Training on

Splunk User & Basic Administration

Splunk Overview

What is Splunk?

Splunk is a big data tool.

It is software that helps in

- Collecting machine data (logs)
- Indexing
- Searching

Helps in identifying patterns, detecting anomalies.

Helps in making better business/investment and operational decisions.

Splunk Products

splunk>enterprise

Bigdata Tool

splunk>cloud

Bigdata Tool on Cloud

Splunk Enterprise Security™

SIEM

splunk> phantom

SOAR

Splunk Architecture

Splunk Basic Architecture

Users

Search Head

1 - 3

Indexer

1 - 5

Forwarder

Heavy Forwarder

1 - 10

Universal Forwarder

. . . . . .

20 - 300

Log Sources

Universal forwarder	Heavy forwarder
Cannot Filter or Mask data	Can Filter and Mask data
Doesn’t have UI	Has UI/Full Splunk instance
Installed on the log sources itself	Standalone server
Can collect performance data	Cannot collect performance data

Splunk Deployment Options

Option 1

Universal Forwarder

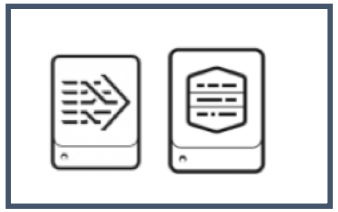
Heavy Forwarder

Indexer

Search Head

Option 2

Universal Forwarder



Indexer
Heavy Forwarder

Search Head

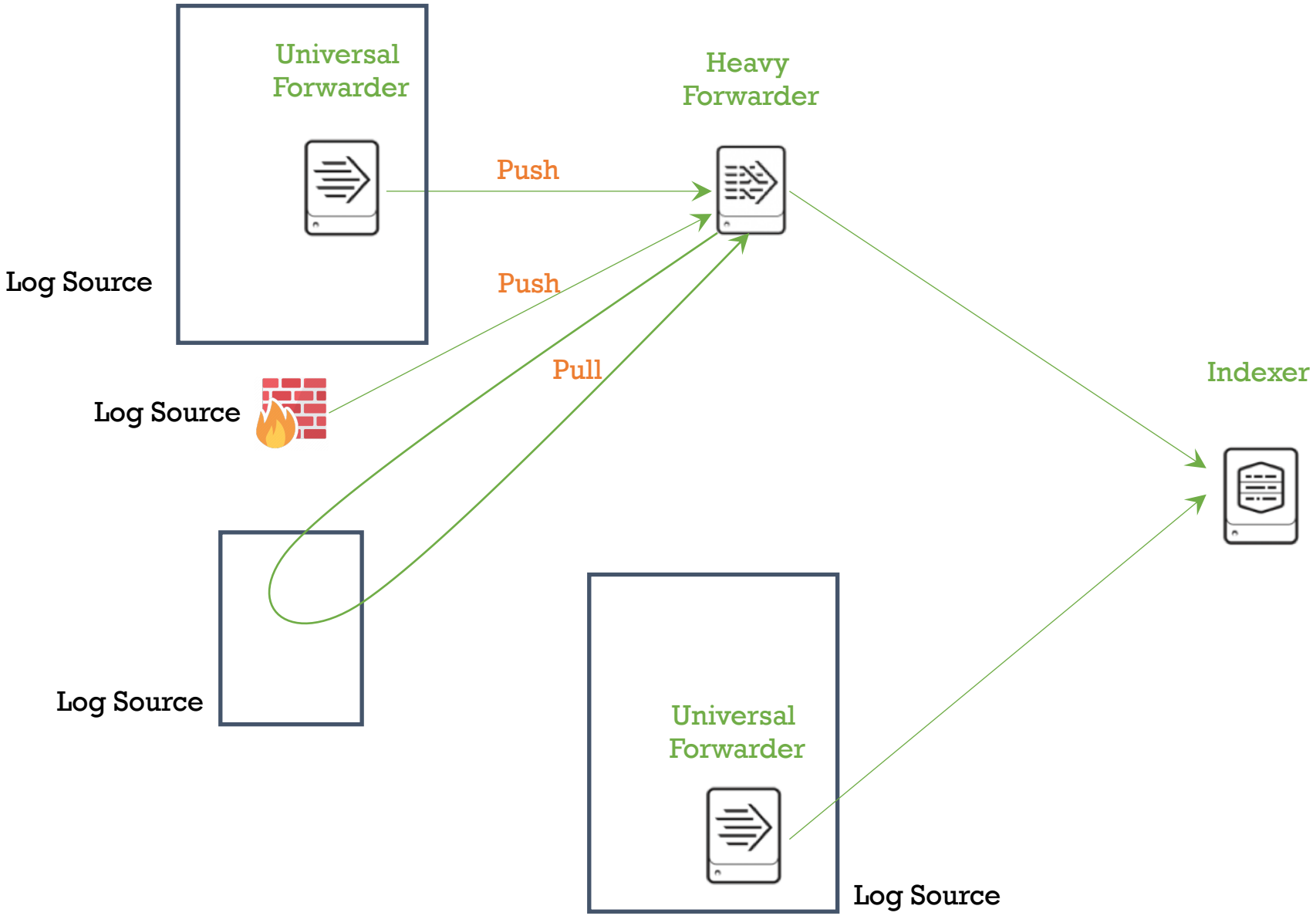
Indexer
Search Head

Option 3

Universal Forwarder

Indexer
Heavy Forwarder
Search Head

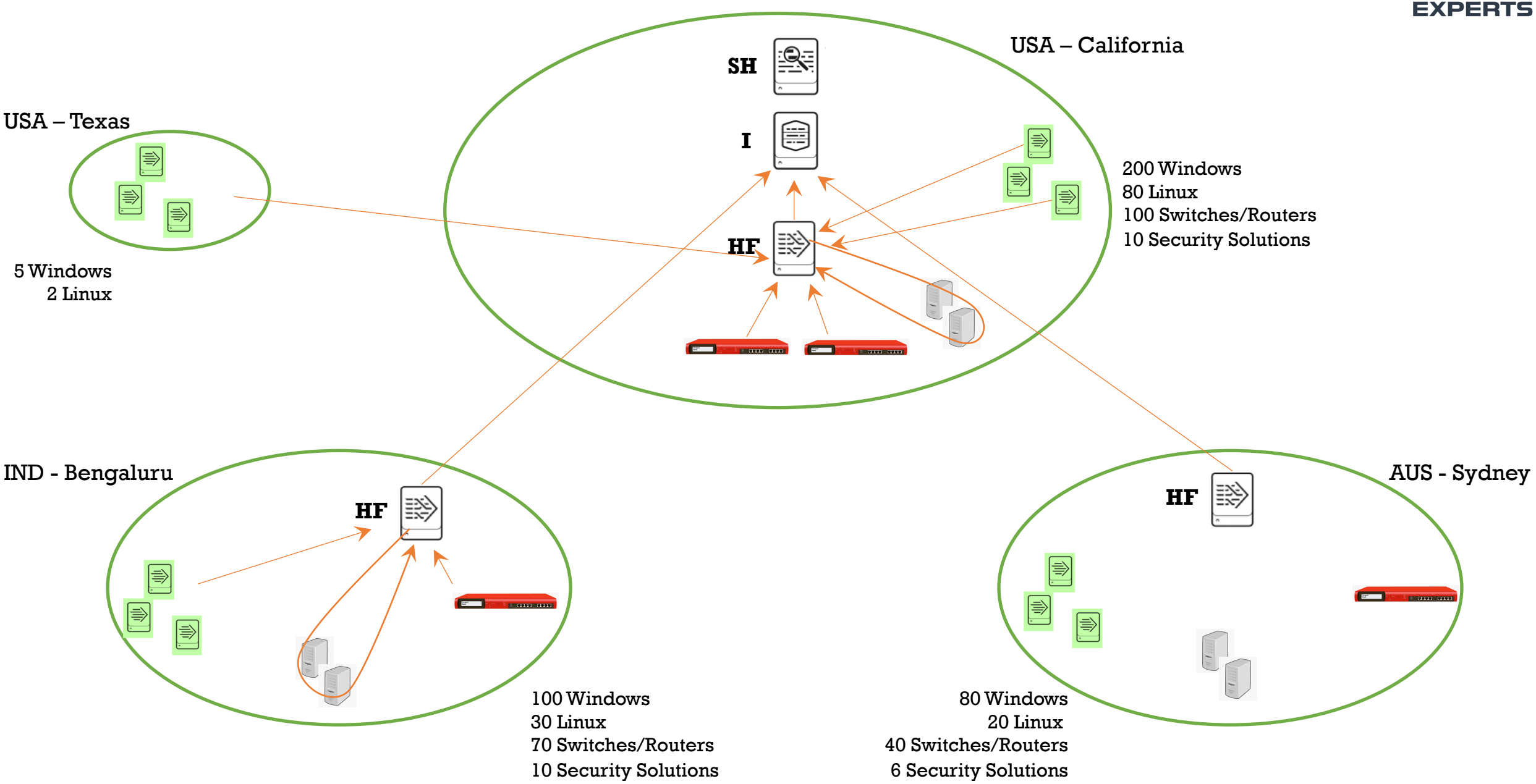
Data Flow


Universal Forwarder
Heavy Forwarder
Push
Log Source
Push
Pull
Log Source
Indexer
Log Source
Universal Forwarder
Log Source

Activity

Example Splunk Deployment

Example Splunk Deployment



Splunk Licensing

Splunk Licensing

Data indexed per day 10 GB/day 50 GB/day 100 GB/day 1 TB/day

EPS – Events Per Second

EPS x 400 bytes x 60 sec x 60 min x 24 hours

2000 x 400 bytes x 60 sec x 60 min x 24 hours = 69120000000 bytes = 69.12 GB

For Windows Event Logs we assume 1 event = 1000 bytes

Storage requirement will be calculated as Data/Day x Retention-Time

Splunk Installation

Splunk Installation Packages

Splunk Enterprise

Heavy Forwarder

Indexer

Search Head

Splunk Universal Forwarder

Universal Forwarder

Splunk Enterprise - Supported Platforms

Windows 10
Windows Server 2016, 2019

Linux
(CentOS/RedHat/Ubuntu/Debian)

Mac OS

Splunk UF - Supported Platforms

Windows Linux Solaris Mac OS FreeBSD AIX

Splunk Enterprise – Port Requirements

Management Port	: **8089**	(Splunk Component to Component)
Indexing Port	: **9997**	(UF/HF to Indexer for log forwarding)
Web Access Port	: **8000**	(User Computer to Search Head)
Syslog Port	: **514**	(Log Source to HF)

Splunk Enterprise - Installation

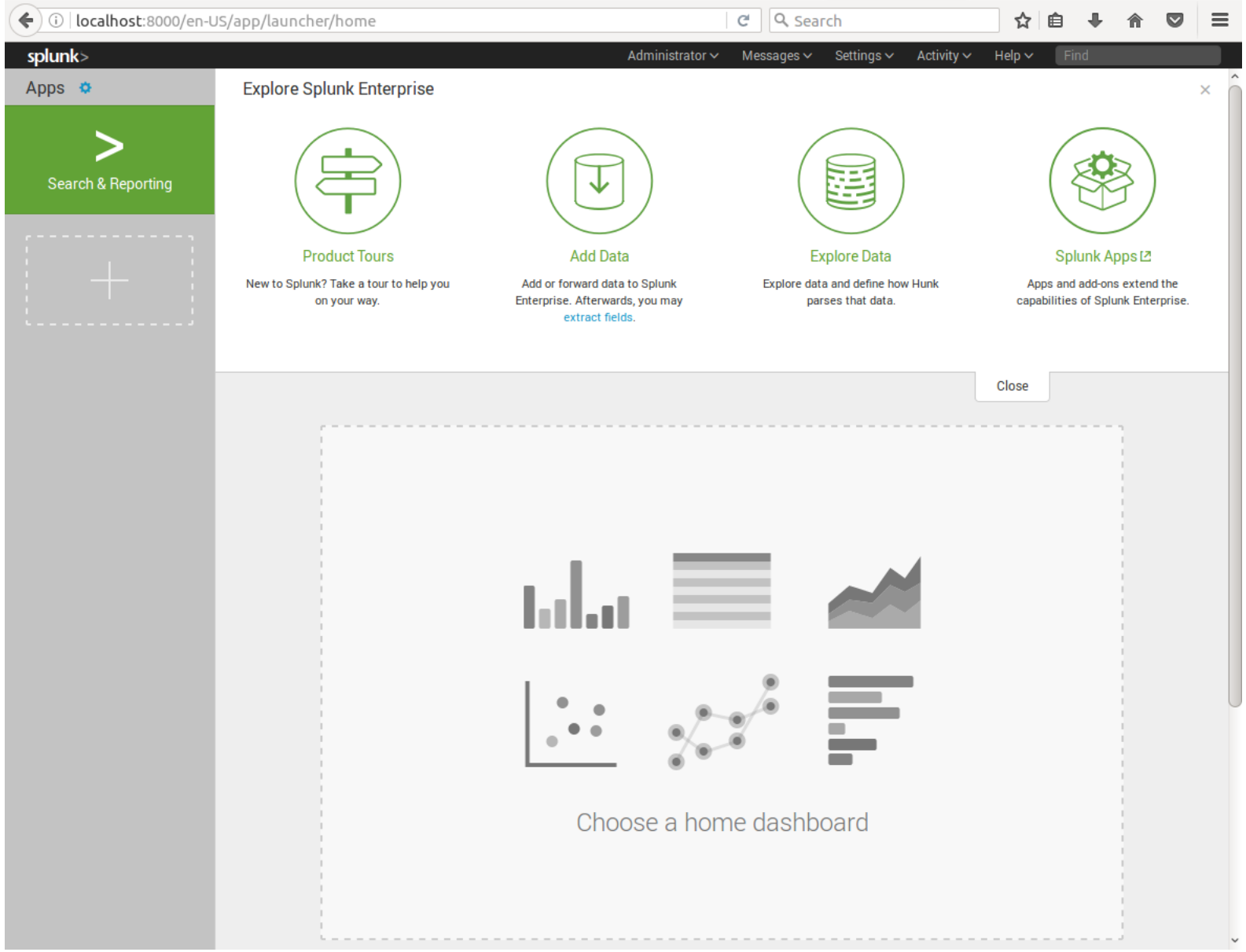
Install Amazon Linux	Install CentOS
Update Amazon Linux	#sudo yum update
Install wget	#sudo yum install wget
Download Splunk Enterprise package	#sudo wget <splunk-package-link>
Extract the package	#sudo tar xvzf splunk-package
Start Splunk service	splunk/bin/# sudo ./splunk start --accept-license*
Open port 8000 on AWS instance	Security → Security Group → Edit Inbound rules
Start Splunk service on boot	splunk/bin/#sudo ./splunk enable boot-start

* Need to set username and password for Splunk Web Access

- Installation is a one-time activity in an org.
- You cannot clear Splunk Admin or Architect interviews with knowledge acquired from our training.
- You might join a company where Splunk is not used.
- We aim to become SOC analysts or Security analysts (No SIEM Admin)
- You might encounter a few interviews regarding the Splunk Admin role – THEY ARE NOTE FOR YOU
- If you are not following Splunk – **IT IS FINE**
- Basic commands are enough (**80-20 rule**)

Accessing the Web Interface

Splunk can be accessed by the URL: http://splunk_server_ip:8000



Demo

Splunk UF – Installation (Collecting logs from Windows)

Splunk UF – Installation (Collecting logs from Ubuntu)

Install Ubuntu	Install CentOS
Update Ubuntu	#apt update
Install wget	#apt install wget
Download Splunk Universal Forwarder package	#wget <splunk-package-link>
Extract the package	#tar xvzf splunk-package
Start Splunk service	splunk/bin/#./splunk start --accept-license*
Start Splunk service on boot	splunk/bin/#./splunk enable boot-start

Demo

Splunk UF – Installation (Collecting logs from Ubuntu)

Splunk UF – Configuration Files

What to collect? # /opt/splunkforwarder/etc/system/local/**inputs.conf**

Where to send? # /opt/splunkforwarder/etc/system/local/**outputs.conf**

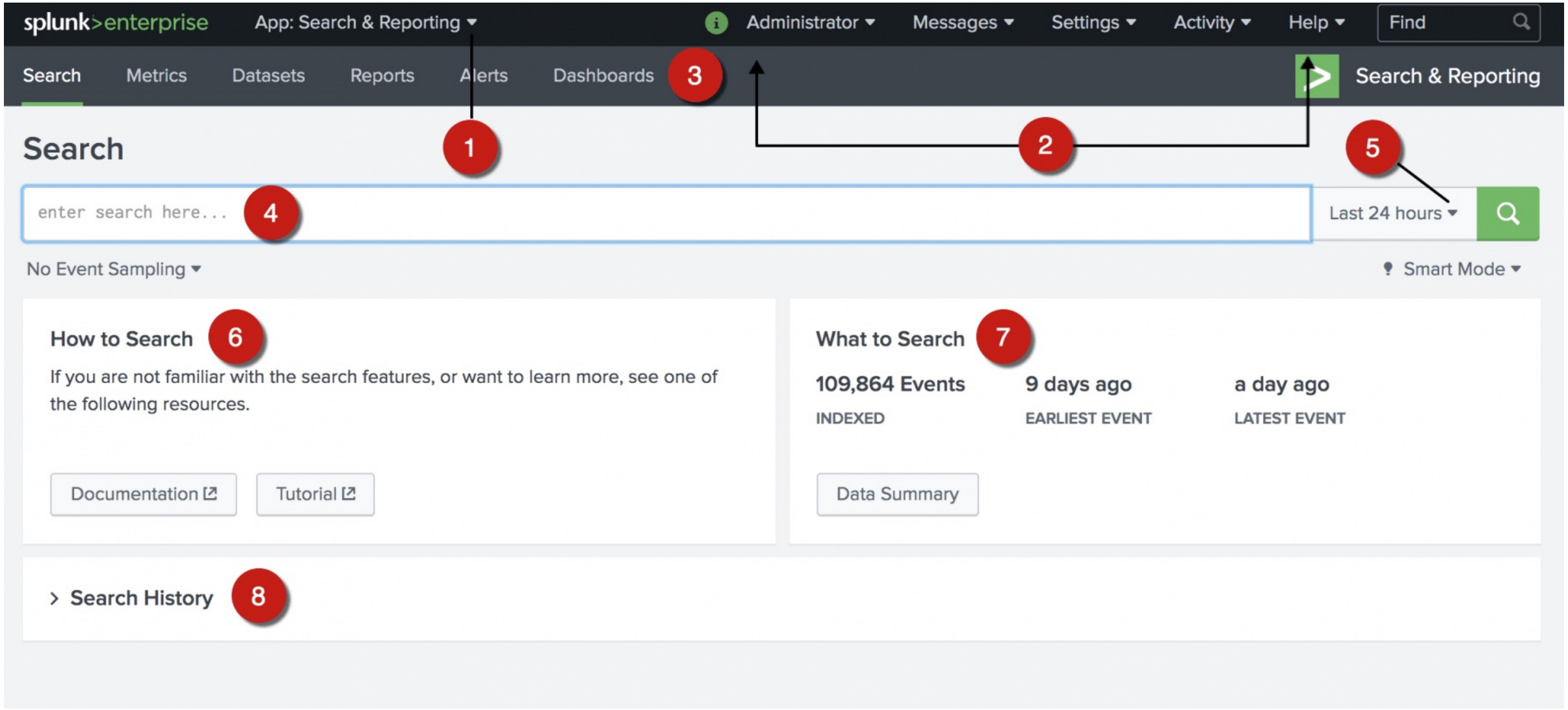
inputs.conf

```
[monitor:///var/log/auth.log]
index = linux
```

outputs.conf

```
[tcpout:send-to-indexer]
server = <indexer_ip>:9997
```

Search and Reporting App



Search UI Options

Number	Element	Description
1	**Applications menu**	Switch between Splunk applications that you have installed. The current application, Search & Reporting app, is listed. This menu is on the Splunk bar.
2	**Splunk bar**	Edit your Splunk configuration, view system-level messages, and get help on using the product.
3	**Apps bar**	Navigate between the different views in the application you are in. For the Search & Reporting app the views are: Search, Metrics, Datasets, Reports, Alerts, and Dashboards.
4	**Search bar**	Specify your search criteria.
5	**Time range picker**	Specify the time period for the search, such as the last 30 minutes or yesterday. The default is **Last 24 hours**.
6	**How to search**	Contains links to the *Search Manual* and the *Search Tutorial*.
7	**What to search**	Shows a summary of the data that is uploaded on to this Splunk instance and that you are authorized to view.
8	**Search history**	View a list of the searches that you have run. The search history appears after you run your first search.

Search Results

Number of search matches

Event Distribution Over Time

Events (384,383) | Patterns | Statistics | Visualization → Result display type

Format Timeline ▾ | − Zoom Out | + Zoom to Selection | × Deselect | 1 day per column

List ▾ | Format | 50 Per Page ▾ | < Prev | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | ... | Next >

< Hide Fields | All Fields

SELECTED FIELDS
- a host 6
- a source 17
- a sourcetype 14

INTERESTING FIELDS
- a attack 100+
- a class 22
- # date_hour 14
- # date_mday 7
- # date_minute 60
- a date_month 3
- # date_second 60

i	Time	Event
>	03/08/2020 12:04:38.000	Aug 3 12:04:38 ip-172-31-18-143 systemd[1]: Stopping Create final runtime dir for shutdown pivot root... host = ip-172-31-18-143 | source = /var/log/syslog | sourcetype = syslog
>	03/08/2020 12:04:38.000	Aug 3 12:04:38 ip-172-31-18-143 systemd[1]: Stopping D-Bus System Message Bus... host = ip-172-31-18-143 | source = /var/log/syslog | sourcetype = syslog
>	03/08/2020 12:04:38.000	Aug 3 12:04:38 ip-172-31-18-143 systemd[1]: Stopping Regular background program processing daemon... host = ip-172-31-18-143 | source = /var/log/syslog | sourcetype = syslog
>	03/08/2020 12:04:38.000	Aug 3 12:04:38 ip-172-31-18-143 systemd[1]: Stopped target Network is Online. host = ip-172-31-18-143 | source = /var/log/syslog | sourcetype = syslog
>	03/08/2020 12:04:38.000	Aug 3 12:04:38 ip-172-31-18-143 systemd[1]: Stopped target Cloud-config availability. host = ip-172-31-18-143 | source = /var/log/syslog | sourcetype = syslog

← Search Results

Field Sidebar

< Hide Fields All Fields

SELECTED FIELDS

a host 2
a source 3
a sourcetype 5

Default Selected Fields

INTERESTING FIELDS

a Ack 100+
a attack 100+
a class 22
Code 5
date_hour 14
date_mday 6
date_minute 60
a date_month 3
date_second 60
a date_wday 5
date_year 1
a date_zone 2
a dest_ip 100+
dest_port 100+

Interesting Fields

a hold string value

hold number value

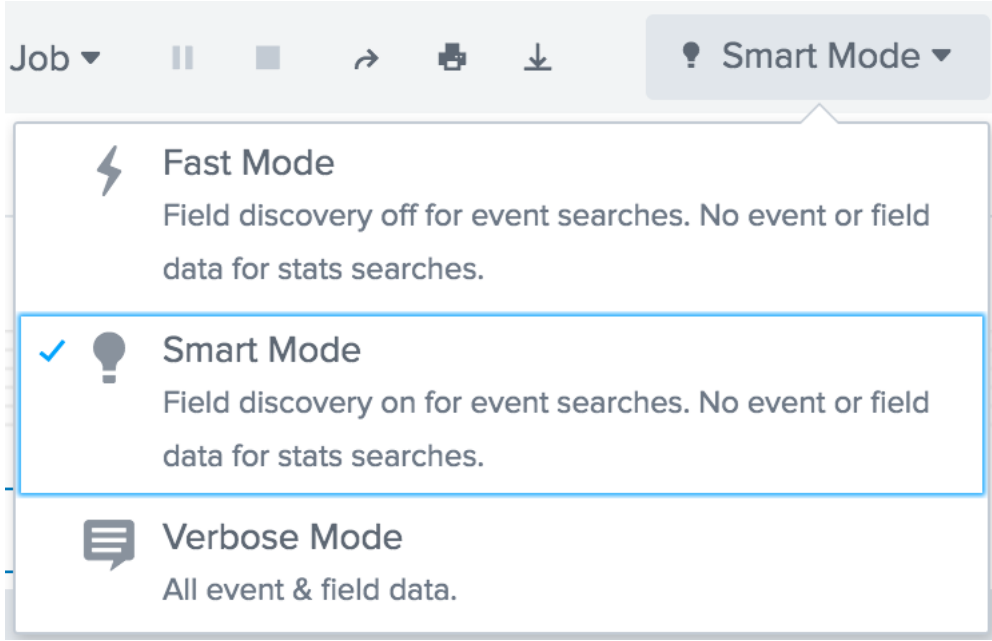
Selected Fields

host: A host is the name of the physical or virtual device where an event originates. It can be used to find all data originating from a specific device.

source: A source is the name of the file, directory, data stream, or other input from which a particular event originates.

sourcetype: Sources are classified into source types, which can be either well known formats or formats defined by the user.

Search Modes



Fast: Returns only Splunk Data like host, source, sourcetype, linecount, splunk_server, index.
and any specifically mentioned fields in search query

Verbose: Returns as much event information as possible, at the expense of slower search performance.

Smart: Toggles search behavior based on whether your search contains transforming commands or not.
For transforming searches, it behaves like Fast mode. For searches without transforming commands, it behaves like Verbose mode.

_time & _raw

The _time field contains an event's timestamp expressed in UNIX time (Epoch Time)

The _raw field contains the original raw data of an event.

```
index=se-assets
| table _time _raw
```

All time

✓ 352,495 events (before 14/08/2020 17:57:56.000) No Event Sampling Job Fast Mode

Events Patterns **Statistics (352,495)** Visualization

100 Per Page Format Preview

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

_time	_raw
2020-08-03 09:14:30	Aug 3 09:14:30 ip-172-31-18-143 sshd[8885]: Received disconnect from 222.186.15.115 port 63711:11: [preauth]
2020-08-03 09:11:56	Aug 3 09:11:56 ip-172-31-18-143 sshd[8883]: Connection closed by 49.235.83.136 port 35968 [preauth]
2020-08-03 09:05:18	Aug 3 09:05:18 ip-172-31-18-143 sshd[8878]: Disconnected from 222.186.175.23 port 10875 [preauth]
2020-08-03 09:05:18	Aug 3 09:05:18 ip-172-31-18-143 sshd[8878]: Received disconnect from 222.186.175.23 port 10875:11: [preauth]
2020-08-03 09:04:37	Aug 3 09:04:37 ip-172-31-18-143 sshd[8874]: Connection reset by 85.209.0.251 port 23522 [preauth]
2020-08-03 09:04:35	Aug 3 09:04:35 ip-172-31-18-143 sshd[8876]: Connection reset by 85.209.0.251 port 23850 [preauth]
2020-08-03 08:56:12	Aug 3 08:56:12 ip-172-31-18-143 sshd[8868]: Disconnected from 222.186.31.166 port 10630 [preauth]
2020-08-03 08:56:12	Aug 3 08:56:12 ip-172-31-18-143 sshd[8868]: Received disconnect from 222.186.31.166 port 10630:11: [preauth]

Basic Searches

SPL - Search Processing Language

Free-form search

Fields & Values - fields are case sensitive, values are not

Time Range Picker

Field Operators

Equal to	field = value
Not Equal to	field != value
Greater Than	field > value
Greater Than or Equal to	field >= value
Less Than	field < value
Less Than or Equal to	field <= value

Boolean Operators

AND – implied between terms, so you do not need to write it.

OR – used to specify that either one of two or more arguments should be true.

NOT – used to filter out events containing a specific word.

Wildcards

You can use the asterisk (*) character as a wildcard to match an unlimited number of characters in a string.

For example, "**fail***" matches **failure, failed, fails** and **failing**

clientip=10* matches all IPs starting with **10** like **100.** **101.** **102.** **10.**

to match 10.x.x.x network **clientip=10.***

Search Best Practices

Narrow the time window

Specify the index, source, or source type

Be specific

Avoid using NOT expressions

Filter as soon as possible

Apache Log Format

Log Format - %h %l %u %t \"%r\" %>s %b \"%{Referer}i\" \"%{User-agent}i\"

127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0" 200 2326 "http://www.example.com/start.html" "Mozilla/4.08 [en] (Win98; I ;Nav)"

Value	Std. Format	Description
127.0.0.1	**%h**	This is the IP address of the client (remote host) which made the request to the server
-	**%l**	The "hyphen" in the output indicates that the requested piece of information is not available. In this case, the information that is not available is the RFC 1413 identity of the client determined by identd on the clients machine.
frank	**%u**	This is the userid of the person requesting the document as determined by HTTP authentication.
[10/Oct/2000:13:55:36 -0700]	**%t**	The time that the server finished processing the request. The format is: [dd/mmm/yyyy:hh:mm:ss zone]
"GET /apache_pb.gif HTTP/1.0"	**\"%r\"**	The request line from the client is given in double quotes. The request line contains a great deal of useful information. First, the method used by the client is GET. Second, the client requested the resource /apache_pb.gif, and third, the client used the protocol HTTP/1.0
200	**%>s**	This is the status code that the server sends back to the client. This information is very valuable, because it reveals whether the request resulted in a successful response (codes beginning in 2), a redirection (codes beginning in 3), an error caused by the client (codes beginning in 4), or an error in the server (codes beginning in 5).
2326	**%b**	The last entry indicates the size of the object returned to the client, not including the response headers. If no content was returned to the client, this value will be "-".
"http://www.example.com/start.html"	**\"%{Referer}i\"**	This gives the site that the client reports having been referred from. (This should be the page that links to or includes /apache_pb.gif).
"Mozilla/4.08 [en] (Win98; I ;Nav)"	**\"%{User-agent}i\"**	The User-Agent HTTP request header. This is the identifying information that the client browser reports about itself.

Transforming Commands

search elements
| **command field1 field2**

search elements
| **command *function* field**

Transforming Commands

Command	Usage	Example
table	Gives the output in the form of a table	| **table** field1 field2 ….
rename	Rename a specific field	| **rename** clientip **as** "source IP"
dedup	Removes duplicates	| **dedup** clientip
sort	Sorts a field default ascending order	| **sort** clientip | **sort –** clientip **sort** clientip **desc**
top	Gives top fieldname default limit is 10	| **top** clientip | **top limit=20** clientip
rare	Gives rare (least) occurring fieldname default limit is 10	| **rare** clientip | **rare limit=5** clientip
head n	Returns latest 'n' results	| **head** 5
tail n	Returns oldest 'n' results	| **tail** 10

Stats Command

Function	Usage	Example
values	Lists the values of a specific field	| stats values(status) by clientip
count	Gives the count of specific field	| stats count BY clientip
dc **distinct_count**	Returns the **number** of unique values present in a field	| stats dc(clientip)
sum	Sum of all the values of a field	| stats sum(bytes)
max	Maximum individual value of a field	| stats max(bytes)
min	Minimum individual value of a field	| stats min(bytes)

Geo Location Commands

Function	Usage	Example
iplocation	Give Geo Location for all public IP s	| iplocation clientip
geostat	Cluster map on map	| geostat count by clientip