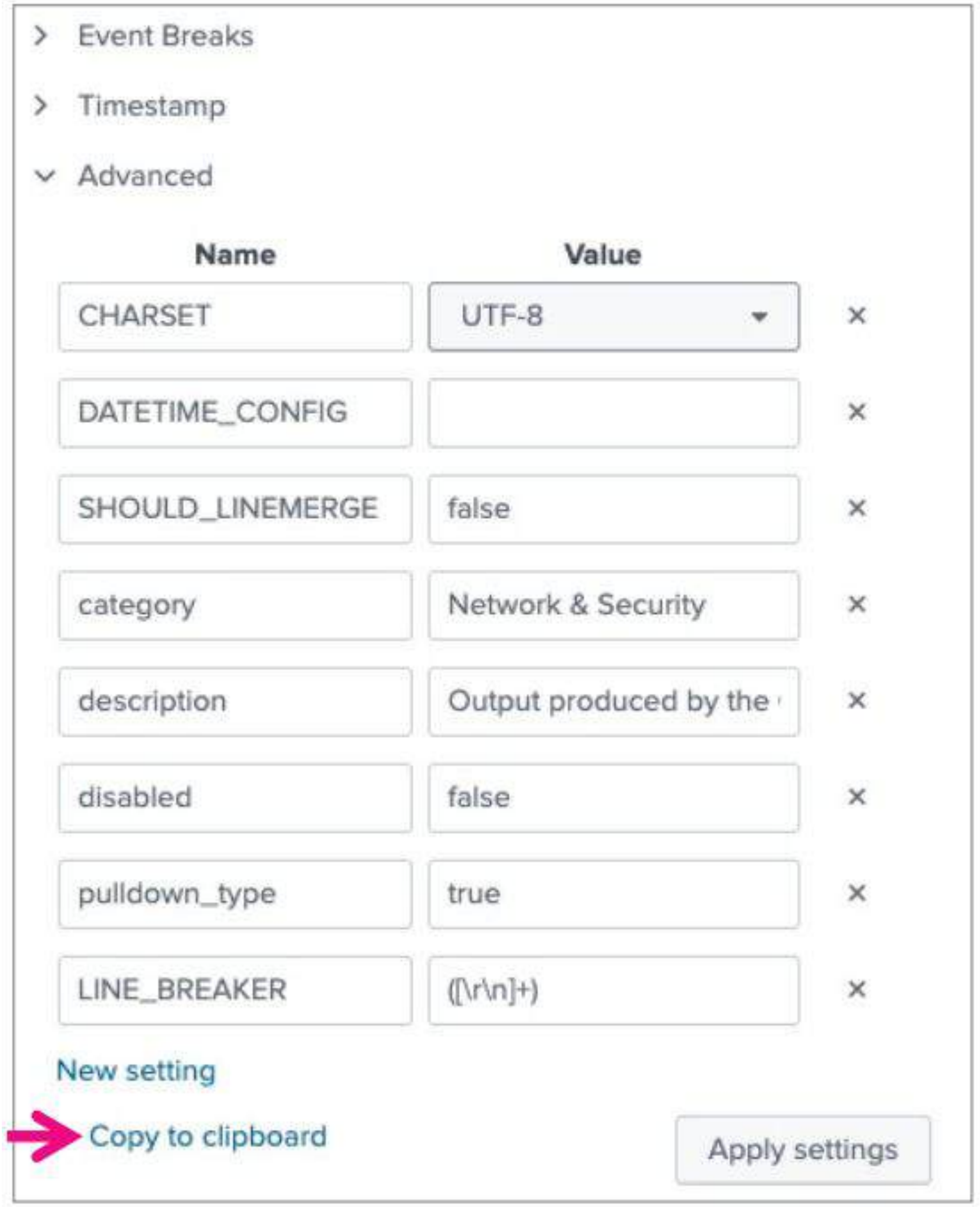
```
LINE_BREAKER
TRUNCATE
SHOULD_LINEMERGE

TIME_PREFIX
MAX_TIMESTAMP_LOOKAHEAD
TIME_FORMAT
```

At the first Merging pipeline

GDI Efficiency in props.conf



1. Load the sample event data
2. Modify the settings iteratively and review until you are satisfied
3. Replace or save as new **sourcetype**
4. Apply the new source type



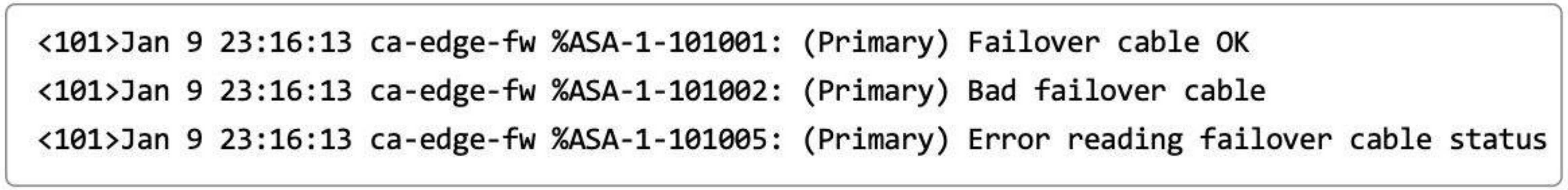
Name	Value
CHARSET	UTF-8
DATETIME_CONFIG	
SHOULD_LINEMERGE	false
category	Network & Security
description	Output produced by the
disabled	false
pulldown_type	true
LINE_BREAKER	([\r\n]+)

https://lantern.splunk.com/Splunk_Platform/Product_Tips/Data_Management/Configuring_new_source_types

Tuning props.conf settings

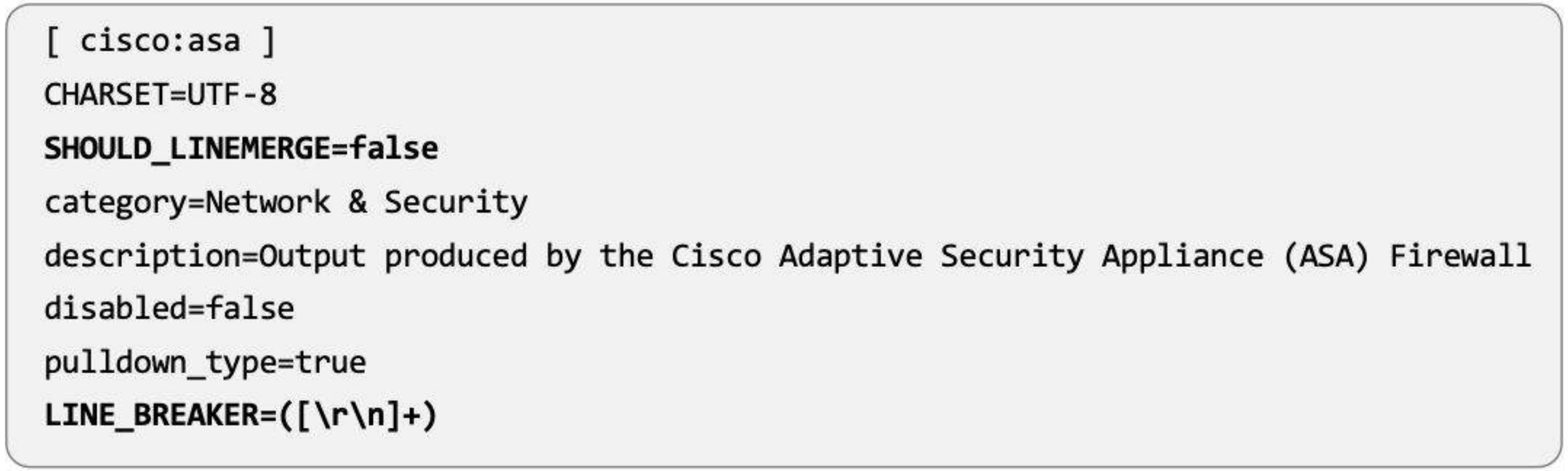
- What values should you set for the remaining attributes?

Examples of cisco.asa syslog data:

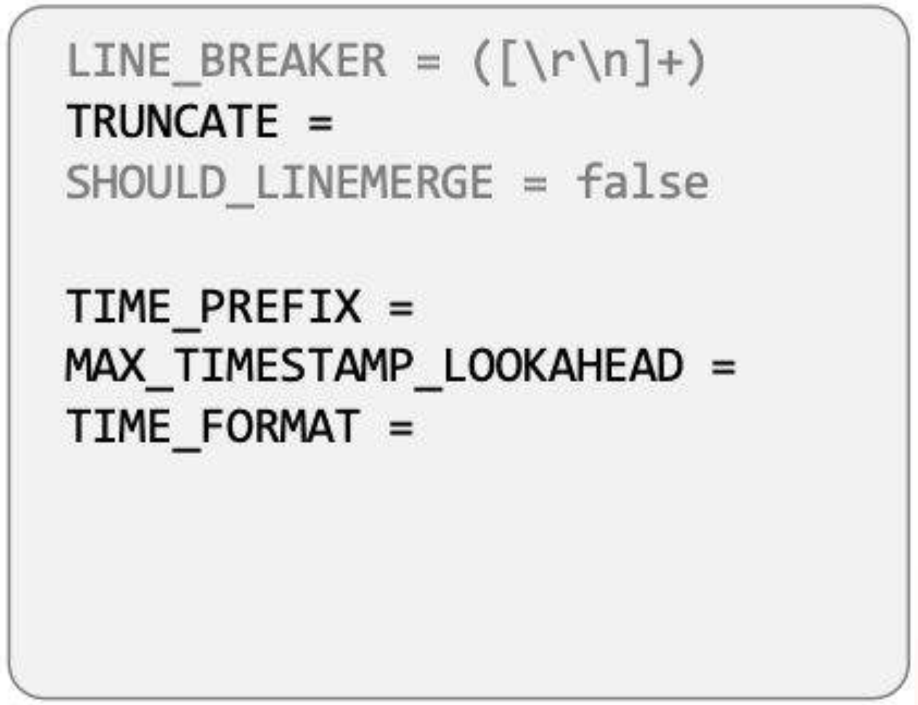


```
<101>Jan 9 23:16:13 ca-edge-fw %ASA-1-101001: (Primary) Failover cable OK
<101>Jan 9 23:16:13 ca-edge-fw %ASA-1-101002: (Primary) Bad failover cable
<101>Jan 9 23:16:13 ca-edge-fw %ASA-1-101005: (Primary) Error reading failover cable status
```

The default sourcetype definition for cisco.asa logs:



```
[ cisco:asa ]
CHARSET=UTF-8
SHOULD_LINEMERGE=false
category=Network & Security
description=Output produced by the Cisco Adaptive Security Appliance (ASA) Firewall
disabled=false
pulldown_type=true
LINE_BREAKER=([\r\n]+)
```



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

TIME_PREFIX =
MAX_TIMESTAMP_LOOKAHEAD =
TIME_FORMAT =
```

Advanced Forwarder Controls

Inputs.conf and **Outputs.conf** quick review

Inputs.conf contains possible settings you can use to configure inputs, distributed inputs such as forwarders, and file system monitoring

Inputs.conf

Outputs.conf determines how the forwarder sends data to receiving Splunk instances, either indexers or other forwarders.

Outputs.conf

https://docs.splunk.com/Documentation/Splunk/latest/admin/Inputsconf

https://docs.splunk.com/Documentation/Splunk/9.4.1/Admin/Outputsconf

Advanced Forwarder Controls (cont.)

Custom Index-time Field with _meta

Inputs.conf

- **Recommended** – Try to extract your fields at search time first
- Adding custom indexed-field with **_meta** can be useful when:
  - You want to capture a transient data
  - It can provide clear performance gain over search-time extraction
- All indexed-fields are stored in a special key called **_meta**

Note Using custom metadata is always a trade-off

https://docs.splunk.com/Documentation/Splunk/latest/admin/Inputsconf

Advanced Forwarder Controls (cont.)

Using _meta with fields.conf and inputs.conf

fields.conf

```
[env]
INDEXED=true

[vm]
INDEXED=true
```

inputs.conf

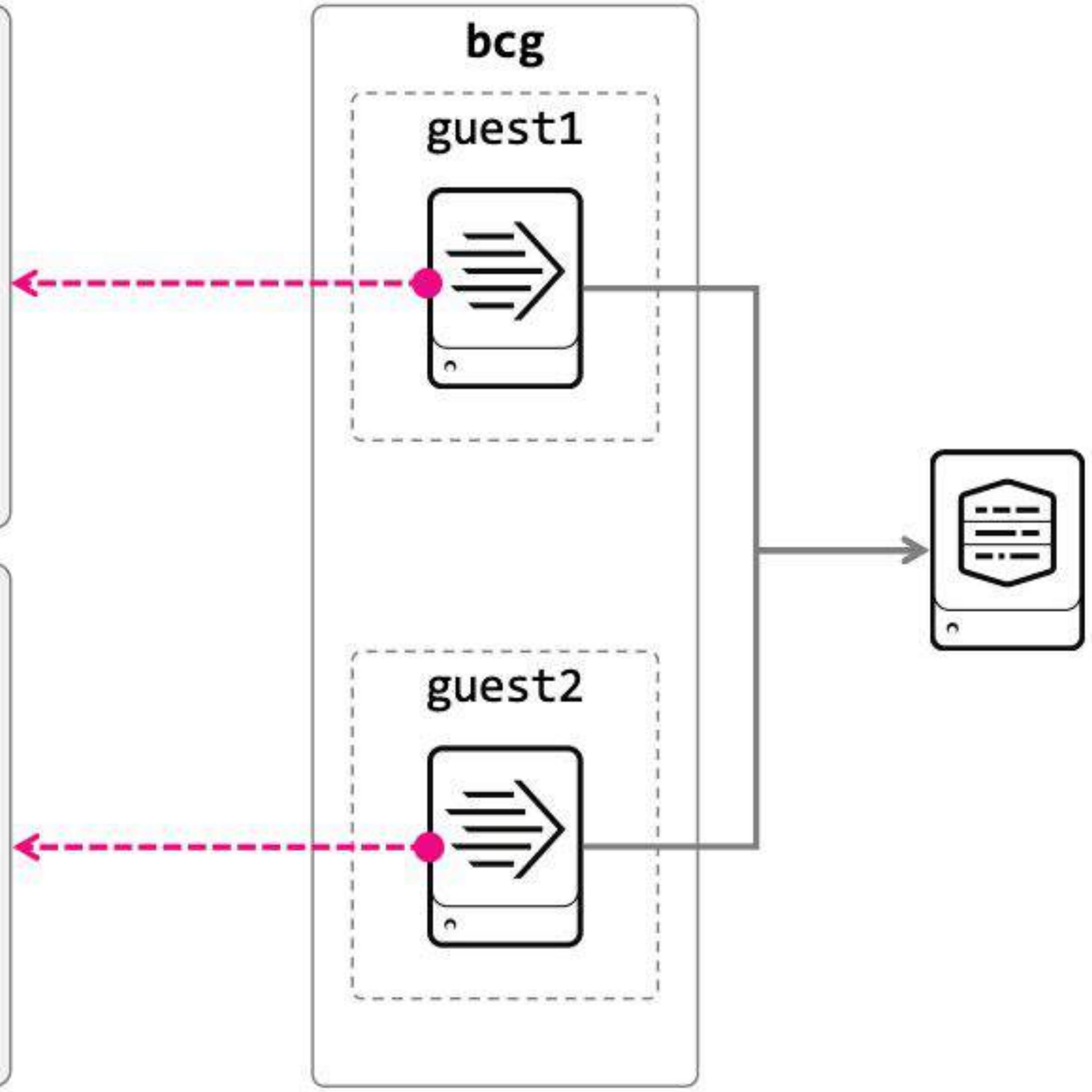
```
[default]
index = network
_meta = env::bcg vm::guest1
[tcp://:5150]
connection_host = ip
sourcetype = log4j
```

fields.conf

```
[env]
INDEXED=true

[vm]
INDEXED=true
```

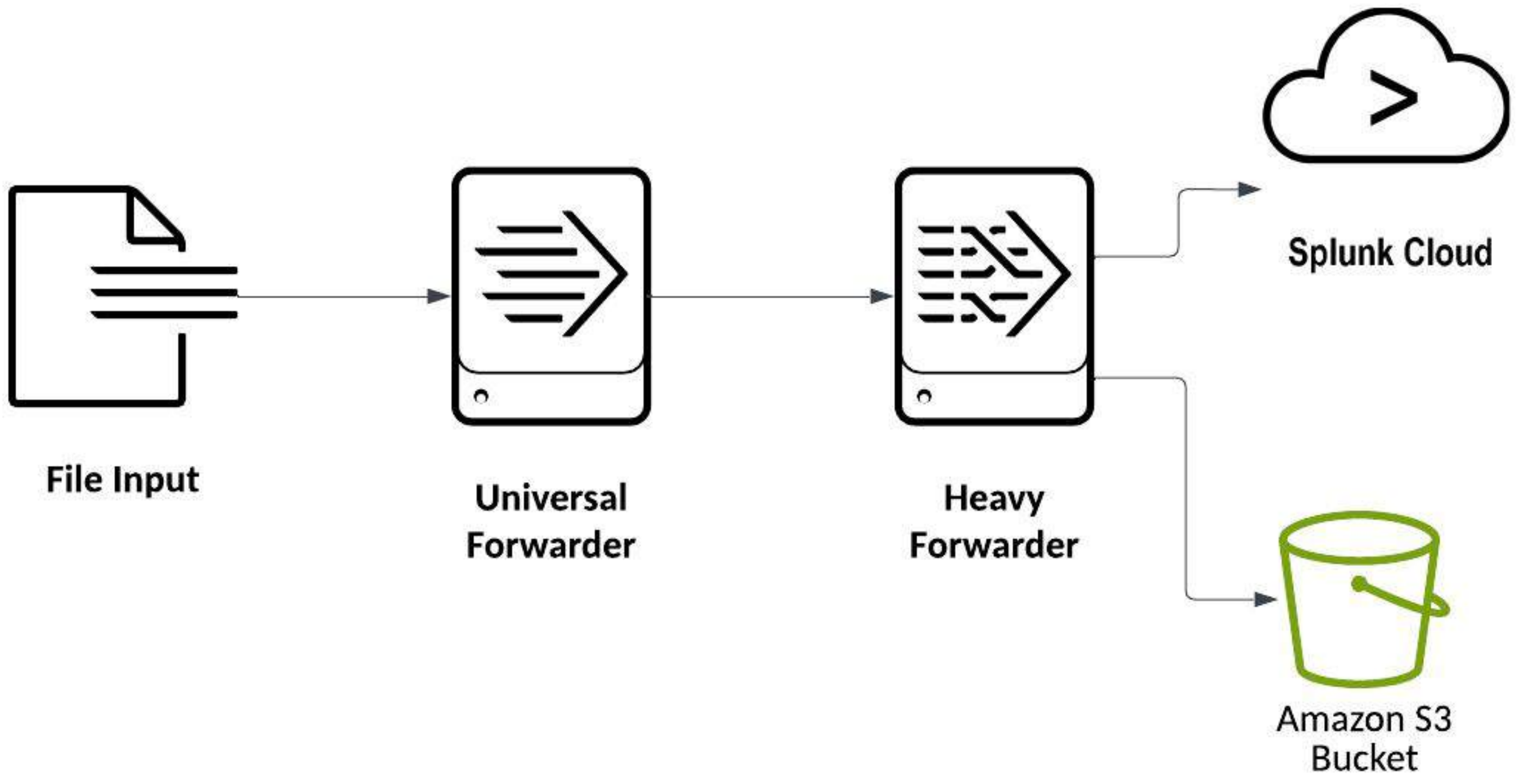
inputs.conf

```
[default]
index = network
_meta = env::bcg vm::guest2
[tcp://:5150]
connection_host = ip
sourcetype = log4j
```



Cloned Forwarding

- Send data simultaneously to multiple Splunk destinations
- Send data to third party destinations (File systems, Splunk cloud, S3, etc.)
- Long-term use is discouraged without consulting Splunk architects



https://docs.splunk.com/Documentation/Splunk/latest/Forwarding/Forwarddatatothird-partysystemsd

Cloned Forwarding (cont.)

Using _TCP_ROUTING to direct traffic

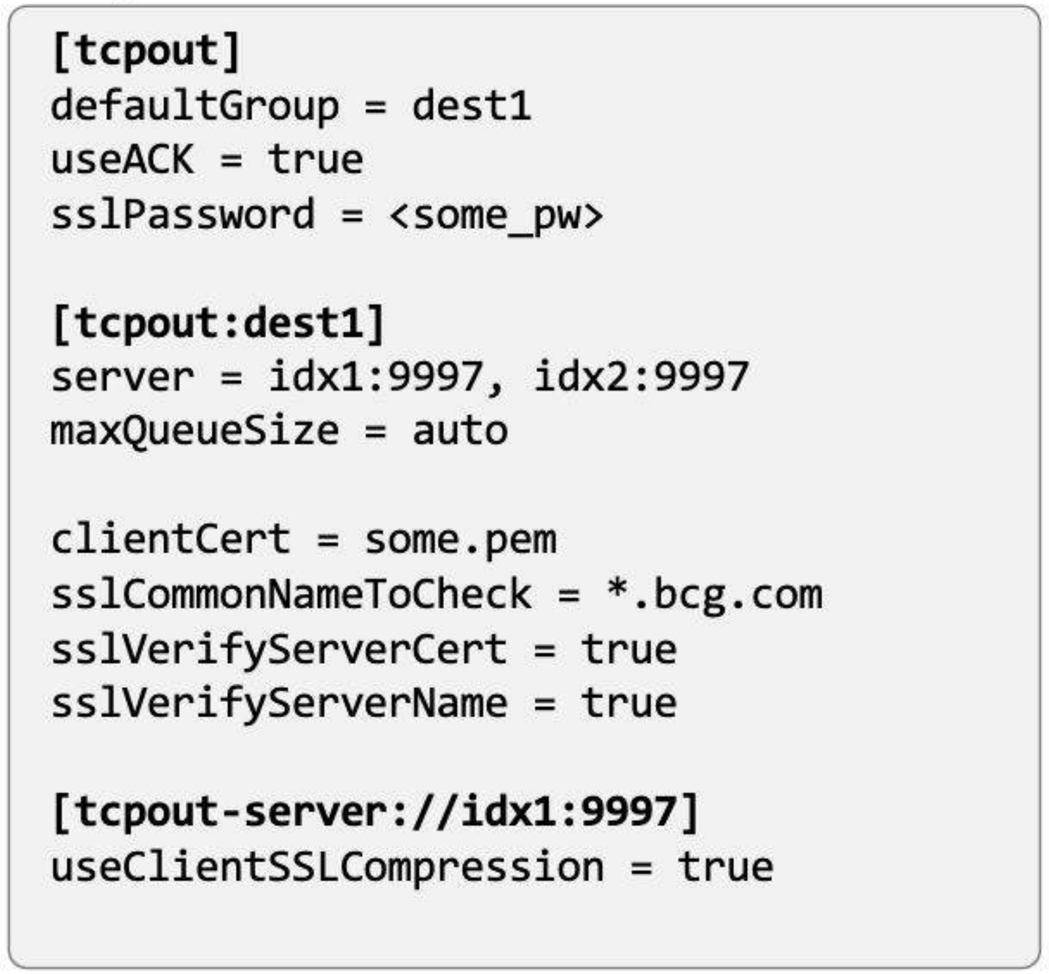
Outputs.conf

Use TCP_ROUTING to selectively route data to specific **tcpout** groups (indexers)

Settings in outputs.conf:

- Defines outgoing traffic types
- Has five output processors
- UF supports httpout and tcpout
- Three tcpout configuration levels:
- TLS settings on Splunk Enterprise is not set by default

outputs.conf



```
[tcpout]
defaultGroup = dest1
useACK = true
sslPassword = <some_pw>

[tcpout:dest1]
server = idx1:9997, idx2:9997
maxQueueSize = auto

clientCert = some.pem
sslCommonNameToCheck = *.bcg.com
sslVerifyServerCert = true
sslVerifyServerName = true

[tcpout-server://idx1:9997]
useClientSSLCompression = true
```

Note Fully understand its consequences before using this attribute

https://docs.splunk.com/Documentation/Splunk/latest/Forwarding/Forwarddatatothird-partysystemsd

https://docs.splunk.com/Documentation/Splunk/9.4.1/Admin/Outputsconf

Cloned Forwarding (cont.)

Additional settings in outputs.conf

Outputs.conf

defaultGroup = <comma-separated list>

- Multiple tcpout names means ***cloning*** to destinations simultaneously
- Set this to blank if you want to manually forward data with **_TCP_ROUTING** from inputs.conf
  - Can override with **props.conf** / **transforms.conf** later
- Each group gets its own **tcpout** queue
  - maxQueueSize=auto and the queue size depends on useACK
    - useACK=true grows the tcpout queue from 500KB to 7MB by default
    - Additionally creates a wait queue of 21MB (3x of tcpout queue)
  - By default, Splunk doesn't drop events but creates backpressure

Recommended Do not use more than one tcpout:<group>

https://docs.splunk.com/Documentation/Splunk/9.4.1/Admin/Outputsconf

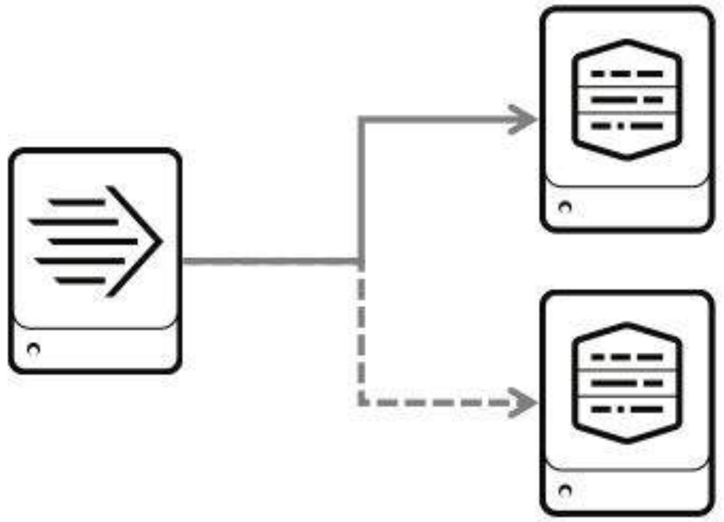
Example Forwarding Topology

.CONF

Outputs.conf

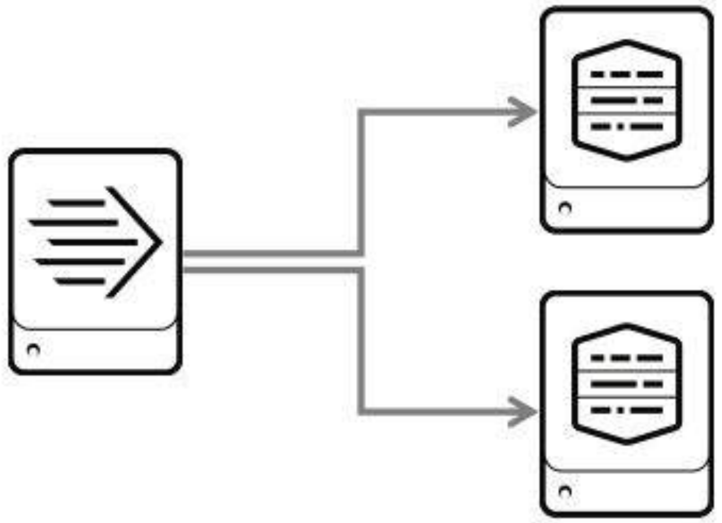
Load balance

```
[tcpout]
defaultGroup = dest1
blockOnCloning = true
dropEventsOnQueueFull = -1
dropClonedEventsOnQueueFull = 5s

[tcpout:dest1]
server = idx1:9997, idx2:9997
```



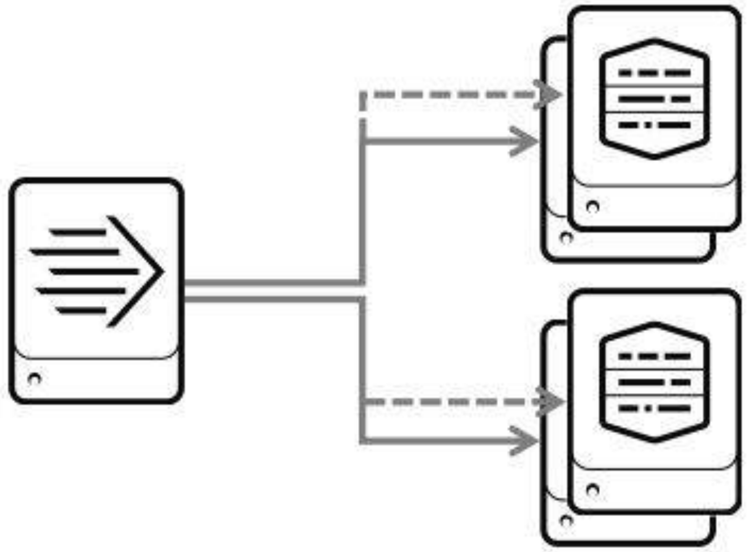
Clone

```
[tcpout]
defaultGroup = dest1,dest2
blockOnCloning = true
dropEventsOnQueueFull = -1
dropClonedEventsOnQueueFull = 5s

[tcpout:dest1]
server = idx1:9997

[tcpout:dest2]
server = idx2:9997
```



Clone and Load balance

```
[tcpout]
defaultGroup = dest1,dest2
blockOnCloning = true
dropEventsOnQueueFull = -1
dropClonedEventsOnQueueFull = 5s

[tcpout:dest1]
server = idx1:9997, idx2:9997

[tcpout:dest2]
server = idx3:9997, idx4:9997
```



Lab 2

Workshop Agenda

- **Introduction**
- **Review of Splunk Indexing**
  - MELT (Metrics, Events, Logs, Traces)
  - Getting Data In Review
  - Overview of GDI Topology
  - HEC
  - Splunk Forwarders (UF, HF)
  - Splunk OTEL Collector
- **Data Processing with Forwarders**
  - What is a Sourcetype
  - GDI efficiency and props.conf
  - Advanced Forwarder Controls
  - Cloned Forwarding
  - Example Forwarding Topology
- **Data Processing with Ingest Actions**
  - Overview of Ingest Actions
  - Ingest Action Examples
  - Benefits of Ingest Actions
  - Ingest Actions RULESET dataflow
  - Deploy a Ruleset
  - IA Pipeline Operations and Metrics

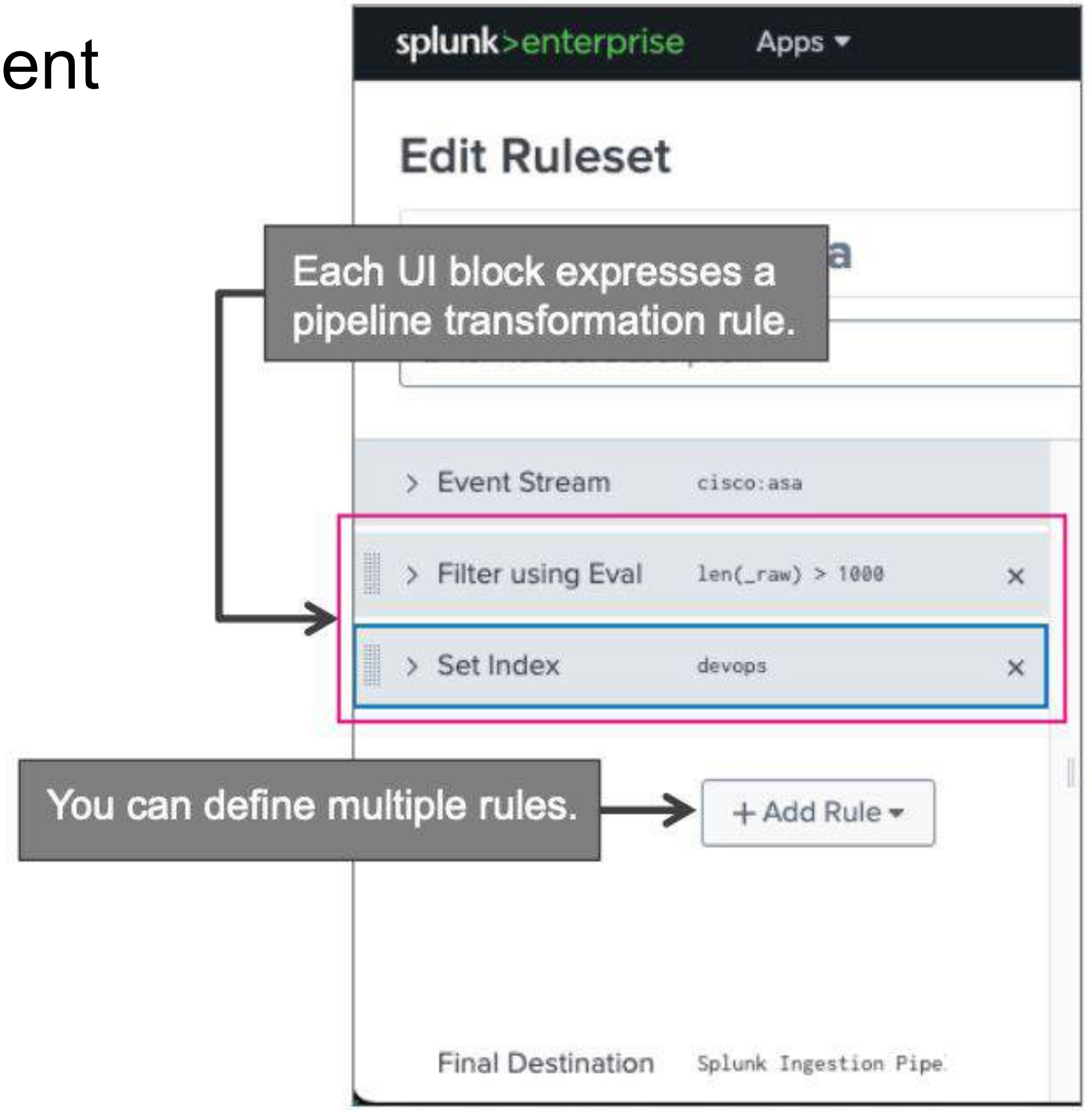
Overview of Ingest Actions

What is Ingest Actions?

Ingest Actions is an interface in Splunk to author and deploy event transformation rules before indexing

Additional details about Ingest Actions:

- Define multiple rules
- Create multiple pipeline transformations
- Saves the combined rules as a ruleset
- Deploys the saved rulesets as an app
- Works with its underlying .conf files
  - **props.conf**
  - **transforms.conf**
  - **outputs.conf**



Edit Ruleset

Each UI block expresses a pipeline transformation rule.

Event Stream cisco:asa

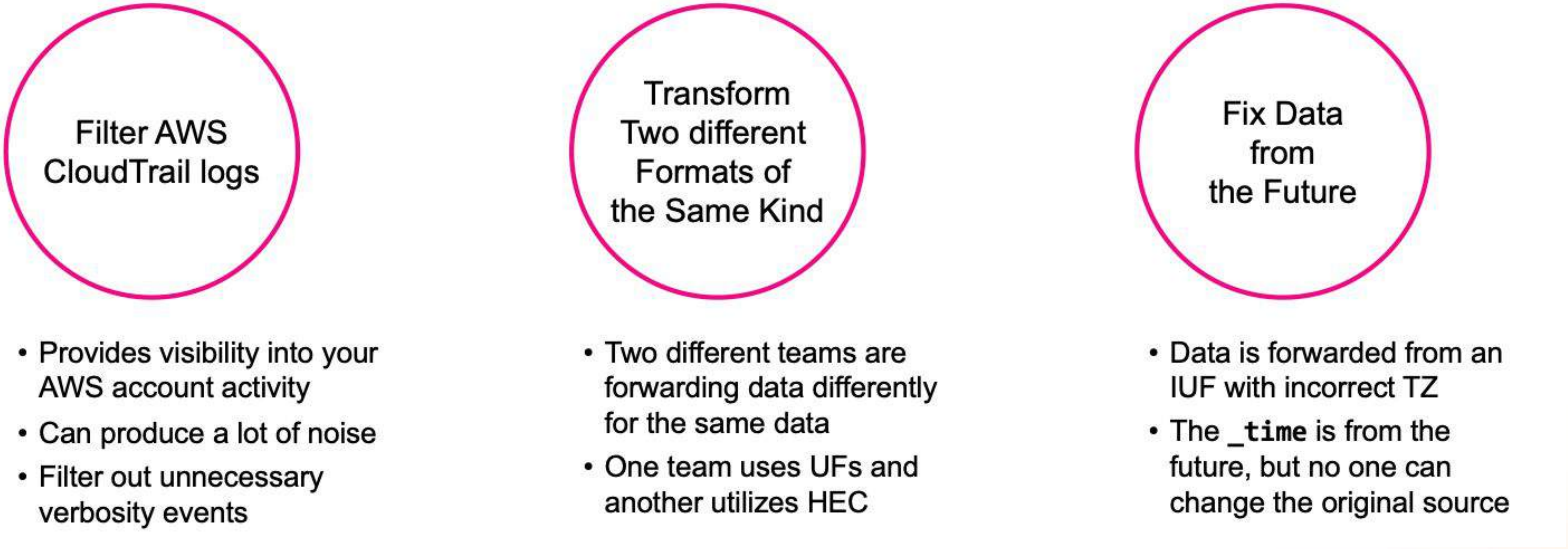
Filter using Eval len(_raw) > 1000

Set Index devops

You can define multiple rules.

Final Destination Splunk Ingestion Pipe

Ingest Actions Examples



Filter AWS CloudTrail logs

- Provides visibility into your AWS account activity
- Can produce a lot of noise
- Filter out unnecessary verbosity events

Transform Two different Formats of the Same Kind

- Two different teams are forwarding data differently for the same data
- One team uses UFs and another utilizes HEC

Fix Data from the Future

- Data is forwarded from an IUF with incorrect TZ
- The `_time` is from the future, but no one can change the original source

https://docs.splunk.com/Documentation/SVA/current/Architectures/IngestActions#Rulesets

Benefits of Ingest Actions

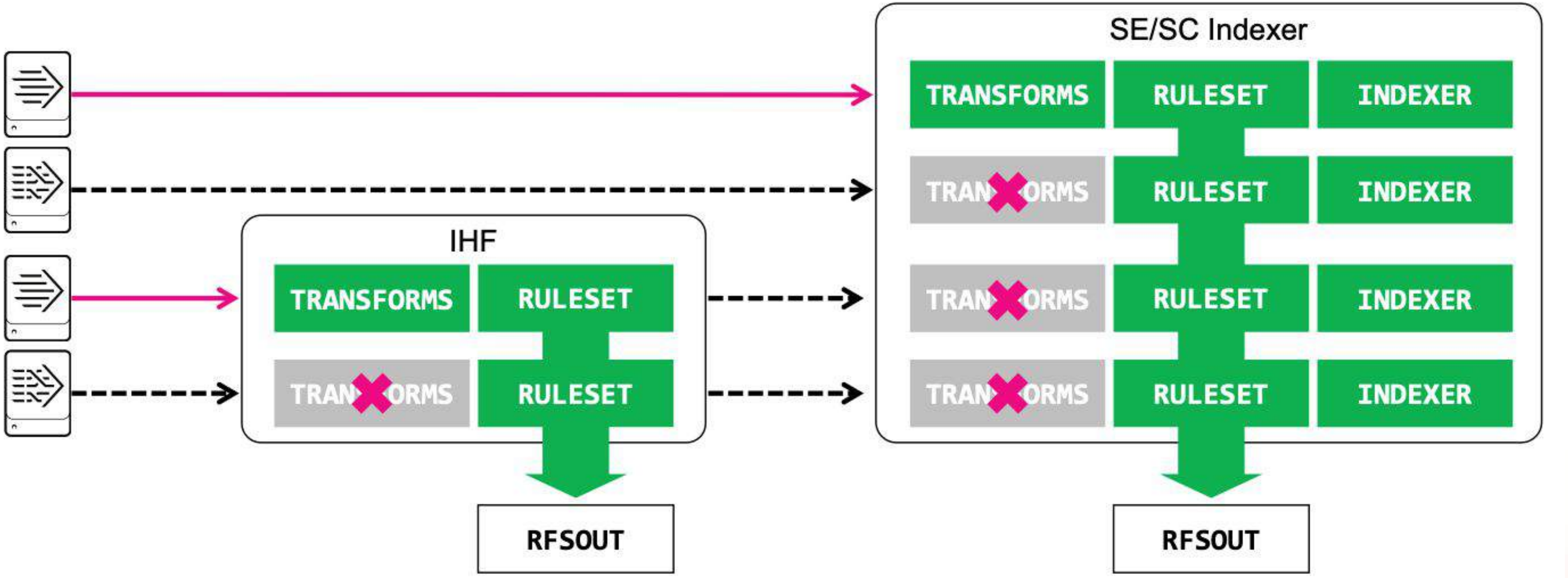
Primary Benefits:

- Mask, redact, remove, or otherwise change raw data before indexing
- Tag, add, lookup, or otherwise augment raw data before indexing
- Filter entire events from being indexed
- Have tighter control over which indexes data is sent to

Additional Benefits

- Send some or all data to third-party storage either independently or concurrently with data going in to the Splunk platform
- Your GDI topologies can remain largely unchanged
- Can develop, preview and validate rules on-the-fly before deploying
- Available as part of the existing licensing, at no additional cost

Ingest Actions RULESET Dataflow



https://docs.splunk.com/Documentation/SVA/current/Architectures/IngestActions#Rulesets

Deploy a Ruleset

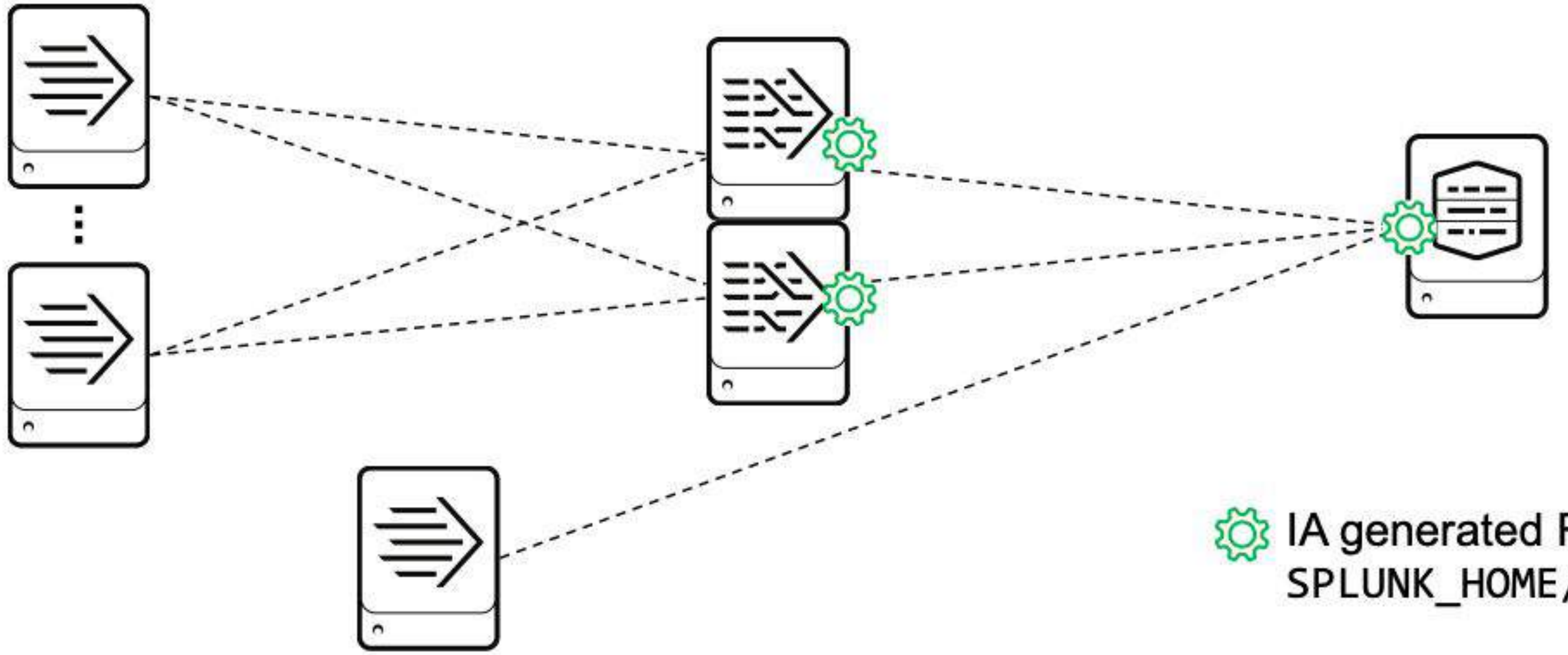
The **splunk_ingest_actions** app deployments depend on where you've authored your rulesets

- **Standalone**
- **Splunk Enterprise**
- **Deployment Server**
- **Cluster Manager**
- **Splunk Cloud**

Ensure that search-time artifacts are compatible with the resulting transformations

Deploy a Ruleset - Standalone

- Good for development and a small-scale deployment



IA generated Rulesets:
`SPLUNK_HOME/etc/apps/splunk_ingest_actions`

Deploy a Ruleset - Deployment Server

The Deployment Server must be dedicated

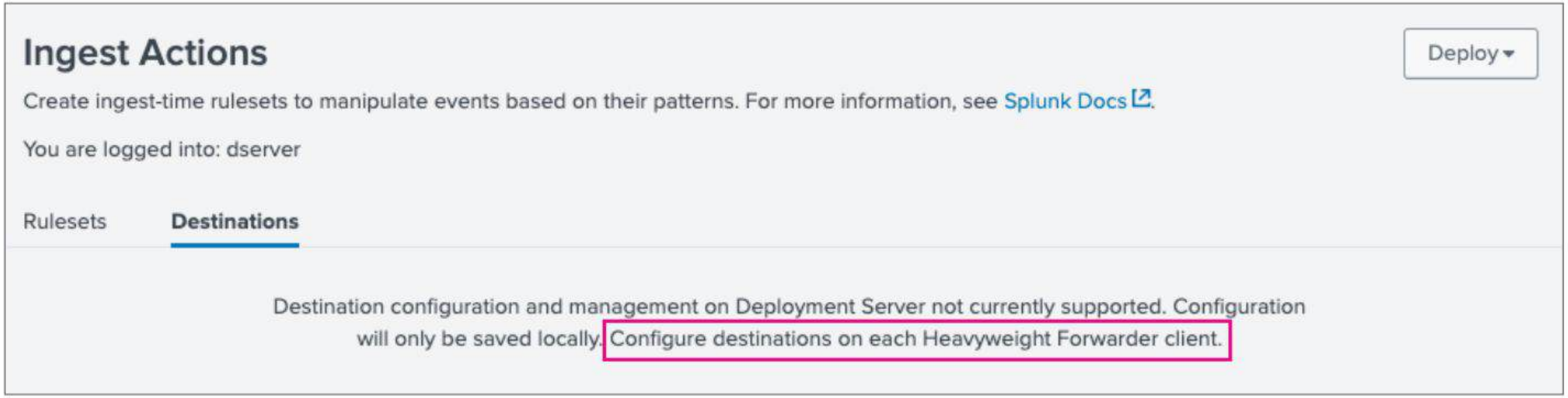
Cannot service any other deployment Clients (DCs)

Configure the HFs processing the rulesets as its DCs

Supports a maximum of 10 HFs (as of 9.3.1)

Can't add the destinations

Manually add from each HF



Ingest Actions

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

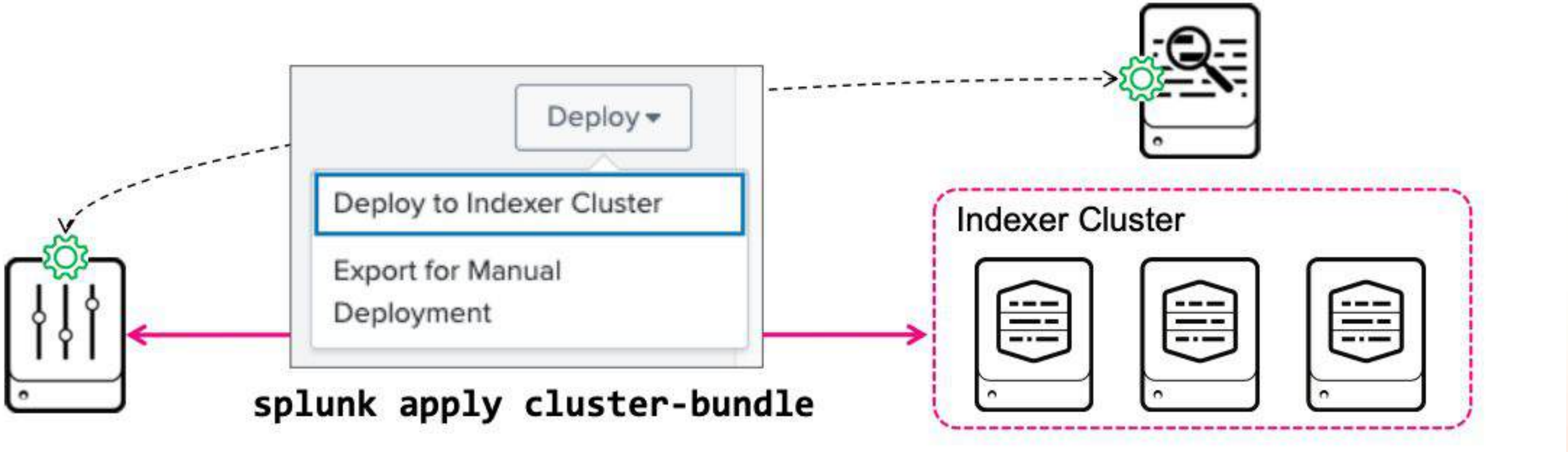
You are logged into: dserver

Destination configuration and management on Deployment Server not currently supported. Configuration will only be saved locally. Configure destinations on each Heavyweight Forwarder client.

Deploy a Ruleset - Index Cluster

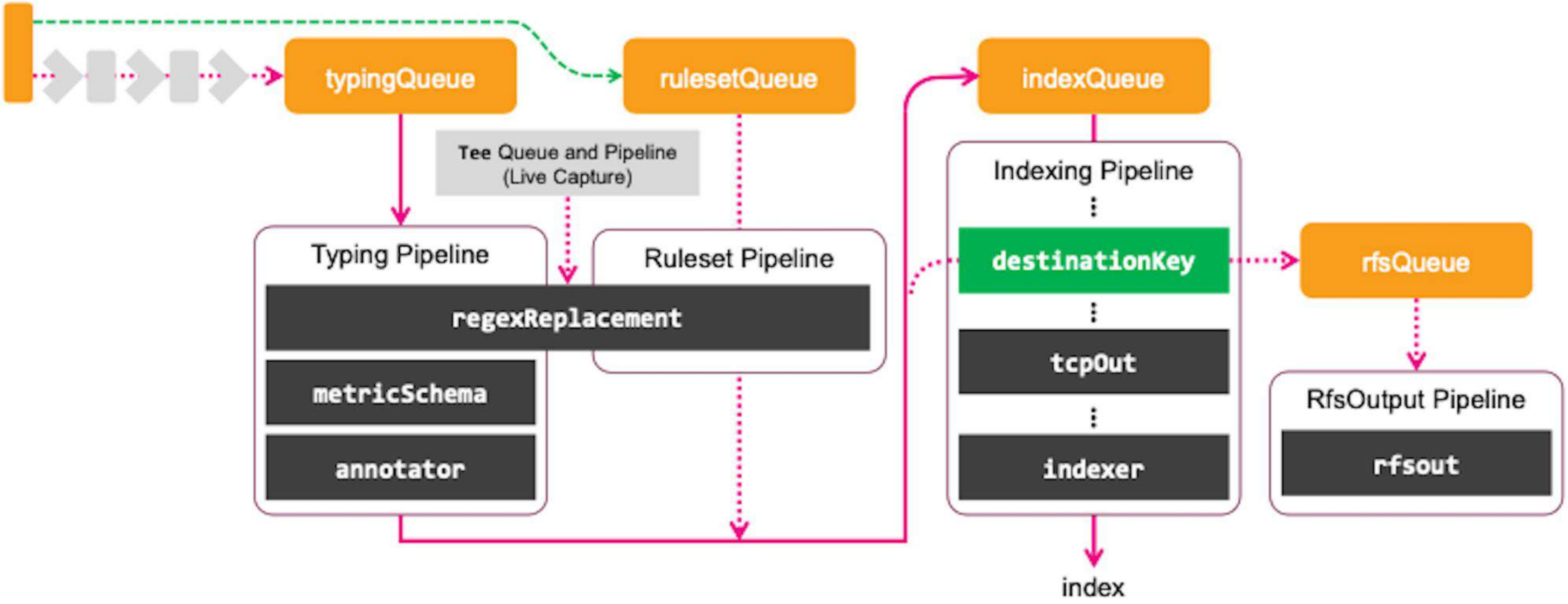
- Use the UI from Cluster Manager or a cluster-connected Search Head
  - On Cluster Manager, the ruleset is saved in the following location:

    SPLUNK_HOME/etc/manager-apps/splunk_ingest_actions
- Search Head proxy-saves the rulesets to the Cluster Manager



```
splunk apply cluster-bundle
```

IA Pipeline Operations and Metrics



Lab 3

Appendix

Title	Link
Configuring New Source Types	https://lantern.splunk.com/Splunk_Platform/Product_Tips/Data_Management/Configuring_new_source_types
HTTP Event Collector	https://docs.splunk.com/Documentation/Splunk/latest/Data/UsetheHTTPEventCollector
About the Universal Forwarder	https://docs.splunk.com/Documentation/Forwarder/9.4.0/Forwarder/Abouttheuniversalforwarder
Types of Forwarders	https://docs.splunk.com/Documentation/Splunk/9.4.1/Forwarding/Typesofforwarders
Splunk Open Telemetry	https://docs.splunk.com/observability/en/gdi/opentelemetry/opentelemetry.html
Splunk OTel with UF	https://docs.splunk.com/observability/en/gdi/opentelemetry/collector-with-the-uf.html#collector-with-the-uf
Inputs.conf	https://docs.splunk.com/Documentation/Splunk/latest/admin/Inputsconf
Outputs.conf	https://docs.splunk.com/Documentation/Splunk/9.4.1/Admin/Outputsconf
Forwarding to Third Party systems	https://docs.splunk.com/Documentation/Splunk/latest/Forwarding/Forwarddatatothird-partysystemsd
Ingest Actions Rulesets	https://docs.splunk.com/Documentation/SVA/current/Architectures/IngestActions#Rulesets
MELT	https://www.splunk.com/en_us/blog/learn/melt-metrics-events-logs-traces.html
Why Source types matter	https://docs.splunk.com/Documentation/Splunk/9.4.1/Data/Whysourcetypesmatter

If you would like to learn more, please consider looking into the EDU course: Flexing Pipeline Processing Techniques with Splunk

Thank you

splunk>
a CISCO company