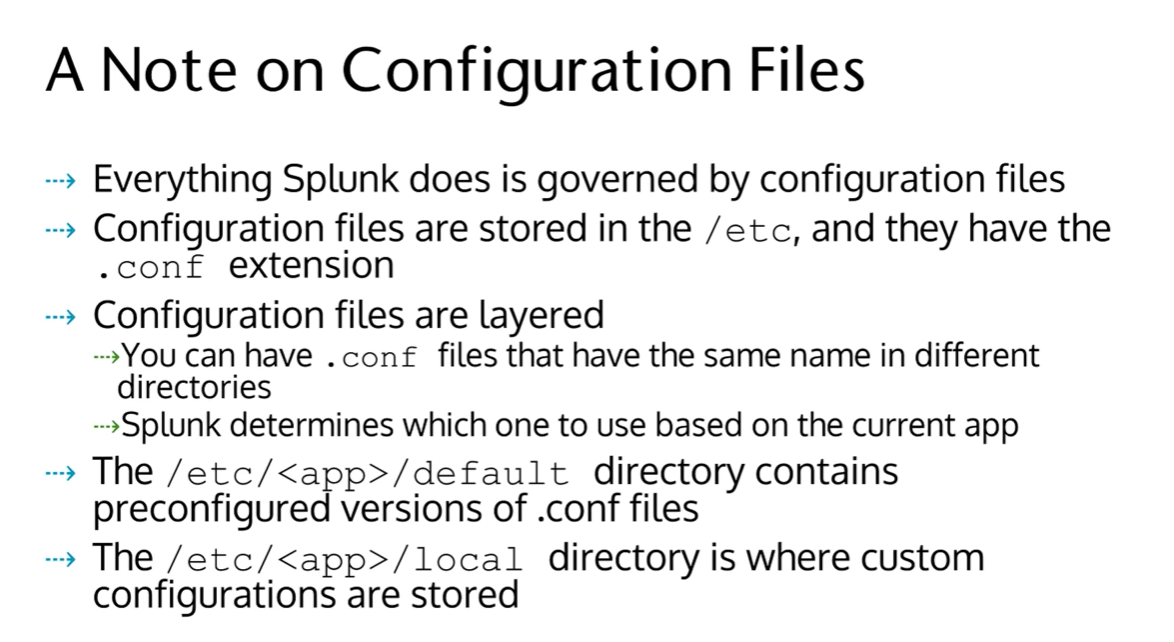
**Splunk basics:**

Splunk is primarily a business intelligence, event management and security incident tool which helps to gain insights from various types of data. It consumes data (mostly machine data) in real time to extract valuable insights for data-driven decision making

Splunk deployment models:



What’s inside a conf file?

[Stanza]

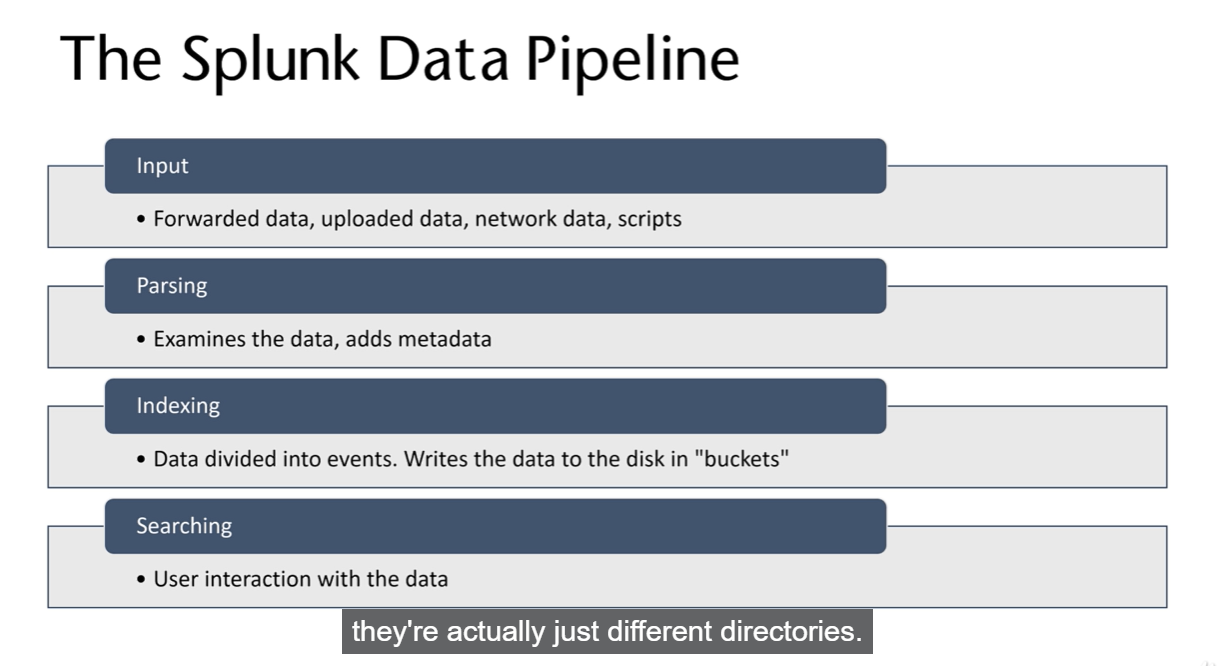
Attribute = Value

[Stanza]

Attribute = Value

Splunk deployment models:

1. Cloud
2. On-premise



Physical components that make up any splunk deployment model. These components can be configured in order to serve any of the above phases in the splunk data pipeline –



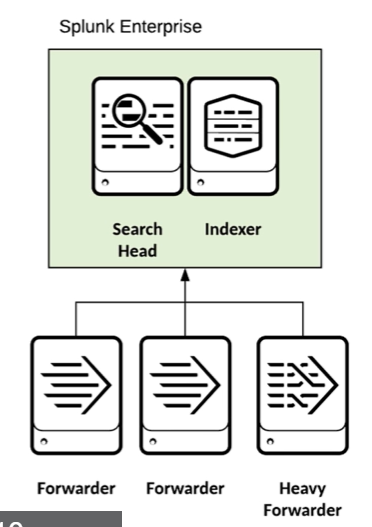
We need to install the universal forwarder explicitly.

The universal forwarder contains only the components that are necessary to forward data. A heavy forwarder is a full Splunk Enterprise instance that can index, search, and change data as well as forward it. The heavy forwarder has some features disabled to reduce system resource usage.

**Types of deployment models:**

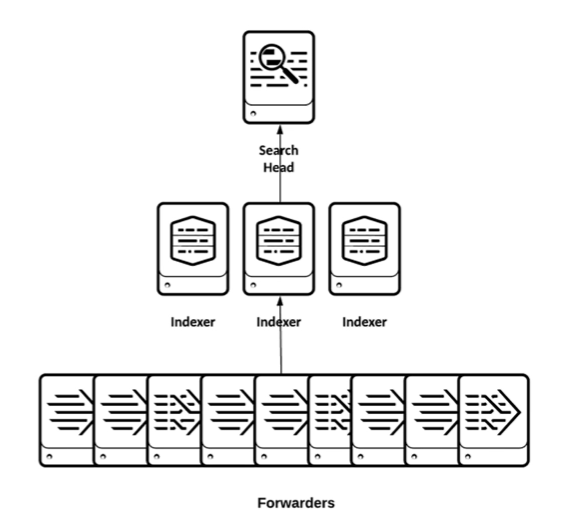
1. Departmental

* Single search head / indexer
* Suitable for 10 users as it has 10 forwarders



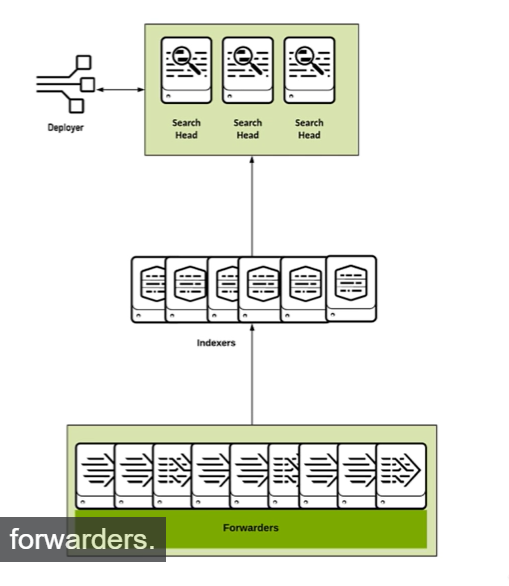
1. Small enterprise

* Independent search head
* 2-3 indexers
* 100-200 forwarders. Ideal for 100 users



1. Distributed search head

* Search head cluster – group of search heads which serves as a central resource for searching. Users can log into any member of the cluster and look for the same result set, view dashboards and run the same queries and enjoying the performance boost of the clustering. The search head cluster is governed by a deployer which distributes files, configurations and updates to all members of the cluster via configuration bundle.



1. Large enterprise

* Search head cluster
* Index head cluster – a group of indexers used for indexing data. This is failsafe and the data is replicated across all the nodes in the cluster in order to avoid loss of data.

**Data storage**

1. Splunk stores data by converting them into events (a single row of data) and storing them in indexes.
2. There are default indexes in splunk like (main and \_internal (stores logs)) and custom indexed created by users
3. It stores data in buckets which are essentially just directories. Each event is stored in key = value format
4. Depending on the age of the data, they are stored in various directories viz. –
5. Hotpath – for new indexed data. This is the place where users interact with data (Read / write)
6. WarmPath – it contains the data rolled from Hot bucket which has no active writing.
7. ColdPath – Contains data rolled from warm buckets and these data are stored in different location.
8. FrozenPath – Contains data rolled from cold buckets and the indexer deletes / archives the data rolled into them after a certain point of time.
9. Thawed – It contains the data restored from an archive.

Splunk app is basically a bunch of conf files which extends the splunk’s functionality. It may be created by individuals, third-party companies and splunk itself. These apps provide UI to the users in order to interact with splunk. The apps are downloaded in splunkbase.com

Splunk add-on enrichs the data with tags and data models.

**Splunk data pipeline deep – dive:**

When splunk consumes data, it divides the data into 64Kb blocks. These blocks are annotated with some metadata keys such as host, source, source type and other metadata associated with further processing of the data like index, character encoding etc.

The keys are applied to entire source and not events.

Types of data input:

1. File and directory input
2. Monitor files and directories

* Locally or remote monitor a directory or file for any update
* Monitor compressed files. The compressed files are uncompressed first and then uploaded into splunk

1. Upload

* Directly upload files for one time analysis

1. MonitorNoHandle

* Monitor rotating files. Only applicable for windows host

1. Network input
2. Capture data files sent over TCP, UDP and SNMP.
3. Windows input
4. Window event logs, registry, active directory, WMI (windows management interface) and perfmon
5. It is recommended to use splunk forwarders than WMI
6. Other sources
7. Metrics input
8. Scripted data – inputs from message queues and APIs
9. Modular inputs – build unique functionality of ingesting specific data from varying sources like querying a database
10. HTTP event collector

**Types of forwarders:**

1. Universal – It collects data from a source and sends to a receiver. It needs to be installed separately.
2. Heavy – It is an extension of universal forwarder where it not only collects data but also parses and routes under certain criteria

**Searching in splunk:**

SPL (splunk search processing language) – set of commands to search data in splunk. It is the primary way users interact with data like query, search, transform, visualize, manipulate.

This feature is primarily handled with search and reporting app that comes with splunk enterprise default installation

The search and reporting app can be used to –

1. Index data
2. Configure data
3. Build dashboards
4. Create visualization and reports

Searching basically happens in a pipeline format where we slowly filter out from a glob of data and use the previously processed data using pipe operator | in SPL.

Handling time in splunk:

Time is a mandatory field in splunk and it assigns timestamp to every single data even if there isn’t. \_time attribute is added by default. This is in Unix time. In UI / web, it is displayed as human-readable form.

How splunk determines which timestamp value to assign?

1. Look for timestamp value in the events itself
2. Searches the filename or source name for timestamp values
3. Look for the file modification time
4. As a last resort, assign current timestamp value.

**Specifying absolute time ranges using SPL –**

Earliest = < %m/ %d/ %y : %H:%M:%S > latest = < %m/ %d/ %y : %H:%M:%S >

For e.g: Earliest = 12/25/2021:00:00:00 latest = 04/30/2024:16:30:00

**Specifying relative time ranges using SPL –**

Time units – years (y), months (mon), days (d), weeks (w), hours (h), minutes (m), seconds (s)

-30m -> 30 mins ago

-7d -> 7 days ago

+1d -> 1 day from now

**Converting time using strftime**

| eval <new\_field\_name> = strftime (<time\_field>, “format”)

For eg : | eval new\_time = strftime (\_time, “%I:%M %p)

**Basic Searching:**

1. Broad search - metadata

* Index

Index = main

* Host

Host = server.com, Host = 192.158.0.3

* Source, Sourcetype

Source = /var/lib, sourcetype = csv

1. Broad search terms

Keywords – failed, error

Wildcards – \*, fail\*, \*success\*, user = \*

Phrases – “failed login”

Booleans – OR, NOT, AND [blank space means AND by-default]

Fields – key-value pairs ( serverA = abc.com )

1. Basic search terms

* Chart / timechart – returns results in tabula format
* Rename – renames a specific field
* Sort – sort results by specific fields
* Stats – statistics (like min, max, avg etc)
* eval – evaluates an expression
* dedup – deletes the duplicate values
* table – builds a table with specific fields

**Field discovery**

1. Splunk auto. discovers default fields like host, source, sourcetype, \_time etc.
2. Splunk looks for obvious key-value pairs in the first 100 events. NOTE: It sometime considers key value instead of key = value valid as well.
3. Field can also be extracted from user / app customizations. These are stored configuration files.

There are 3 levels of field discovery also known as search modes.

1. Fast – recommended to be used when you know all the fields to be searched
2. Smart (default)
3. Verbose – recommended to be used when you have little knowledge about the data

Custom field extraction can be built using Splunk field extractor. It uses regex to find discover field with the specified pattern.