Splunk - Beginner to Architect

Use-Case - Demo Corp Bank

Demo Corp Bank operates hundreds of servers in its datacenter, along with various applications, networking devices, and firewalls.

Challenge: Monitoring the application and security logs for suspicious activity requires manually logging into each server to review the logs.

Potential Solution

An organization can push logs from all the applications and devices to the centralized platform.

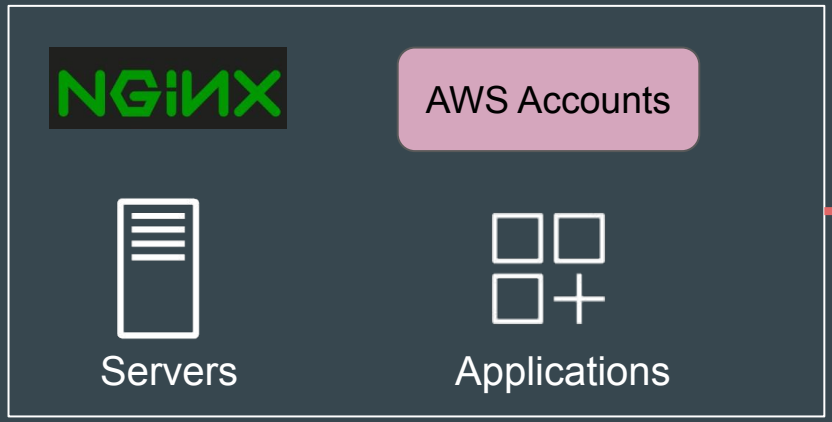
This centralized platform can be used for monitoring and analysis.

Central Analysis Platform

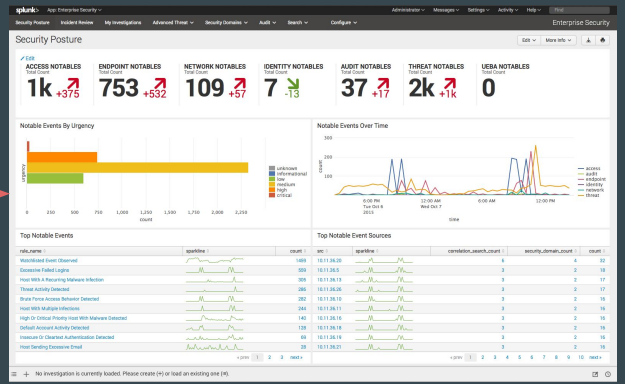
Setting the Base

Splunk is one of the most popular log analyzing and monitoring tools.

Splunk allows us to search, analyze and visualize data gathered from a wide variety of devices.



Push Logs



Reference Screenshot - Logs Events in Splunk

i	Time	Event
>	21/06/2025 04:59:03.000	66.249.68.128 - - [21/Jun/2025:04:59:03 +0000] "GET / HTTP/1.1" 302 145 "-" "Mozilla/5.0 (Linux; Android 6.0.1; Nexus 5X Build/MMB29P) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/137.0.7151.103 Mobile Safari/537.36 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)" "-" host = kp_labs source = access-logs.zip:./access-logs/access.log sourcetype = access_combined
>	21/06/2025 04:56:53.000	65.49.1.144 - - [21/Jun/2025:04:56:53 +0000] "GET /geoserver/web/ HTTP/1.1" 404 153 "-" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10.14; rv:62.0) Gecko/20100101 Firefox/62.0" "-" host = kp_labs source = access-logs.zip:./access-logs/access.log sourcetype = access_combined
>	21/06/2025 04:55:00.000	65.49.1.148 - - [21/Jun/2025:04:55:00 +0000] "GET /favicon.ico HTTP/1.1" 404 555 "-" "Mozilla/5.0 (Macintosh; Intel Mac OS X 14_4) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/123.0.0.0 Safari/537.36" "-" host = kp_labs source = access-logs.zip:./access-logs/access.log sourcetype = access_combined
>	21/06/2025 04:53:57.000	40.124.176.81 - - [21/Jun/2025:04:53:57 +0000] "GET /developmentserver/metadatauploader HTTP/1.1" 404 153 "-" "Mozilla/5.0 zgrab/0.x" "-" host = kp_labs source = access-logs.zip:./access-logs/access.log sourcetype = access_combined
>	21/06/2025 04:53:45.000	65.49.1.146 - - [21/Jun/2025:04:53:45 +0000] "GET /webui/ HTTP/1.1" 404 153 "-" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/605.1.15 (KHTML, like Gecko) Version/16.3 Safari/605.1.15" "-" host = kp_labs source = access-logs.zip:./access-logs/access.log sourcetype = access_combined
>	21/06/2025 04:52:16.000	65.49.1.150 - - [21/Jun/2025:04:52:16 +0000] "GET / HTTP/1.1" 200 82 "-" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10.14; rv:62.0) Gecko/20100101 Firefox/62.0" "-" host = kp_labs source = access-logs.zip:./access-logs/access.log sourcetype = access_combined
>	21/06/2025 04:49:13.000	198.55.98.76 - - [21/Jun/2025:04:49:13 +0000] "GET /.env HTTP/1.1" 404 555 "-" "Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/536.6 (KHTML, like Gecko) Chrome/20.0.1092.0 Safari/536.6" "-" host = kp_labs source = access-logs.zip:./access-logs/access.log sourcetype = access_combined

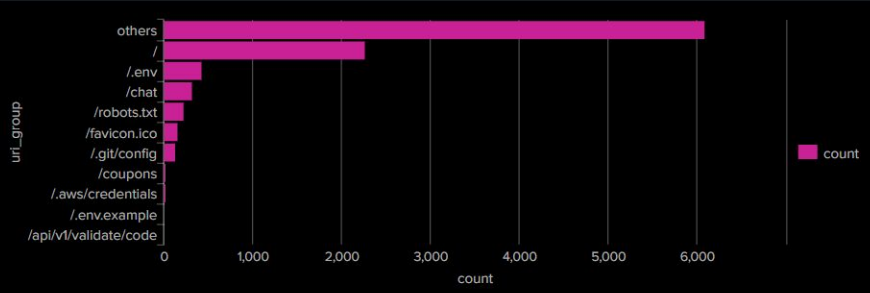
Reference Screenshot - Dashboards Created

Request the server getting every day

Total number of request the server getting every day

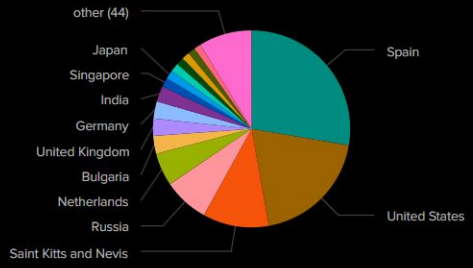
uri request

URI requests (/chat, /coupons and /) and others



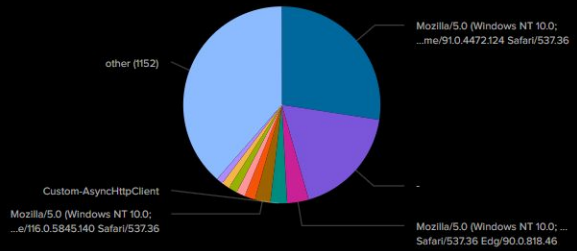
geolocation

Countries from where request are coming



user agent details

User agent details



3 Important Benefits

1. All the metrics, and logs from all devices can be analyzed centrally.

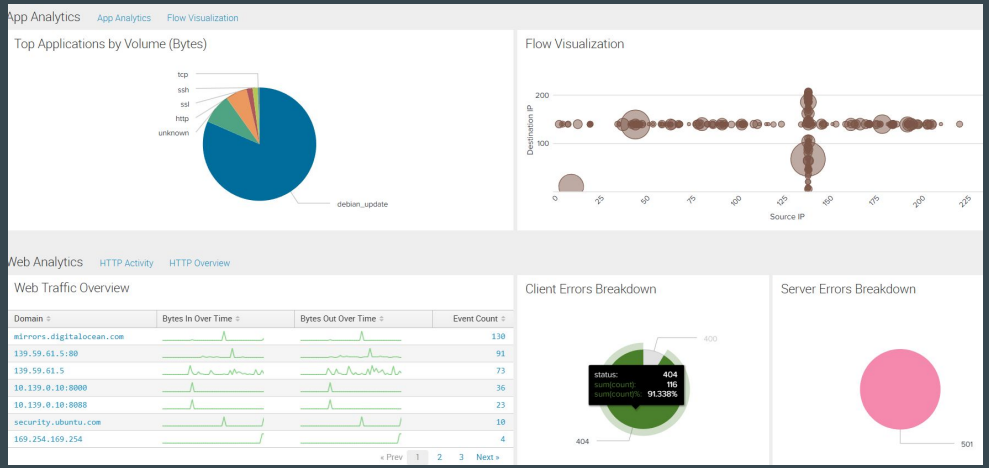
2. Generate dashboards, alarms, and alerts on suspicious activity.

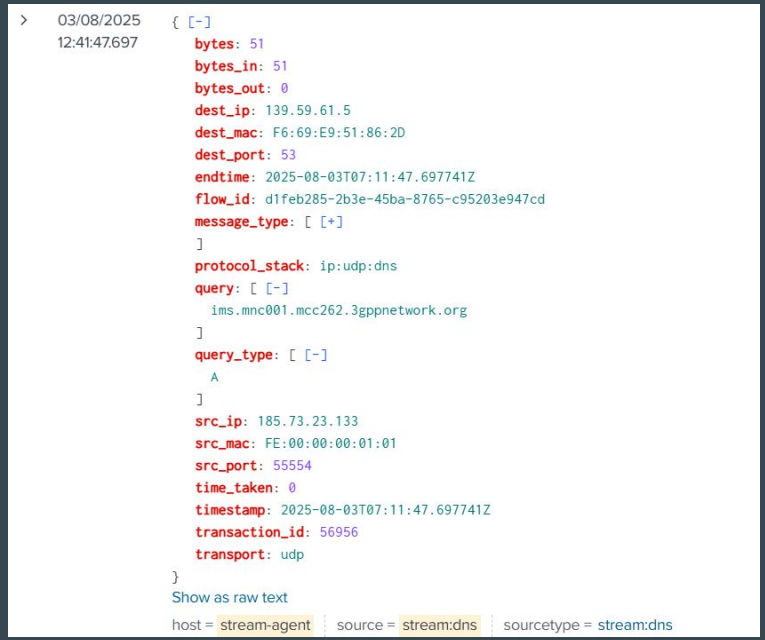
3. Correlate all the logs to find any attack pattern.

Figure 1: Magic Quadrant for Security Information and Event Management

Interesting Part - Network Events

Splunk can also monitor network traffic at server level, such as DNS traffic, TCP/UDP traffic, HTTP packet captures, etc..

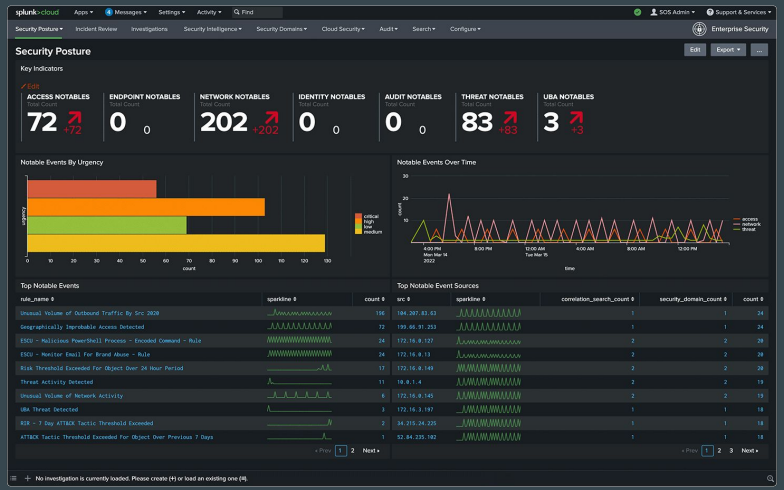




My Journey with Splunk - Part 1

We had 100+ AWS accounts, 1000+ servers, WAF, IPS etc across organization.

I led the efforts to build a highly-available, distributed Splunk cluster to support heavy ingestion of logs to detect security attacks for the entire organization.

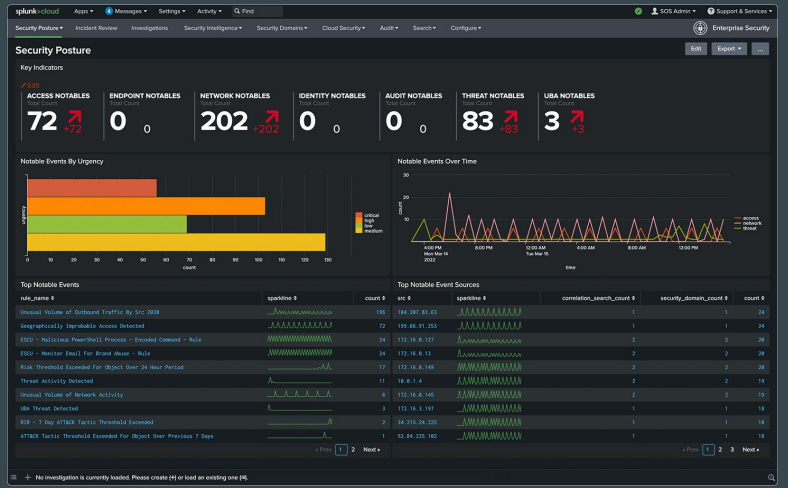


My Journey with Splunk - Part 2

A dedicated SOC team used to monitor Splunk dashboards and alerts 24/7.

Splunk License Type we used: 500 GB per day.

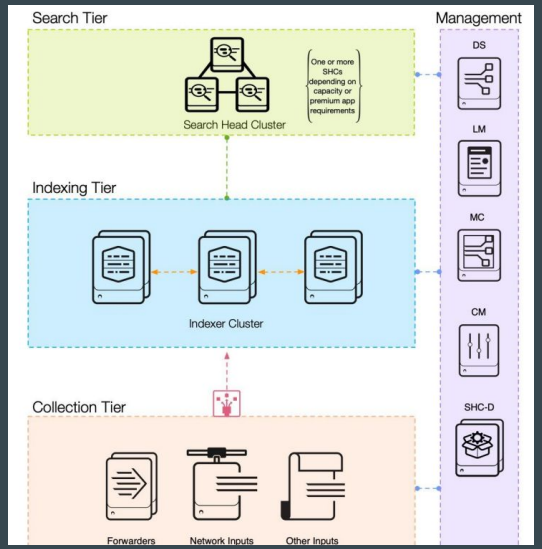
SOC Team



What is this course all about?

This course will help you build a very strong foundation in Splunk from an architectural perspective.

You will learn how to ingest logs, analyse attack vectors, build dashboards, set up and deploy a highly available Splunk for your organisation.



About Me

- DevSecOps Engineer - Defensive Security.
- Teaching is one of my passions.
- I have total of 16+ courses, and around 420,000+ students now.

Something about me :-

- RedHat Certified Architect, Certified Ethical Hacker.
- AWS Certified [Advanced Networking, Security Specialty, DevOps Pro, SA Pro, …]
- HashiCorp Certified [Terraform Professional [Vault and Consul Associate]
- Part time Security Consultant + 13 more Certifications

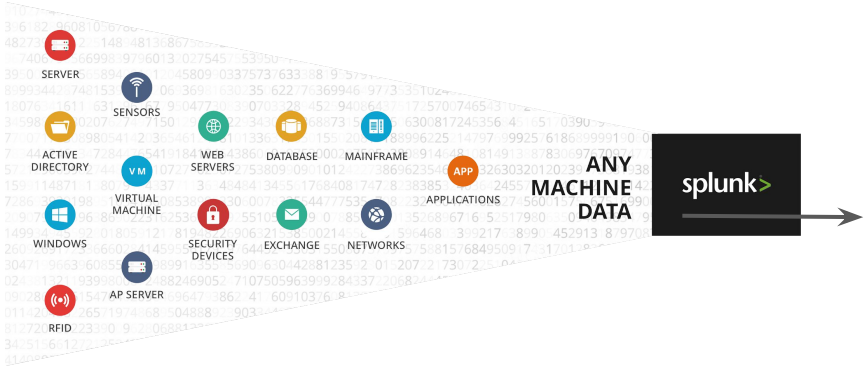
Introduction to Splunk

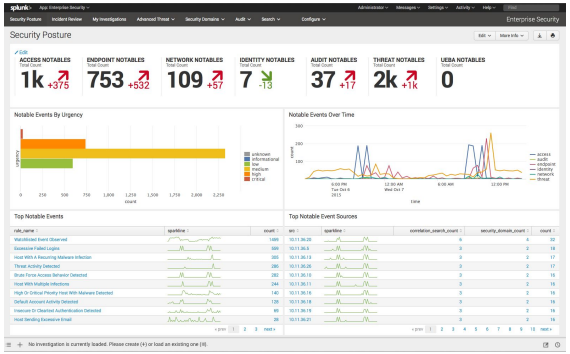
It’s just awesome!

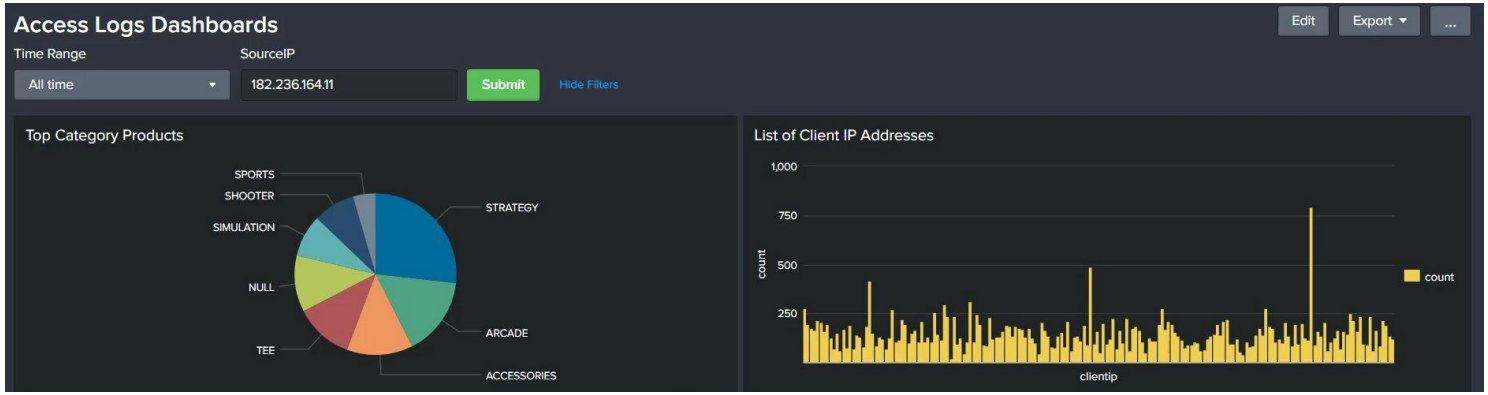
Overview of Splunk

Splunk is one of the most popular log analyzing and monitoring tools.

Splunk is a software platform to search, analyze and visualize the machine-generated data gathered from wide variety of devices.







>	10/5/14 9:47:41.000 PM	Sun Oct 05 2014 21:47:41 mailsv1 sshd[2174]: Failed password for invalid user operator from 69.175.97.11 port 3401 ssh2 host = mailsv1 source = /opt/log/mailsv1/secure.log sourcetype = linux_secure
>	10/5/14 9:44:19.000 PM	Sun Oct 05 2014 21:44:19 www3 sshd[4904]: Failed password for invalid user operator from 10.2.10.163 port 1679 ssh2 host = www3 source = /opt/log/www3/secure.log sourcetype = linux_secure
>	10/5/14 9:40:18.000 PM	Sun Oct 05 2014 21:40:18 mailsv1 sshd[28961]: pam_unix(sshd:session): session closed for user myuan by (uid=0) host = mailsv1 source = /opt/log/mailsv1/secure.log sourcetype = linux_secure
>	10/5/14 9:36:31.000 PM	Sun Oct 05 2014 21:36:31 mailsv1 sshd[3910]: Failed password for invalid user proxy from 10.2.10.163 port 2572 ssh2 host = mailsv1 source = /opt/log/mailsv1/secure.log sourcetype = linux_secure
>	10/5/14 9:36:04.000 PM	Sun Oct 05 2014 21:36:04 mailsv1 sshd[89792]: pam_unix(sshd:session): session opened for user myuan by (uid=0) host = mailsv1 source = /opt/log/mailsv1/secure.log sourcetype = linux_secure
>	10/5/14 9:24:24.000 PM	Sun Oct 05 2014 21:24:24 www1 sshd[3145]: Failed password for invalid user proxy from 10.3.10.46 port 3175 ssh2 host = www1 source = /opt/log/www1/secure.log sourcetype = linux_secure
>	10/5/14 9:21:08.000 PM	Sun Oct 05 2014 21:21:08 mailsv1 sshd[2784]: Failed password for invalid user operator from 10.2.10.163 port 4797 ssh2 host = mailsv1 source = /opt/log/mailsv1/secure.log sourcetype = linux_secure
>	10/5/14 9:11:54.000 PM	Sun Oct 05 2014 21:11:54 www2 sshd[4787]: Failed password for invalid user operator from 86.9.190.90 port 1062 ssh2 host = www2 source = /opt/log/www2/secure.log sourcetype = linux_secure
>	10/5/14 9:09:40.000 PM	Sun Oct 05 2014 21:09:40 www2 sshd[4939]: Failed password for invalid user operator from 10.2.10.163 port 1677 ssh2 host = www2 source = /opt/log/www2/secure.log sourcetype = linux_secure
>	10/5/14 9:02:11.000 PM	Sun Oct 05 2014 21:02:11 www2 sshd[5623]: Failed password for invalid user operator from 210.192.123.204 port 3584 ssh2 host = www2 source = /opt/log/www2/secure.log sourcetype = linux_secure

Powerful Marketplace

Splunk has its own marketplace referred as splunkbase were people can submit their apps and addons.

This allows customers to use out of box solution for wide variety of use-cases.

Splunk is More than Log Monitoring Solution

When a software platform is powerful in searching, analyzing and visualizing, it can be used for much more wider areas.

Splunk has been promoting new apps in various niche specific areas like:

- Security information and event management (SIEM)
- Splunk IT Service Intelligence
- Splunk User Behavior Analytics

Figure 1. Magic Quadrant for Security Information and Event Management

Source: Gartner (December 2017)

Creating Splunk Account

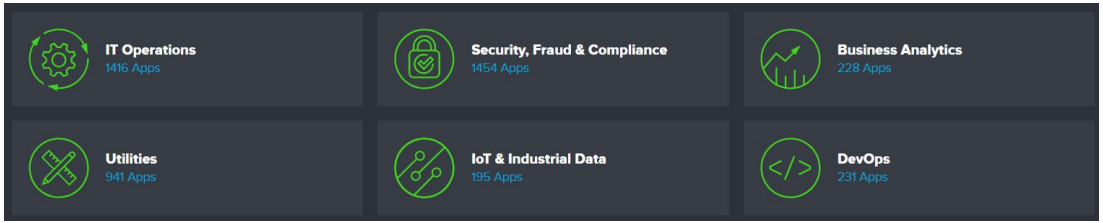
Let's get started!

Overview of Splunk Account

Creating a Splunk account is an important first step and it allows users to perform various operations.

Some of these include:

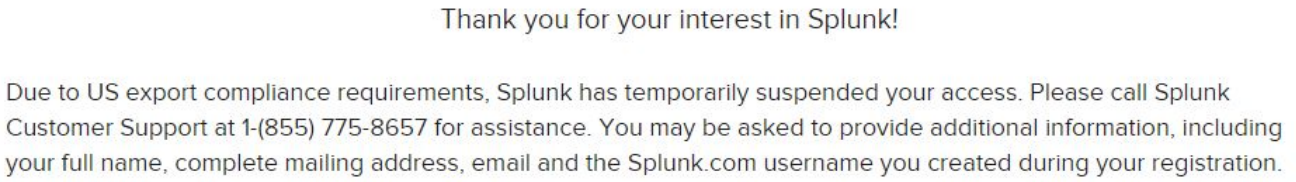
- Free Trials and Downloads
- Download Apps and Add-Ons from Splunk Marketplace.



Important Note - SignUp Process

Sometimes, Splunk signup process might fail due to restrictions based on name and countries part of a consolidated list.

The Consolidated List is a list of all persons and entities who are subject to targeted financial sanctions

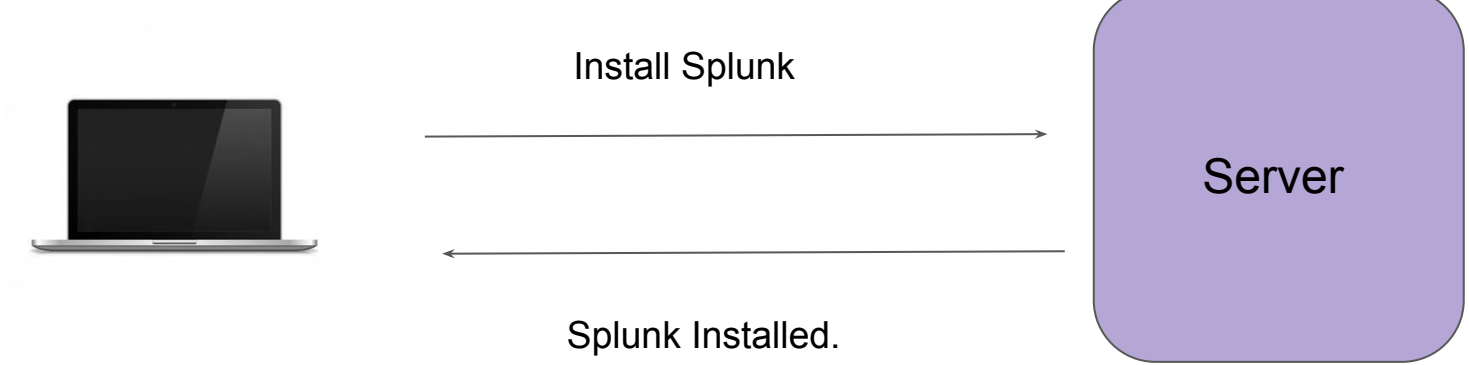


Infrastructure for Splunk

Let's get started!

Creating Infrastructure for Splunk

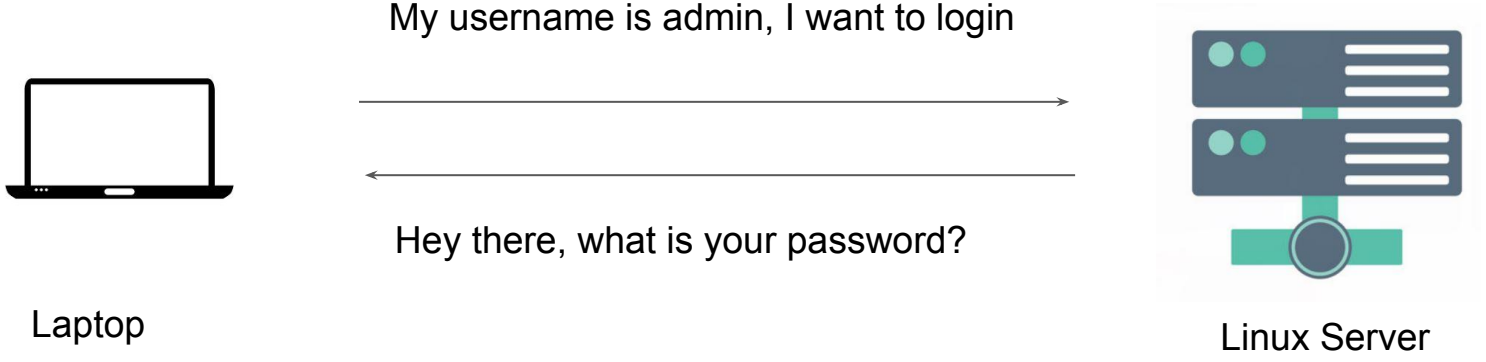
To begin with Splunk installation process, we need one server hosting Ubuntu OS.



Password Based Authentication

There can be multiple methods for authentication against a system.

Password based authentication is the simplest form.

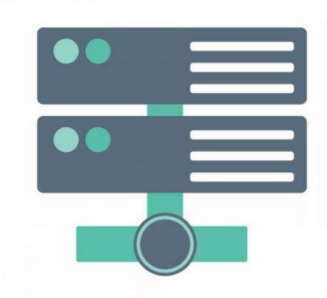


Login with Credentials

Laptop

My username is admin, my password is 12345. I want to login.

Login successful.



Linux Server

Challenges with Password Based Authentication

Password based authentication is generally considered to be less-secure.

Many users write down the passwords in notepad files or as part of sticky notes.

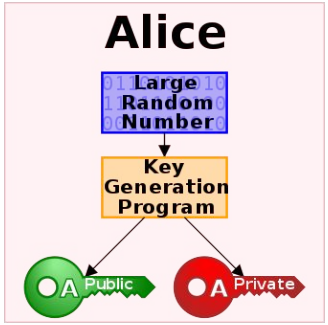
Most users would not create a complex password that is difficult to hack.

Key Based Authentication

In this type of authentication, there are two special keys that are generated.

One key is called as Public Key and second key is called as Private key.

If public key is stored in server and is used as authentication mechanism, only the corresponding private key can be used to successfully authenticate.

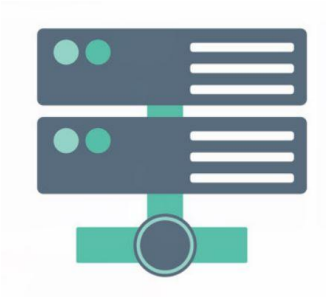


Key Based Authentication

Laptop

My username is admin, I want to login

Hey there, password is not allowed.
You need to authenticate via the key.



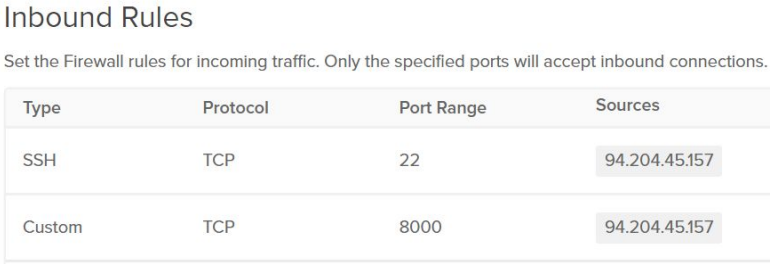
Linux Server

Firewall Rules

We do not want the entire internet to connect to our server.

With the help of Firewall, you can restrict the connection to your Splunk instance.

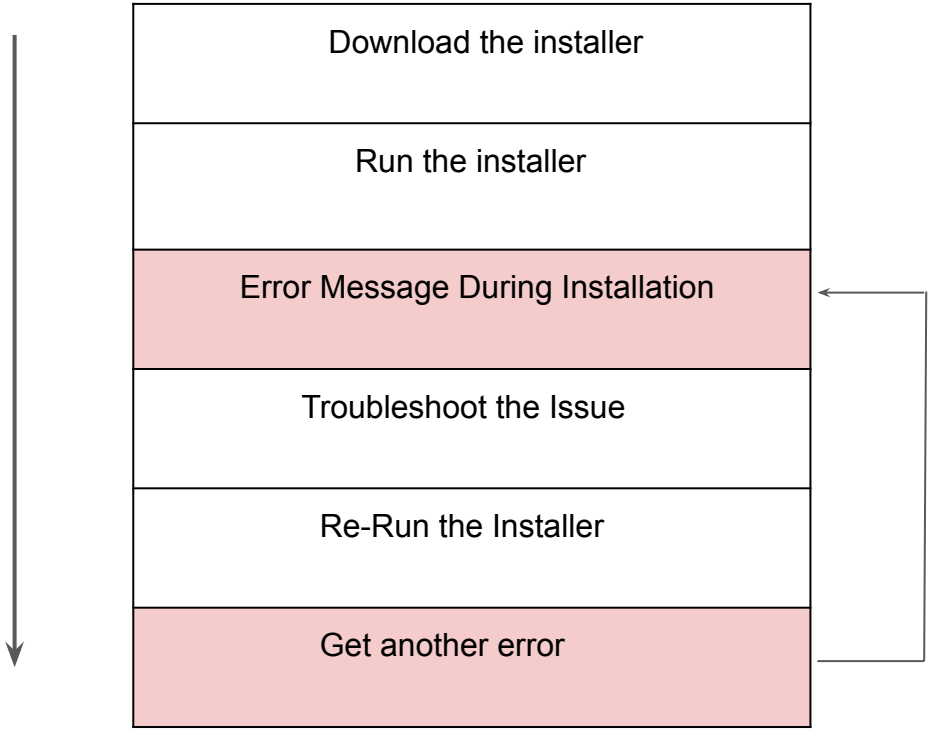
Ports	Description
22	Connection to SSH.
8000	Connection to Splunk.



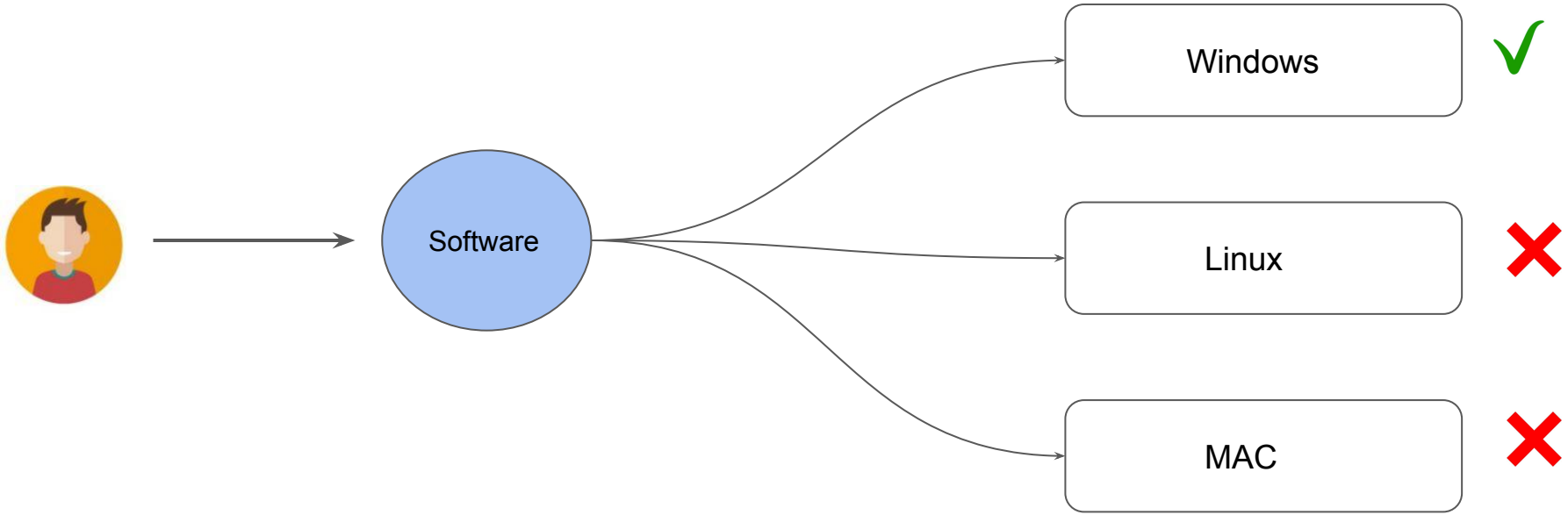
Introduction to Docker

Build once, use anywhere

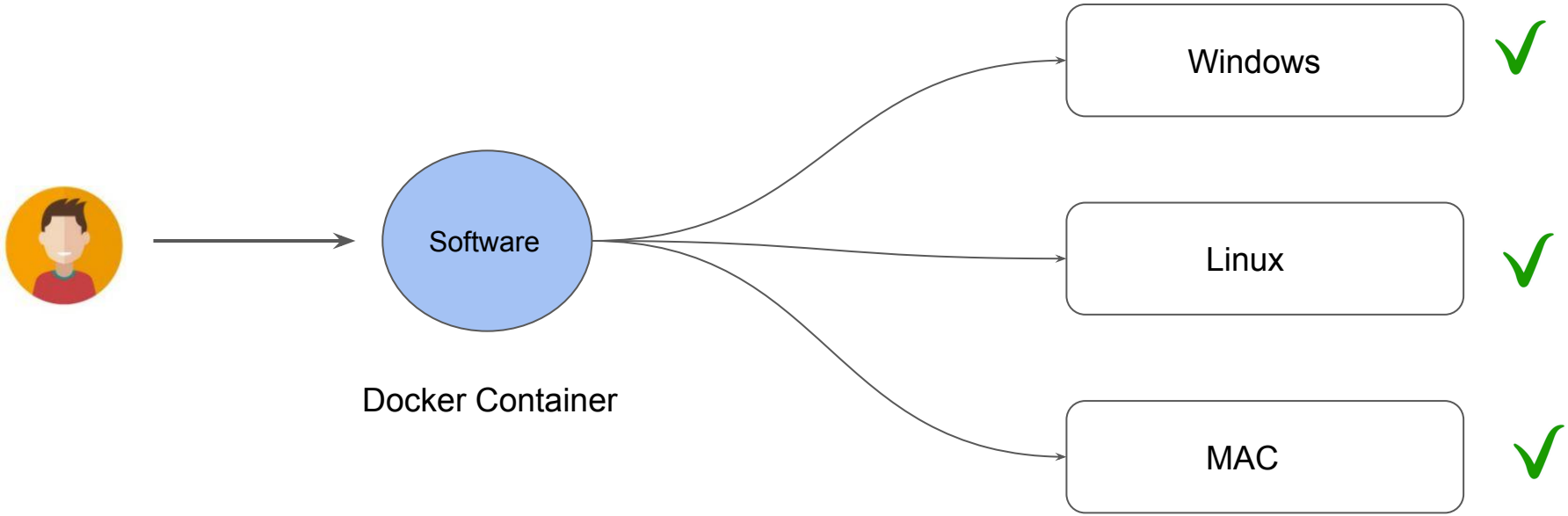
Installation of Software Workflow



What is Docker Trying to Achieve?



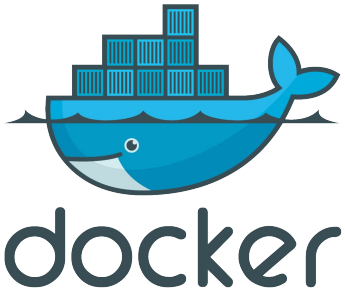
What is Docker Trying to Achieve?



Introduction

Docker is an open platform, once we build a docker container, we can run it anywhere, say it windows, linux, mac whether on laptop, data center or in cloud.

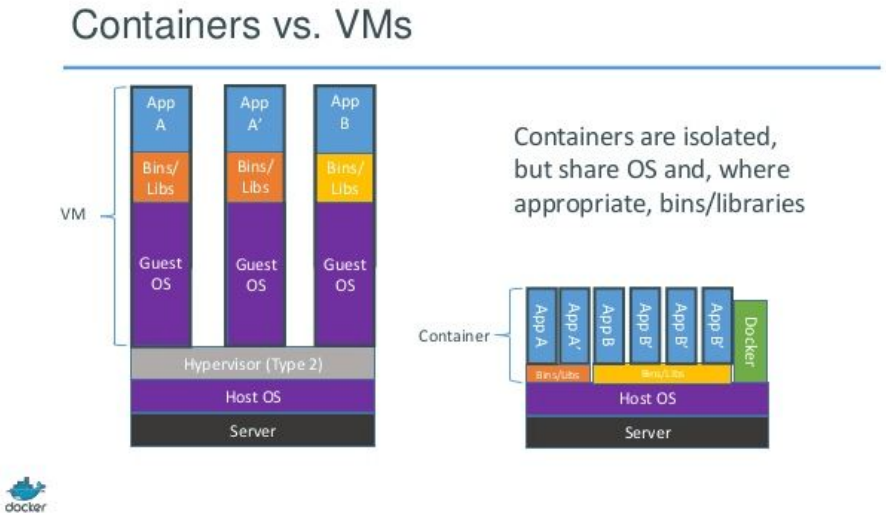
It follows the build once, run anywhere approach.



Containers vs Virtual Machines

Virtual Machine contains entire Operating System.

Container uses the resource of the host operating system

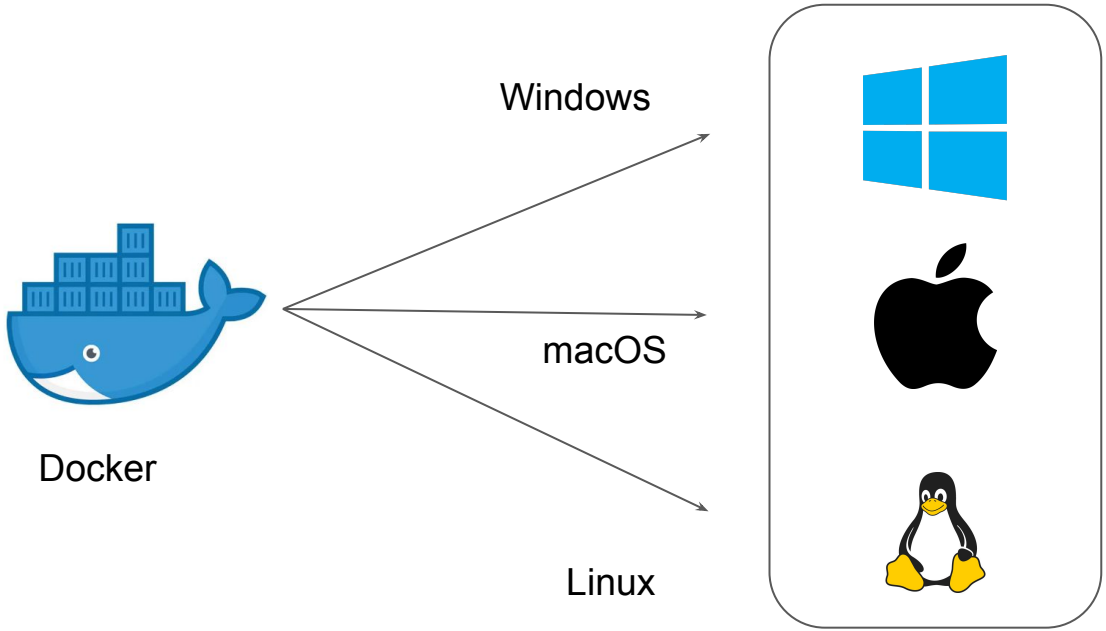


Installation Methods of Docker

Let's Install

Installing Docker

Docker can be installed in wide variety of operating systems.



Preferred Choice for Docker Installation Method

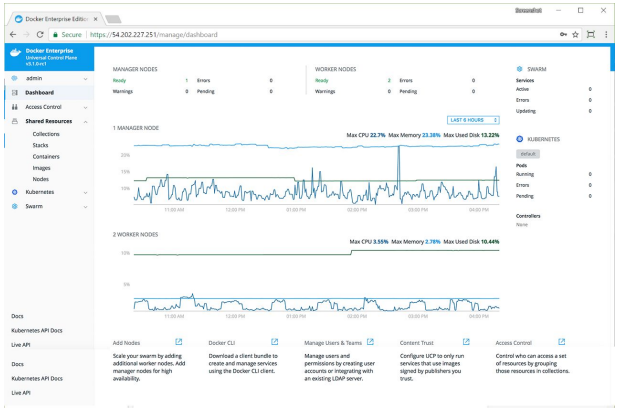
To begin with, you can install Docker Desktop directly within your laptop.

The preferred OS for Docker installation would be Linux.

Why Linux is Preferred Method?

In this course, we will be testing many Docker features and enterprise features.

- Basic Docker Features
- Docker Swarm
- Docker UCP/DTR
- Kubernetes



Revising the Choices

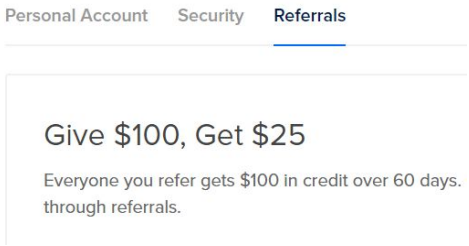
Following diagram illustrates the architecture that we will be following in this course:

Criteria	Choices
Operating System	Ubuntu
Cloud Provider	Digital Ocean

Why Digital Ocean?

1. They provide multiple coupon codes which gives great amount of credits ranging from $50-100 USD for new users.

2. Provides Managed K8s cluster which will be required in the later section of the course.

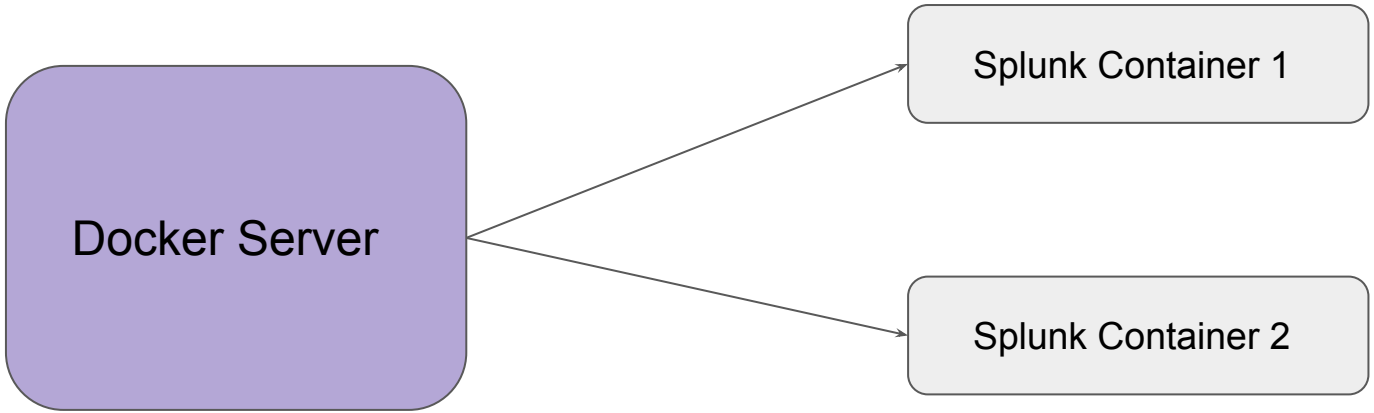


Launching Splunk in Docker

Let's get started!

Splunk Infrastructure via Docker

Once we have Docker up and running, we can launch Splunk Docker container to get started.



Basic Docker Commands

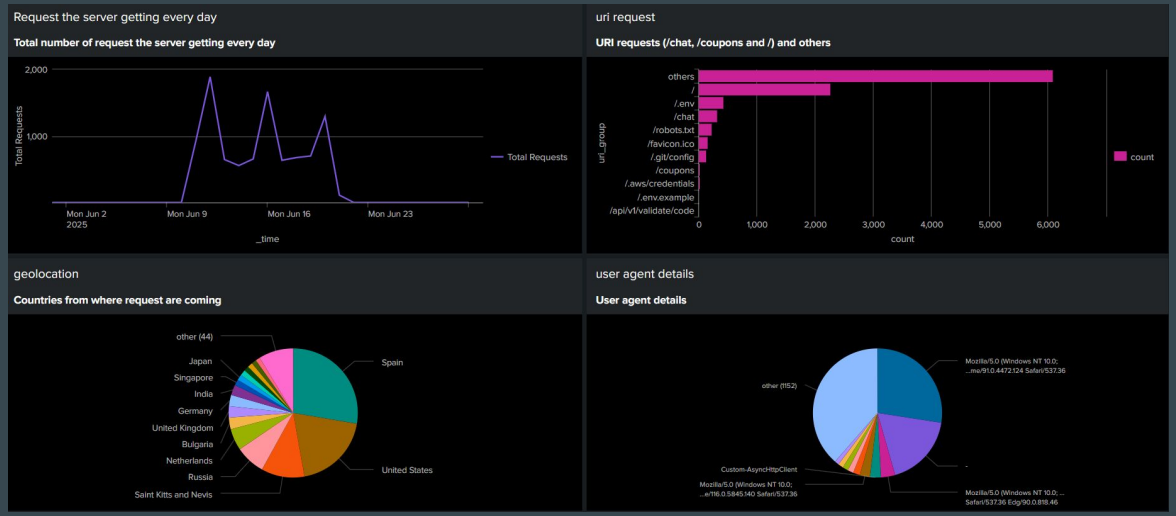
Let us explore the most important Docker commands that will often be used.

Basic Commands	Description
docker ps	Shows list of running containers.
docker ps -a	Show all containers
docker stop [container-name]	Stops Docker Container
docker start [container-name]	Starts Docker container
docker rm [container-name]	Remove Docker container

Onboarding Data to Splunk

Setting the Base

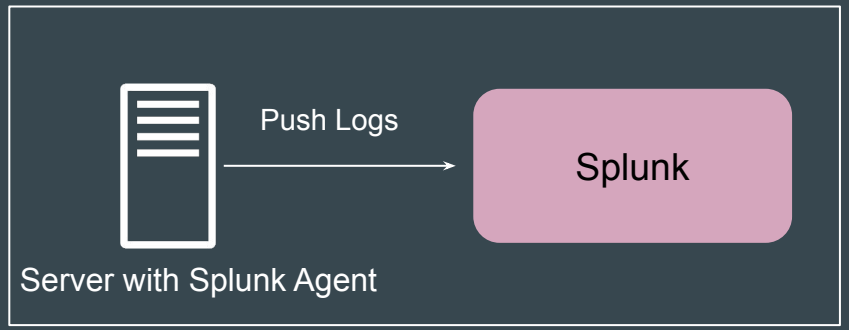
Before we start building amazing dashboards in Splunk, we first need to add some data to Splunk.



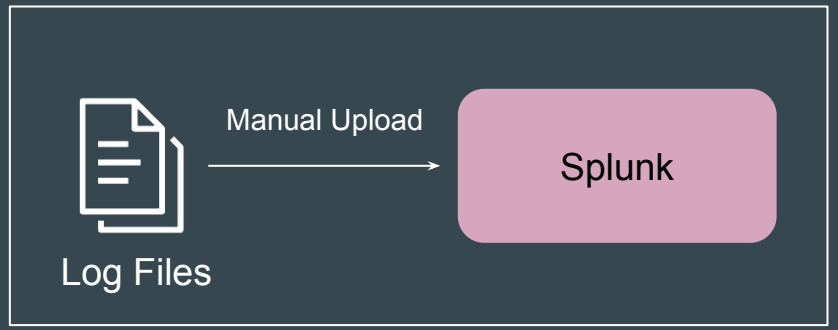
Adding Data to Splunk

There are multiple ways to add data to Splunk.

1. Install Splunk agent on Servers. Agent pushes logs to Splunk.
2. Manually upload log files to Splunk instance.
3. Splunk can connect to 3rd party platforms like AWS to fetch data.



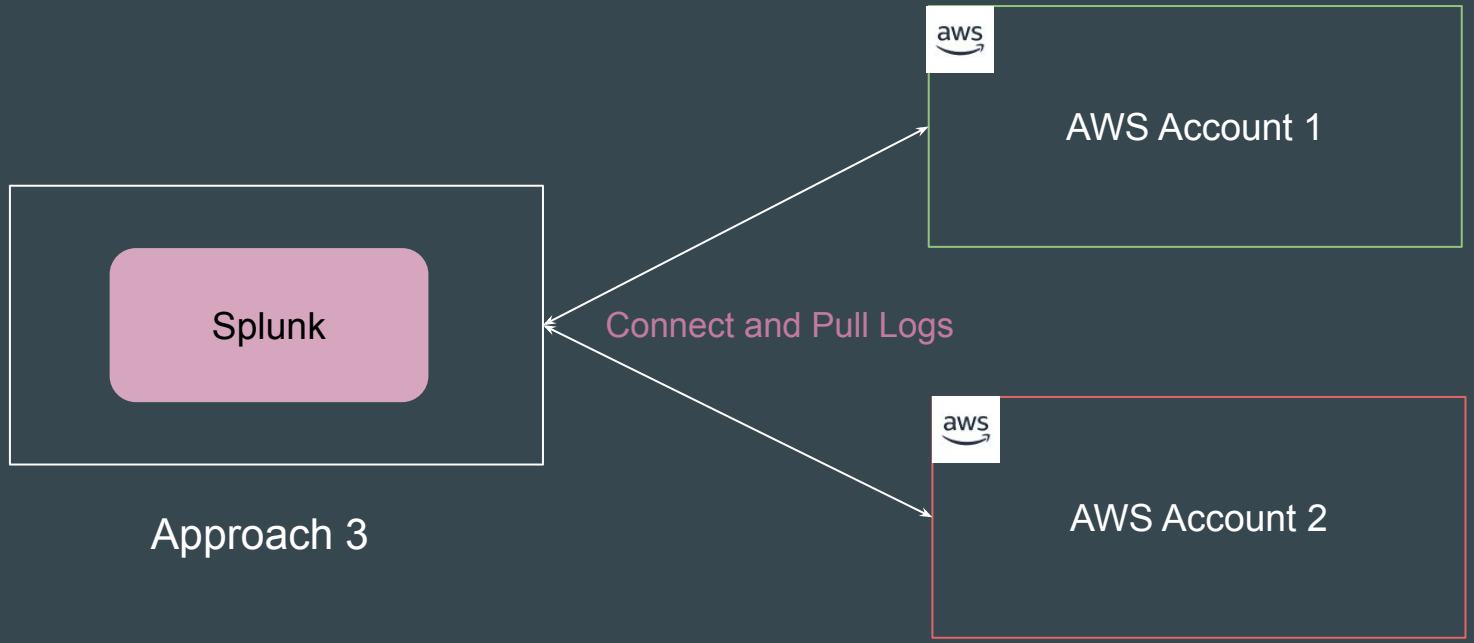
Approach 1



Approach 2

Approach 3

Splunk can connect to 3rd party platforms like Cloud platforms to fetch data.



Approach 3

Source Types in Splunk

Importance of Field Extractions

Field extraction refers to the process of identifying specific pieces of information (fields) from log entries, such as timestamps,user IDs, etc.

182.236.164.11 - - [25/July/2025:18:20:56] "GET /cart.do?action=addtocart&productId=BS-AG-G09 HTTP 1.1" 200 2252 "Mozilla/5.0 (Macintosh; Intel Mac OS X)" 506

Field Extraction

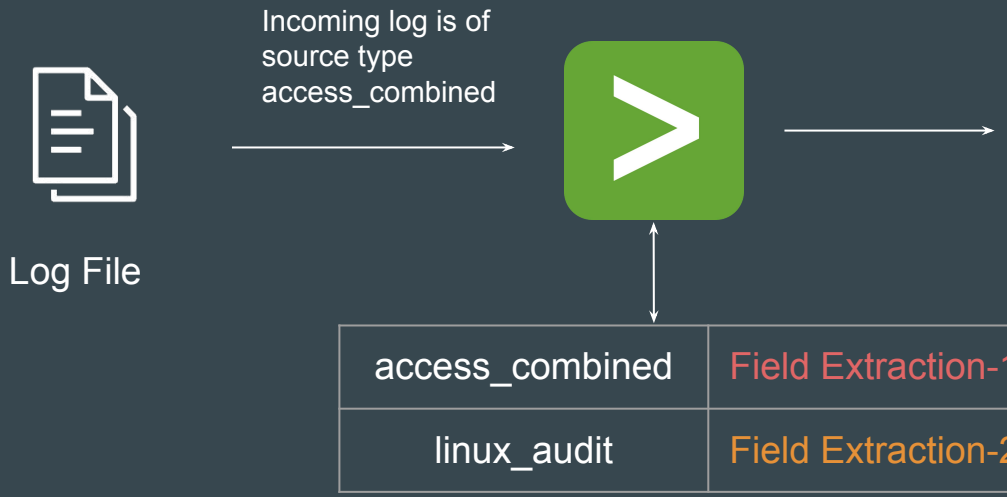
Field	Value
clientip	182.236.164.11
request_time	25/July/2025:18:20:56
product_id	BS-AG-G09
action	addtocart

Important Question

How does Splunk know what individual fields are from a log file?

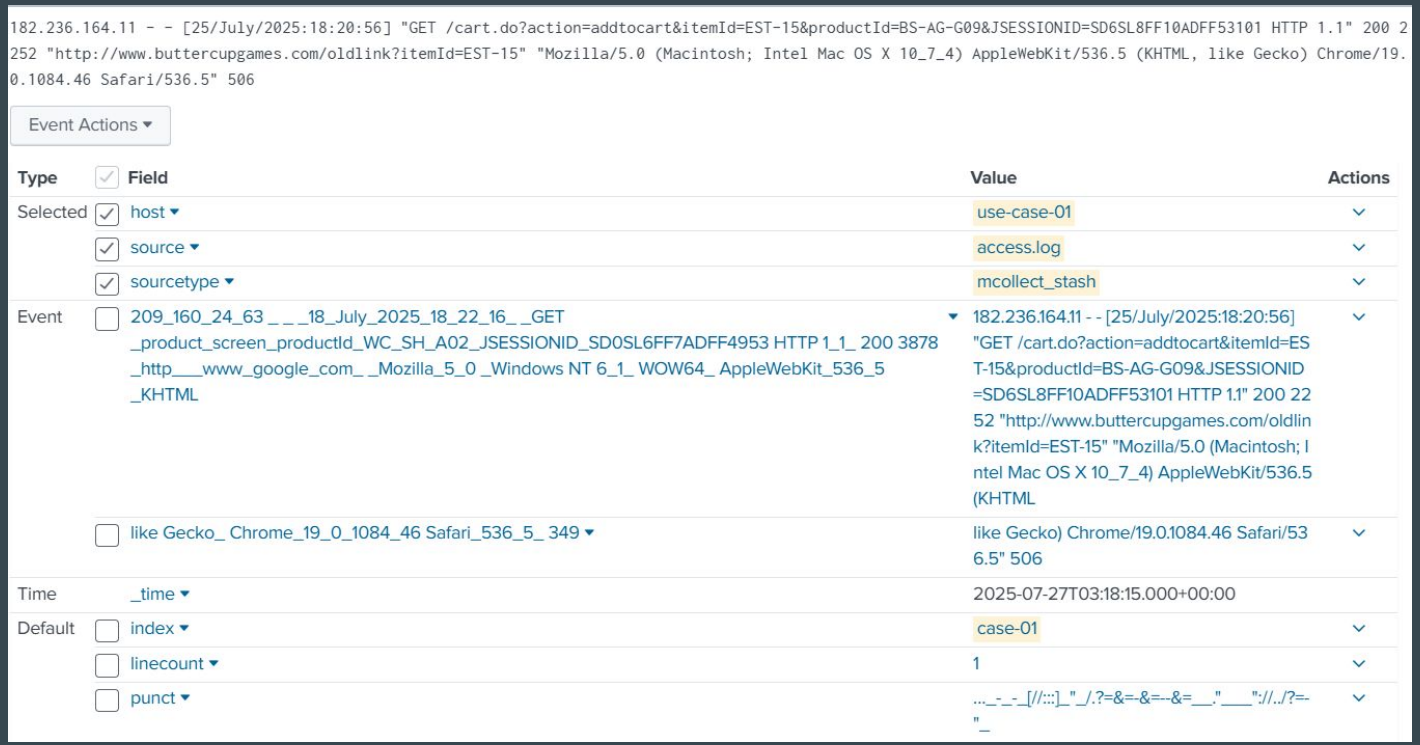
Basics of Source Types

A source type is Splunk's way of identifying the format of your data, so that Splunk can parse and extract fields automatically.

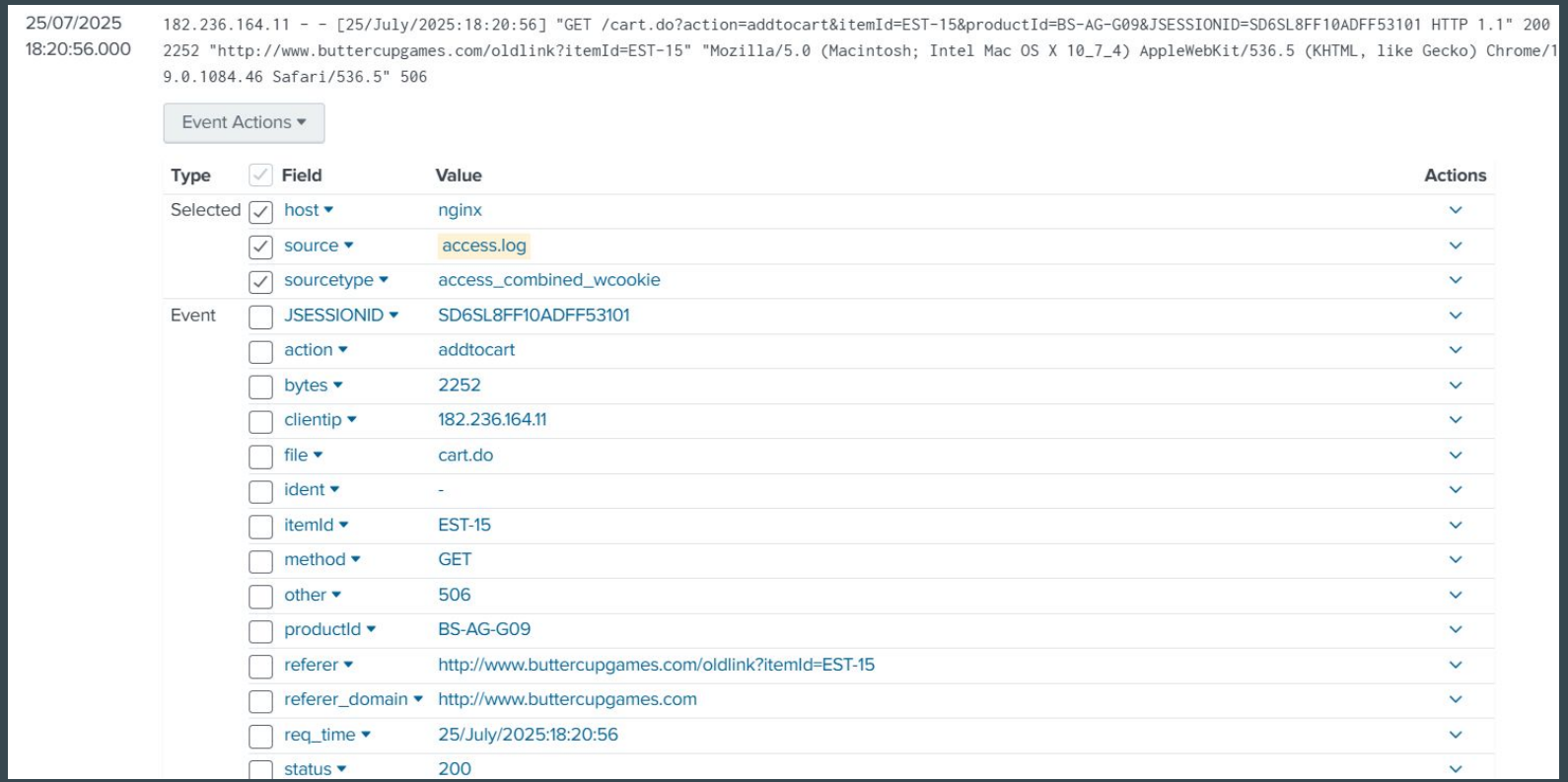


Type		Field	Value
Selected	✓	host ▾	DESKTOP-LE3E08O
	✓	source ▾	access.log
	✓	sourcetype ▾	access_combined
Event		JSESSIONID ▾	SD6SL8FF10ADFF53101
		action ▾	addtocart
		bytes ▾	2252
		clientip ▾	182.236.164.11
		file ▾	cart.do
		ident ▾	-
		itemId ▾	EST-15
		method ▾	GET
		other ▾	506
		productId ▾	BS-AG-G09
		referer ▾	http://www.buttercupgames.com/oldlink?itemId=EST-15
		referer_domain ▾	http://www.buttercupgames.com
		req_time ▾	25/Oct/2018:18:20:56
		status ▾	200

Example 1 - Data with Wrong Source Type



Example 2 - Data with Correct Source Type



Information about Field Extraction

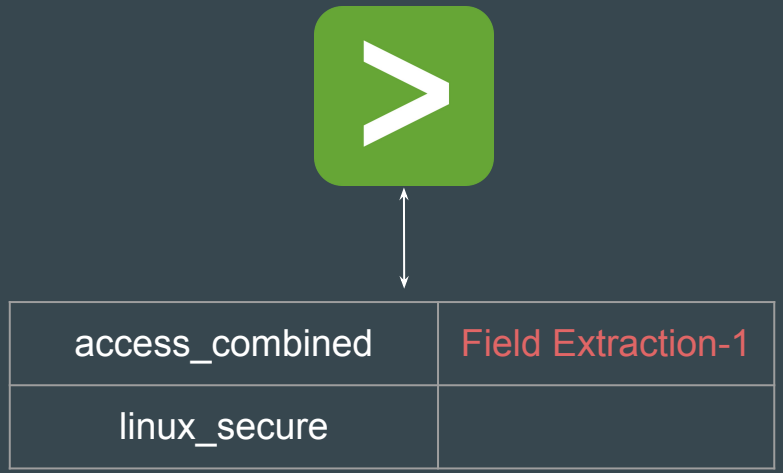
Each sourcetype in Splunk must internally be associated with the appropriate field extractions; otherwise, Splunk will not automatically extract the relevant fields from the log events.

access_combined	Field Extraction-1
linux_audit	Field Extraction-2
mysqld	Field Extraction-3

Default Configuration

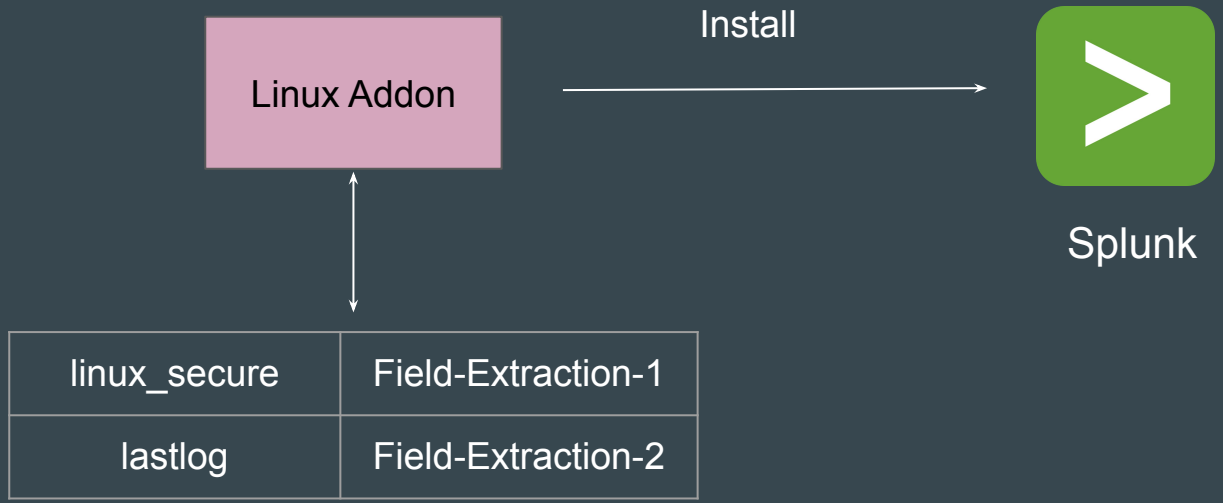
Assigning the correct sourcetype to your data in Splunk does not guarantee that fields in log events will be extracted automatically.

In many cases, you may need to install relevant Splunk Add-ons that provide the necessary field extractions for specific source types.

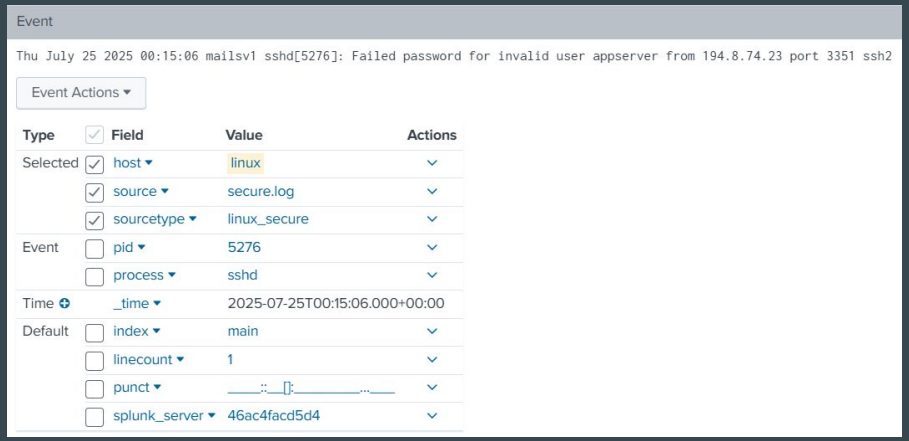


Splunk Addon to the Rescue

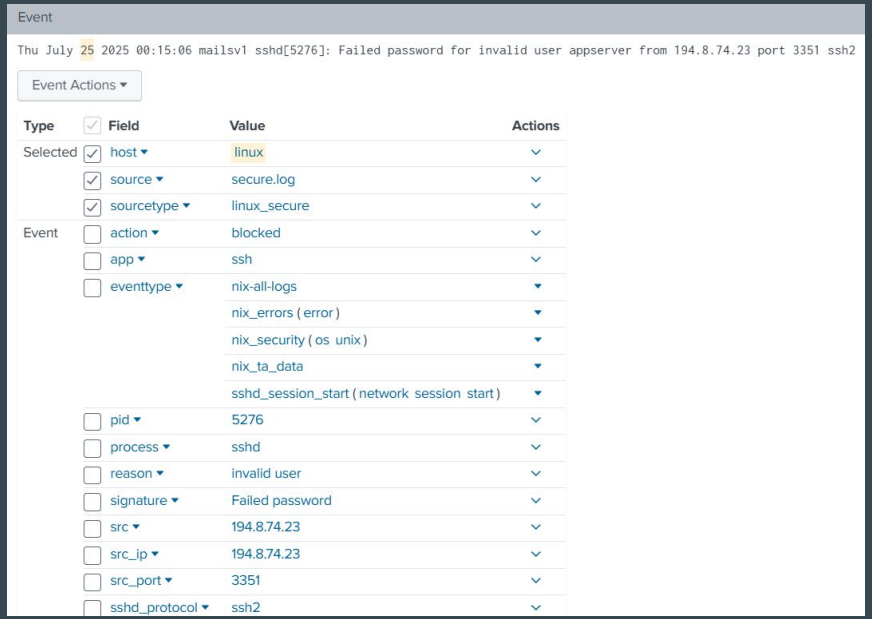
For logs where field extraction does not happen properly, you can install the relevant Splunk add-ons from marketplace that has the necessary field extractors.



Reference - Before and After Add-On Installation



Before Linux Addo-On

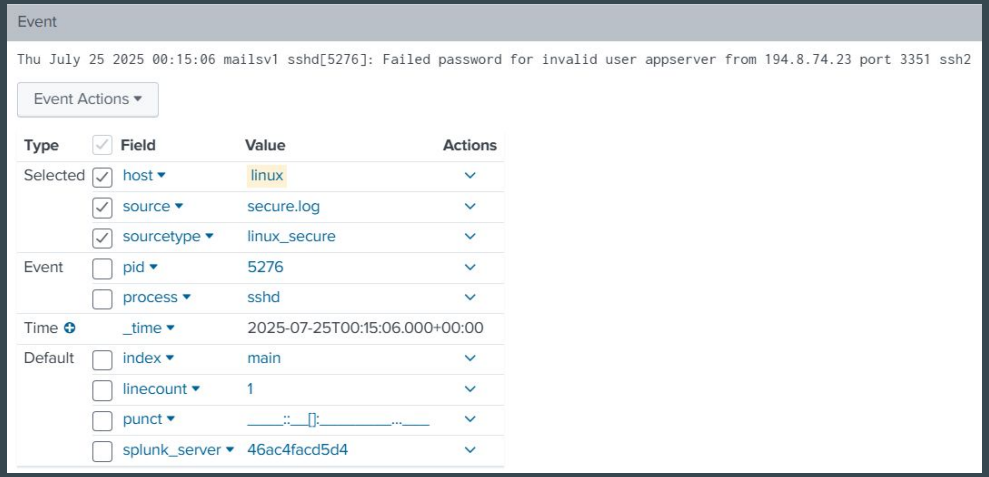


After Linux Addon-On

Field Extraction - Linux Secure Logs

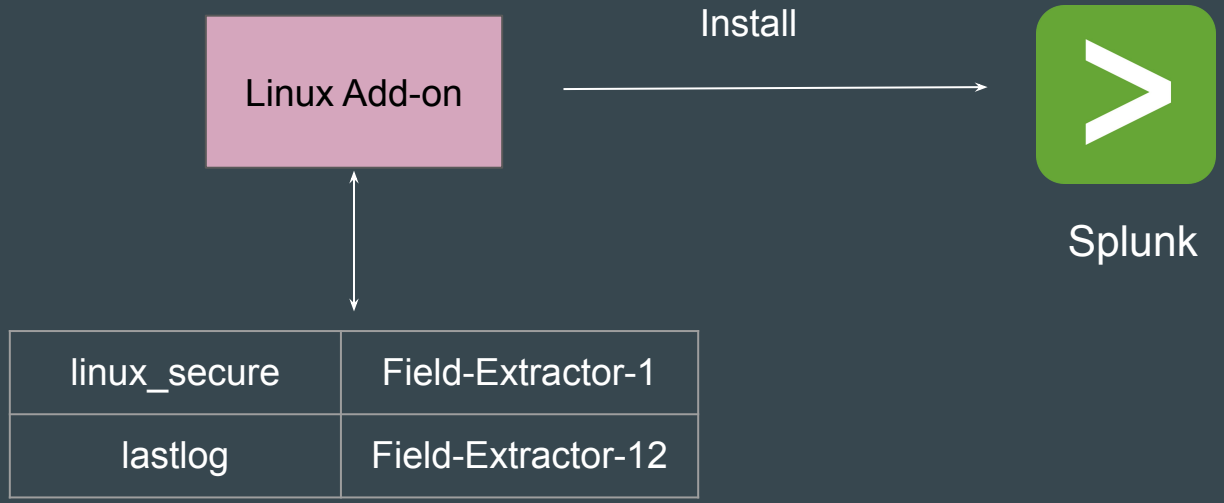
The Current Challenge

We had uploaded log events from /var/log/secure file to Splunk, but proper field extraction is not happening.



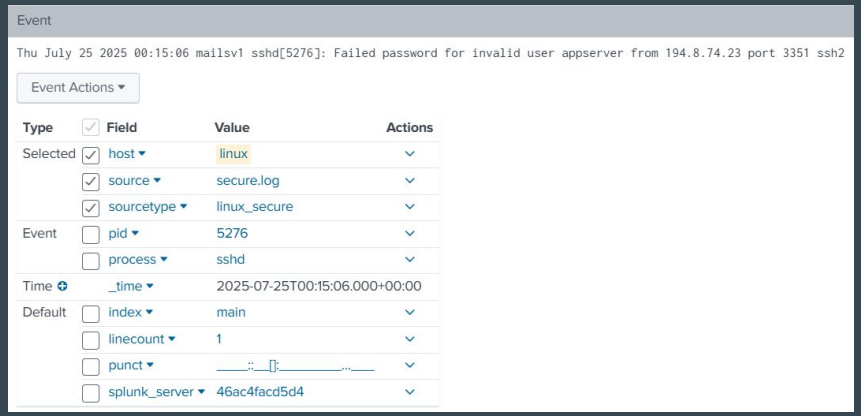
Installing Splunk Addon

For logs where field extraction does not happen automatically, you can install various Splunk Add Ons from marketplace that has the necessary field extractors.

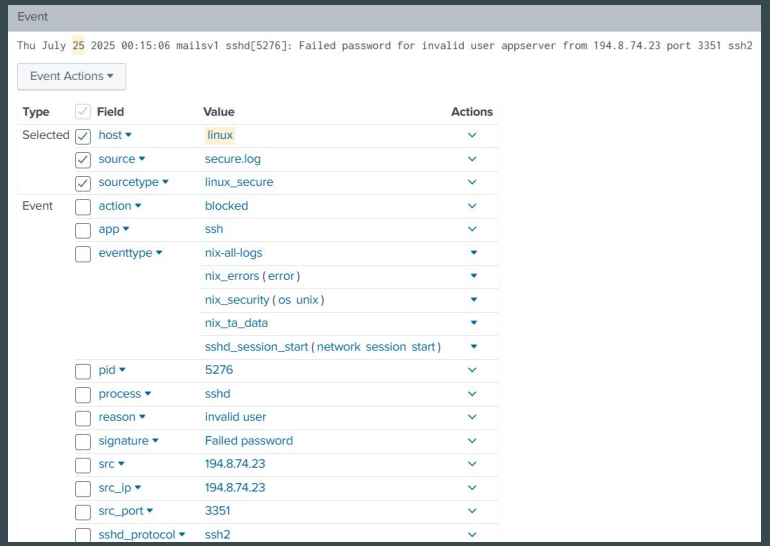


Splunk Addon for Linux

When you install the Splunk Add-on for Linux, it comes with a variety of prebuilt field extractions for multiple types of Linux logs.



Before Addon-On

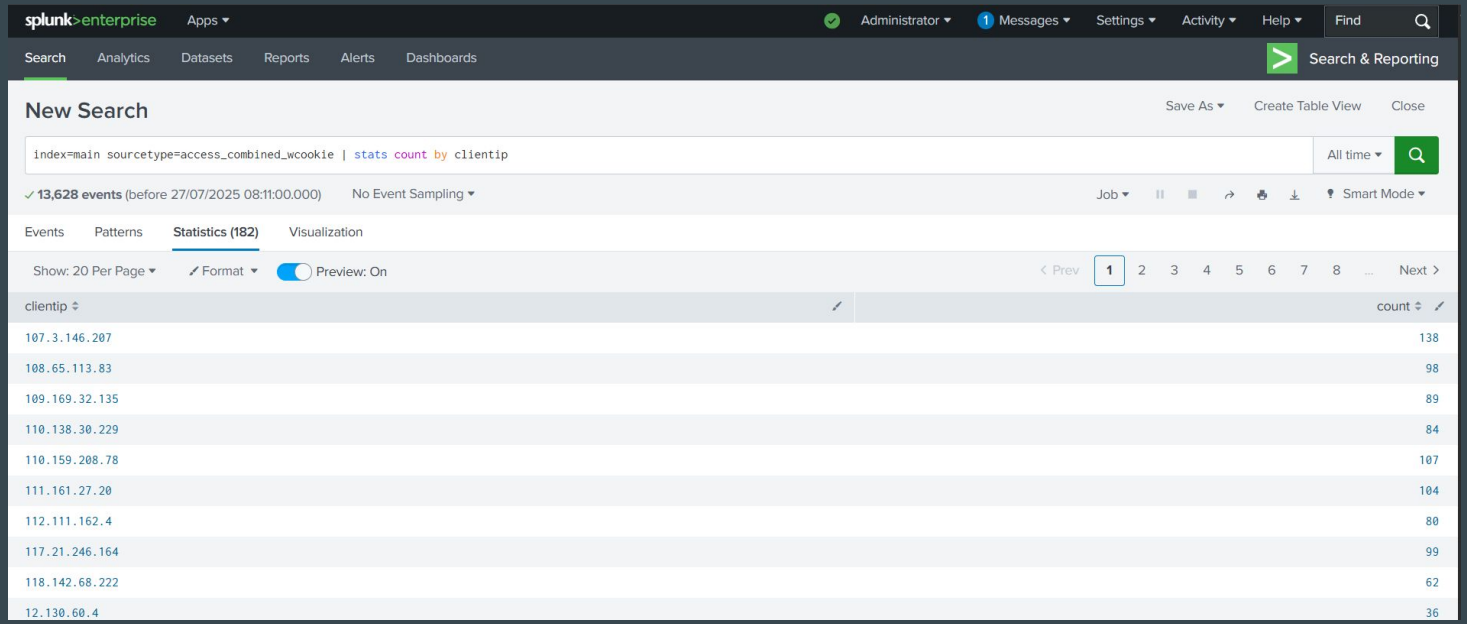


After Linux Addon-On

Basics of Search

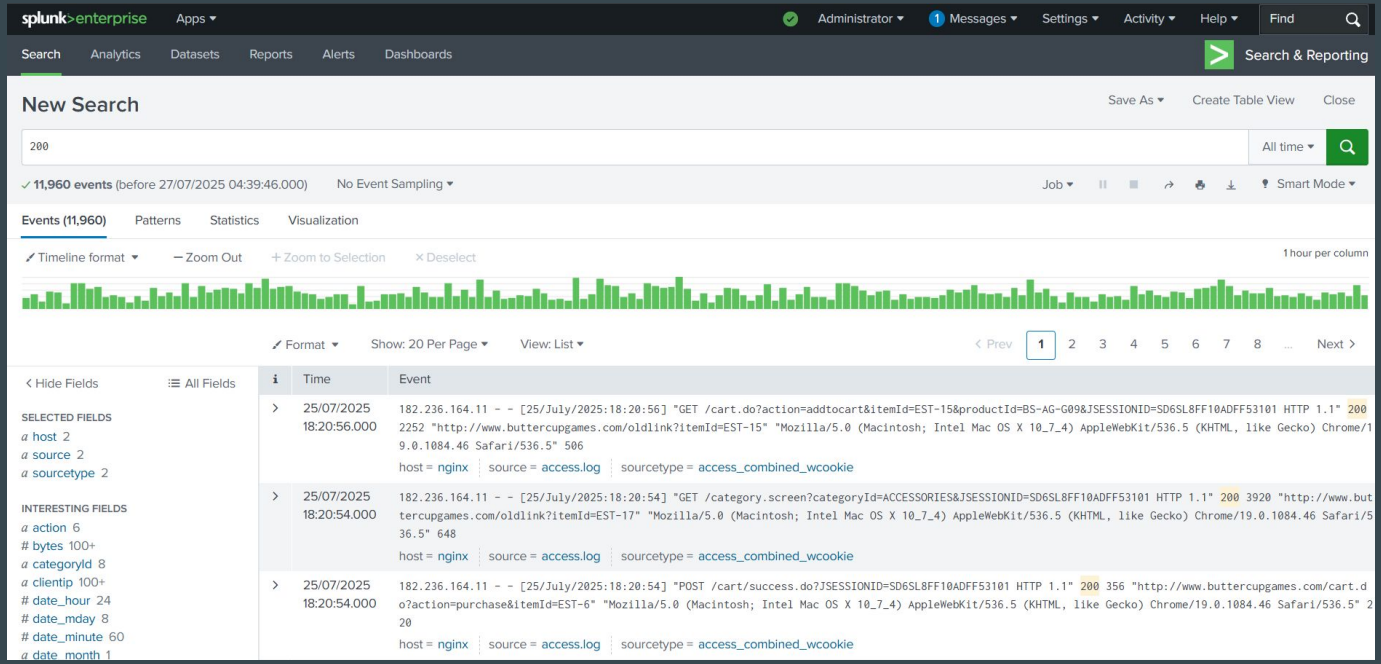
Search Processing Language (SPL)

The Search Processing Language (SPL) allows you to search, analyze, and visualize data within the Splunk platform.



Simple Search

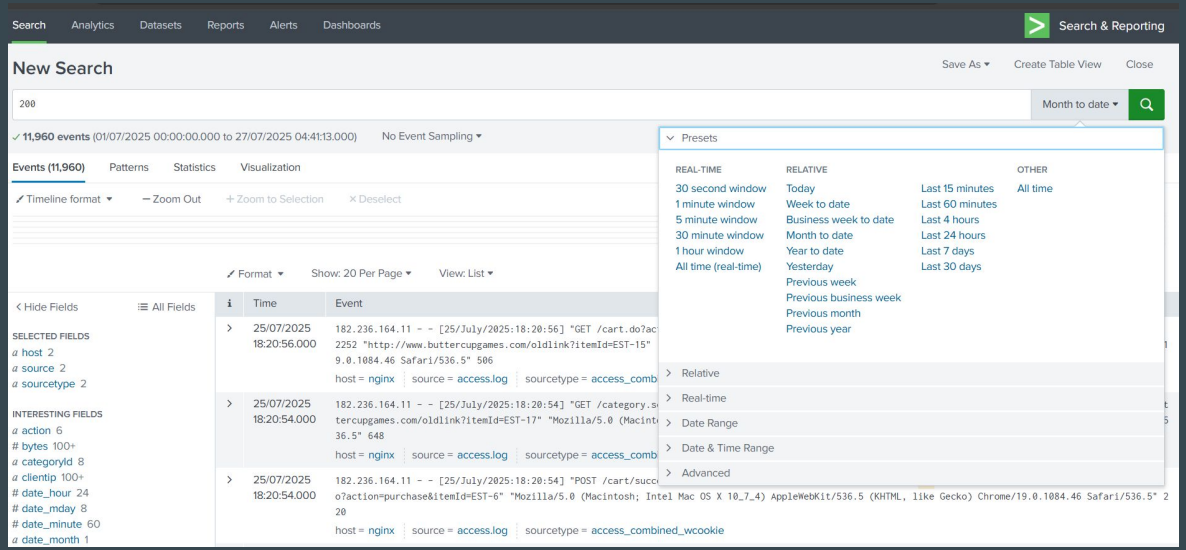
One of the easiest ways to search for specific data is to type the data value in the search string.



Time Range Picker

Restricting, or filtering, your search criteria using a time range is the easiest and most effective way to optimize your searches.

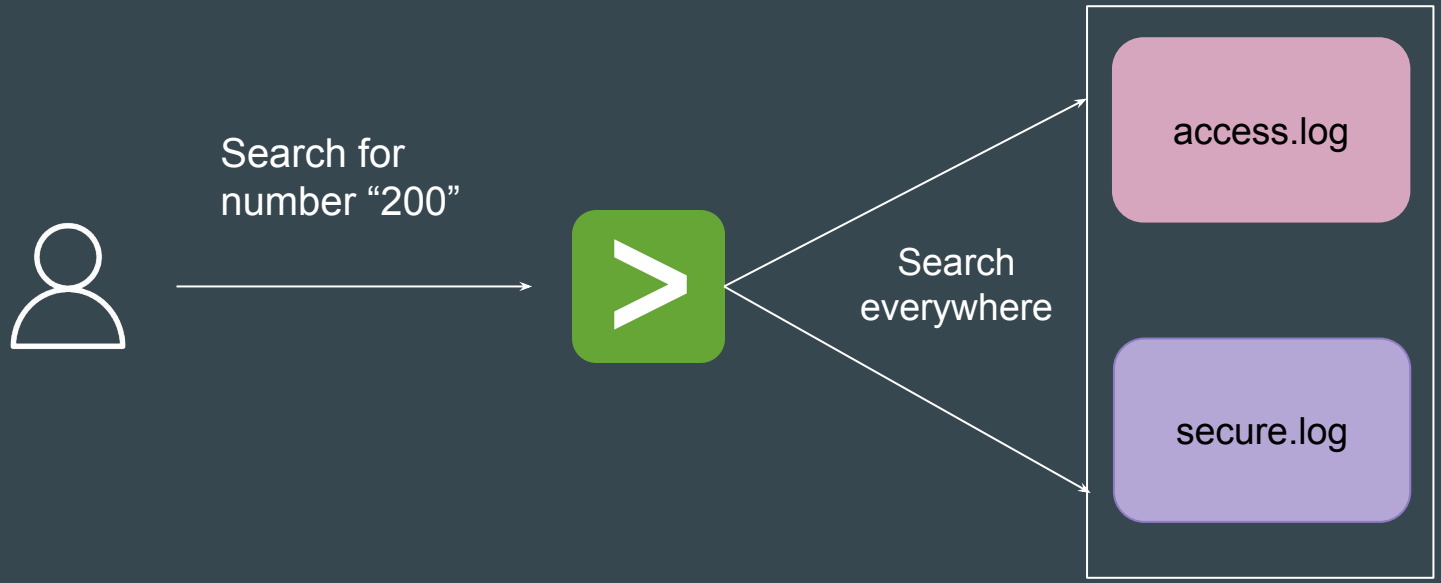
You can use time ranges to troubleshoot an issue, if you know the approximate time frame when the issue occurred.



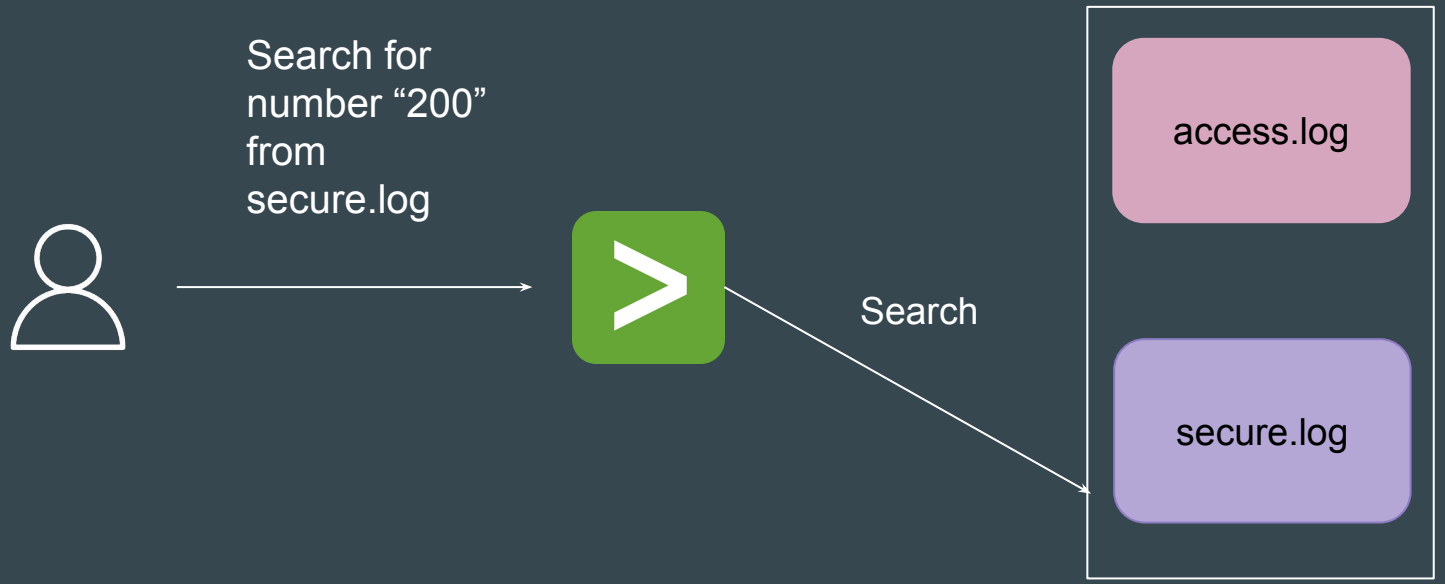
Filtering to Right Source

Entering a value for the field you are searching in the search bar causes Splunk to scan all data.

You can optimize your query by specifying the relevant log type, which narrows down the search and speeds up results.



Filtering to Right Source



Boolean Expressions

The Splunk search processing language (SPL) supports the Boolean operators: AND, OR, and NOT.

Use-Case	SPL
Search for all failed login attempts for user root	root AND failed
Search for failed logins for all user except root	failed NOT root
Search failed logins for user admin OR root	failed admin OR root

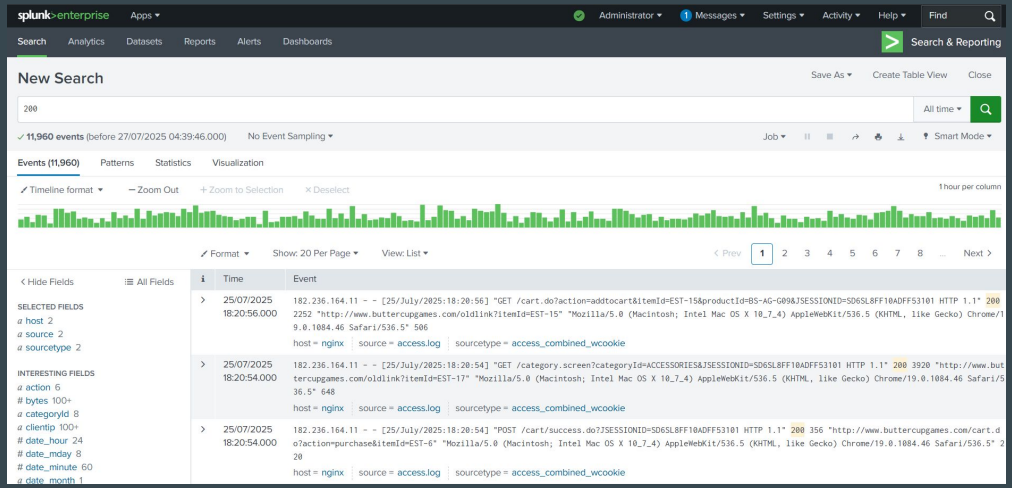
The AND operator is always implied between terms, that is: failed root is the same as failed AND root. So unless you want to include it for clarity reasons, you should not need to specify the AND operator.

SPL Commands

Search Processing Language (SPL)

A basic search based on a text string is a form of SPL—it's the simplest SPL query

However, the true power of SPL lies in its extensive set of commands, functions, and clauses that allow you to transform, analyze, and visualize your data.



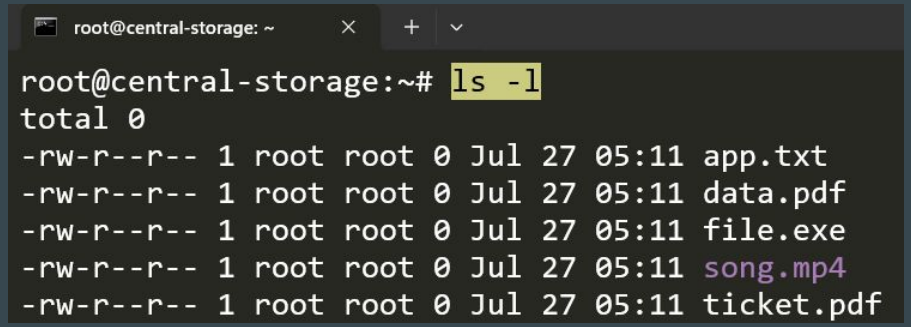
SPL Commands

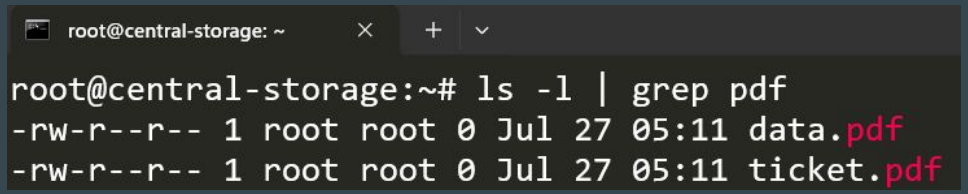
Commands in Splunk tell Splunk what to do with the data retrieved from an index.

Sample Category of Commands	Description
Filtering Commands	These commands, like search and where, are used to filter the result set based on certain criteria
Transforming Commands	These commands, such as stats, chart, timechart, and top, transform the data into a statistical summary or a visualization.
Field Manipulation Commands	Commands like eval and rename allow you to create, modify, or rename fields.

Pipe Structure

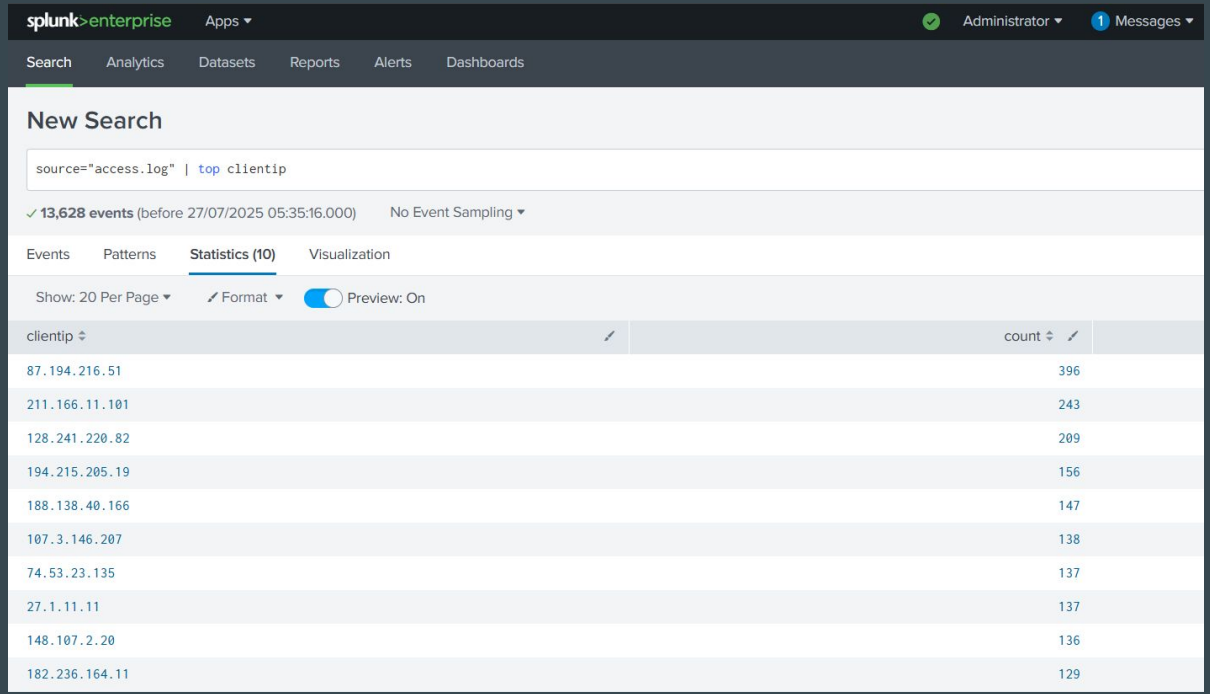
The pipe (|) in Linux is used to connect the output of one command directly to the input of another command.





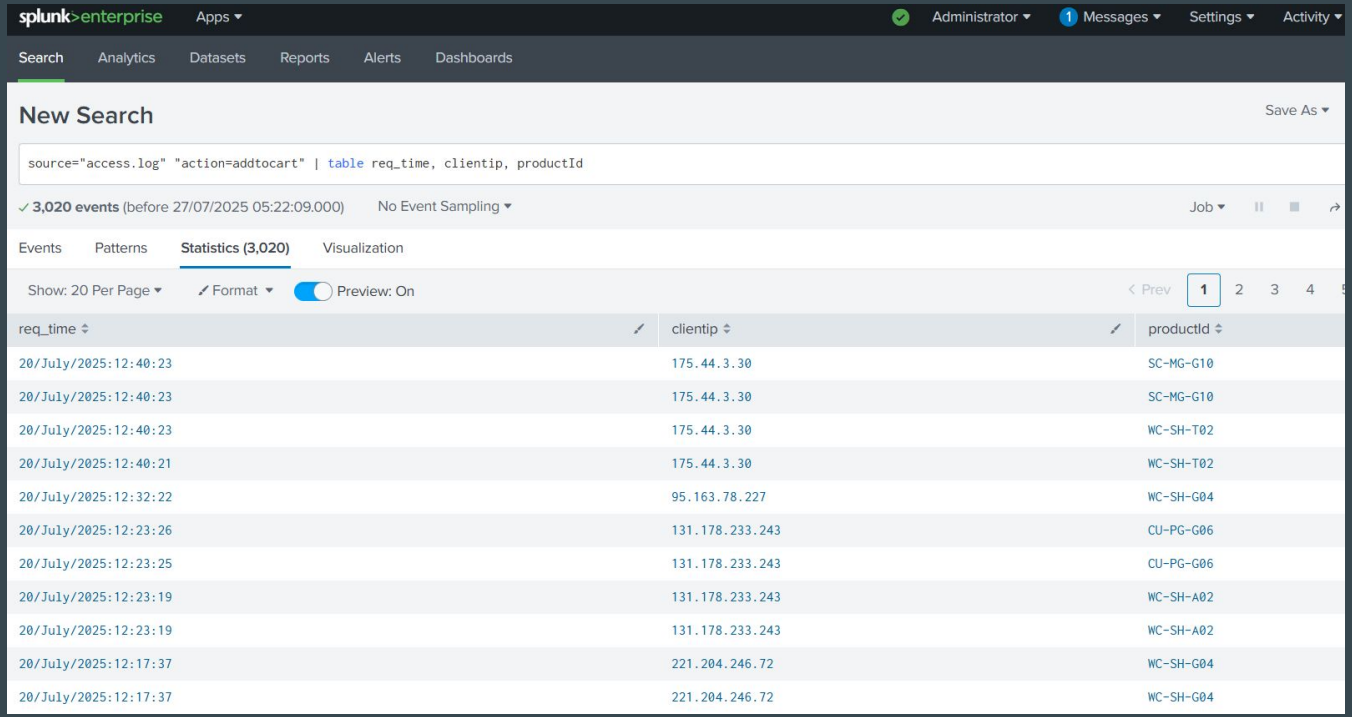
Pipe Structure in Splunk

The pipe | character tells Splunk software to use the output or result of one command (to the left of the pipe) as the input for the next command (to the right of the pipe)



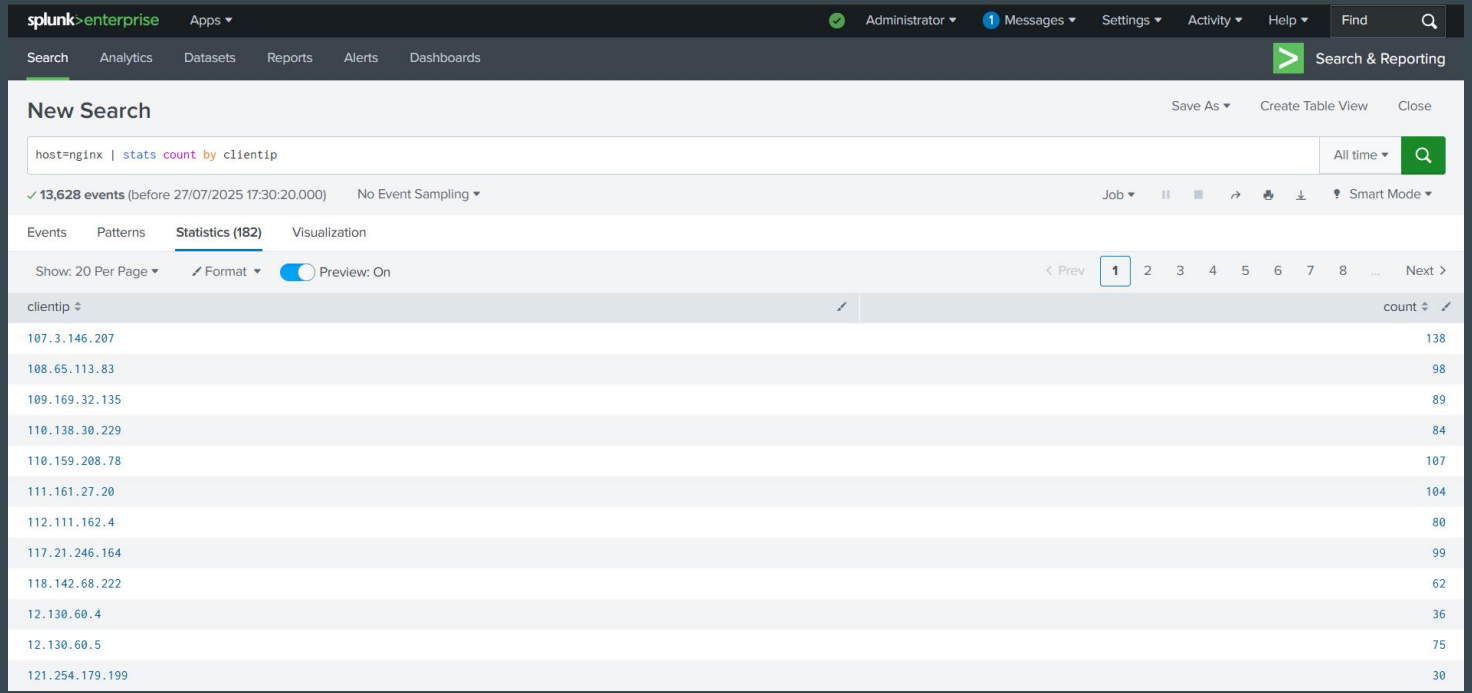
Command - Table

The table command returns a table that is formed by only the fields that you specify in the arguments.



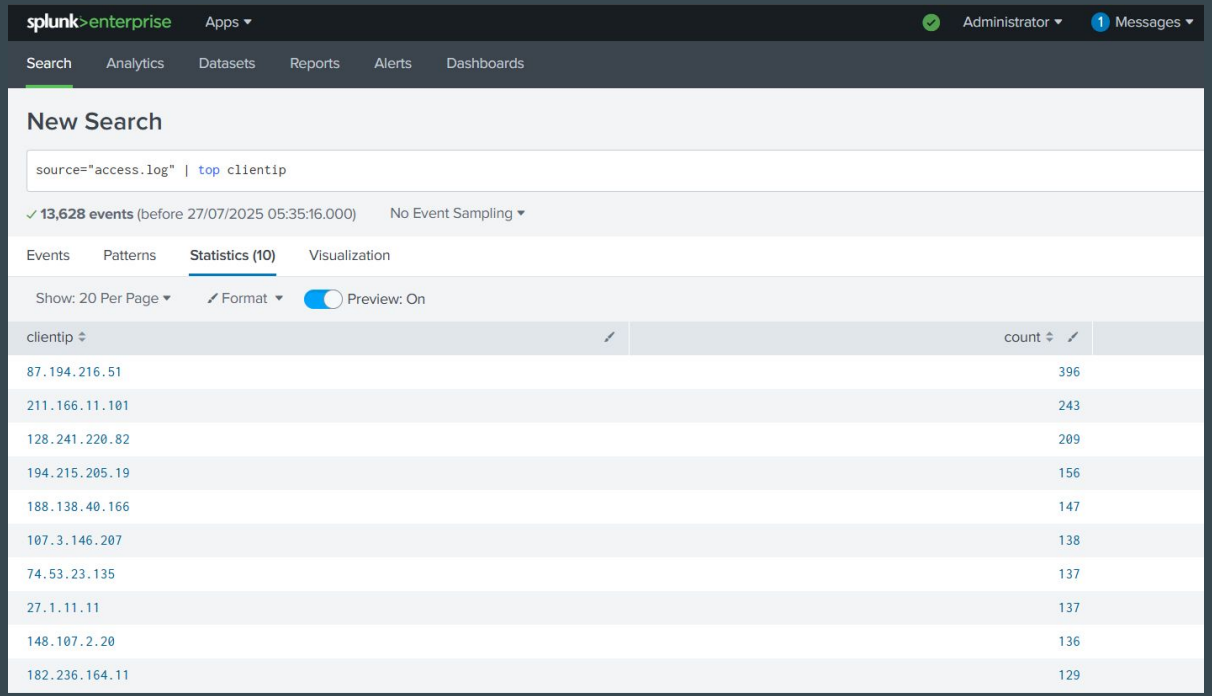
Command - stats

stats command calculates aggregate statistics, such as average, count, and sum, over the incoming search results set.



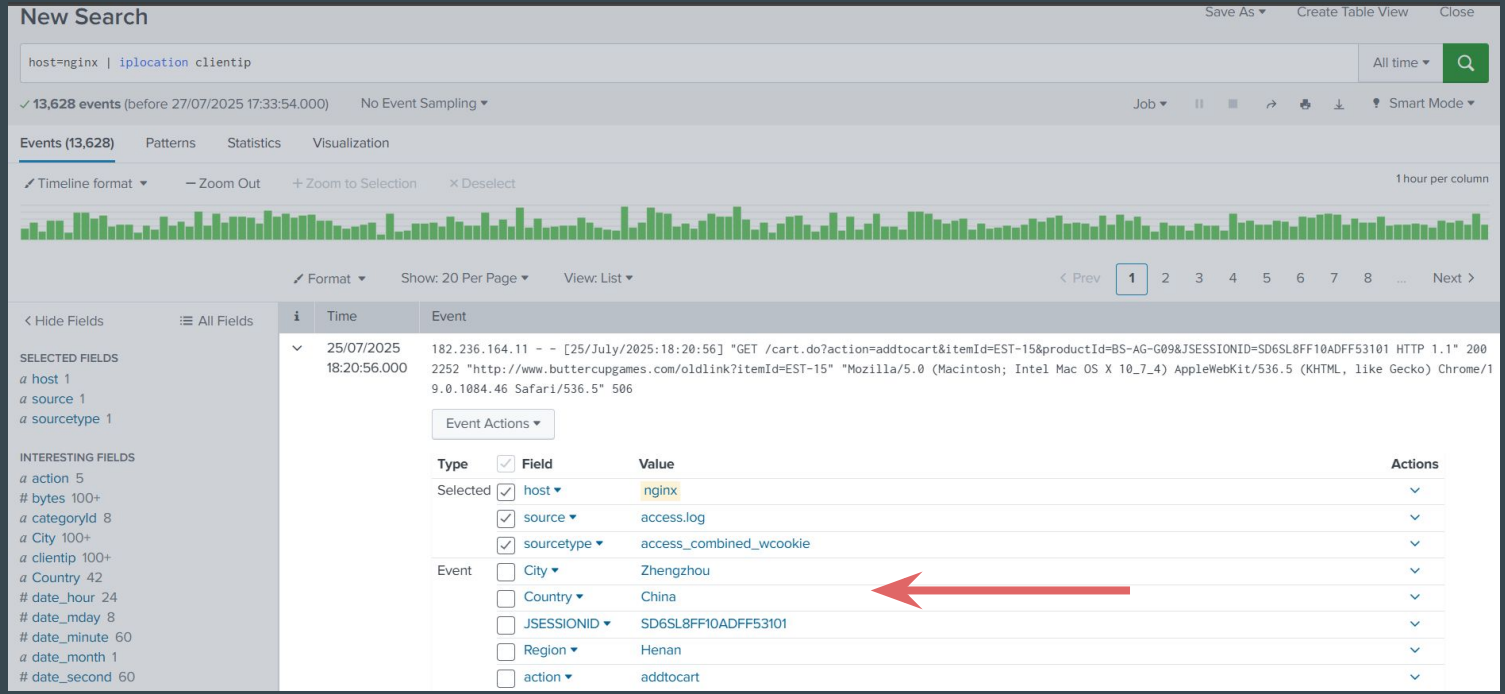
Command - top

The top command finds the most common values for the fields in the field list.



Command - iplocation

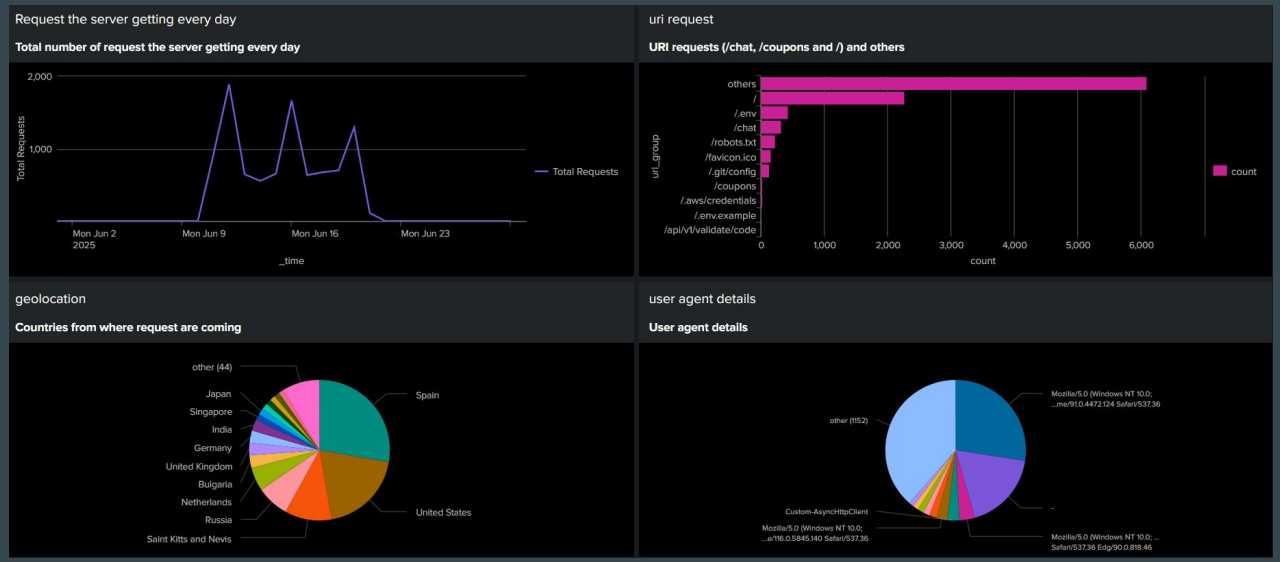
The iplocation command is used to enrich your events with geographical information (like country, city) based on an IP address field.



Basics of Visualization

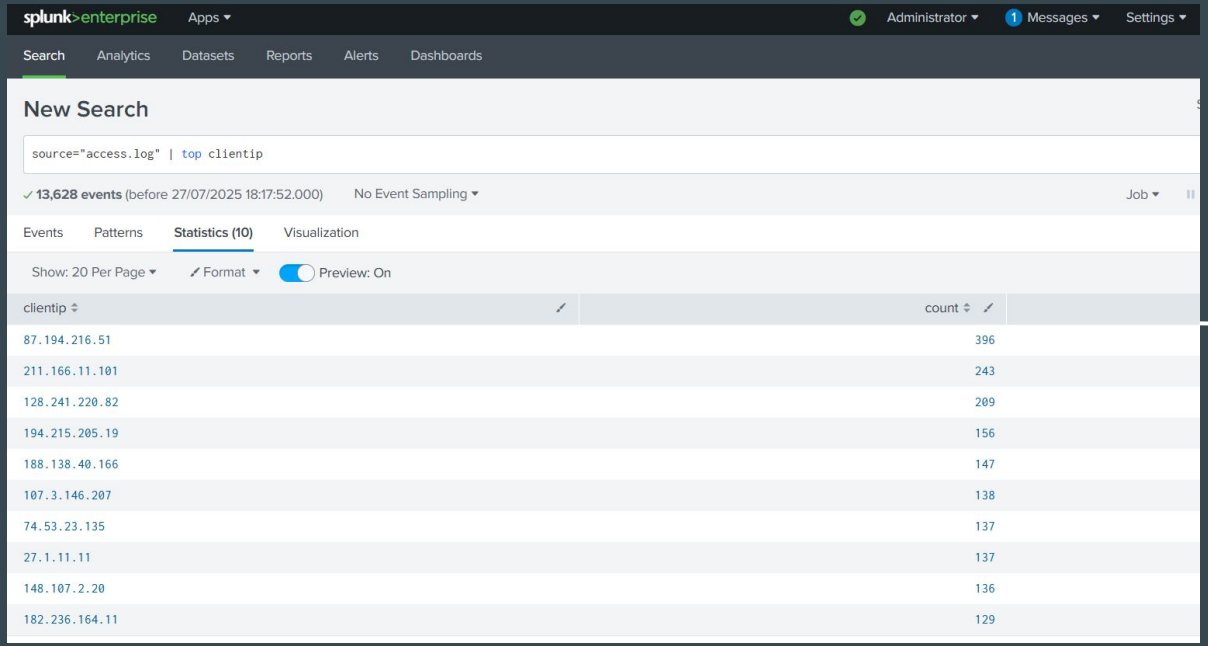
Setting the Base

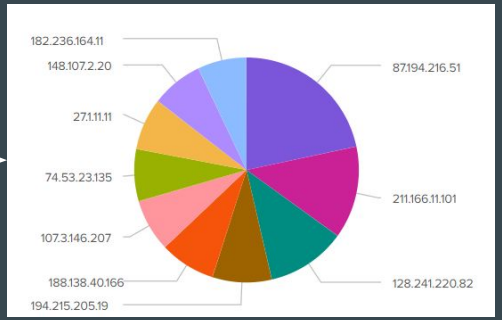
There are many visualization types and configurations available in Splunk to choose from.



Point to Note

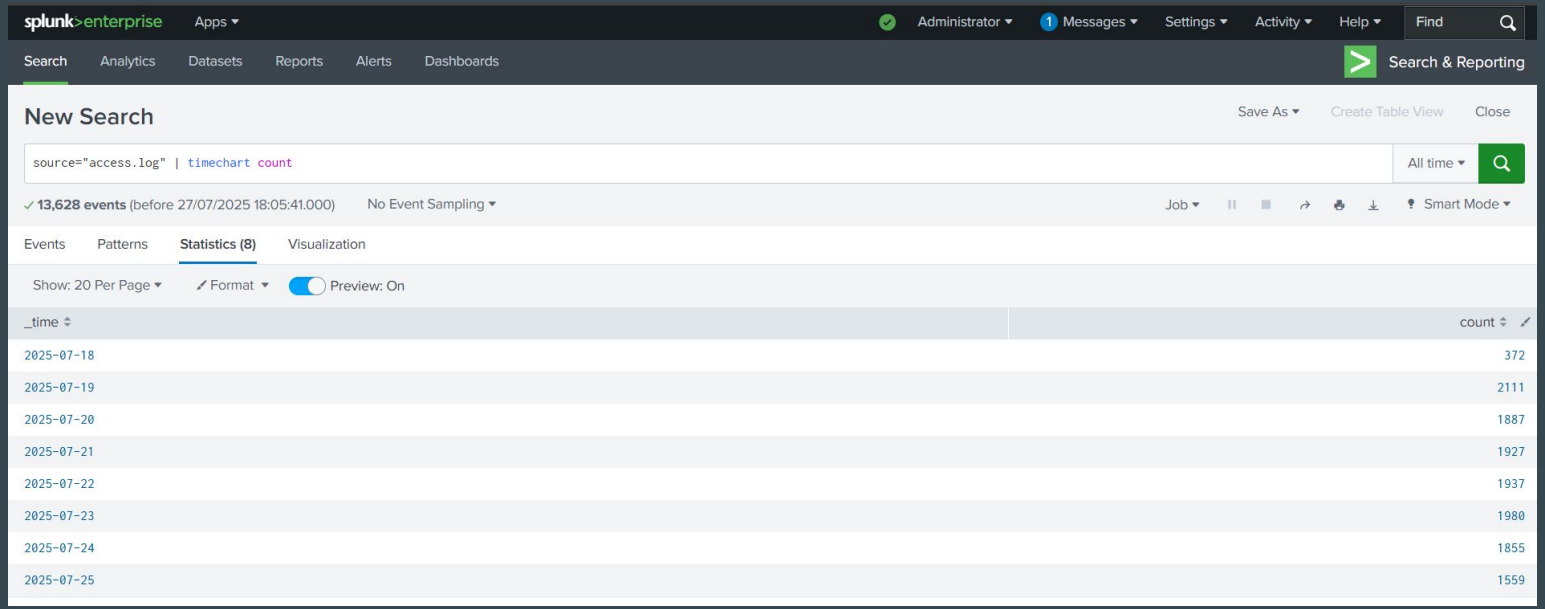
To create charts visualizations, your search must transform event data into statistical data tables.



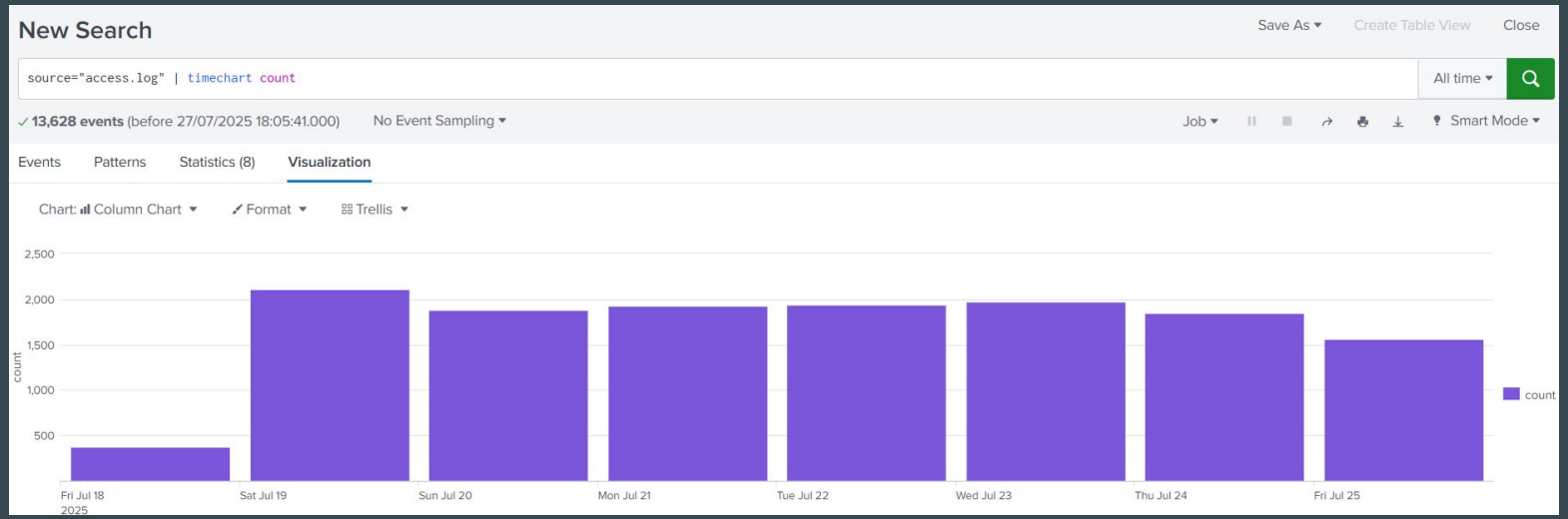


Timechart

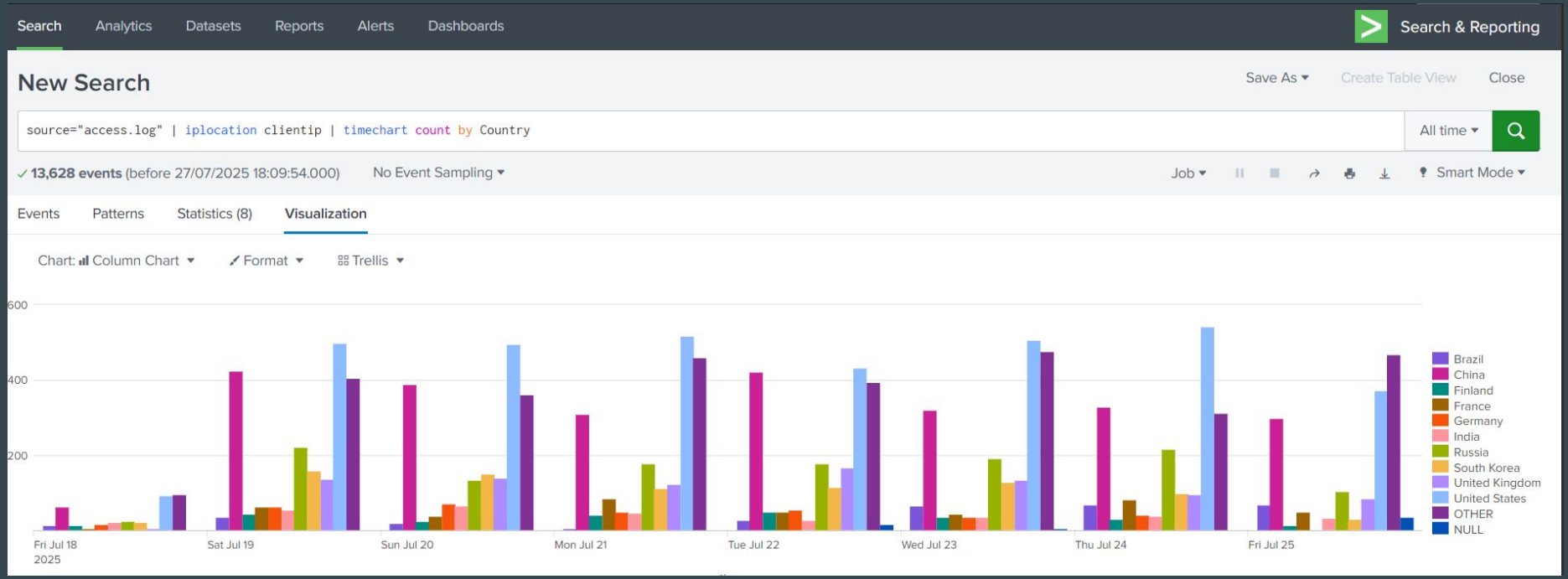
The timechart command takes your results and groups them into time-based buckets depending on your search time range and data.



Example 1 - Timechart Based Visualization



Example 2 - Timechart Based Visualization



Challenge - Finding Attack Vectors

Understanding the Use-Case

1. The Linux Secure logs are indexed in Splunk.

2. You are working as a Splunk Administrator.

3. The Auditor has asked you for a specific set of data based on requirements.

Requirements by Auditor

Sr No	Requirements
1	Find the total number of SSH failed login attempts.
2	Top Attacking IP Addresses
3	All Targeted Failed Usernames
4	Geographic Location of Attackers
5	Successful SSH Logins by User and IP
6	Failed Login Attempts Over Time (Days)
7	Geolocation of ALL Countries from which requests coming from (Map)

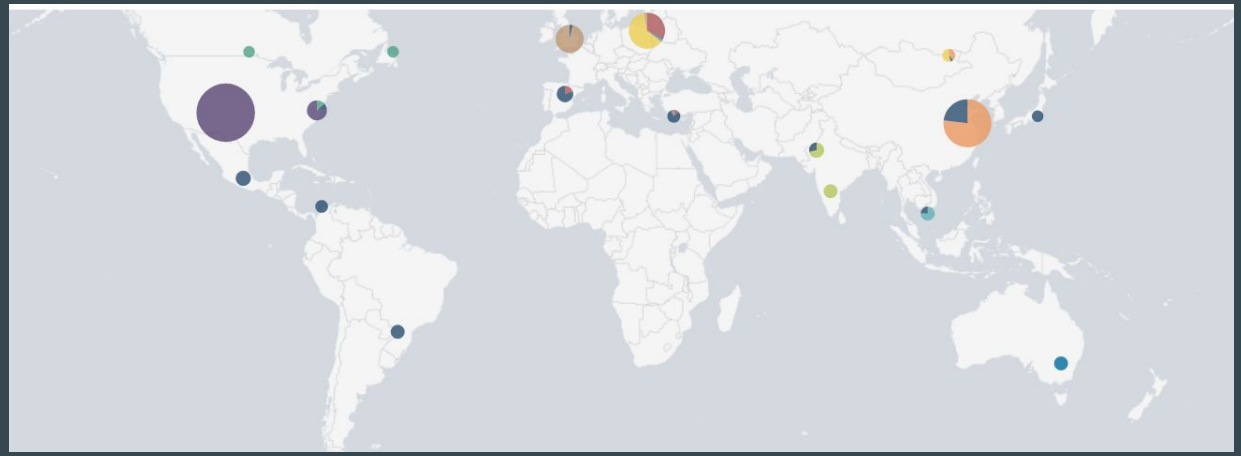
Solution - Part 1

Requirement	Solution
Find the total number of SSH failed login attempts.	source="secure.log" "Failed password"
Top Attacking IP Addresses	source="secure.log" "Failed password" | top src
All Targeted Failed Usernames	source="secure.log" "Failed password" | stats count by user
Geographic Location of Attackers	source="secure.log" "Failed password" | iplocation src | stats count by Country
Successful SSH Logins by User and IP	source="secure.log" "Accepted password" | stats count by user, src
Failed Login Attempts Over Time (Days)	source="secure.log" "Failed password" | timechart count
Geostats	source="secure.log" | iplocation src | geostats count by Country

Geolocation Information

geostats command is used to generate statistics to display geographic data and summarize the data on maps.

The command generates statistics which are clustered into geographical bins to be rendered on a world map.

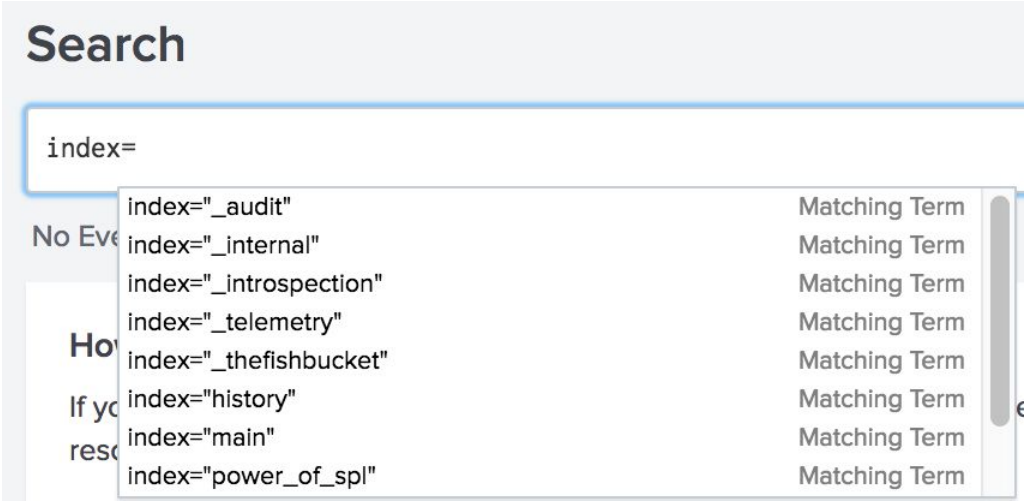


Splunk Search Assistant

Reports are awesome

Overview of Splunk Search Assistant

When you begin typing certain letters or term into the search bar, the search assistant will begin to show you terms and searches that matches what you are typing.



Search Assistant Modes

Splunk Search Assistant has three modes:

- Full
- Compact
- None

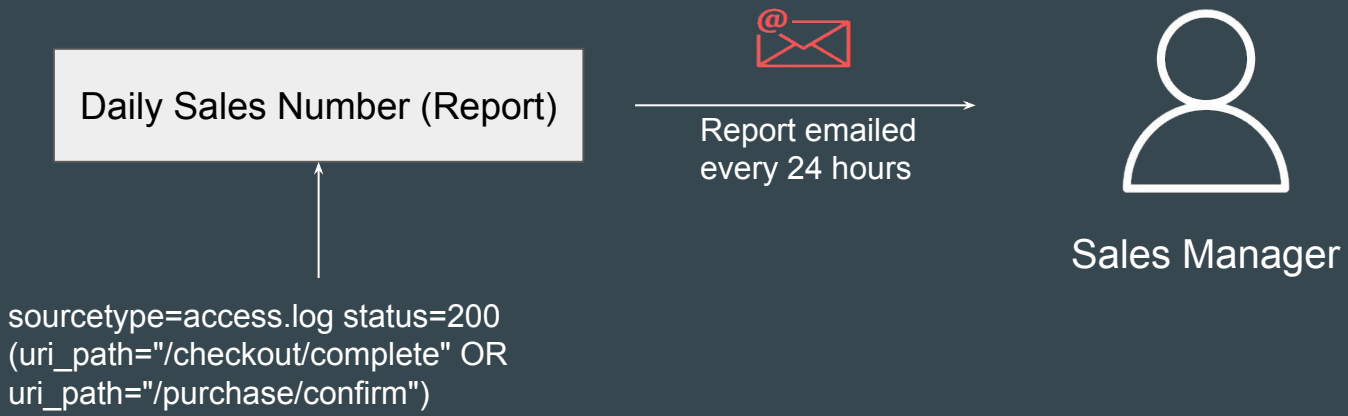
By default, the compact mode is selected but can be changed from Account Settings.

Splunk Reports

Setting the Base

A Splunk report is a saved search query that runs on your data to generate summarized results.

Splunk Report can be scheduled to run at intervals and distribute results via email, Slack, webhooks, or saved to a dashboard.



Use-Case for Today's Practical

The Security Manager should receive an email every 24 hours containing a list of IP addresses and their corresponding country locations from which failed SSH login attempts have occurred.

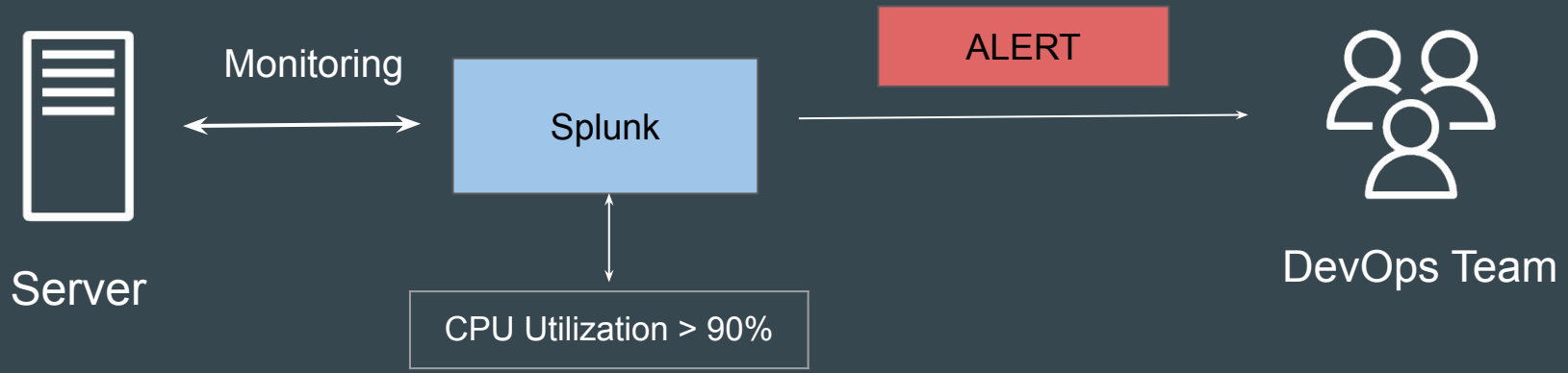
Country	count
Argentina	24
Australia	132
Bahamas	39
Barbados	31
Belgium	76
Brazil	197
Canada	134
China	1178
Czechia	41
Denmark	65
Finland	238
France	266

Splunk Alerts

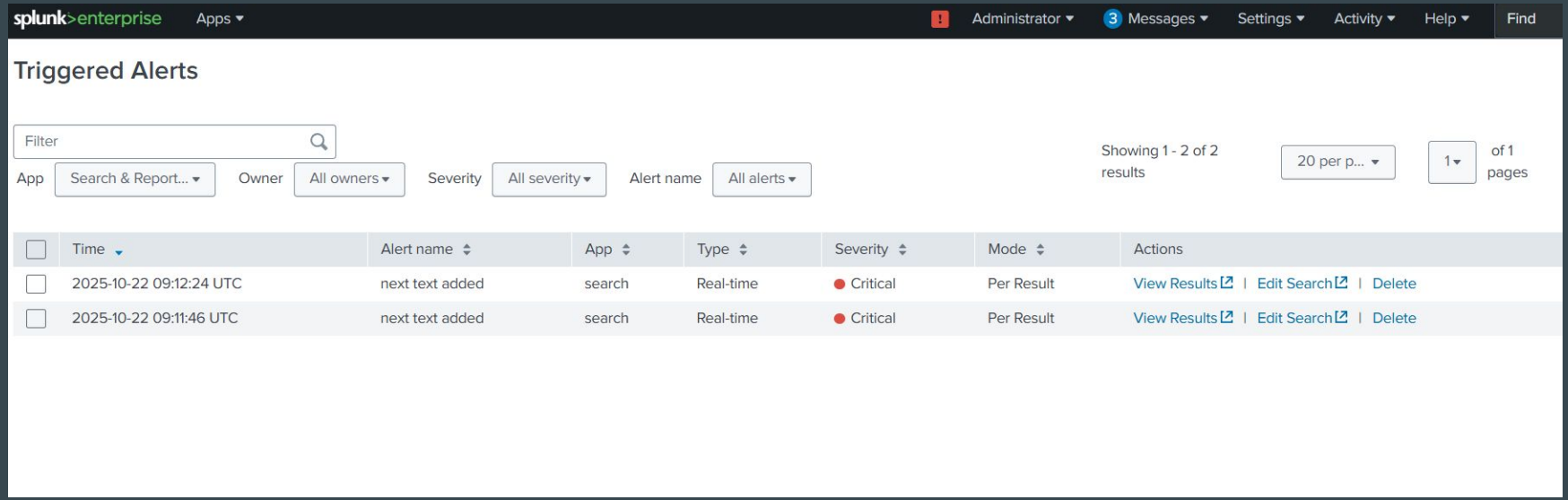
Setting the Base

Alerts use a saved search to look for events in real time or on a schedule.

Alerts trigger when search results meet specific conditions.



Screenshot of Triggered Alerts



Practical Use-Case to Implement

The file /tmp/readonly.txt is intended to be read-only and should remain unaltered.

Configure an alert that triggers whenever a new write operation occurs on this file.

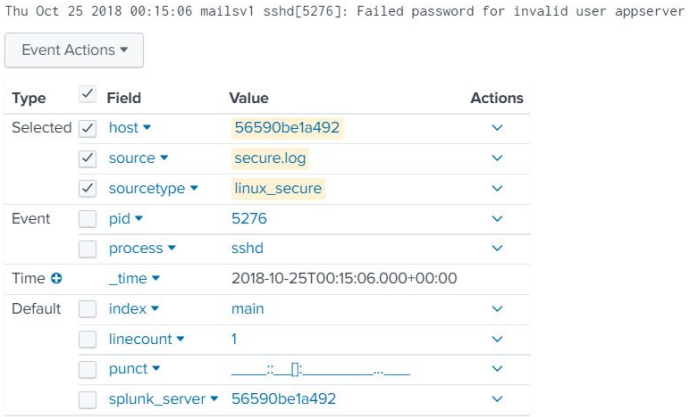
Add-ons and Apps

True Extensibility

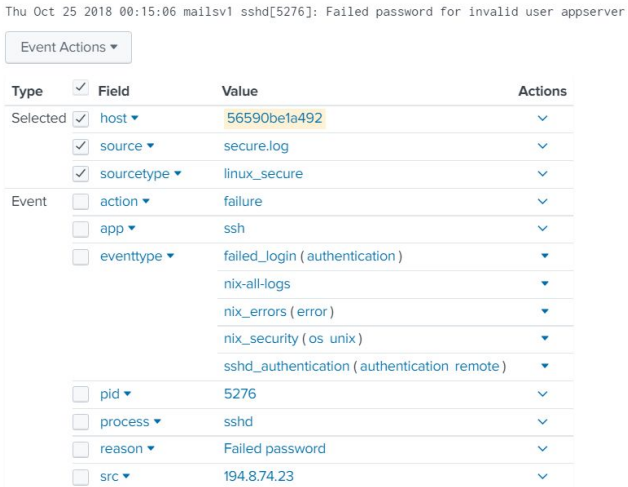
Use-Case: Linux Authentication Logs

In our previous section, where we had uploaded the Linux authentication logs, we observed that the logs were not parsed by default.

However, after installing the Linux Add-On, the log was automatically parsed.



Before AddOn

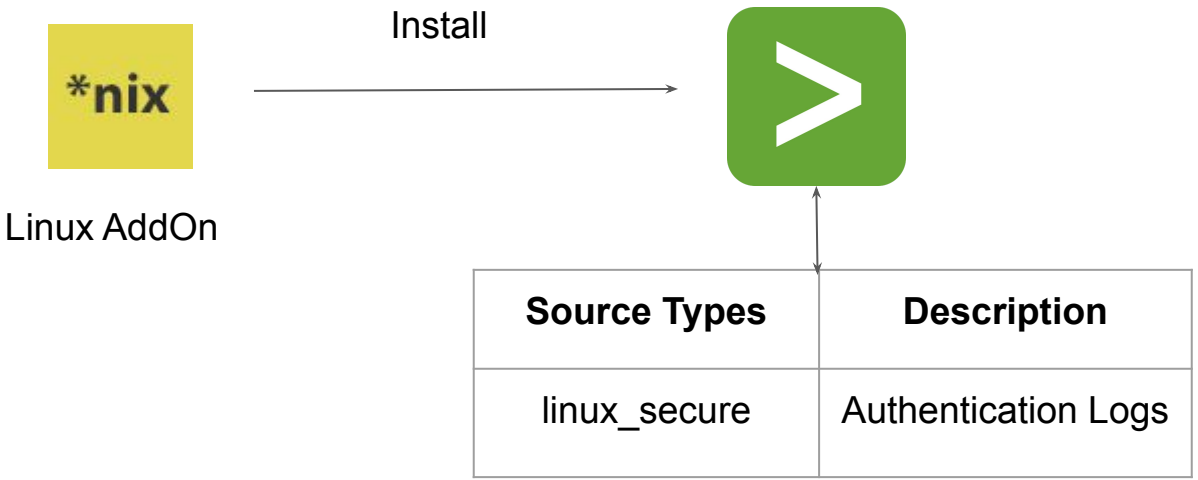


After AddOn

Overview of Splunk Add-ons

Add-ons extend the functionality of the Splunk platform and provide required field extractions, lookups, saved searches, and others.

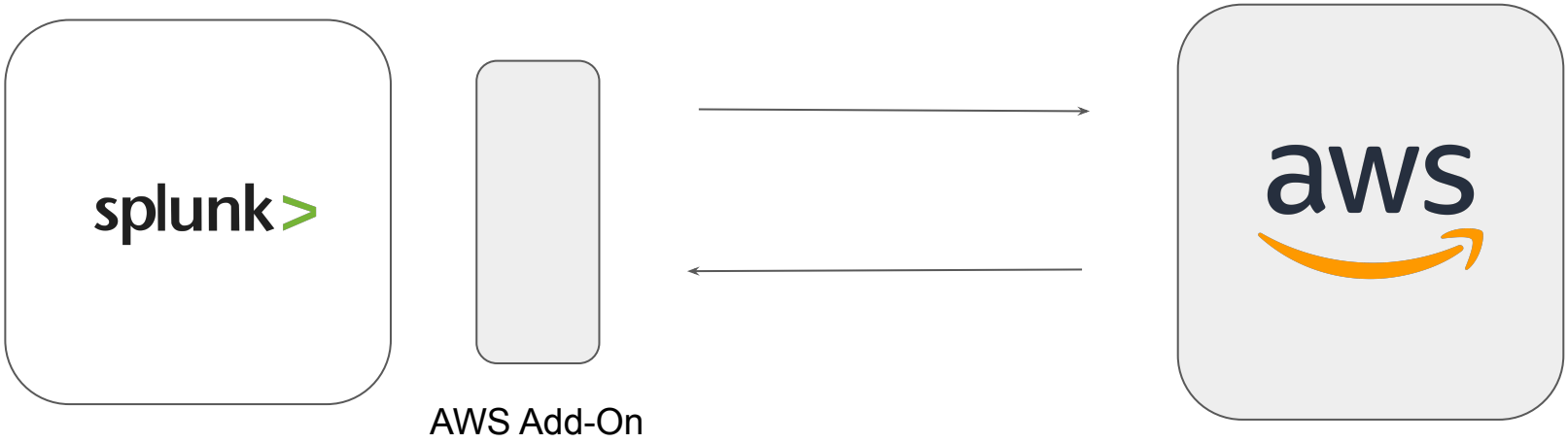
For the logs that Splunk does not parse by default, you can install various AddOns from the Splunk marketplace to do the parsing for us.



Source Types	Description
linux_secure	Authentication Logs

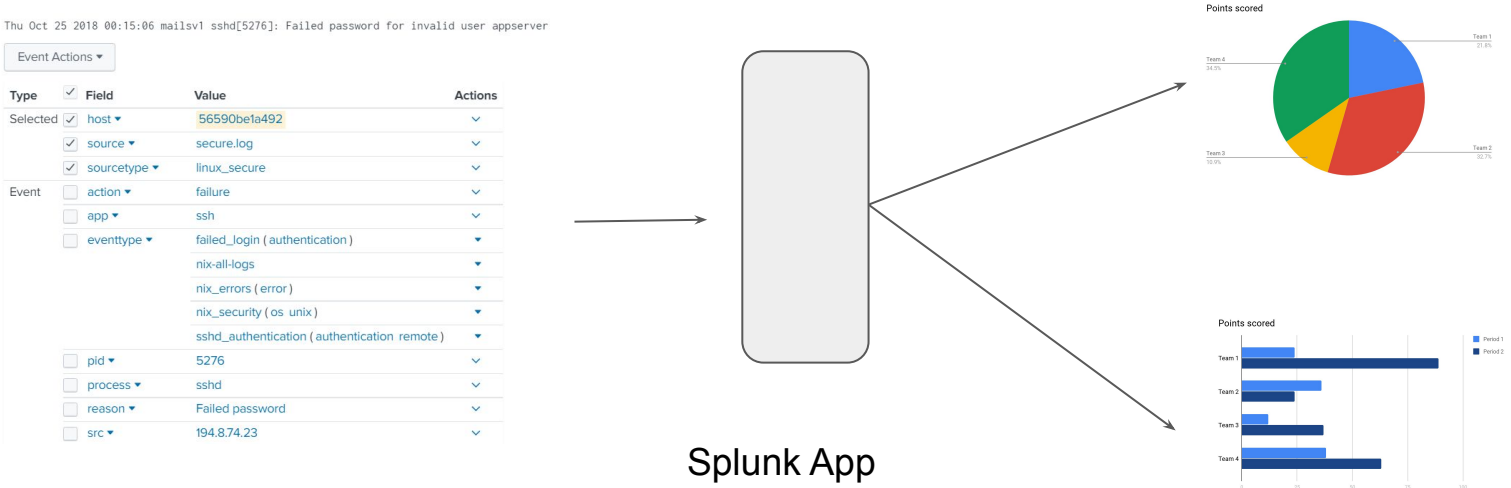
Splunk Add-on For Data Input

Add-ons can also be used to extract data from a specific destination.



Overview of Splunk Apps

Apps delivers user experience that makes data immediately useful typically with pre-built dashboards that makes data easy to analyze.



Apps and Addon Support

There are three types of support criteria that you will generally see in splunkbase:

- Splunk Supported
- Developer Supported
- Community Supported

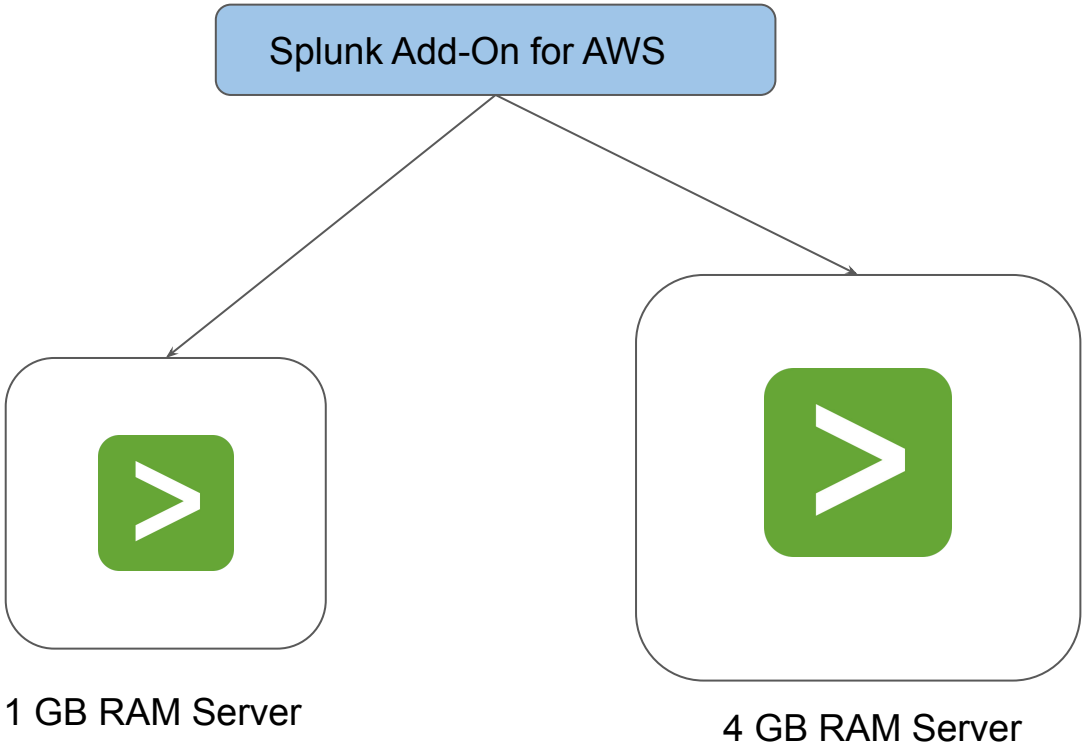
By default, Splunk platform includes one basic app that allows us to work with our data; the Search and Reporting app.

Important Note - Hardware Considerations

Some of the Splunk Add-ons and Apps are resource heavy.

Running them in instance with lower hardware will lead to Splunk/Server going down or becoming unresponsive.

Demo - Different Hardware for Add-On



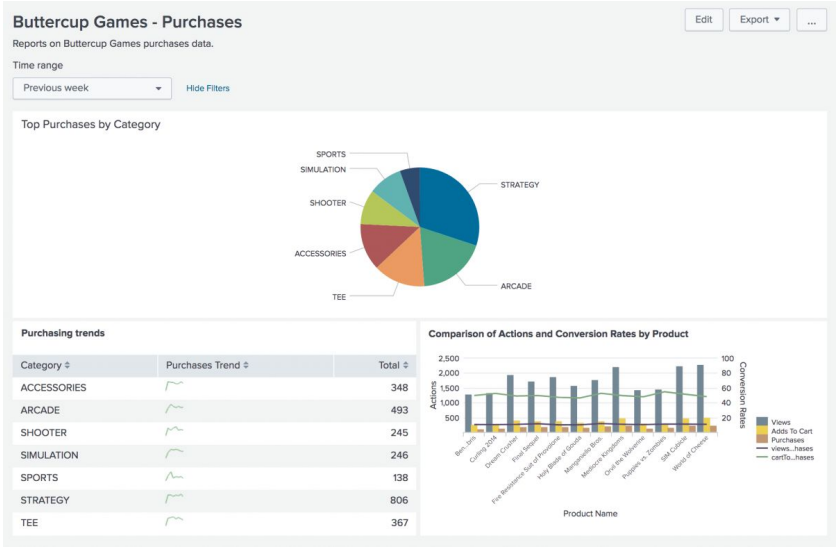
Dashboards and Panels

Everyone likes Good Visualization

Dashboards and Panels

A dashboard contains one or more panels.

Dashboard panels use searches to generate visualizations

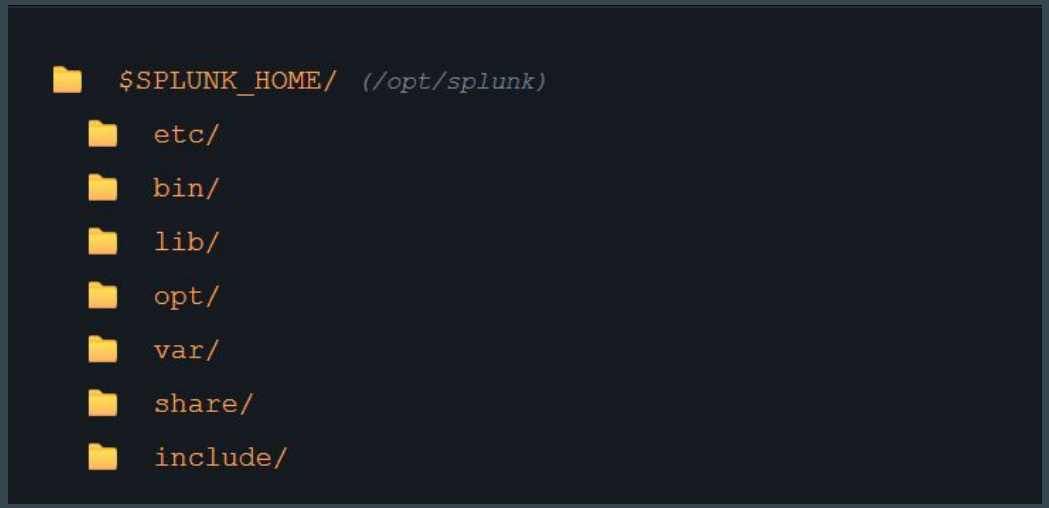


Splunk Directory Structure

Setting the Base

Splunk is installed in a root directory often referred to as $SPLUNK_HOME

$SPLUNK_HOME = (/opt/splunk on Linux or C:\Program Files\Splunk on Windows)



Reference Screenshot - $SPLUNK_HOME

```
[splunk@78ae6bb5c2a4 ~]$ ls -l /opt/splunk
total 5092
-r--r--r--  1 splunk splunk    1090 May 10 06:52 LICENSE.txt
-r--r--r--  1 splunk splunk     517 May 27 18:39 README-splunk.txt
drwxr-xr-x  1 splunk splunk    4096 Jun  5 18:18 bin
-r--r--r--  1 splunk splunk      57 May 27 18:36 copyright.txt
drwxr-xr-x 18 splunk splunk    4096 Oct 22 09:21 etc
drwxr-xr-x  4 splunk splunk    4096 May 27 19:01 include
drwxr-xr-x  1 splunk splunk    4096 Jun  5 18:18 lib
-r--r--r--  1 splunk splunk   59708 May 27 18:36 license-eula.txt
drwxr-xr-x  1 splunk splunk    4096 Oct 22 09:08 openssl
drwxr-xr-x  6 splunk splunk    4096 May 27 19:01 opt
drwxr-xr-x  1 splunk splunk    4096 Oct 22 09:09 quarantined_files
drwxr-xr-x  1 splunk splunk    4096 Jun  5 18:17 share
-r--r--r--  1 splunk splunk 5085411 May 27 19:05 splunk-9.4.3-237ebbd22314-linux-amd64-manifest
drwxr-xr-x  2 splunk splunk    4096 May 27 19:04 swidtag
drwxr-xr-x  8 splunk splunk    4096 Oct 22 09:08 var
```

Important Directories to Know

Important Directories	Description
bin	Contains all the Splunk executables, command-line tools, and scripts. This is the home of the Splunk CLI (Command Line Interface).
etc	This is the most important directory for administrators. It contains all the configuration files (.conf), authentication settings, app/add-on configurations, and license files. Your day-to-day configuration work happens here.
var	This directory stores all the data that changes during Splunk's operation. This includes the indexed data itself, Splunk's internal logs, search job artifacts, and information about running processes.

Point to Note

For 99% of your administrative tasks, you will be interacting with only two main directories of /etc and /var

Configuration file precedence

Setting the Base

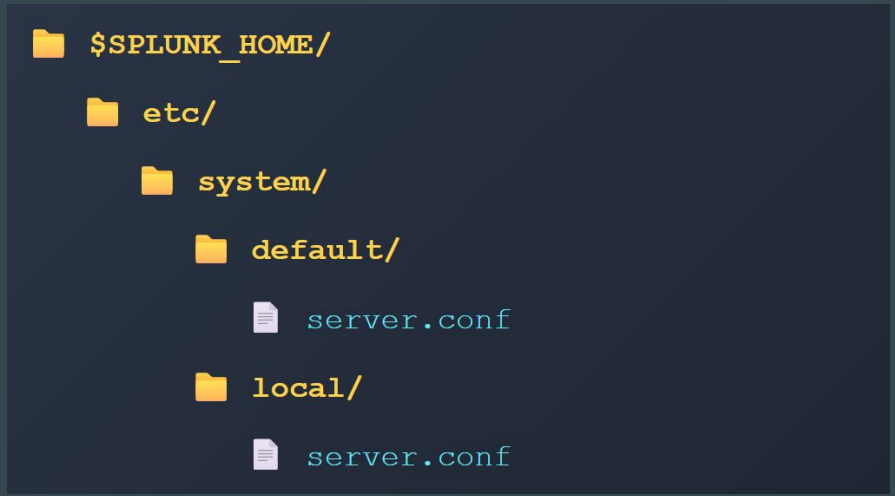
Splunk software uses configuration files (.conf) to determine nearly every aspect of its behavior

```
[splunk@78ae6bb5c2a4 ~]$ ls -l /opt/splunk
total 5092
-r--r--r--  1 splunk splunk    1090 May 10 06:52 LICENSE.txt
-r--r--r--  1 splunk splunk     517 May 27 18:39 README-splunk.txt
drwxr-xr-x  1 splunk splunk    4096 Jun  5 18:18 bin
-r--r--r--  1 splunk splunk      57 May 27 18:36 copyright.txt
drwxr-xr-x 18 splunk splunk    4096 Oct 22 09:21 etc
drwxr-xr-x  4 splunk splunk    4096 May 27 19:01 include
drwxr-xr-x  1 splunk splunk    4096 Jun  5 18:18 lib
-r--r--r--  1 splunk splunk   59708 May 27 18:36 license-eula.txt
drwxr-xr-x  1 splunk splunk    4096 Oct 22 09:08 openssl
drwxr-xr-x  6 splunk splunk    4096 May 27 19:01 opt
drwxr-xr-x  1 splunk splunk    4096 Oct 22 09:09 quarantined_files
drwxr-xr-x  1 splunk splunk    4096 Jun  5 18:17 share
-r--r--r--  1 splunk splunk 5085411 May 27 19:05 splunk-9.4.3-237ebbd22314-linux-amd64-manifest
drwxr-xr-x  2 splunk splunk    4096 May 27 19:04 swidtag
drwxr-xr-x  8 splunk splunk    4096 Oct 22 09:08 var
```

Point to Note

A Splunk platform deployment can have many copies of the same configuration file.

When editing configuration files, it is important to understand how Splunk software evaluates these files and which ones take precedence.



System - Local and Default

system/local	system/default
Highest Priority	Lowest Priority

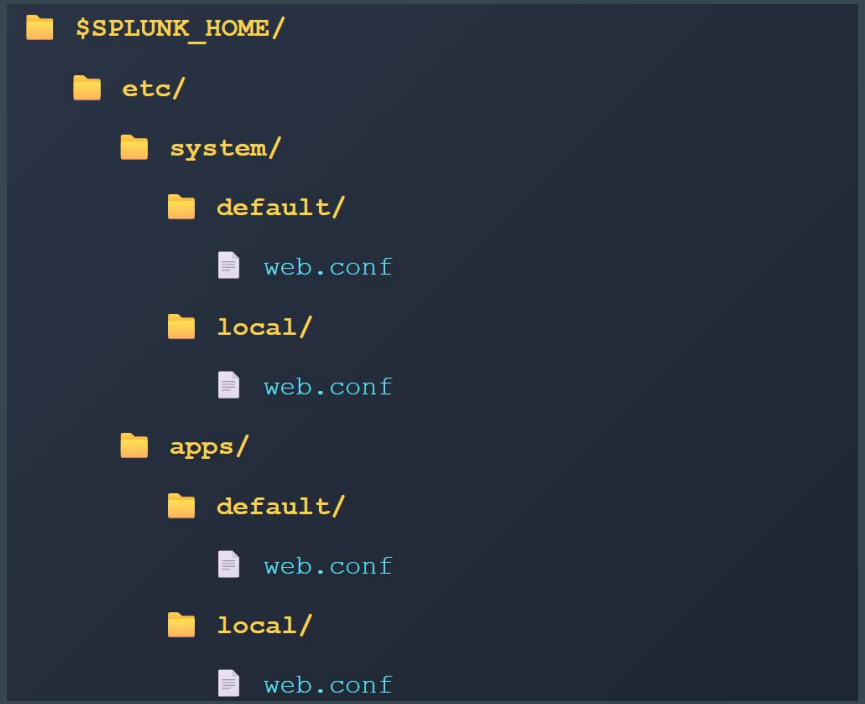
Example Use-Case - Priority

System Local Directory	System Default Directory
serverName = splunk-local	serverName = splunk-default

Final Value That Splunk will Use = splunk-local

One Step Further -

The configuration files can also be present in apps/ folder and overall precedence slightly changes.



Overall Precedence

System local directory -- highest priority
App local directories
App default directories
System default directory -- lowest priority

How app directory names affect precedence

When determining priority in the global context, Splunk software uses lexicographical order to determine priority among the collection of apps directories.

For example, files in an apps directory named "A" have a higher priority than files in an apps directory named "B", and so on.

Also, all apps starting with an uppercase letter have precedence over any apps starting with a lowercase letter, due to lexicographical order. ("A" has precedence over "Z", but "Z" has precedence over "a", for example.)

Indexes

Backend Storage

Overview of Indexes

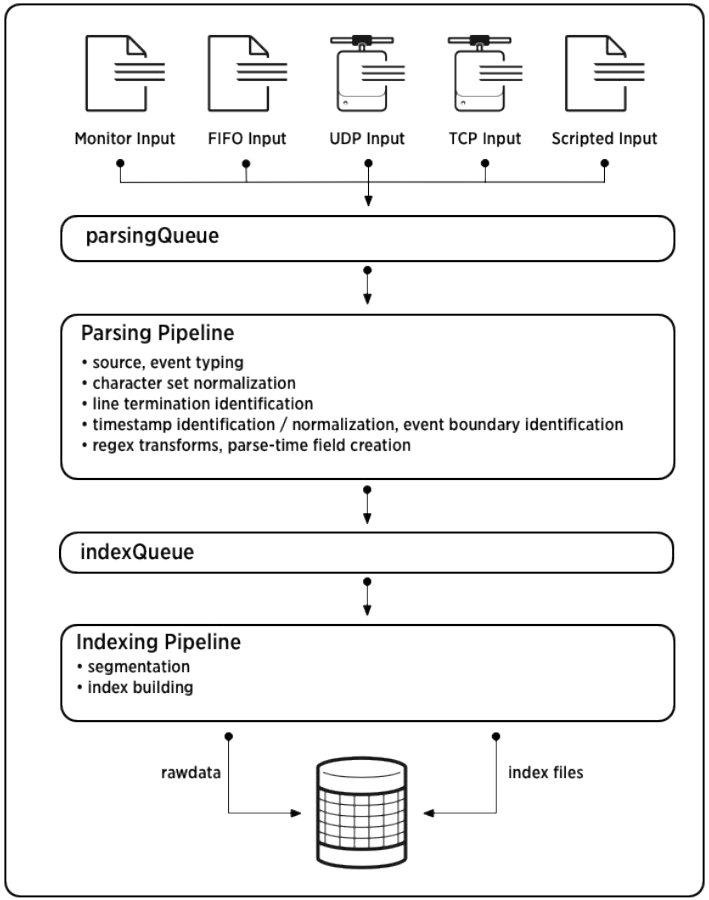
The index is a repository of Splunk data.

Splunk transforms incoming data into events, which it stores in the indexes.

When Splunk indexes your data, it creates number of files. These files falls into two main categories:

- The raw data in compressed form (rawdata)
- Indexes that points to raw data (tsidx files), plus some meta-data files.

These files resides in set of directories organized by age.


Monitor Input
FIFO Input
UDP Input
TCP Input
Scripted Input
parsingQueue
Parsing Pipeline
• source, event typing
• character set normalization
• line termination identification
• timestamp identification / normalization, event boundary identification
• regex transforms, parse-time field creation
indexQueue
Indexing Pipeline
• segmentation
• index building
rawdata
index files

Default Set of Indexes

Splunk Enterprise comes with number of pre-configured indexes, including:

main	This is default index. All data gets stored here unless specified.
_internal	Stores Splunk's internal logs.
_audit	Contain events related to user search history, file system change monitor and auditing specific.

Buckets Lifecycle

Index Architecture

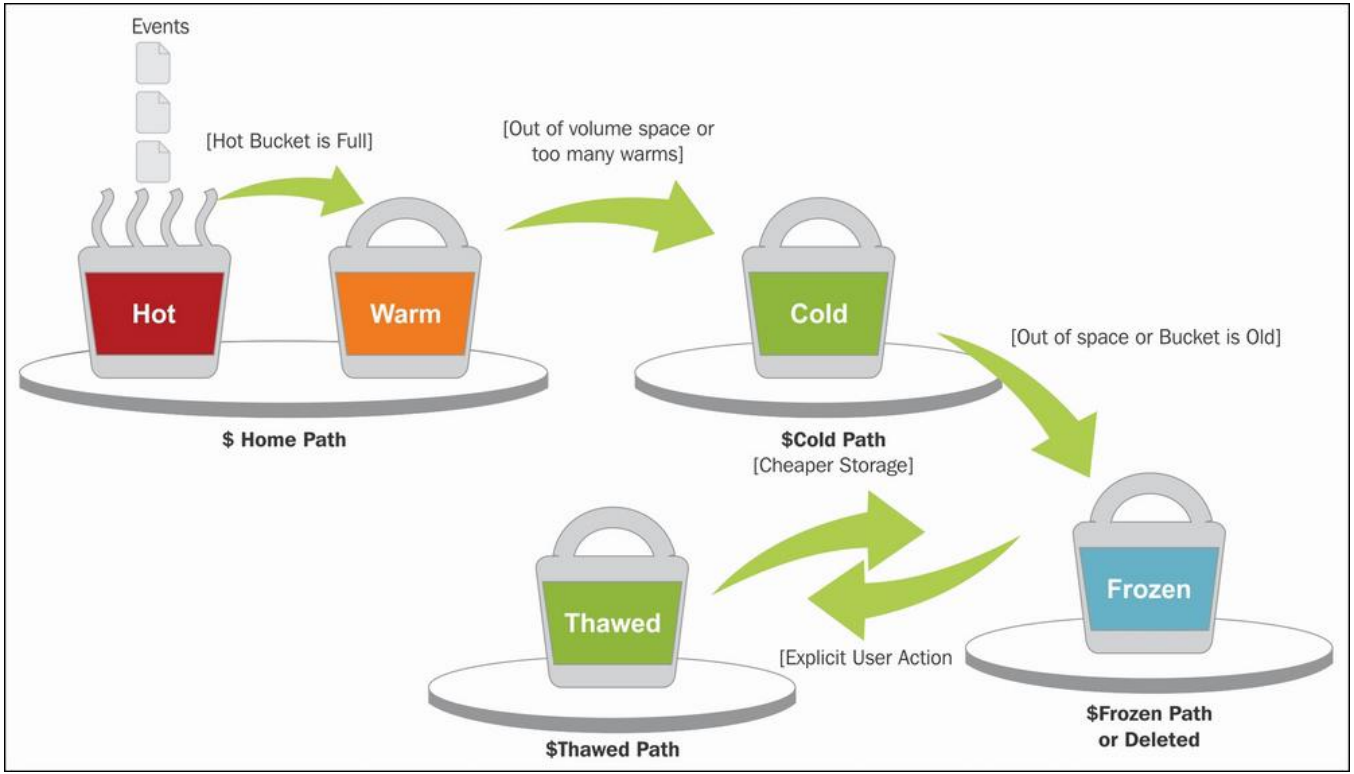
Overview of Bucket Lifecycle

Splunk stores all its data in directories on server called buckets

A bucket moves through several stages as it ages – hot, warm, cold, frozen

hot	All of the new data is written here and most recent data is kept here.
warm	Data rolled from hot. Data is not actively written to warm buckets.
cold	Data rolled from warm. Rarely searched data as it has aged / archived.
frozen	Data rolled from cold. The data is deleted, but can be archived.
thawed	If data in frozen bucket is archived, it can be indexed again by thawing it.

Bucket Lifecycle



Hot Bucket to Warm Bucket

Buckets are rolled from hot to warm in following condition:

- We get too many hot buckets [maxHotBuckets]
- Hot bucket has not received data since a while
- Timespan of buckets is too large.
- Bucket meta-data files have grown large.
- Index clustering replication error.
- Splunk is restarted

Warm to Cold Buckets

Ideally, historical data should go here.

Allows us to keep older data on slower (cheaper) storage.

Buckets are rolled from warm to cold when there are too many warm buckets.

```
[index_name]
coldPath = $SPLUNK_DB/$_index_name/colddb
maxWarmDBCount = 300
```

Cold to Frozen

Data in frozen is no longer searchable.

Data rolls from cold to frozen bucket when:

- Total size of index (hot+warm+cold) grows too large.
- Oldest event in bucket exceeds specific age.

Config: coldToFrozenDir

In default process, tsidx file is removed and bucket is specified to destination we specify.

Thawing Process

This is generally manual process for restoring archived data.

Overall Steps:

i) mv /tmp/frozendb/db* $SPLUNK_HOME/var/lib/splunk/defaultdb/thaweddb/

ii) splunk rebuild $SPLUNK_HOME/var/lib/splunk/defaultdb/thaweddb/db*

iii) splunk restart

Splunk Workflow Actions

Analyst Functionality

Overview of Workflow Actions

Splunk WorkFlow Actions allows us to add interactivity between the indexed fields and other web resources.

Example:

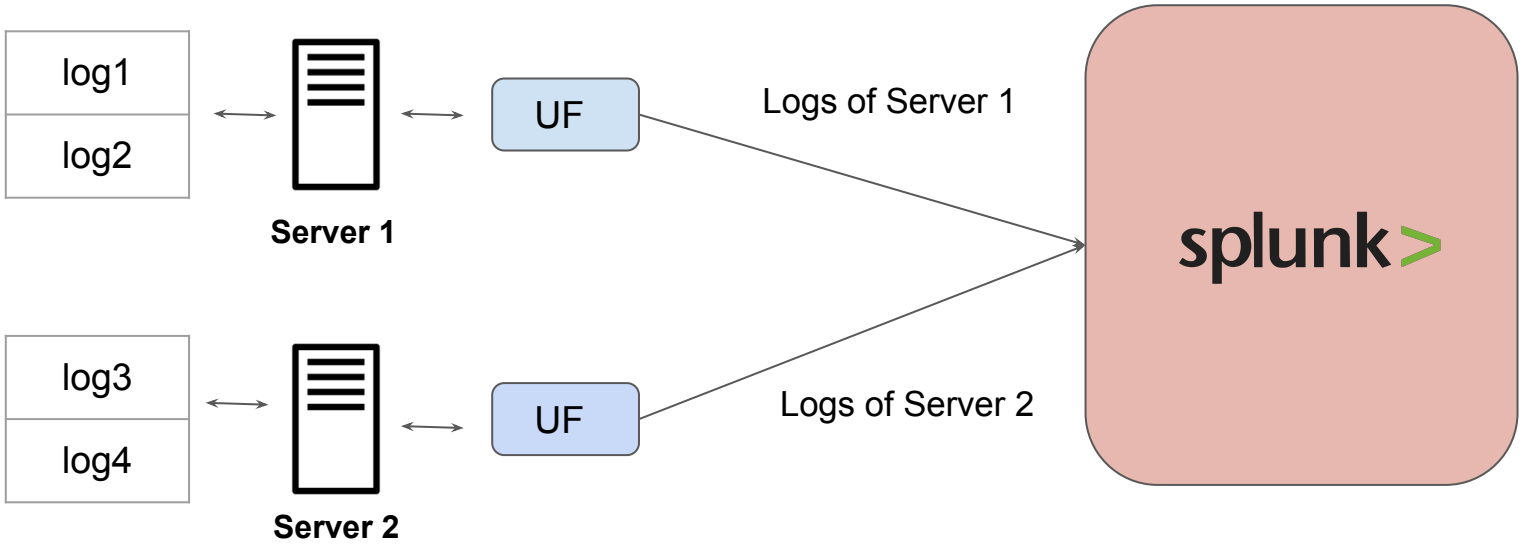
There is a field called as clientip in access_combined log file.
You can add option for “Whois Lookup” based on the IP address in clientip field.

Universal Forwarder

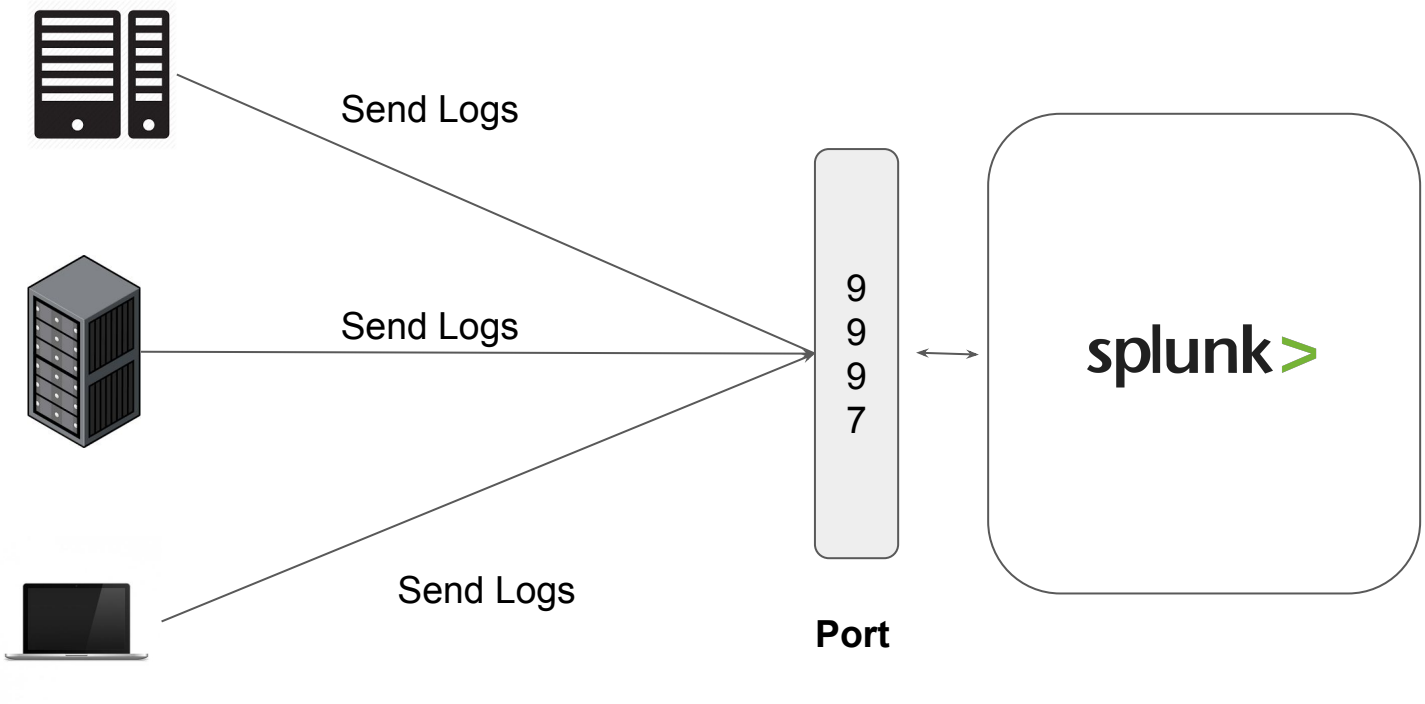
Let's get started!

Basics of Universal Forwarder

Universal Forwarder is an agent that collects data from a server and send it to Splunk.



Basic Architecture



Installation Options

Splunk offers universal forwarder agent for various operating systems, including:

- Linux
- Windows
- MAC
- Solaris
- FreeBSD
- AIX

Agent Management

Understanding the Challenge

When manually configuring the Universal Forwarder, we ran two key commands:

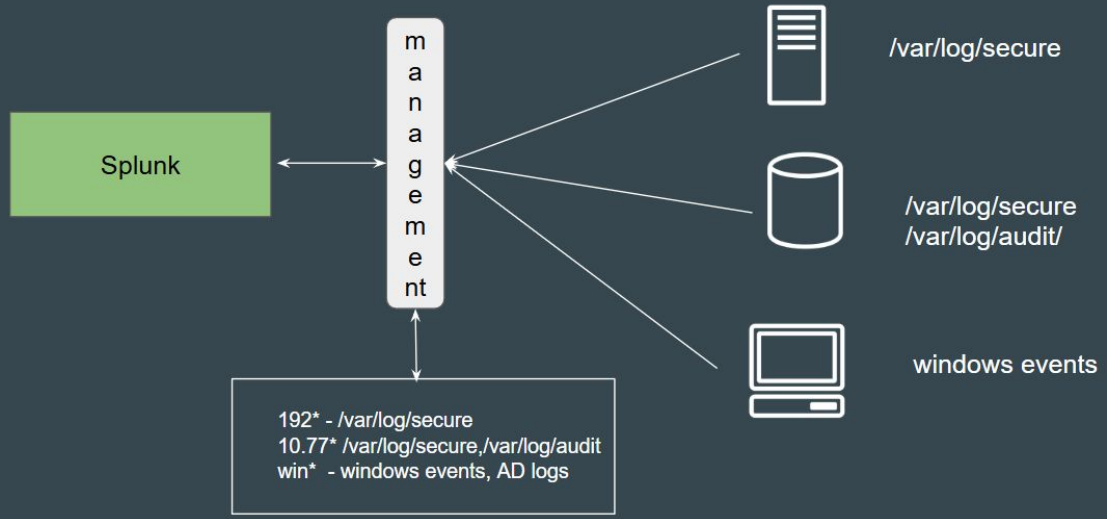
Commands to Setup Universal Forwarder
/opt/splunkforwarder/bin/splunk add monitor /var/log
/opt/splunkforwarder/bin/splunk add forward-server [IP-OF-SPLUNK]:9997

Agent Management allows us to centrally deploy changes to all universal forwarders, eliminating the need to manually execute commands on each individual forwarder.

Setting the Base

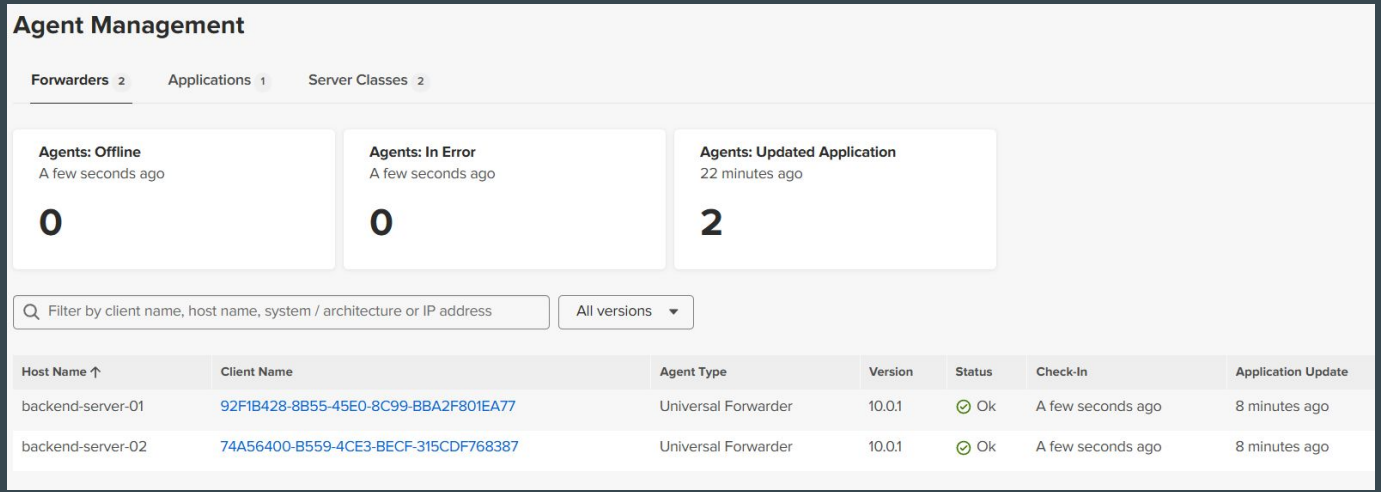
The agent management is a tool for managing data collection agents.

Its primary purpose is to manage large groups of forwarders,



Agent and Agent Management

Agents can be universal forwarders, heavy forwarders, indexers, search heads, or OTel collectors.



Point to Note

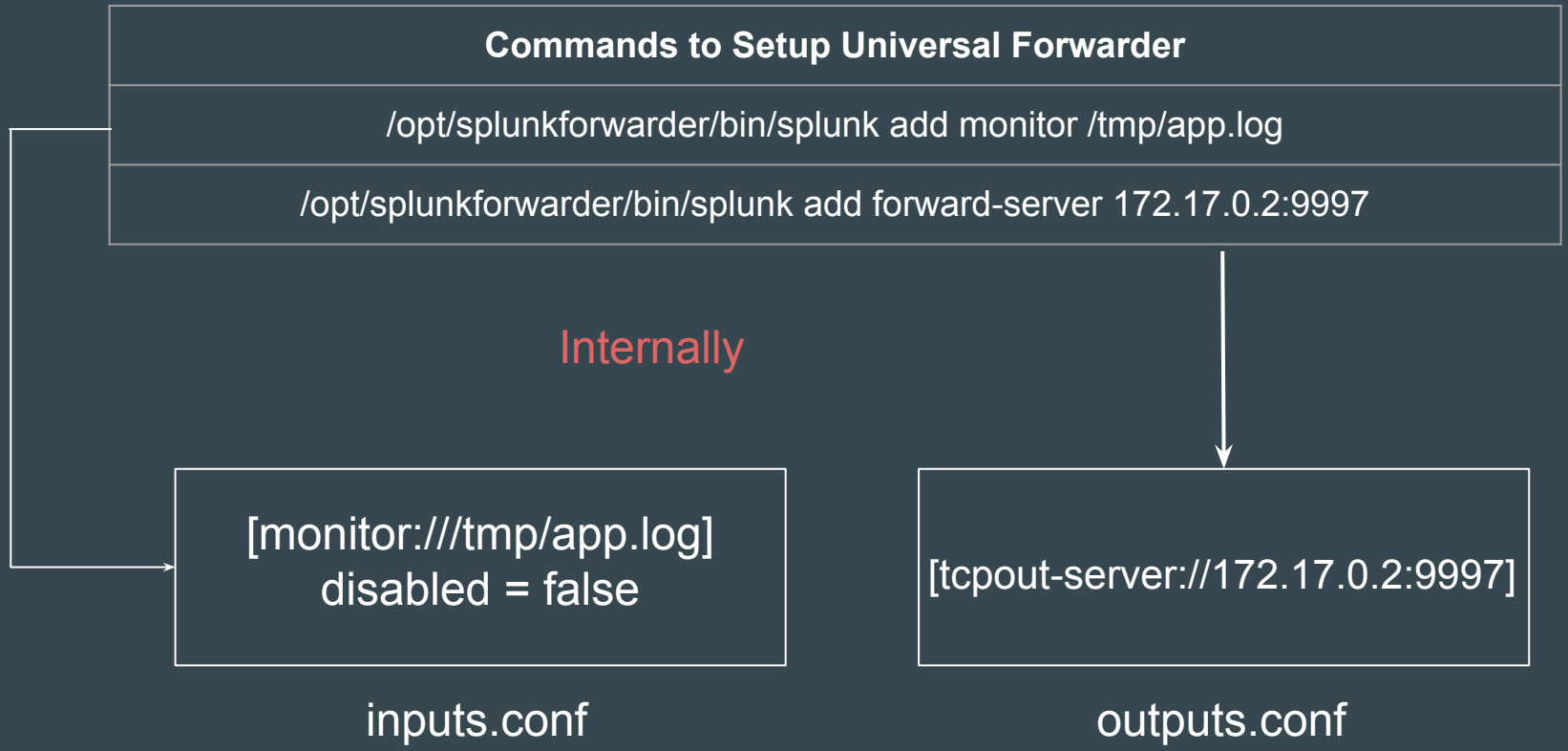
Starting with version 10.0, Splunk has updated the following product names:

1. Deployment Server is now called Agent Management

2. Forwarder Management is now called Agent Management

Server Class and Deployment App

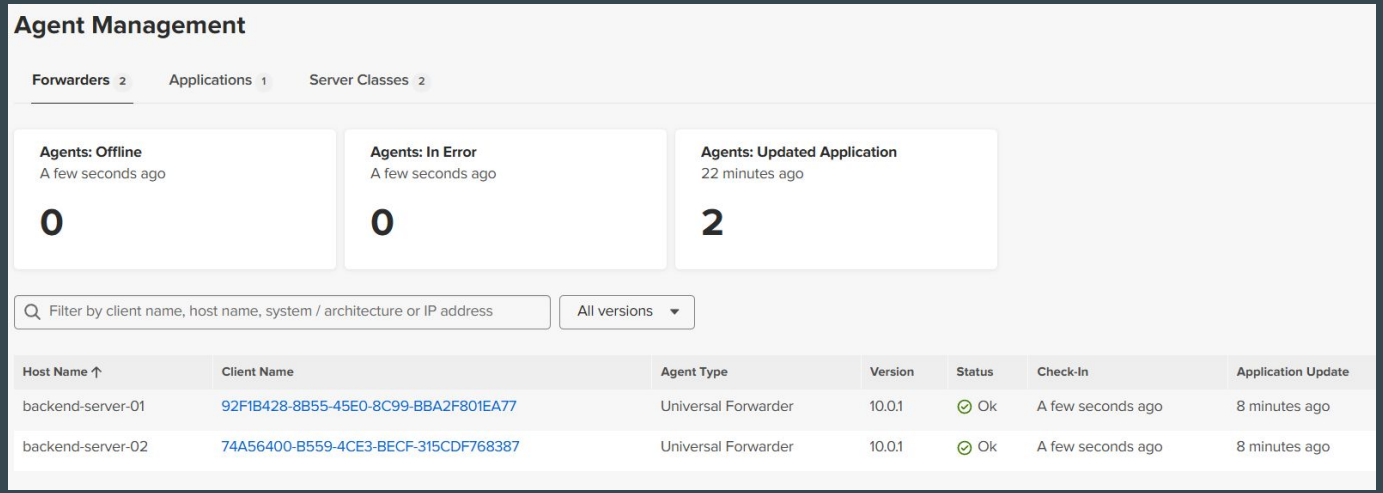
Commands to Setup Universal Forwarder



Setting the Base

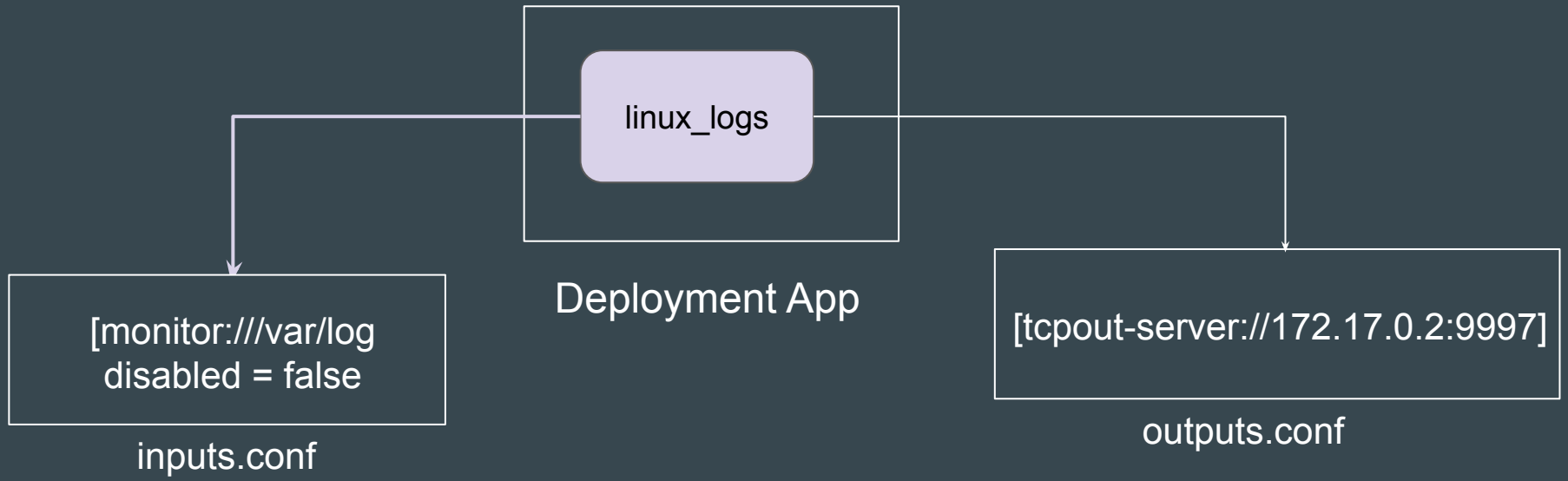
Two Important Terminologies in Agent Management:

Deployment Apps and Server Class



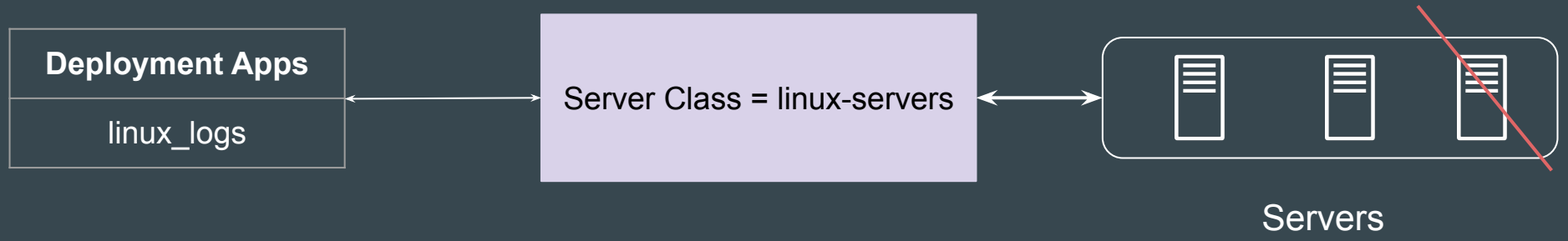
1 - Deployment App

A deployment application (app) is a set of content (including configuration files) maintained on the agent management and deployed as a unit to agents of a server class

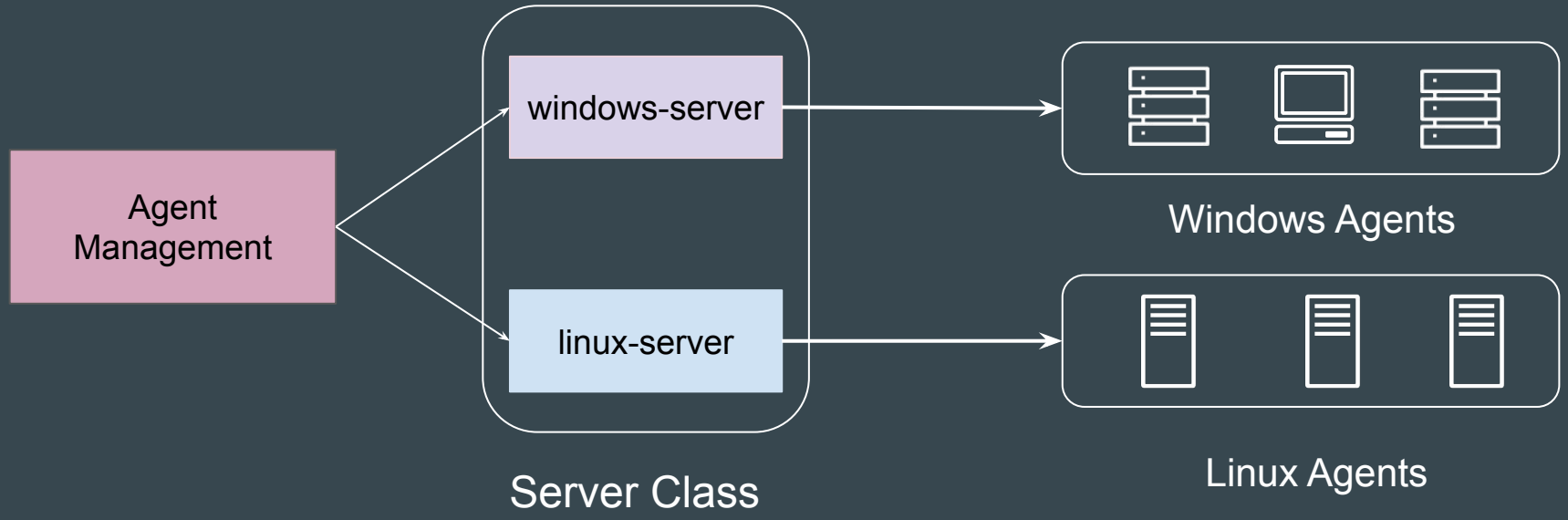


2 - Server Class

You use server classes to map a group of agents to one or more deployment apps.



High-Level Workflow - Multiple Server Class



Regular Expressions

Let's Regex

Basics of Regular Expressions

Regular Expressions (regex) is a sequence of characters that defines a search pattern.

"There is a Rainbow which arises on the south shore of Mumbai"

Rainbow - Literal Character

Meta Character is a character or sequence of character that has special meaning that provides information about the other characters

Meta Characters

\d - Any digit from 0-9
\w - Any word (A-Z, a-z, 0-9]
\s - whitespace

. - Any character.
[] - Matches characters in brackets.
[^] - Matches characters not in brackets.

Example 1: Matching Phone Numbers

Sample Data:

125-450-955
550.220.900

Example 2: Matching List of Names

Sample Regular Expression:

Mr. Zeal
Mr Harsh
Mrs Surekha
Mr L
Ms Alice

Parsing Web Server Logs

Let's Regex

Getting the basics straight

There are two ways to have data parsed in Splunk:

1) Create an addon and write custom regex
2) Use Add-Ons from marketplace which has built in regex for specific log.

Sample Log Data:

93.180.71.3 - - [17/May/2015:08:05:27 +0000] "GET /downloads/product_1 HTTP/1.1" 304 0 "-" "Debian APT-HTTP/1.3 (0.8.16~exp12ubuntu10.21)"

Named Capturing Group

Named Capturing group makes understanding the parsed data in much more easier manner.

It is used extensively in various Splunk Add-Ons available in the marketplace.

Sample Syntax:

(?<name>regex)

Importance of Source Types

Field Extractions

Importance of Source Types

In Splunk, field extractions and regex are generally defined at the source types level.

They can be defined in props.conf as well as transforms.conf

If source type of your log is incorrect then it will not get parsed properly.

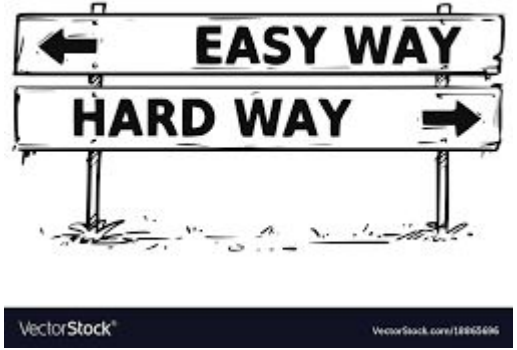
Splunk comes with some built-in source types and it's associated regex for common logs.

Interactive Field Extractor (IFX)

Field Extractions

The Easy way for extracting fields

Interactive Field Extractor allows us to teach splunk on how to extract fields from your data without writing regex.



props.conf and transforms.conf

Working with Source Types

Creating Custom Source Types

- Splunk comes with default source types and field extractions for common log files.
- However we can create our own custom source type as well.
- Every source type has some associated configuration settings.

These configuration parameters and source type details are stored in props and transforms.

Props and Transforms

In props.conf, we define that event with source type XXXX has the extraction of YYYY applied to it during the search time.

transforms.conf contains the actual extractions.

Event Types

Making the search better

Understanding EventTypes

EventTypes are categorization system to help you make sense of your data.

sourcetype=access_combined status=200 action=purchase

If you save the above as an eventtype success_purchase, any event that gets returned by the search gets associated eventtype

Limitation of Event Types

We cannot have event type based on search which has the following aspects:

i) Includes pipe operator after a simple search.

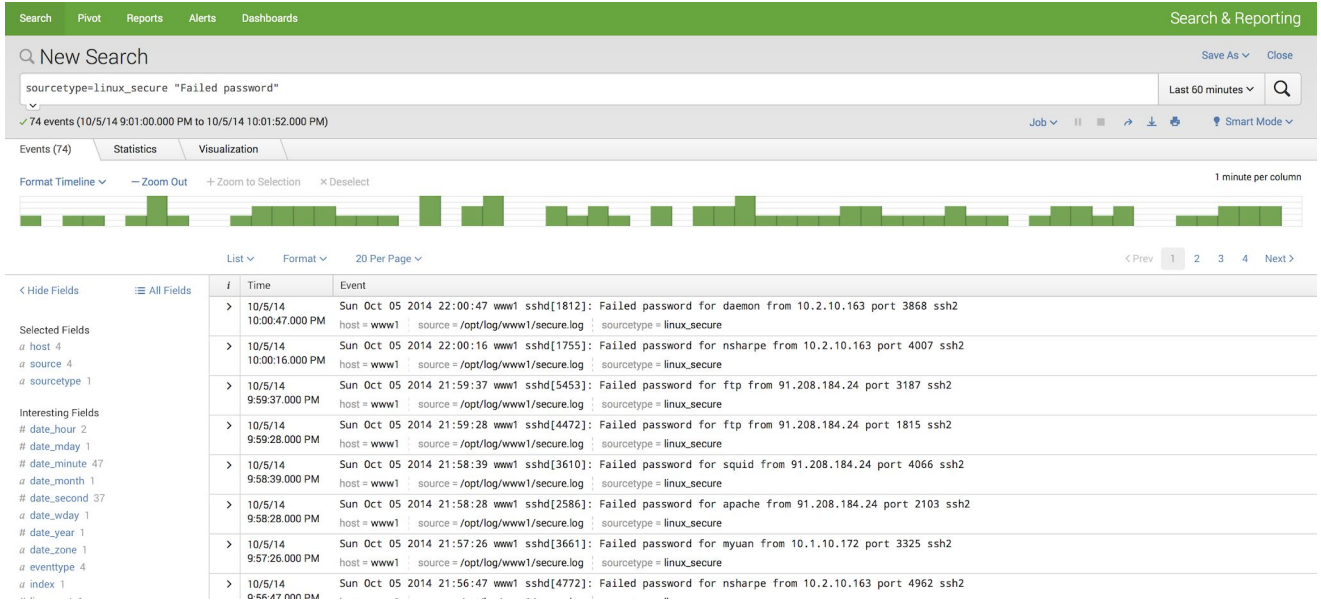
ii) Includes a sub-search.

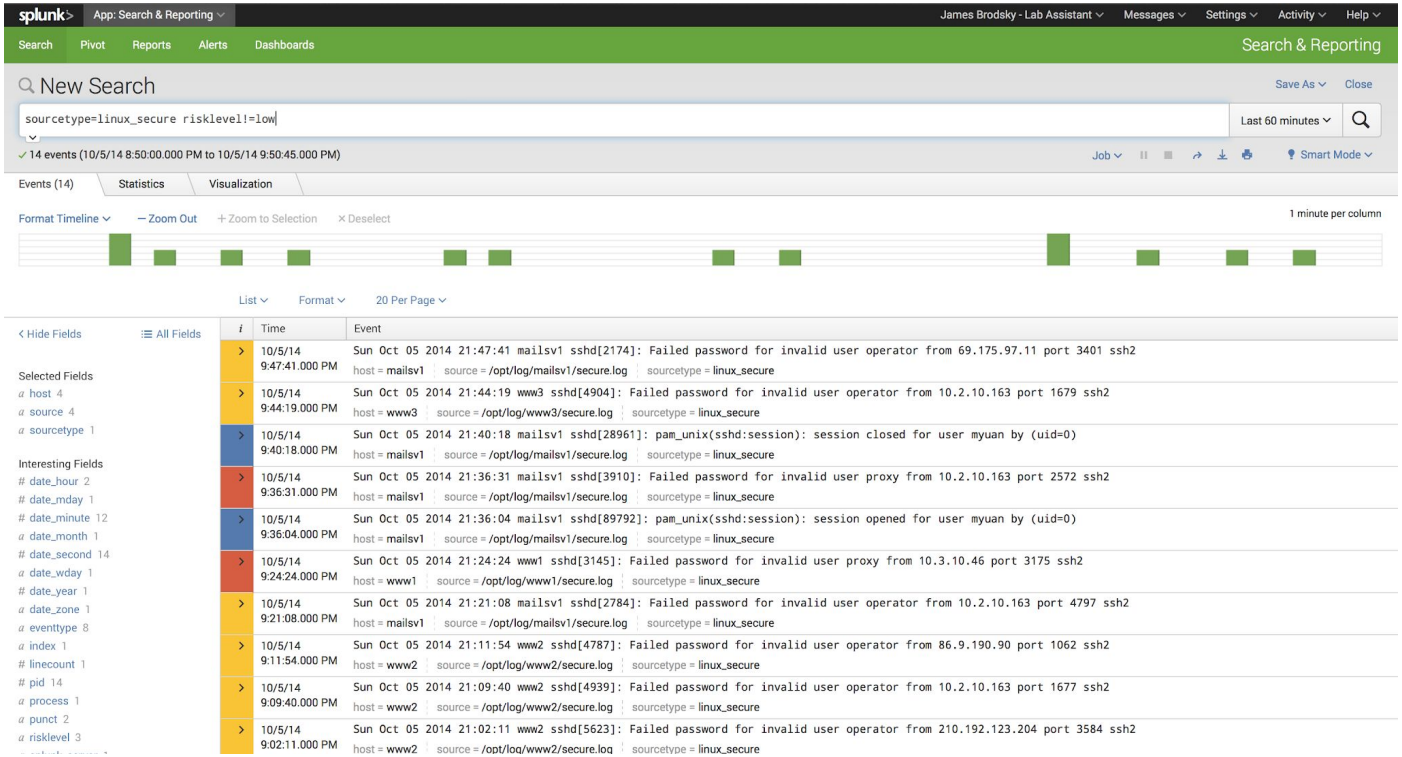
Colored Events

Making the search better

Understanding Colored Event Types

Typically event type field gets attached to the matching events when wildcard search is used.



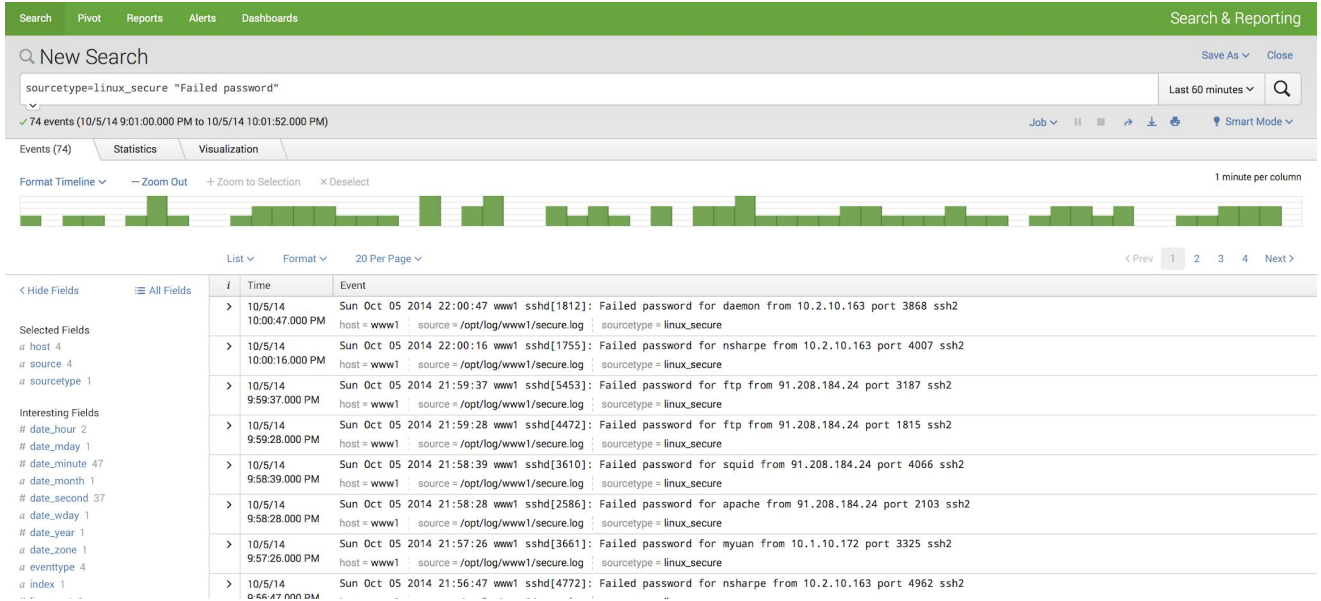

splunk>
App: Search & Reporting
James Brodsky - Lab Assistant
Messages
Settings
Activity
Help
Search
Pivot
Reports
Alerts
Dashboards
Search & Reporting
New Search
Save As
Close
sourcetype=linux_secure risklevel!=low
Last 60 minutes
14 events (10/5/14 8:50:00.000 PM to 10/5/14 9:50:45.000 PM)
Job
Smart Mode
Events (14)
Statistics
Visualization
Format Timeline
Zoom Out
Zoom to Selection
Deselect
1 minute per column
List
Format
20 Per Page
Hide Fields
All Fields
Selected Fields
a host 4
a source 4
a sourcetype 1
Interesting Fields
date_hour 2
date_mday 1
date_minute 12
a date_month 1
date_second 14
a date_wday 1
date_year 1
a date_zone 1
a eventtype 8
a index 1
linecount 1
pid 14
a process 1
a punct 2
a risklevel 3
i
Time
Event
10/5/14 9:47:41.000 PM
Sun Oct 05 2014 21:47:41 mailsv1 sshd[2174]: Failed password for invalid user operator from 69.175.97.11 port 3401 ssh2
host = mailsv1 source = /opt/log/mailsv1/secure.log sourcetype = linux_secure
10/5/14 9:44:19.000 PM
Sun Oct 05 2014 21:44:19 www3 sshd[4904]: Failed password for invalid user operator from 10.2.10.163 port 1679 ssh2
host = www3 source = /opt/log/www3/secure.log sourcetype = linux_secure
10/5/14 9:40:18.000 PM
Sun Oct 05 2014 21:40:18 mailsv1 sshd[28961]: pam_unix(sshd:session): session closed for user myuan by (uid=0)
host = mailsv1 source = /opt/log/mailsv1/secure.log sourcetype = linux_secure
10/5/14 9:36:31.000 PM
Sun Oct 05 2014 21:36:31 mailsv1 sshd[3910]: Failed password for invalid user proxy from 10.2.10.163 port 2572 ssh2
host = mailsv1 source = /opt/log/mailsv1/secure.log sourcetype = linux_secure
10/5/14 9:36:04.000 PM
Sun Oct 05 2014 21:36:04 mailsv1 sshd[89792]: pam_unix(sshd:session): session opened for user myuan by (uid=0)
host = mailsv1 source = /opt/log/mailsv1/secure.log sourcetype = linux_secure
10/5/14 9:24:24.000 PM
Sun Oct 05 2014 21:24:24 www1 sshd[3145]: Failed password for invalid user proxy from 10.3.10.46 port 3175 ssh2
host = www1 source = /opt/log/www1/secure.log sourcetype = linux_secure
10/5/14 9:21:08.000 PM
Sun Oct 05 2014 21:21:08 mailsv1 sshd[2784]: Failed password for invalid user operator from 10.2.10.163 port 4797 ssh2
host = mailsv1 source = /opt/log/mailsv1/secure.log sourcetype = linux_secure
10/5/14 9:11:54.000 PM
Sun Oct 05 2014 21:11:54 www2 sshd[4787]: Failed password for invalid user operator from 86.9.190.90 port 1062 ssh2
host = www2 source = /opt/log/www2/secure.log sourcetype = linux_secure
10/5/14 9:09:40.000 PM
Sun Oct 05 2014 21:09:40 www2 sshd[4939]: Failed password for invalid user operator from 10.2.10.163 port 1677 ssh2
host = www2 source = /opt/log/www2/secure.log sourcetype = linux_secure
10/5/14 9:02:11.000 PM
Sun Oct 05 2014 21:02:11 www2 sshd[5623]: Failed password for invalid user operator from 210.192.123.204 port 3584 ssh2
host = www2 source = /opt/log/www2/secure.log sourcetype = linux_secure

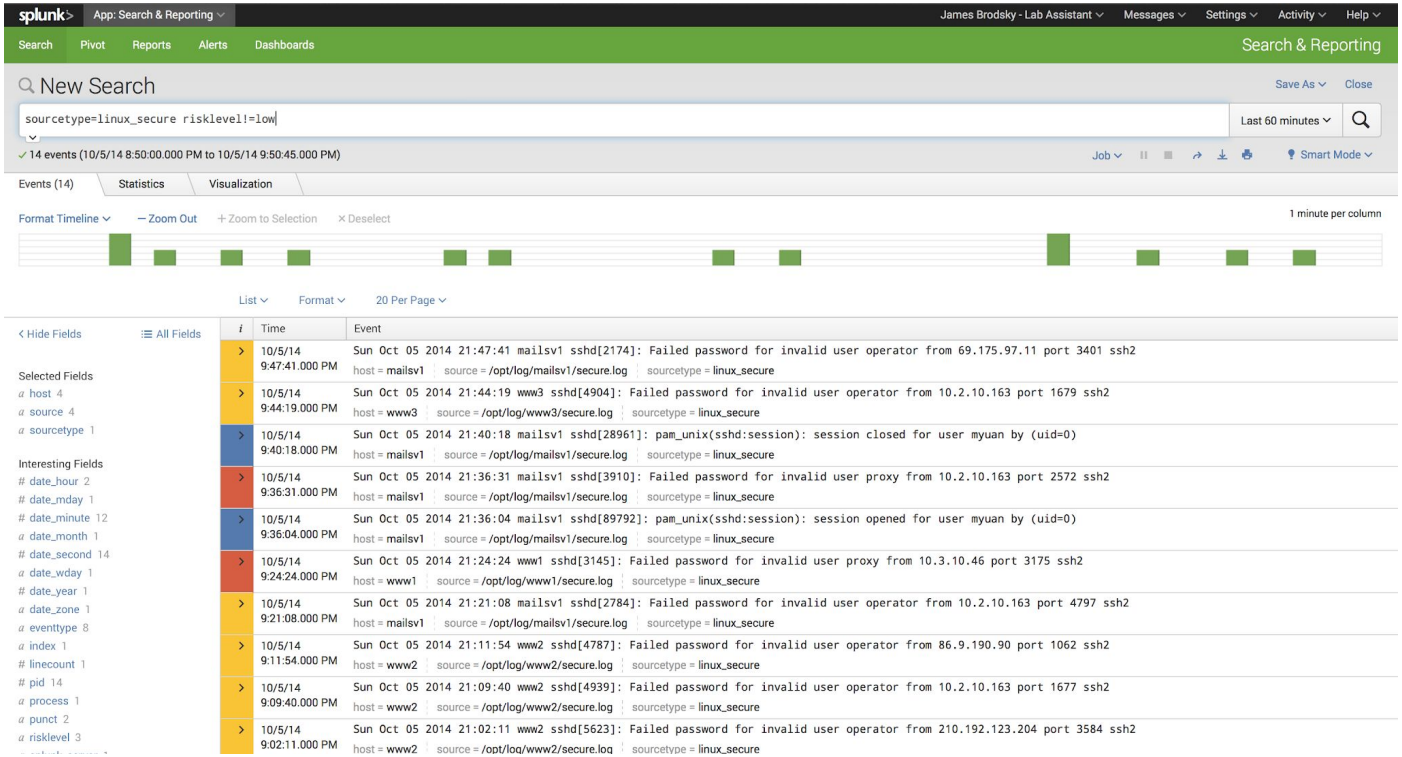
Colored Events

Making the search better

Understanding Colored Event Types

Typically event type field gets attached to the matching events when wildcard search is used.




splunk>
App: Search & Reporting
James Brodsky - Lab Assistant
Messages
Settings
Activity
Help
Search
Pivot
Reports
Alerts
Dashboards
Search & Reporting
New Search
Save As
Close
sourcetype=linux_secure risklevel!=low
Last 60 minutes
14 events (10/5/14 8:50:00.000 PM to 10/5/14 9:50:45.000 PM)
Job
Smart Mode
Events (14)
Statistics
Visualization
Format Timeline
Zoom Out
Zoom to Selection
Deselect
1 minute per column
List
Format
20 Per Page
Hide Fields
All Fields
Selected Fields
a host 4
a source 4
a sourcetype 1
Interesting Fields
date_hour 2
date_mday 1
date_minute 12
a date_month 1
date_second 14
a date_wday 1
date_year 1
a date_zone 1
a eventtype 8
a index 1
linecount 1
pid 14
a process 1
a punct 2
a risklevel 3
i
Time
Event
10/5/14 9:47:41.000 PM
Sun Oct 05 2014 21:47:41 mailsv1 sshd[2174]: Failed password for invalid user operator from 69.175.97.11 port 3401 ssh2
host = mailsv1 source = /opt/log/mailsv1/secure.log sourcetype = linux_secure
10/5/14 9:44:19.000 PM
Sun Oct 05 2014 21:44:19 www3 sshd[4904]: Failed password for invalid user operator from 10.2.10.163 port 1679 ssh2
host = www3 source = /opt/log/www3/secure.log sourcetype = linux_secure
10/5/14 9:40:18.000 PM
Sun Oct 05 2014 21:40:18 mailsv1 sshd[28961]: pam_unix(sshd:session): session closed for user myuan by (uid=0)
host = mailsv1 source = /opt/log/mailsv1/secure.log sourcetype = linux_secure
10/5/14 9:36:31.000 PM
Sun Oct 05 2014 21:36:31 mailsv1 sshd[3910]: Failed password for invalid user proxy from 10.2.10.163 port 2572 ssh2
host = mailsv1 source = /opt/log/mailsv1/secure.log sourcetype = linux_secure
10/5/14 9:36:04.000 PM
Sun Oct 05 2014 21:36:04 mailsv1 sshd[89792]: pam_unix(sshd:session): session opened for user myuan by (uid=0)
host = mailsv1 source = /opt/log/mailsv1/secure.log sourcetype = linux_secure
10/5/14 9:24:24.000 PM
Sun Oct 05 2014 21:24:24 www1 sshd[3145]: Failed password for invalid user proxy from 10.3.10.46 port 3175 ssh2
host = www1 source = /opt/log/www1/secure.log sourcetype = linux_secure
10/5/14 9:21:08.000 PM
Sun Oct 05 2014 21:21:08 mailsv1 sshd[2784]: Failed password for invalid user operator from 10.2.10.163 port 4797 ssh2
host = mailsv1 source = /opt/log/mailsv1/secure.log sourcetype = linux_secure
10/5/14 9:11:54.000 PM
Sun Oct 05 2014 21:11:54 www2 sshd[4787]: Failed password for invalid user operator from 86.9.190.90 port 1062 ssh2
host = www2 source = /opt/log/www2/secure.log sourcetype = linux_secure
10/5/14 9:09:40.000 PM
Sun Oct 05 2014 21:09:40 www2 sshd[4939]: Failed password for invalid user operator from 10.2.10.163 port 1677 ssh2
host = www2 source = /opt/log/www2/secure.log sourcetype = linux_secure
10/5/14 9:02:11.000 PM
Sun Oct 05 2014 21:02:11 www2 sshd[5623]: Failed password for invalid user operator from 210.192.123.204 port 3584 ssh2
host = www2 source = /opt/log/www2/secure.log sourcetype = linux_secure

Access Control

Permissions Matrix

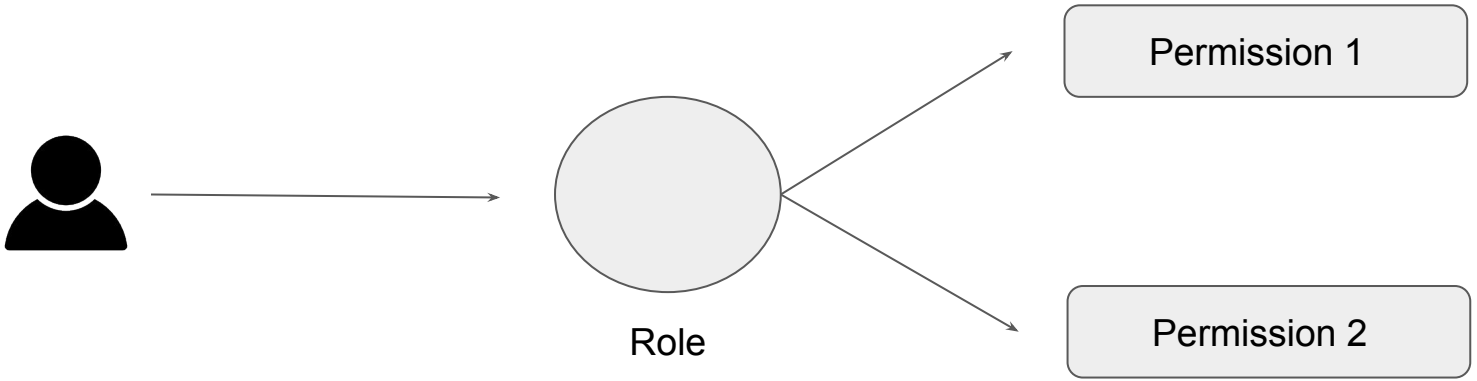
Need of Access Control

i) Extremely Sensitive Data were having access to system might involve legal risk, consider using a isolated separate Splunk instance for relevant audience.

ii) When you have sensitive data, then you can restrict access based on Index to the users.

User Authentication

Splunk Enterprise Authentication allows us to create users, add them to roles (groups) and assign permissions to those roles.



Distributed Splunk Architecture

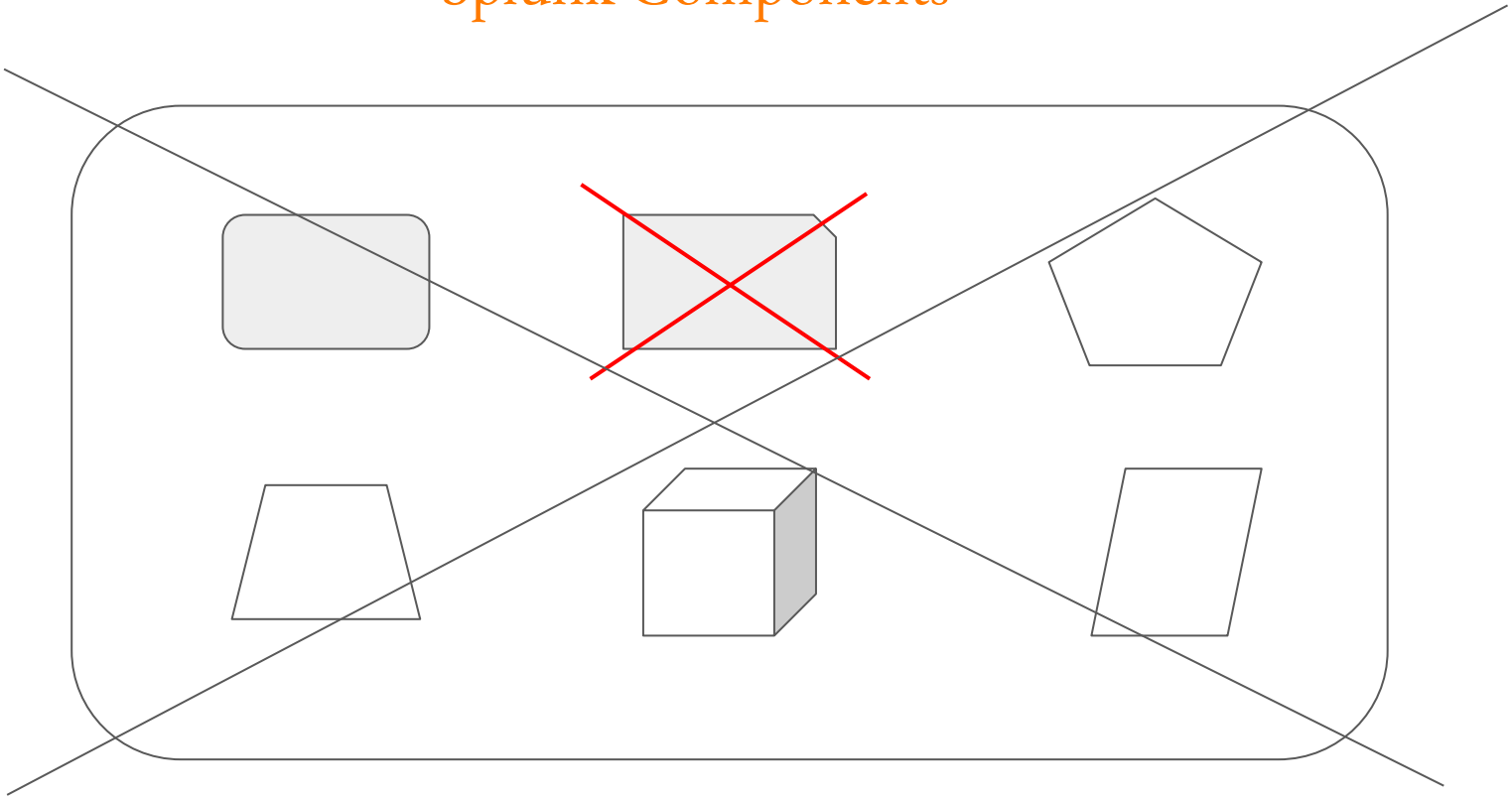
Building Base for Clustering

Understanding Splunk Components

Splunk Enterprise consist of various other sub-components. These includes:

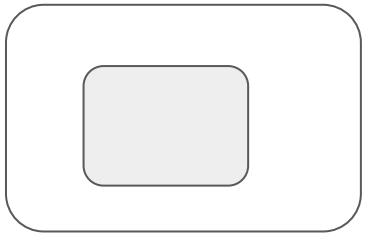
- Indexer
- Search Head
- Deployment Server
- Forwarders
- License Master
- Monitoring Console

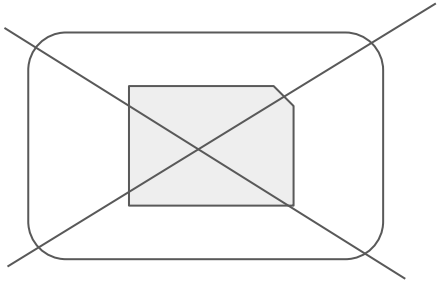
Splunk Components

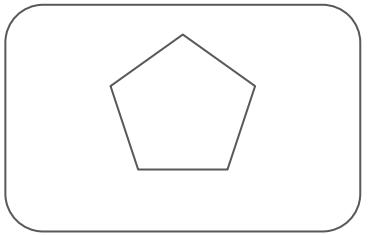


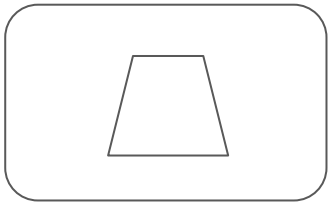
Splunk Enterprise

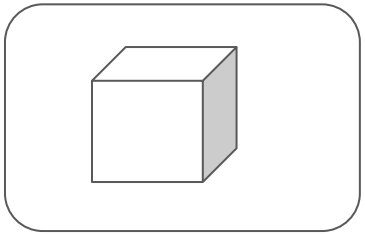
Distributed Splunk Components

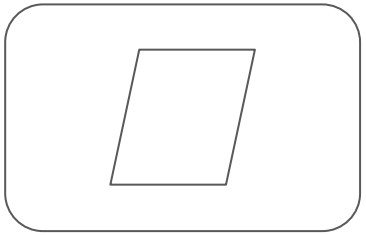




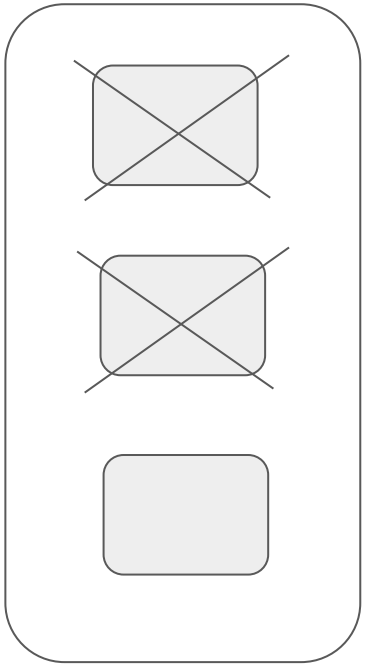




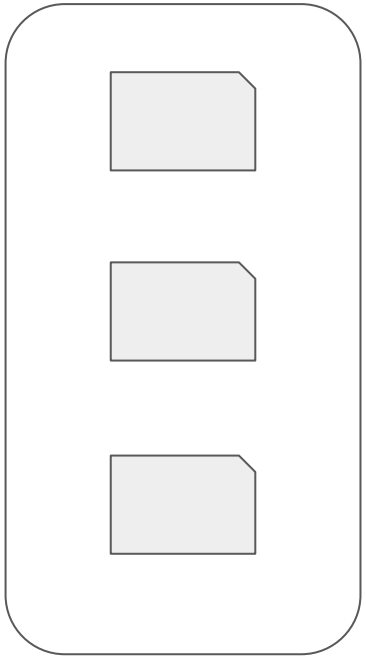




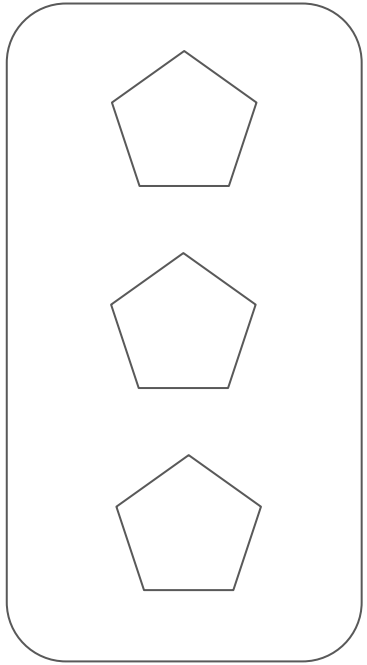
Clustered Setup



Component 1 - Cluster

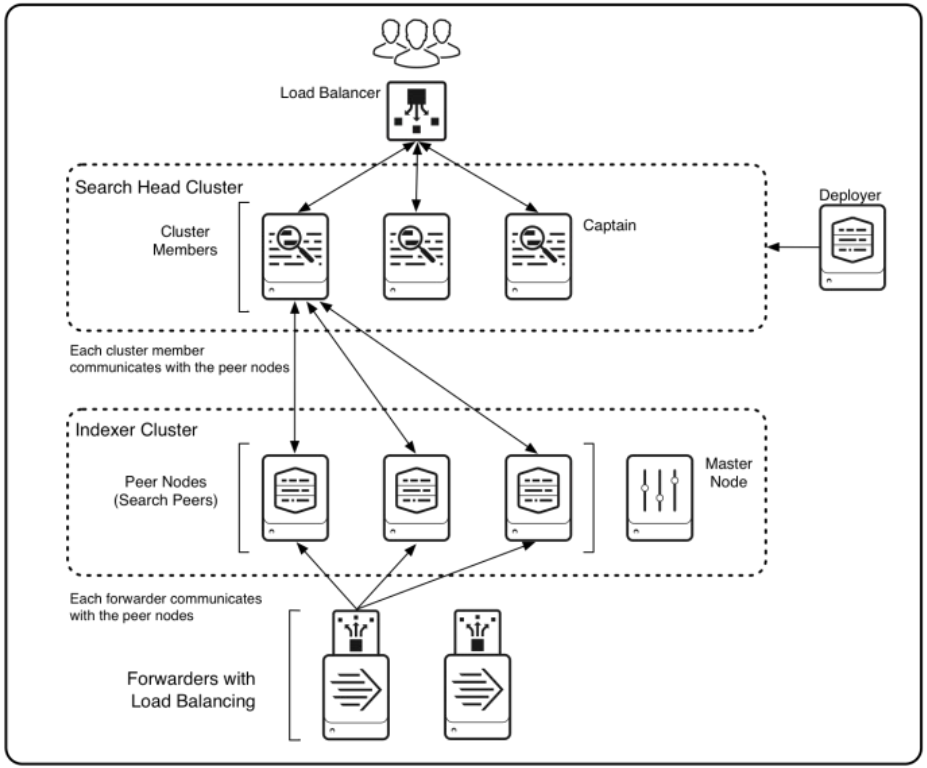


Component 2- Cluster



Component 3- Cluster

Clustered Setup



It is important to understand components

Splunk Enterprise consist of various other sub-components. These includes:

- Indexer
- Search Head
- Deployment Server
- Forwarders
- License Master
- Monitoring Console

Splunk Clustering Setup

Splunk Enterprise supports clustering features for two major components:

- Indexer
- Search Heads

For components like license master, heavy forward etc, Splunk supports Active-Passive failover.

License Master

Understanding Licensing Aspects

Understanding how licensing works!

- Splunk Enterprise ingests external data, indexes it, and stores it on disk.
- Licenses specify how much external data you can index per day.
- All Splunk Enterprise instances require a license.
- If you have a standalone indexer, you can install the license locally.
- In-case of distributed env, we need to configure a license master.

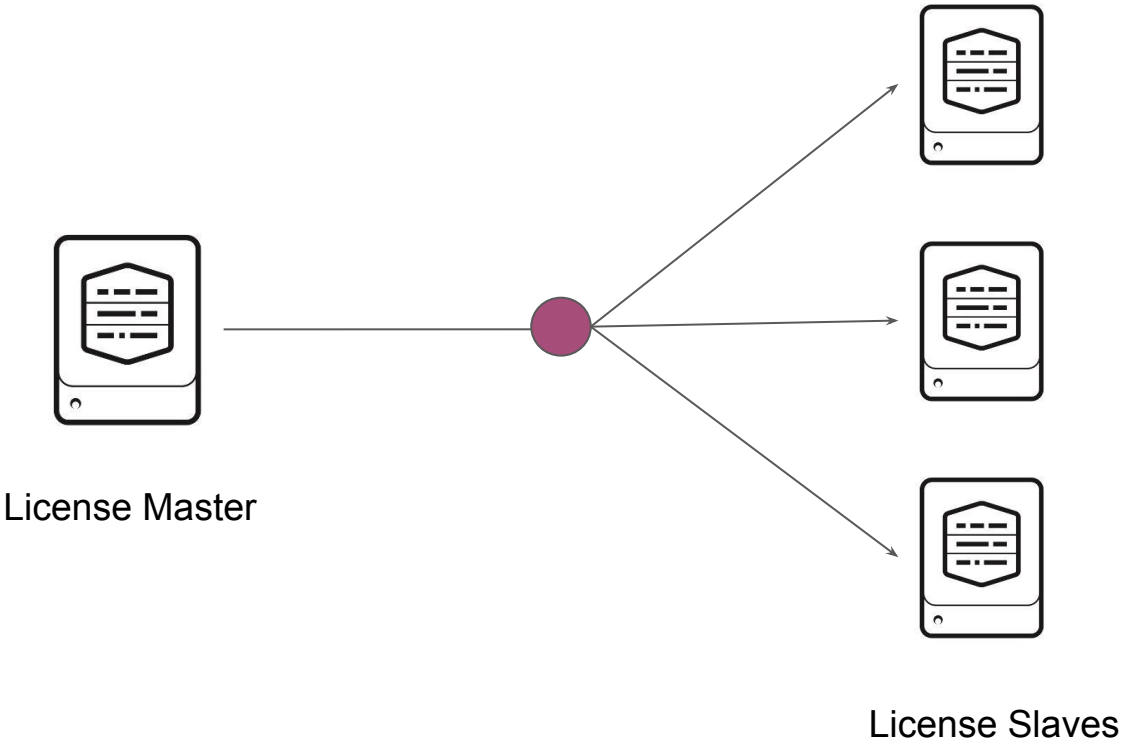
License Architecture

Splunk Instance 01

Splunk Instance 02

Splunk Instance 03

License Master Architecture



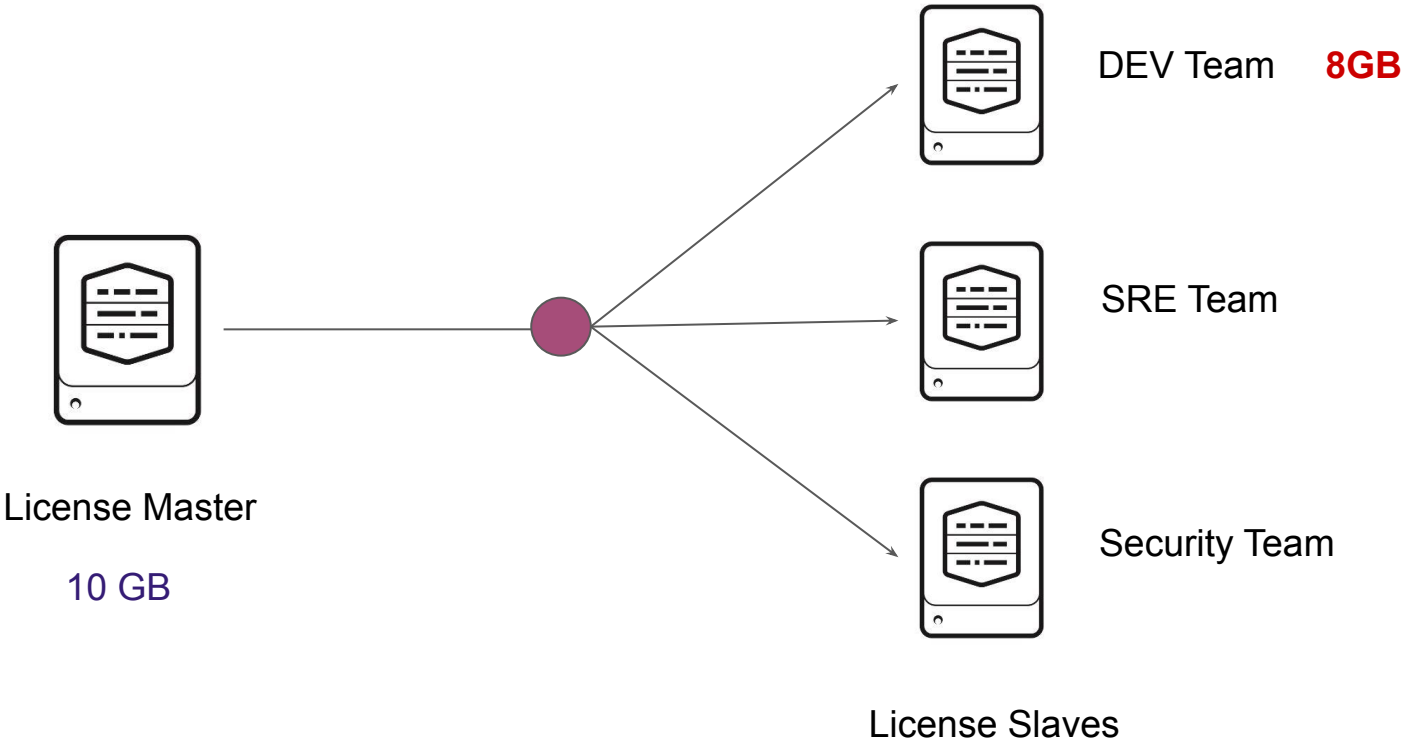
License Pool

Understanding Licensing Aspects

Overview of License Pool

- Splunk License resides in license stack called as Splunk Enterprise Stack.
- The stack has default license pool called auto_generated_pool_enterprise
- Any license slave that connects to this license master has access to the default pool.

License Master Architecture



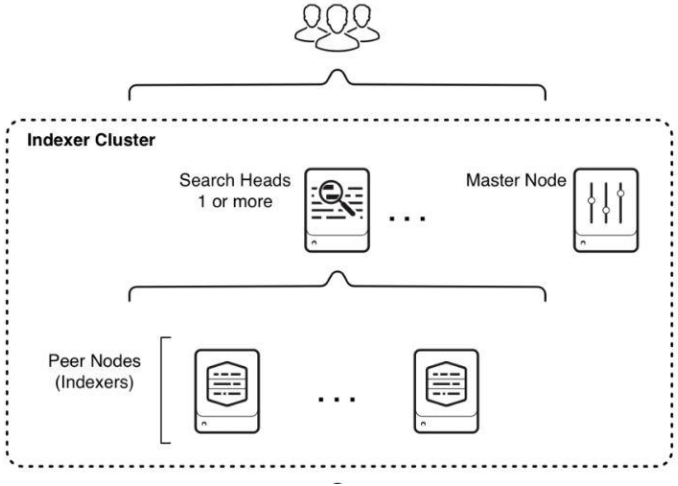
Indexer

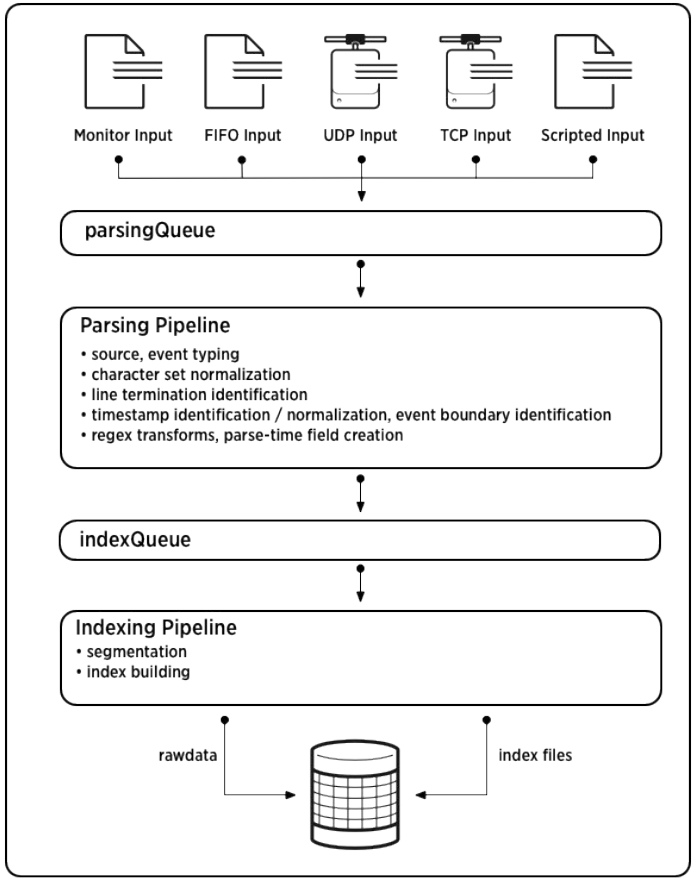
Indexer Component

Overview of Indexer

Indexer is a component in Splunk Enterprise whose responsibility is to index data, transform data into searchable events and placing results into an index.

To ensure high availability of data, we can deploy indexer cluster.





Parsing Stage

Indexing Stage

Parsing and Indexing Stage

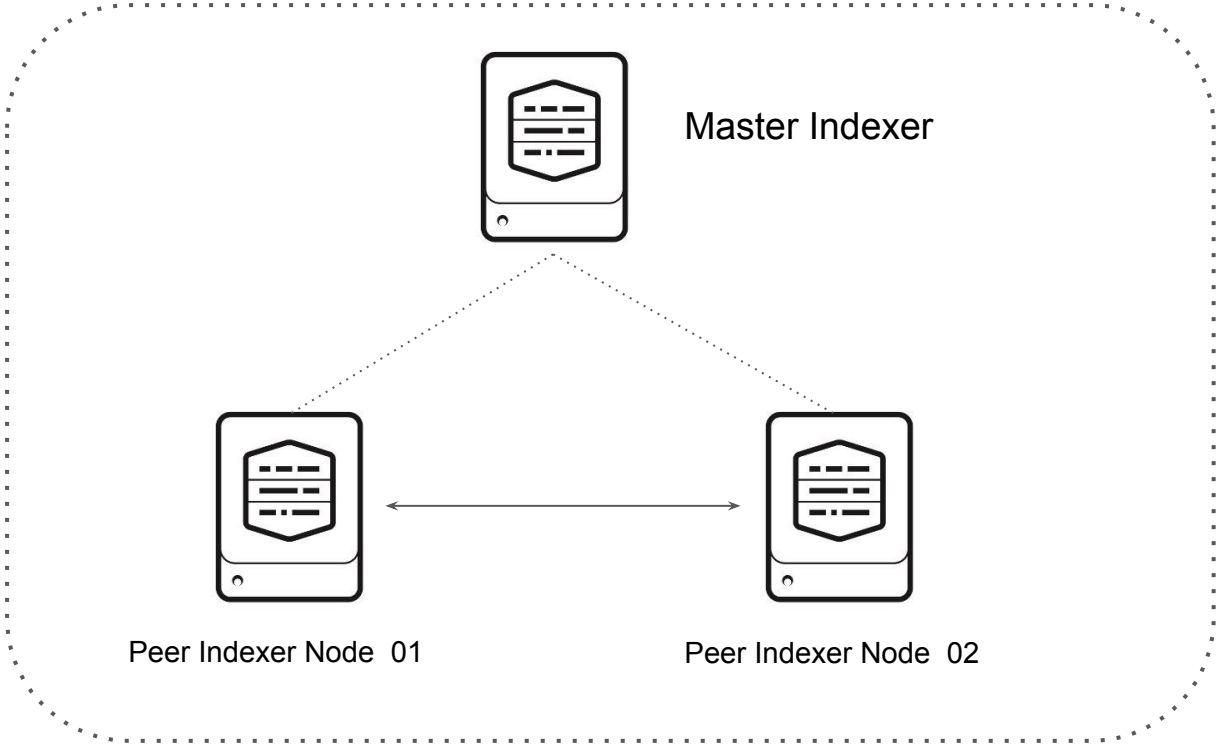
While parsing, Splunk Enterprise performs number of actions, including:

- Extracting the set of default fields for each event, including host, source and sourcetype.
- Configuring character set encoding.
- Identifying line termination using line break rules.
- Mask sensitive details in data like credit card numbers.

During indexing pipeline, splunk performs

- Breaking events into segments that can be searched upon.
- Building index data structures.
- Writing raw data and index files to disks.

Indexer Cluster Architecture



Masking Sensitive Data

Indexer Component

Importance of Masking Data

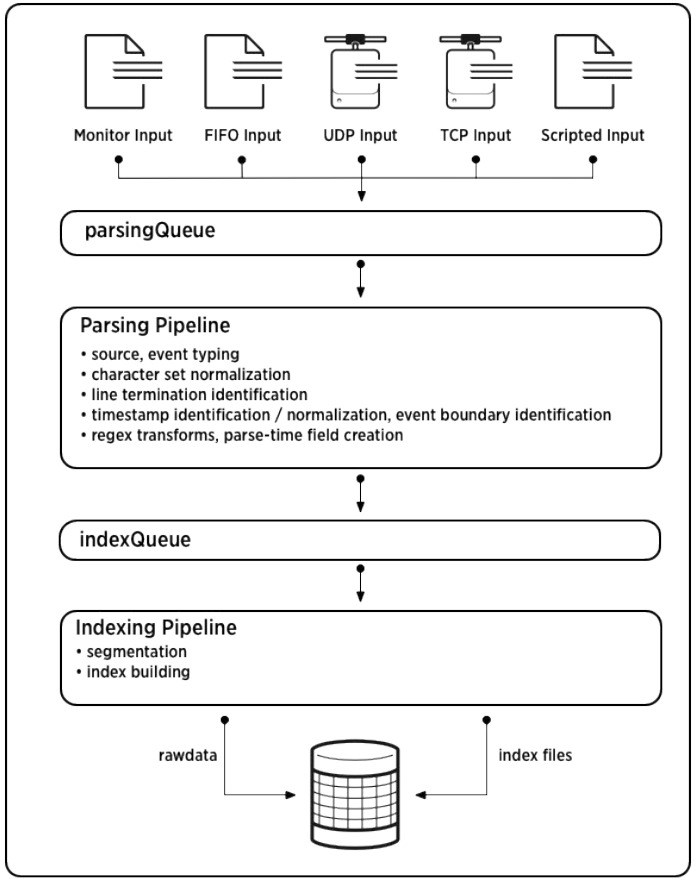
It might happen that log files would have sensitive information like Credit Cards, SSN etc.

In such use-cases, you might want to mask such information;

ss=123456789, cc=1234-5678-9012-3456

TO

ss=123456789, cc=xxxx-xxxx-xxxx-3456



Parsing Stage

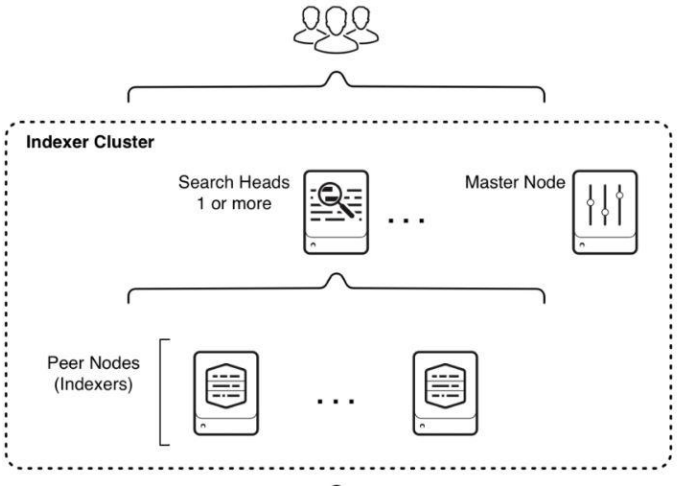
Indexing Stage

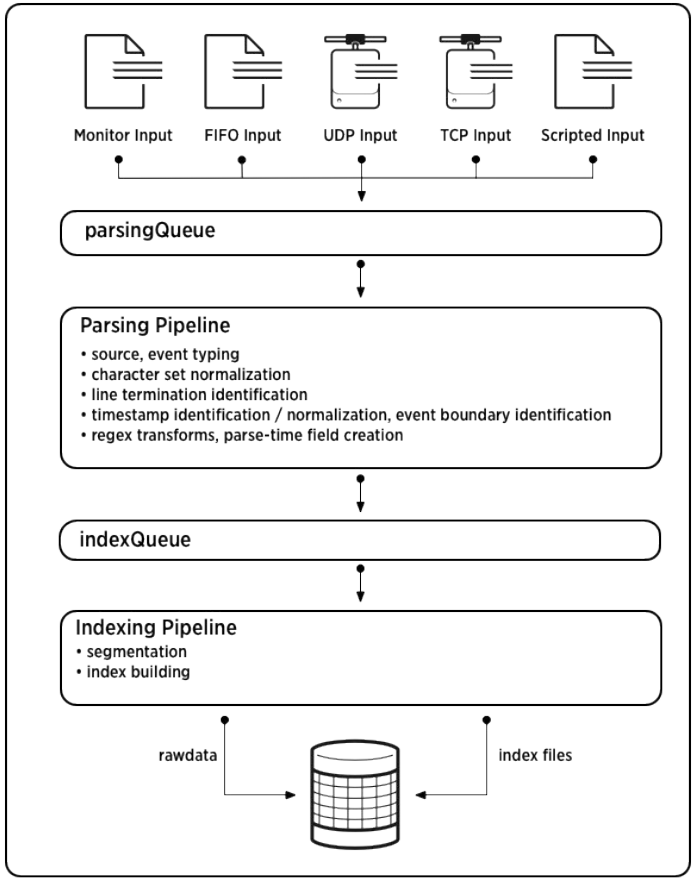
Search Head

Searching Component

Overview of Search Head

Search Head is a component in Splunk Enterprise whose responsibility is to handle the search management functions, directing search requests to search peers and then merging the results back to the users.





Parsing Stage

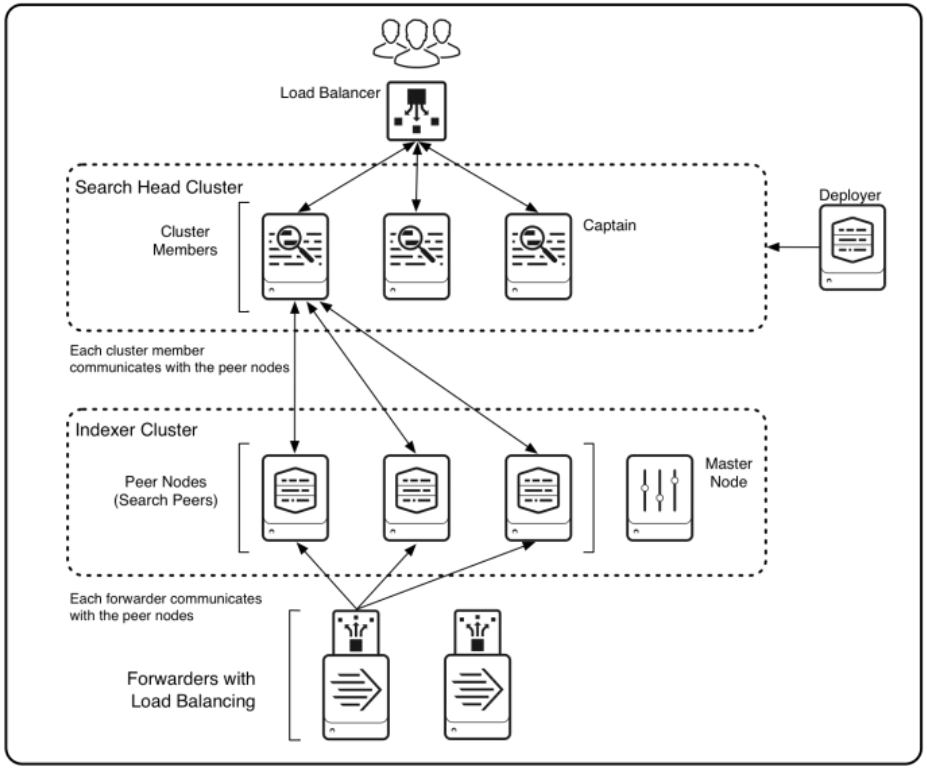
Indexing Stage

Typical Functions of Search Head

Search Heads are used for number of functions, some of the primary one includes::

- Search Related Functions.
- Building Dashboards and Reports.
- Data Models.
- Alerting Related Functionality.

Clustered Setup



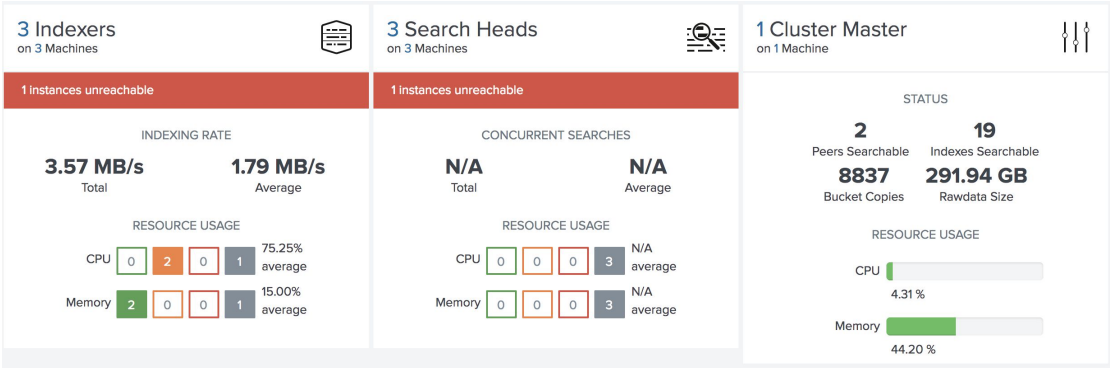
Monitoring Console

Monitoring Component

Overview of Monitoring Console

Monitoring Console allows us to view is detailed information about the topology and performance of your Splunk Enterprise deployment.

The Monitoring Console provides pre-built dashboards that give you visibility into many areas of your deployment, including search and indexing performance, resource usage, license usage, and more.



Available Dashboards in Monitoring Console

Pre-Built available dashboards in Monitoring Console provide insights into following areas:

- search performance and distributed search framework
- indexing performance
- operating system resource usage
- Splunk app key value store performance
- search head and indexer clustering
- index and volume usage
- forwarder connections and Splunk TCP performance
- HTTP Event Collector performance
- and license usage.

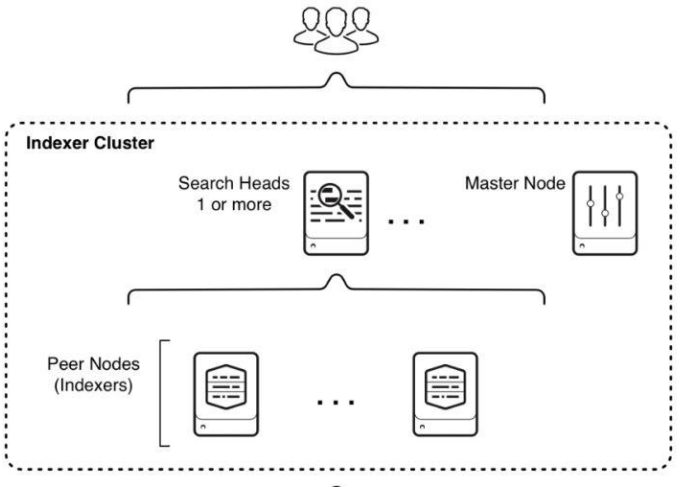
Indexer Clustering

Index Architecture

Overview of Indexer

Indexer is a component in Splunk Enterprise whose responsibility is to index data, transform data into searchable events and placing results into an index.

To ensure high availability of data, we can deploy indexer cluster.

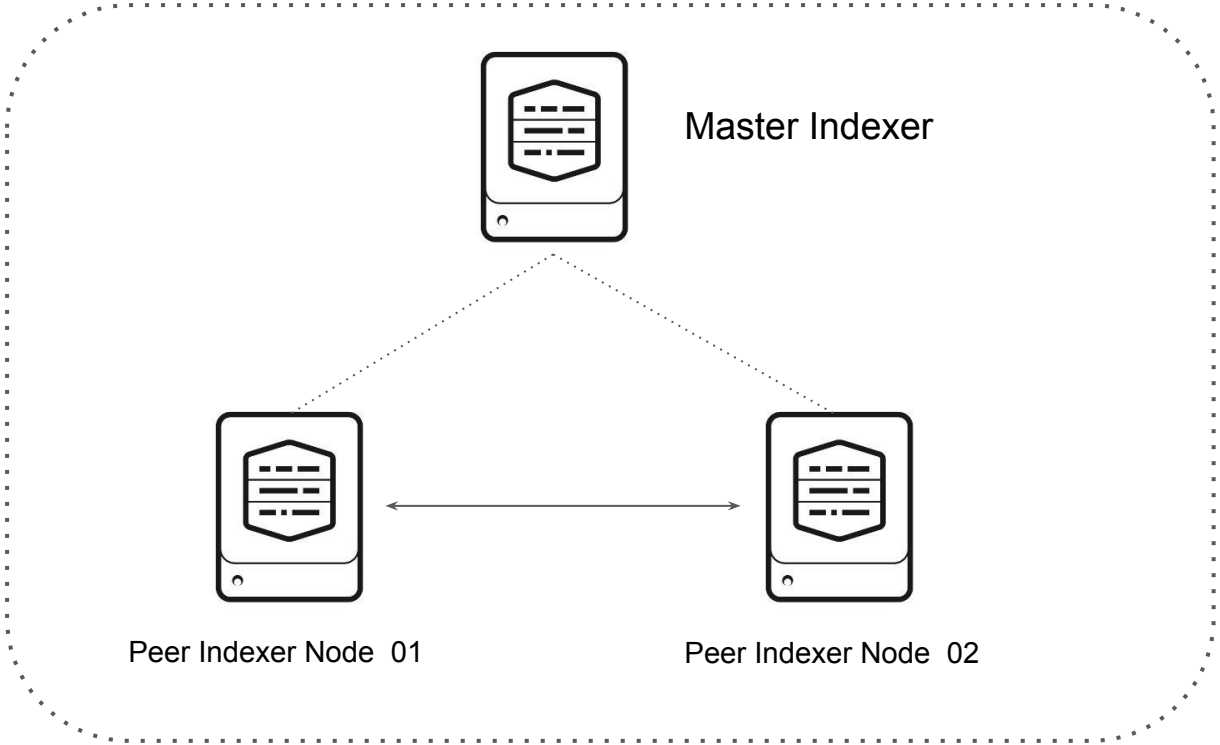


Master Node and Peer Node

Master Node coordinates the activities of the peer nodes.

Peer Nodes are the nodes which has the actual data and performs the replication related activities. Any search related activity requests are sent to the peer nodes.

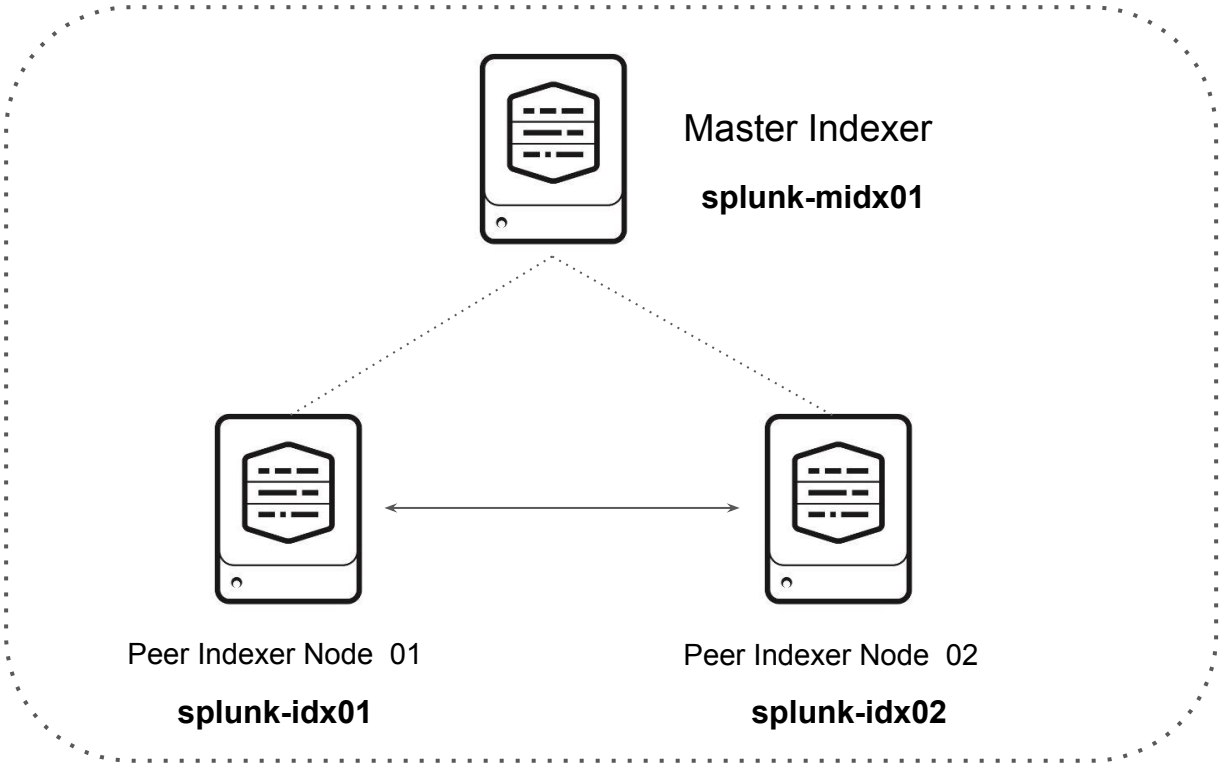
Indexer Cluster Architecture



Infrastructure for IDX Cluster

Let's Start Exploring

Indexer Cluster Architecture



Things to Do

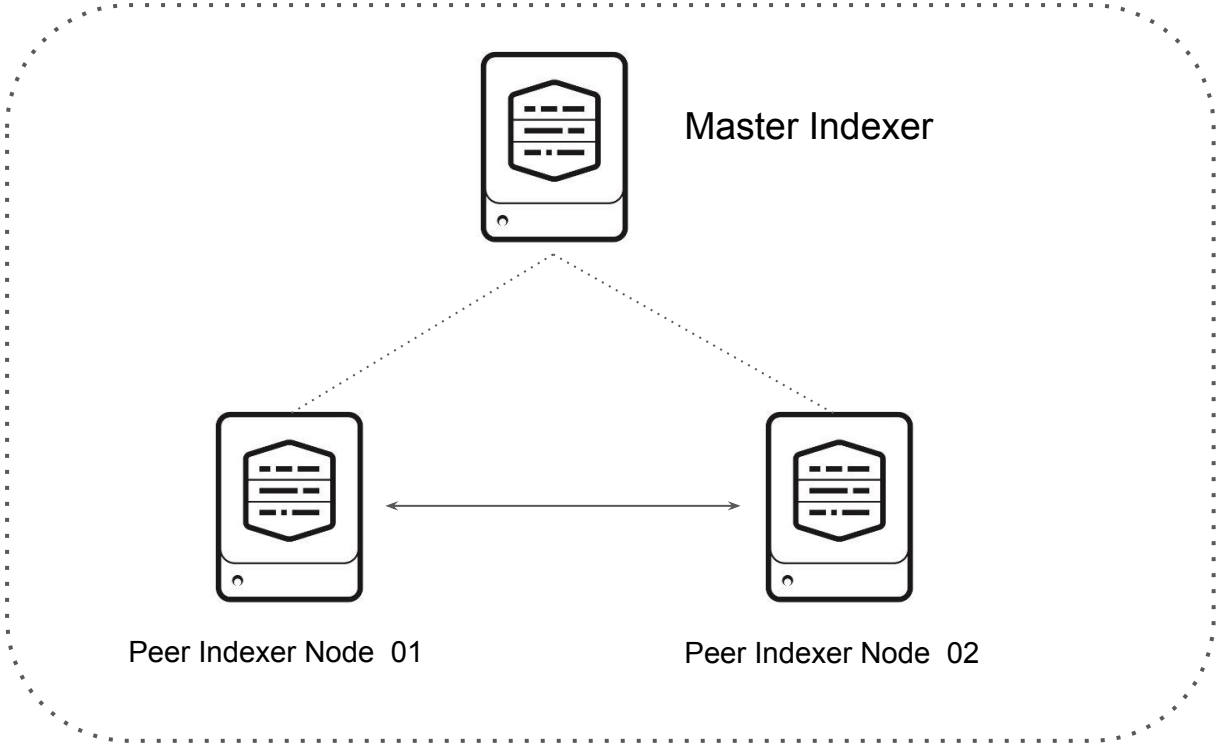
1. Launch 3 Servers representing 1 Master Indexer and 2 Peer Indexer Nodes.

2. Install and Start Splunk in all of them.

Master Indexer

Index Architecture

Indexer Cluster Architecture



Master Indexer

A cluster has one, and only one master node.

Master node coordinates activities of the peer nodes.

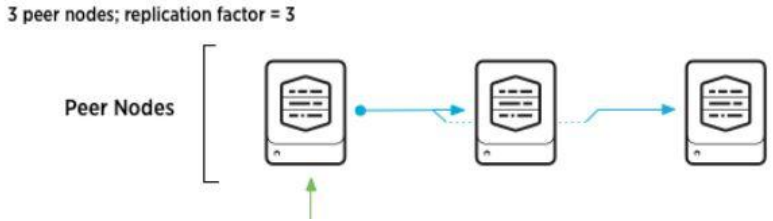
It does not itself store or replicate data.

Replication Factor

Replication Factor determines how many copies of data the cluster maintains.

This is a key factor since it determines the cluster's fault tolerance.

For example, if we want to ensure that system can handle failure of two peer nodes, we must configure replication factor of 3, which means cluster will store 3 identical copies of data on separate ndes.



Search Factor

Search Factor determines the number of immediately searchable copies of data the cluster maintains.

Searchable copies of data requires more storage than non-searchable copies.

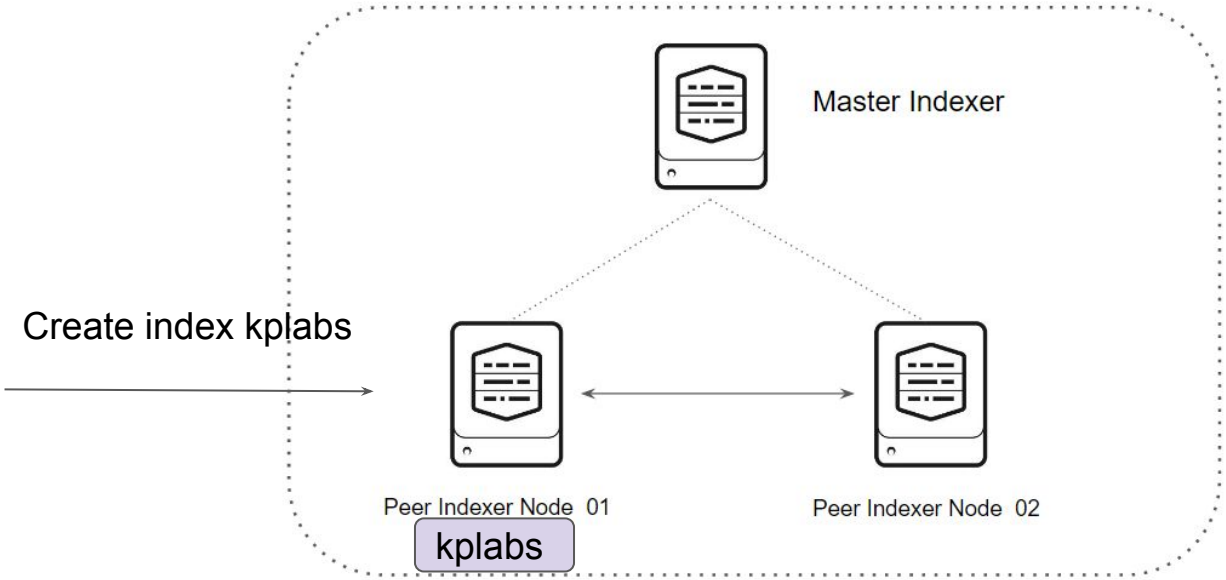
A non-searchable copy is basically plain raw data without any index files.

Configuration Bundle

Deploying Updates to Peer Indexers

Avoid Configuring Things in Peer Nodes

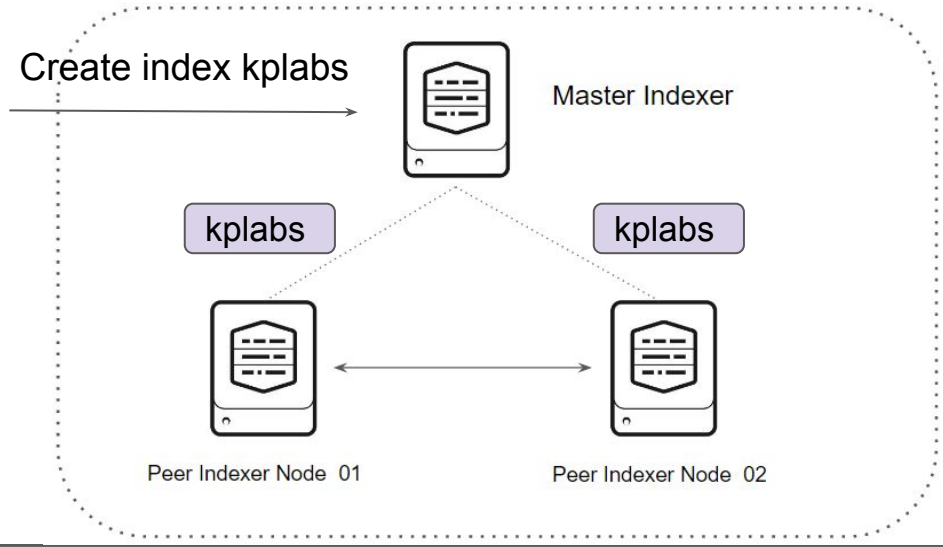
Manual changes made to the Indexer Peer nodes would not always be replicated.



Pushing Things Centrally

It is recommended to the set of configurations that is common to all peers via Master Indexer.

This common set of configuration is referred to as Configuration Bundle.



Location of Configuration Bundle

Component	Location
Master Indexer	$SPLUNK_HOME/etc/master-apps
Peer Indexer	$SPLUNK_HOME/etc/slave-apps

Forwarding Logs to Indexer Cluster

Send Logs to Indexers

Approach to Send Data to Indexer Cluster

Once we have the Indexer cluster built, universal forwarder would need to start sending logs to the peer indexer nodes.

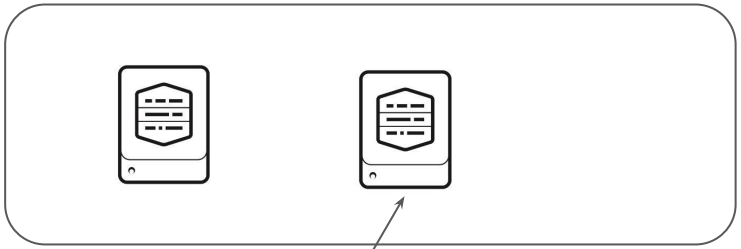
There are two ways to connect forwarders to indexer nodes:

i) Using the Indexer Discovery Feature [recommended]

ii) Connect forwarders directly to the peer nodes.

Approach: Connect Forwarders to Peer Nodes

In this approach, the Universal forwarders will have IP addresses of all the peer nodes configured and the logs will be uploaded accordingly.



Upload Data

Peer Indexers
192.168.0.5:9997
192.168.0.6:9997

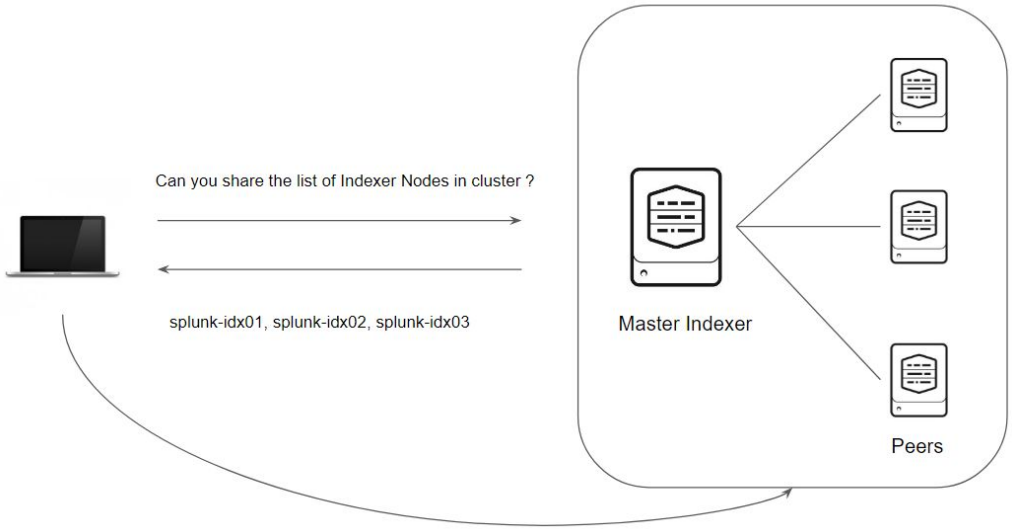
Indexer Discovery

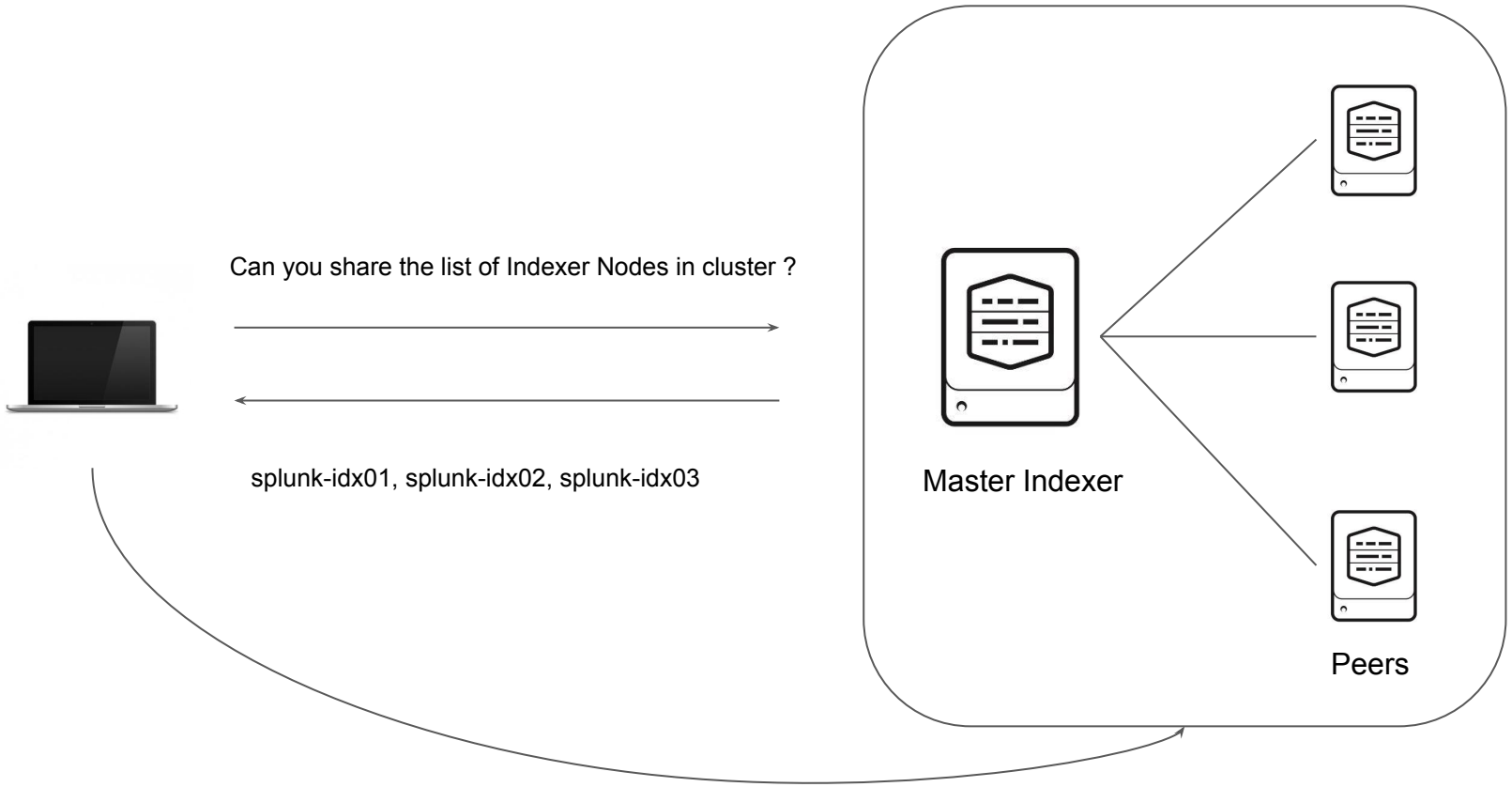
Indexer Architecture

Overview of Indexer Discovery

With Indexer Discovery method, each forwarder queries master node for list of peer nodes within the cluster.

It then uses load balancing to forward data to set of peer nodes.




Can you share the list of Indexer Nodes in cluster ?
splunk-idx01, splunk-idx02, splunk-idx03
Master Indexer
Peers

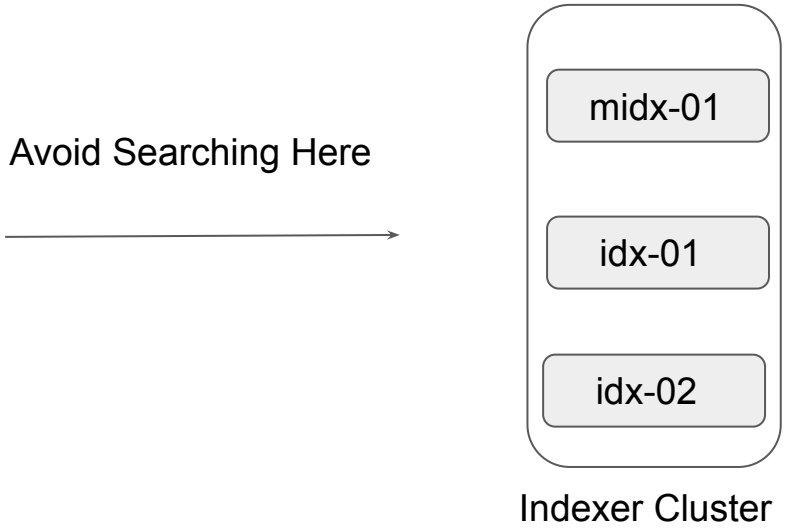
Search Head Clustering

Search Head Architecture

Avoid Searching from Indexer Nodes

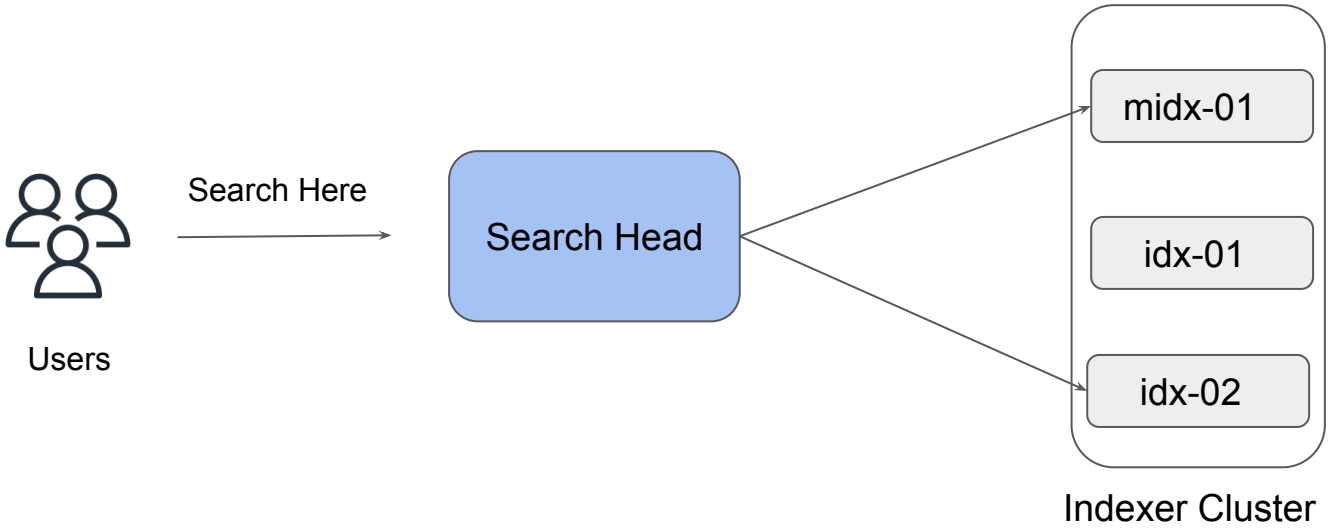
Indexer nodes are primarily designed to index data from multiple sources.

Search and Reporting is a resource intensive application and can slow down a indexer node if larger queries are run.

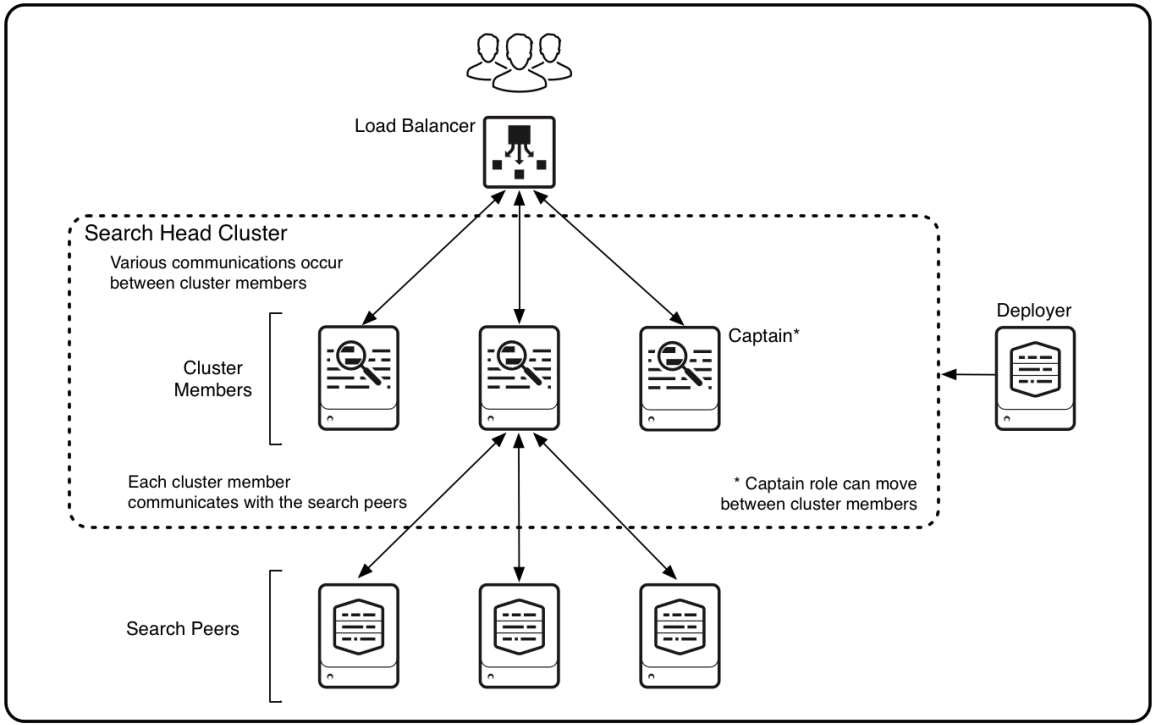


Search Head Component

Search Head is a component whose responsibility is to handle the search management functions, directing search requests to search peers and then merging the results back to the users.



Architecture of Search Head Clustering



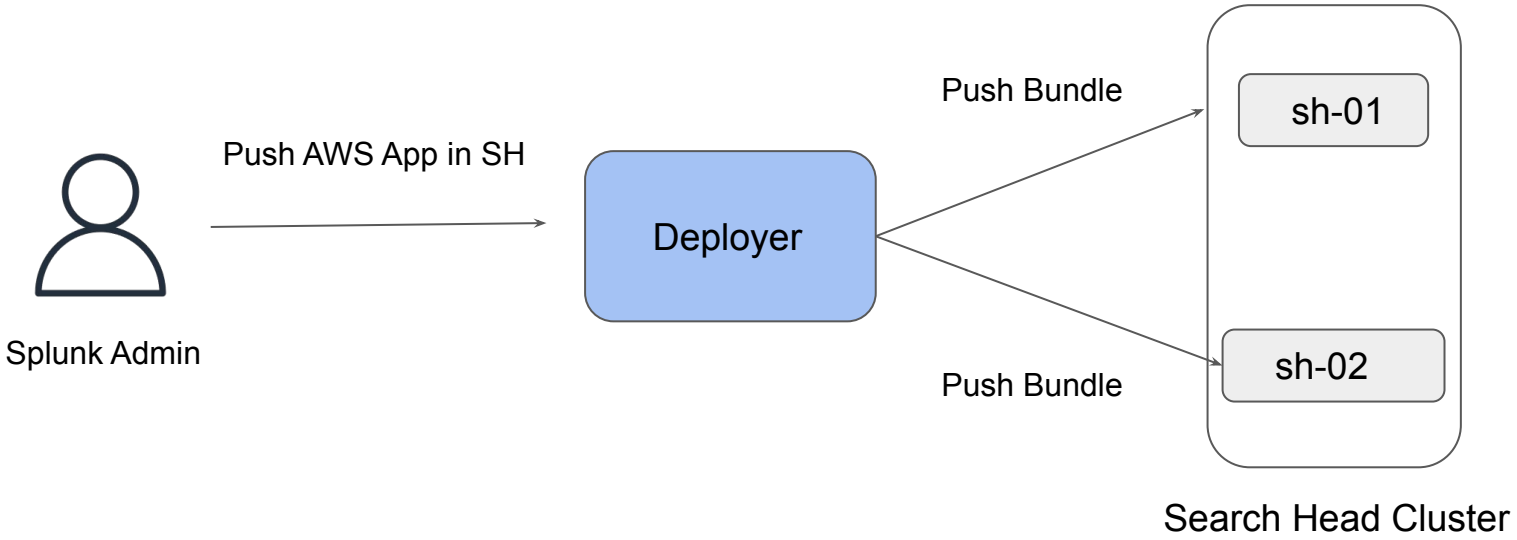
What gets replicated by default?

There are various aspects which gets replicated within the members:

- Alert Actions
- Data Models
- Workflow actions
- Users
- Saved Searches
- Macros
- Lookups
- Event Types
- Many Many more

Deployer Component

The deployer is used to distribute apps and certain other configuration updates to search head cluster members.



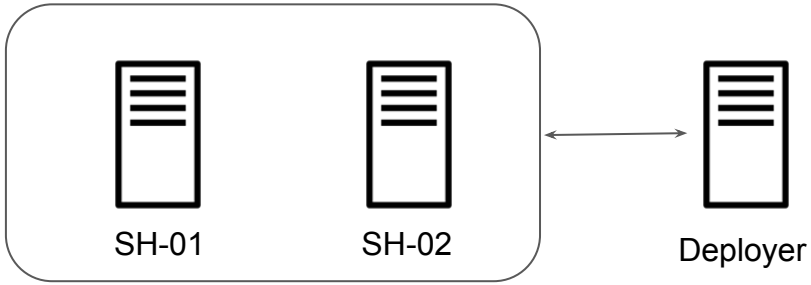
Infrastructure for Search Head

Let's get started!

Setup for Search Head Clustering

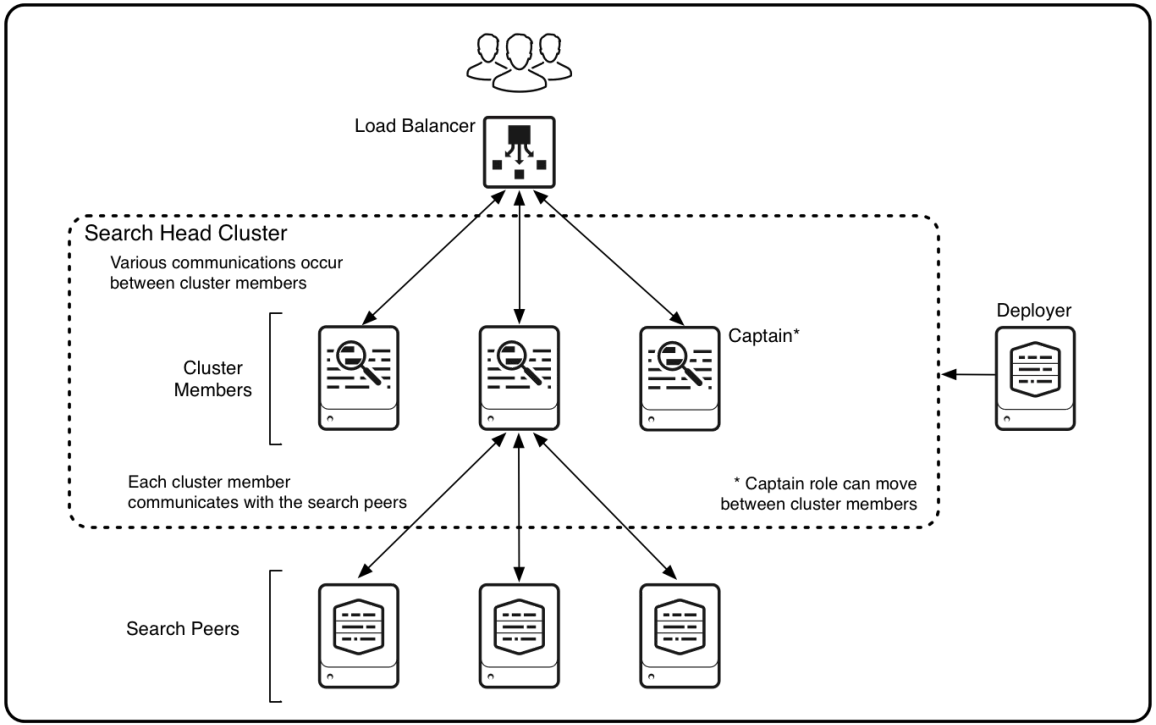
Total Servers Needed: 3

- 2 for Search Head Members
- 1 for Deployer



Search Head Cluster Members

Architecture of Search Head Clustering



btool

Troubleshooting Configuration Issues

Need of btool

The Splunk Enterprise configuration file system supports many overlapping configuration files in many different locations.

This flexibility can make it hard to figure out exactly which configuration value Splunk Enterprise is using.

Overview of btool

btool is a command line tool that can help us troubleshoot configuration file issues or see what values are being used by your Splunk Enterprise installation.

btool shows you the merged settings in the .conf files

Data Model

Making things easier.

Basic Problem Overview

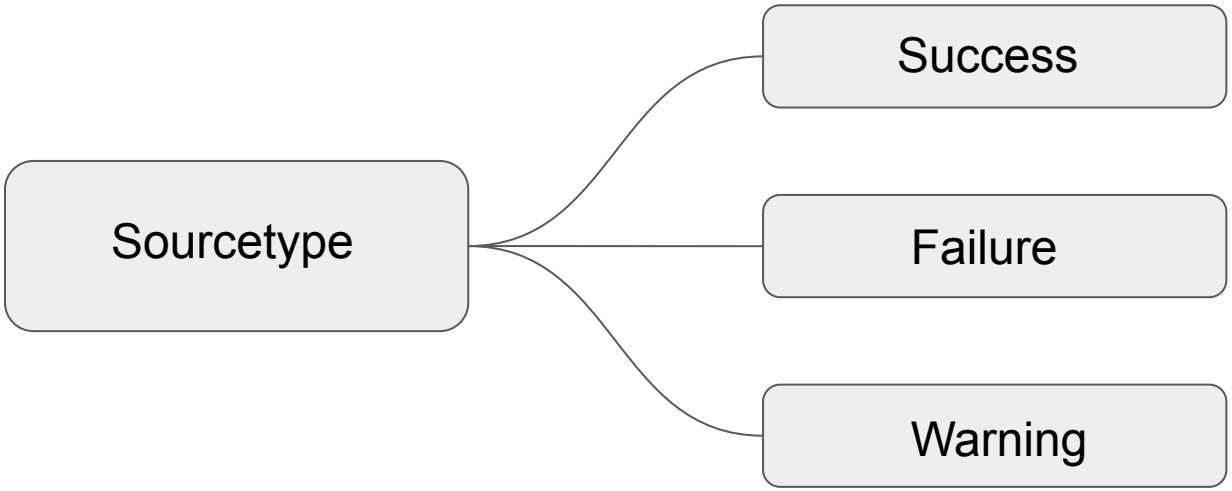
- You have data which contains certain critical business indicators.
- The data structure is multi-layered and complicated.
- The users who wants to see business indicators are less technical.
- You don’t want to be point of contact for everything they want out of data.

Searches can be long and complex

```
sourcetype=aws:cloudtrail eventName=RunInstances errorCode=success | bucket span=10m
_time | stats count AS instances_launched by _time userName | eventstats
avg(instances_launched) as total_launched_avg, stdev(instances_launched) as
total_launched_stdev | eval threshold_value = 4 | eval isOutlier=if(instances_launched >
total_launched_avg+(total_launched_stdev * threshold_value), 1, 0) | search isOutlier=1 AND
_time >= relative_time(now(), "-10m@m") | eval num_standard_deviations_away =
round(abs(instances_launched - total_launched_avg) / total_launched_stdev, 2) | table _time,
userName, instances_launched, num_standard_deviations_away, total_launched_avg,
total_launched_stdev
```

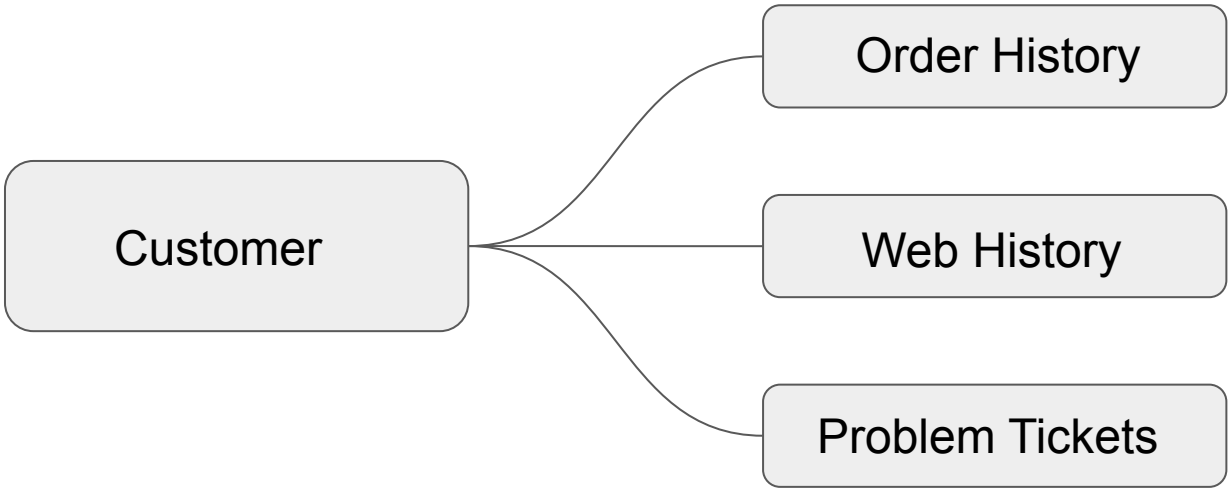
Basics of Data Model

Data Model allows us to provide meaningful representation of underlying raw data.

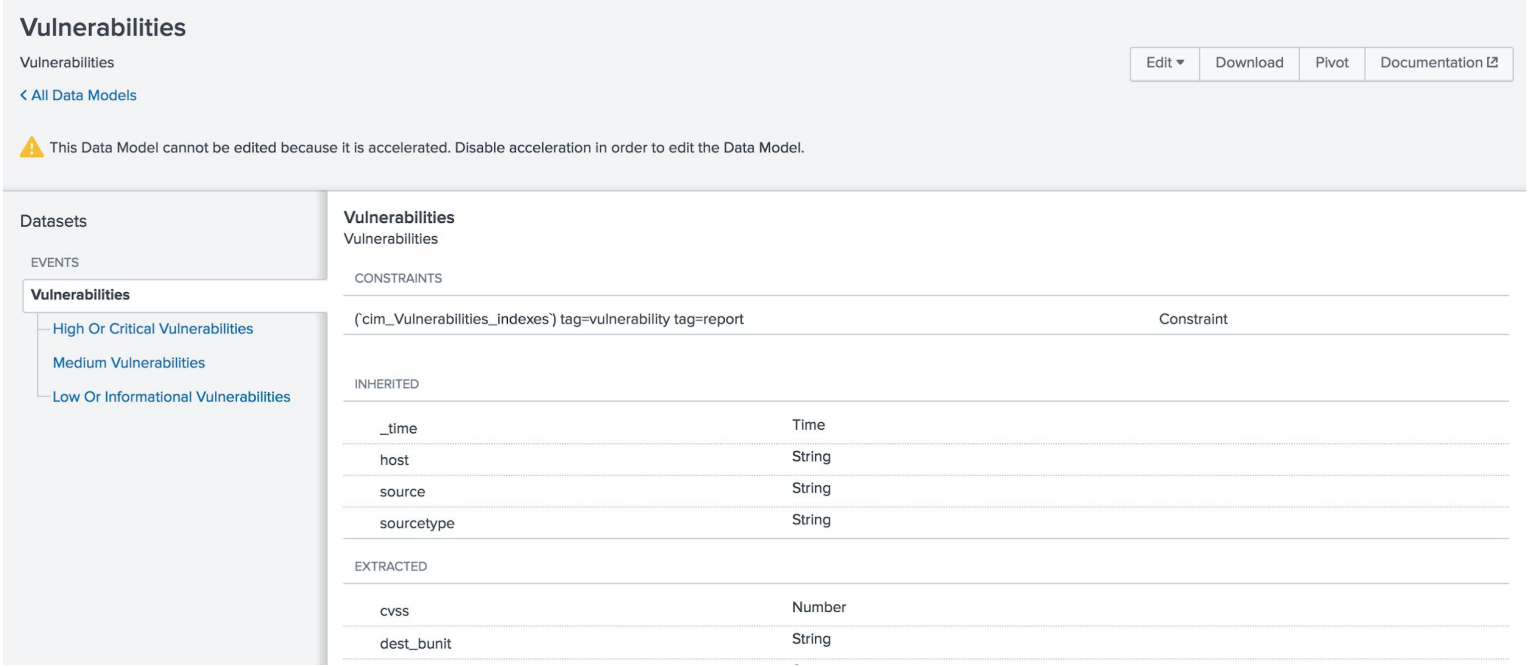


Basics of Data Model

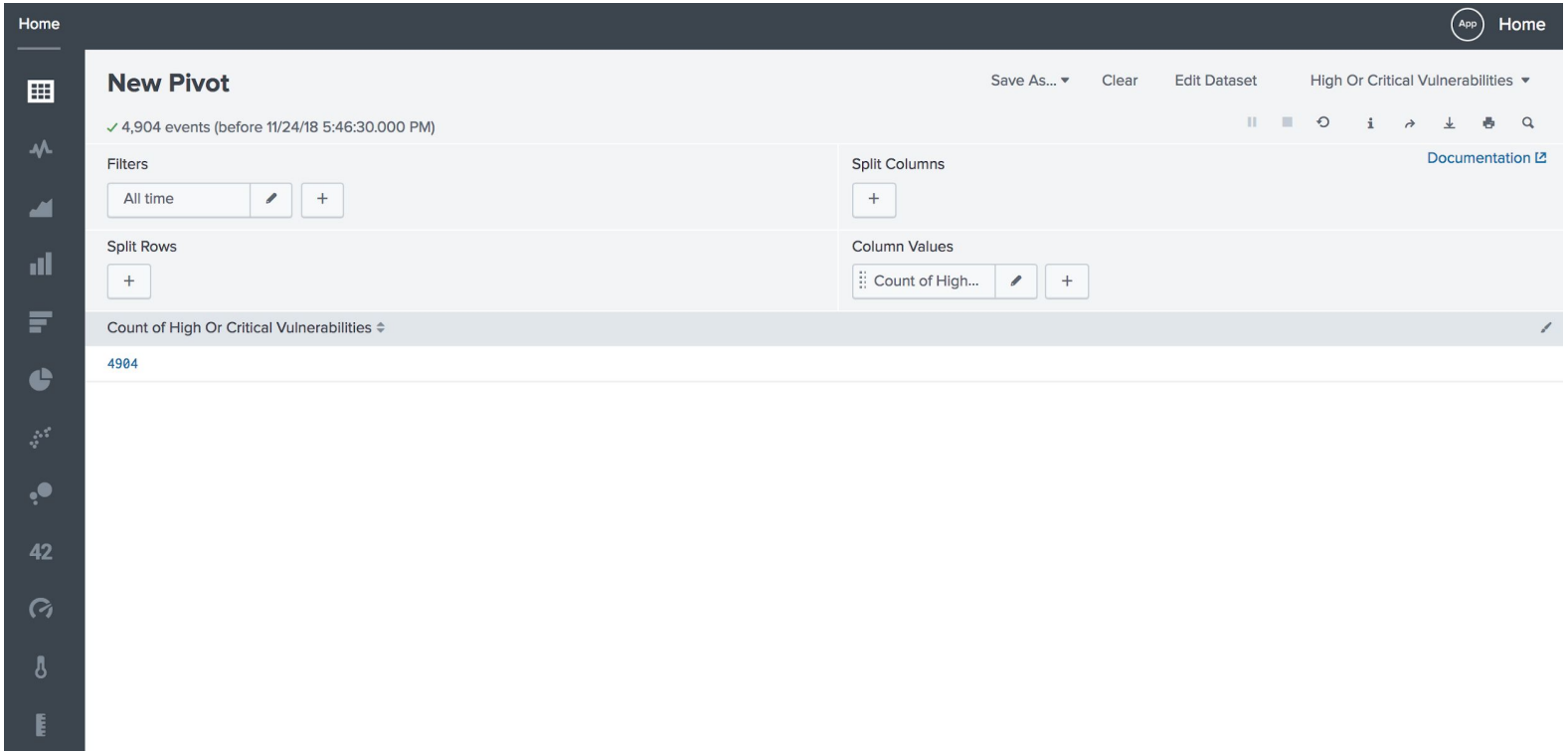
Data Model allows us to provide meaningful representation of underlying raw data.



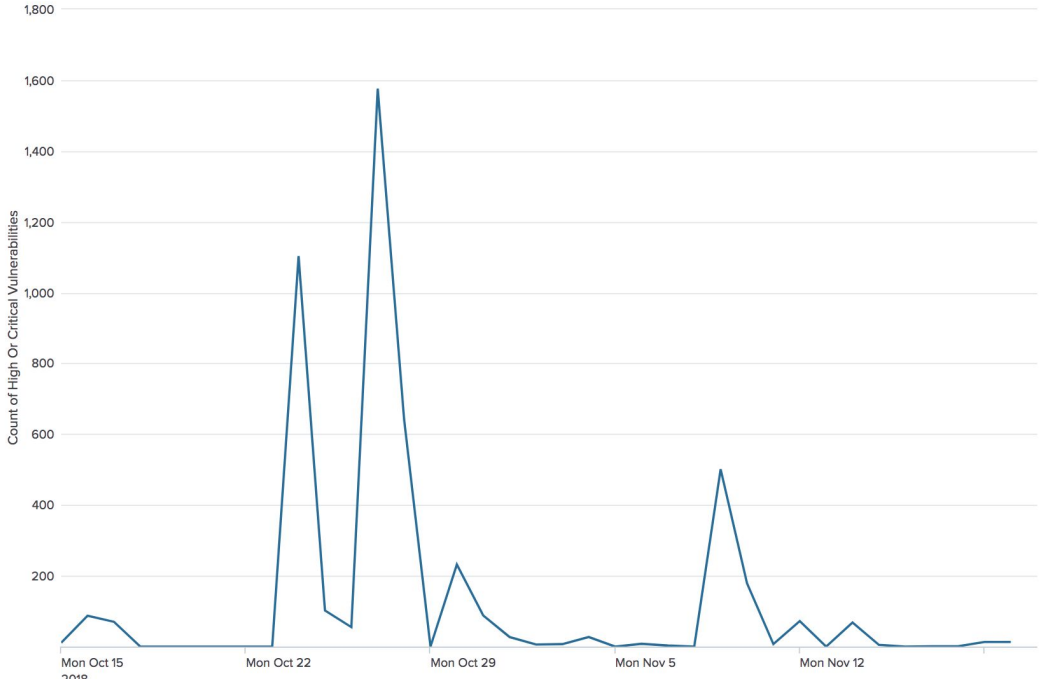
Vulnerability Data Model



Vulnerability Data Model



Vulnerability Data Model



Splunk - Official Support

One course for all

Overview of Splunk Support Programs

Splunk offers variety of support plans for customers.

These are primarily divided into following categories:

1. Community
2. Base
3. Standard
4. Premium