**Splunk tool**- it’s big data management tool, it is a software for searching, monitoring, and analysing [machine-generated](https://en.wikipedia.org/wiki/Machine-generated_data) [big data](https://en.wikipedia.org/wiki/Big_data), via a Web-style interface.

**Founder of Splunk**- Rob Das and Erik Swan

**Why name Splunk**- Splunking is a standard English word which means digging to issue.

**Splunk components**- Forwarder, indexer, search-head, licence master, and management component.

**Management component**- Deployment server, deployer and cluster master.

**Splunk Story**: Splunk founders are generally focus on how to detect a problem, they keep asking this question to everyone, and their next question was if data says there is an issue then how you can resolve it, and finally they ended their story with Splunk.

They had a technical talk with near around 70 companies, that deals in different domain like EAU, banking and telecommunication, and the most common answer is people do google for solving their issue or based on their experience. So, they made google of data by which we can troubleshoot the issue.

**What is Google for IT data?** Answer: **Splunk**- Search Engine for data.

**Splunk**- Splunk is a US based company that focus on data analysis.

**What is data?** – Useful peace of information that can be used for analysis purpose.

**Machine data**- it’s system generated data, which is unstructured, and complex.

**Other similar monitoring tool**- ELK, Sumologic, Graylog and Loggy.

**Programming language behind Splunk**- C programming.

**Which algorithm Splunk works**- big data concept.

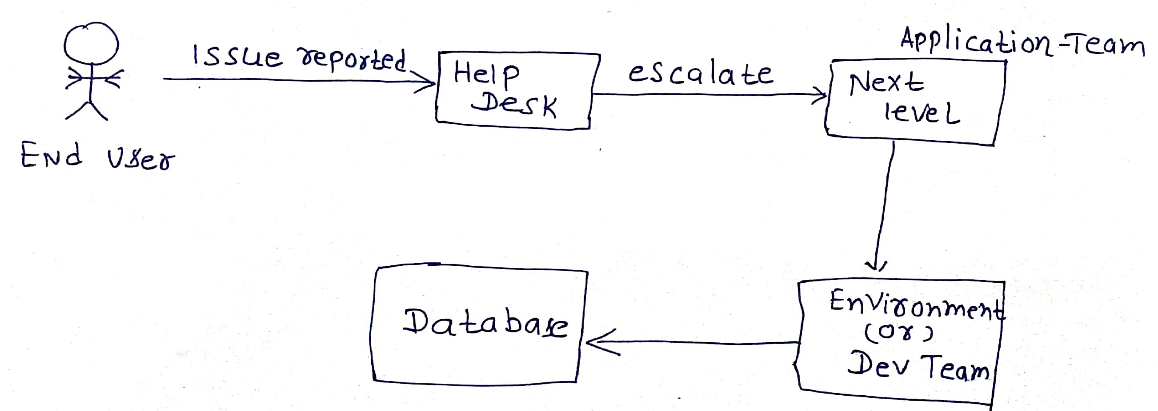
**Disadvantage of Splunk** – Can’t provide technical data like heap dumps and thread dumps (HP Diag. can provide such information). Note- Audio, Video, and Image can’t be put as input for Splunk.

**Splunk Version-**

1) **Splunk Enterprise**- it’s for business purpose, and there is no data limit.

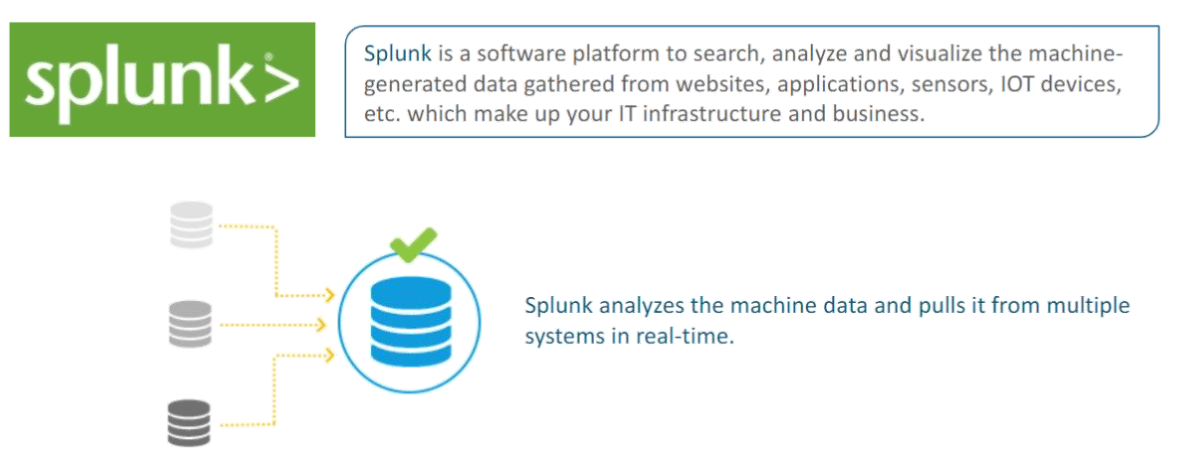
2) **Splunk light**- it’s for medium size organization like 20/30 GB/day.

3) **Splunk cloud**- Splunk as vendor will manage h/w and s/w, they give you console link.

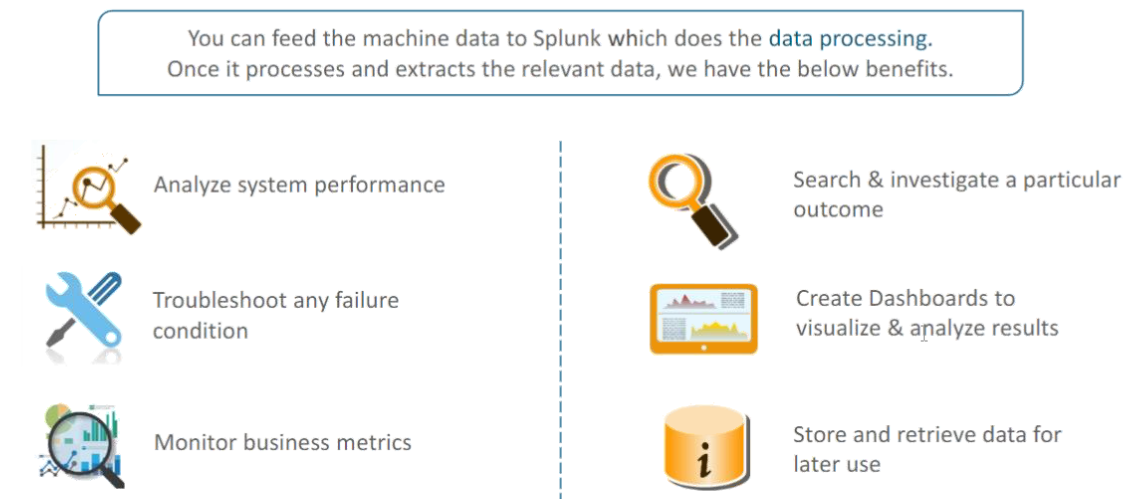
**Note**- if you are having Splunk installed in your system, then you can resolve/detect issue very fast. How? - Splunk deals with logs, and by seeing logs you can engage correct team.

Explanation: If any issue reported in IT, it’s done by helpdesk, then we will be engaging the respective teams for log, but here we can directly do by Splunk search head.

Standard Definition of Splunk



**What Splunk can do?**



**Question**: Can Splunk used as a proactive tool for problem detection? Answer: Yes, we can do by setting alerts.

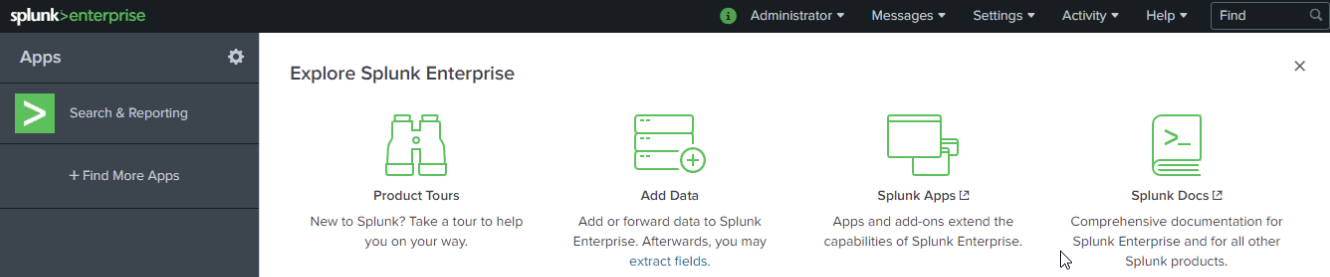
**Splunk Free vs Splunk Enterprise**- Many features will be not available in free Splunk.

Splunk Boot start command- [sudo] $SPLUNK\_HOME/bin/splunk enable boot-start

**Splunk demon**- it’s service name that’s run-in background.

**Difference between Unix and Window’s installation**- You need to start services manually for Unix system.

**Splunk Monitoring console-**

****

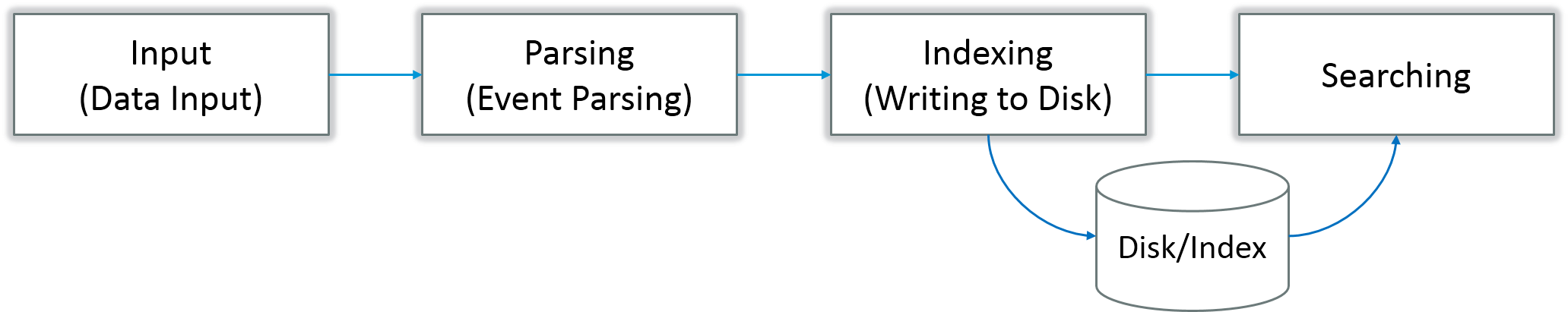
**Core components of Splunk-**

1. **Forwarder, 2) Indexer, 3) Search head.**

**Different Stages in Data Pipeline**

There are primarily 3 different stages in Splunk:

* Data Input stage
* Data Storage stage
* Data Searching stage



**Data Input Stage**

In this stage, Splunk software consumes the raw data stream from its source, breaks it into 64K blocks, and annotates each block with metadata keys. The metadata keys include hostname, source, and source type of the data. The keys can also include values that are used internally, such as character encoding of the data stream and values that control the processing of data during the indexing stage, such as the index into which the events should be stored.

**Data Storage Stage**

Data storage consists of two phases: Parsing and Indexing.

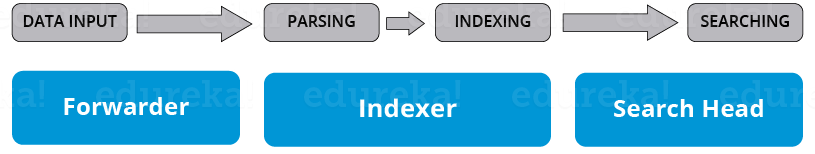
1. In Parsing phase, Splunk software examines, analyses, and transforms the data to extract only the relevant information. This is also known as event processing. It is during this phase that Splunk software breaks the data stream into individual events. The parsing phase has many sub-phases:
   1. Breaking the stream of data into individual lines
   2. Identifying, parsing, and setting timestamps
   3. Annotating individual events with metadata copied from the source-wide keys
   4. Transforming event data and metadata according to regex transform rules
2. In Indexing phase, Splunk software writes parsed events to the index on disk. It writes both compressed raw data and the corresponding index file. The benefit of Indexing is that the data can be easily accessed during searching.

**Data Searching Stage**

This stage controls how the user accesses, views, and uses the indexed data. As part of the search function, Splunk software stores user-created knowledge objects, such as reports, event types, dashboards, alerts and field extractions. The search function also manages the search process.

**Splunk Components**

If you look at the below image, you will understand the different data pipeline stages under which various Splunk components fall under.



There are 3 main components in Splunk:

* Splunk Forwarder, used for data forwarding
* Splunk Indexer, used for Parsing and Indexing the data
* Search Head, is a GUI used for searching, analyzing and reporting

**Splunk Forwarder**

Splunk Forwarder is the component which you have to use for collecting the logs. Suppose, you want to collect logs from a remote machine, then you can accomplish that by using Splunk’s remote forwarders which are independent of the main Splunk instance.

In fact, you can install several such forwarders in multiple machines, which will forward the log data to a Splunk Indexer for processing and storage. What if you want to do real-time analysis of the data? Splunk forwarders can be used for that purpose too. You can configure the forwarders to send data to Splunk indexers in real-time. You can install them in multiple systems and collect the data simultaneously from different machines in real time.

Compared to other traditional monitoring tools, Splunk Forwarder consumes very less cpu ~1-2%. You can scale them up to tens of thousands of remote systems easily, and collect terabytes of data with minimal impact on performance.

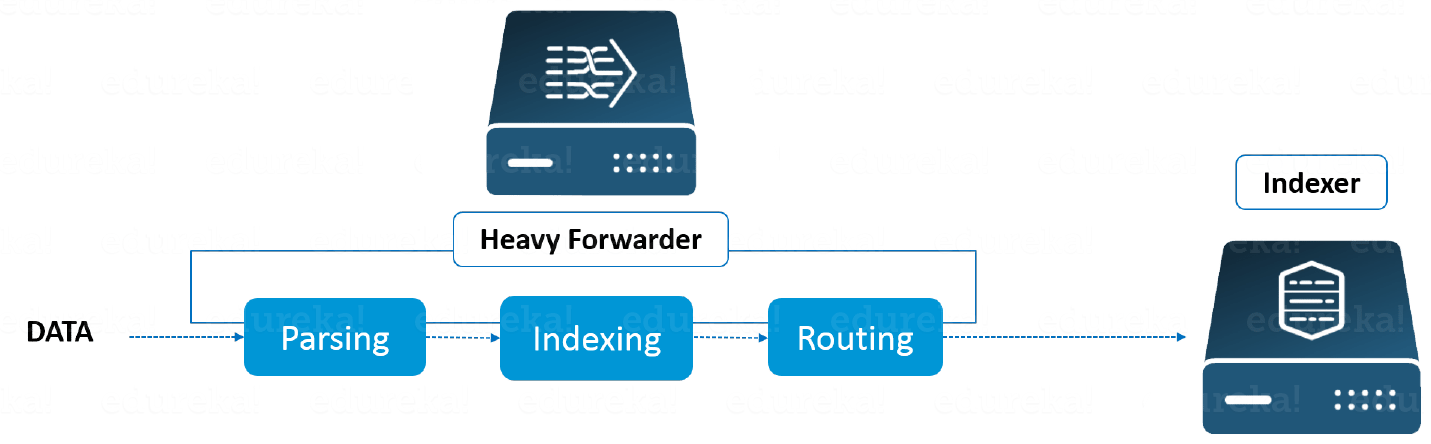
**Now, let us understand the different types of Splunk forwarders.**

**Universal Forwarder** – You can opt for an universal forwarder if you want to forward the raw data collected at the source. It is a simple component which performs minimal processing on the incoming data streams before forwarding them to an indexer.

Data transfer is a  major problem with almost every tool in the market. Since there is minimal processing on the data before it is forwarded, lot of unnecessary data is also forwarded to the indexer resulting in performance overheads.

Why go through the trouble of transferring all the data to the Indexers and then filter out only the relevant data? Wouldn’t it be better to only send the relevant data to the Indexer and save on bandwidth, time and money? This can be solved by using Heavy forwarders which I have explained below.

**Heavy Forwarder** – You can use a Heavy forwarder and eliminate half your problems, because one level of data processing happens at the source itself before forwarding data to the indexer. Heavy Forwarder typically does parsing and indexing at the source and also intelligently routes the data to the Indexer saving on bandwidth and storage space. So, when a heavy forwarder parses the data, the indexer only needs to handle the indexing segment.



**Splunk Indexer**

Indexer is the Splunk component which you will have to use for indexing and storing the data coming from the forwarder. Splunk instance transforms the incoming data into events and stores it in indexes for performing search operations efficiently. If you are receiving the data from a Universal forwarder, then the indexer will first parse the data and then index it. Parsing of data is done to eliminate the unwanted data. But, if you are receiving the data from a Heavy forwarder, the indexer will only index the data.

As the Splunk instance indexes your data, it creates a number of files. These files contain one of the below:

* Raw data in compressed form
* Indexes that point to raw data (index files, also referred to as tsidx files), plus some metadata files

These files reside in sets of directories called buckets.

Let me now tell you how Indexing works.

Splunk processes the incoming data to enable fast search and analysis. It enhances the data in various ways like:

* Separating the data stream into individual, searchable events
* Creating or identifying timestamps
* Extracting fields such as host, source, and source type
* Performing user-defined actions on the incoming data, such as identifying custom fields, masking sensitive data, writing new or modified keys, applying breaking rules for multi-line events, filtering unwanted events, and routing events to specified indexes or servers

This indexing process is also known as event processing.

Another benefit with Splunk Indexer is data replication. You need not worry about loss of data because Splunk keeps multiple copies of indexed data. This process is called Index replication or Indexer clustering. This is achieved with the help of an Indexer cluster, which is a group of indexers configured to replicate each other’s’ data.

**Splunk Search Head**

Search head is the component used for interacting with Splunk. It provides a graphical user interface to users for performing various operations. You can search and query the data stored in the Indexer by entering search words and you will get the expected result.

You can install the search head on separate servers or with other Splunk components on the same server. There is no separate installation file for search head, you just have to enable Splunk web service on the Splunk server to enable it.

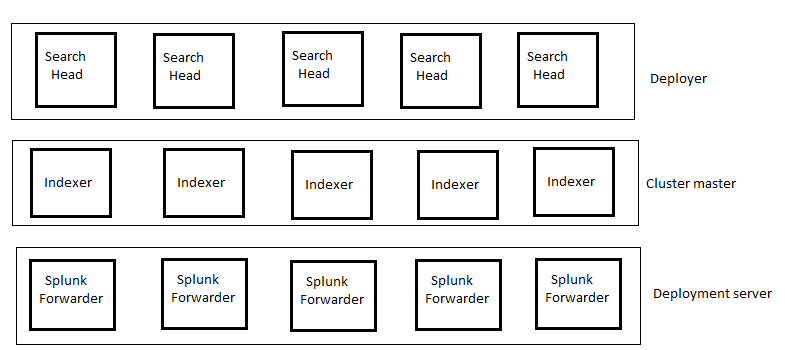
A Splunk instance can function both as a search head and a search peer. A search head that performs only searching, and not indexing is referred to as a dedicated search head. Whereas, a search peer performs indexing and responds to search requests from other search heads.

In a Splunk instance, a search head can send search requests to a group of indexers, or search peers, which perform the actual searches on their indexes. The search head then merges the results and sends them back to the user. This is a faster technique to search data called distributed searching.

Search head clusters are groups of search heads that coordinate the search activities. The cluster coordinates the activity of the search heads, allocates jobs based on the current loads, and ensures that all the search heads have access to the same set of knowledge objects.

Advantage of HF- it does index of data, so we will have less load on indexer.

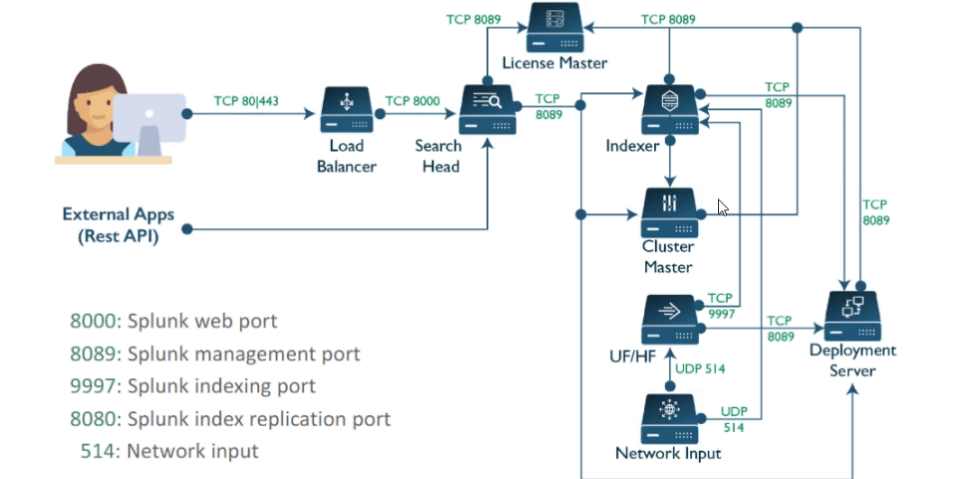
Management component of Splunk- 1) Deployment server, 2) Cluster master, 3) deployer.



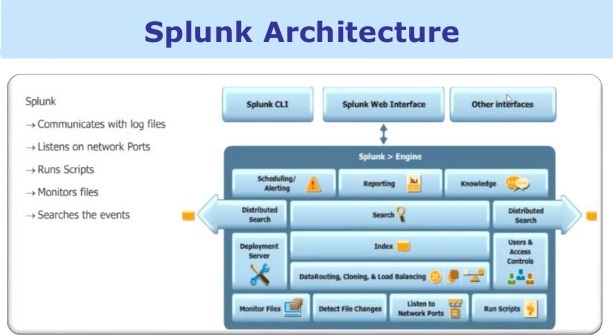
**Multiple Splunk forwarders can we managed by deployment server.**

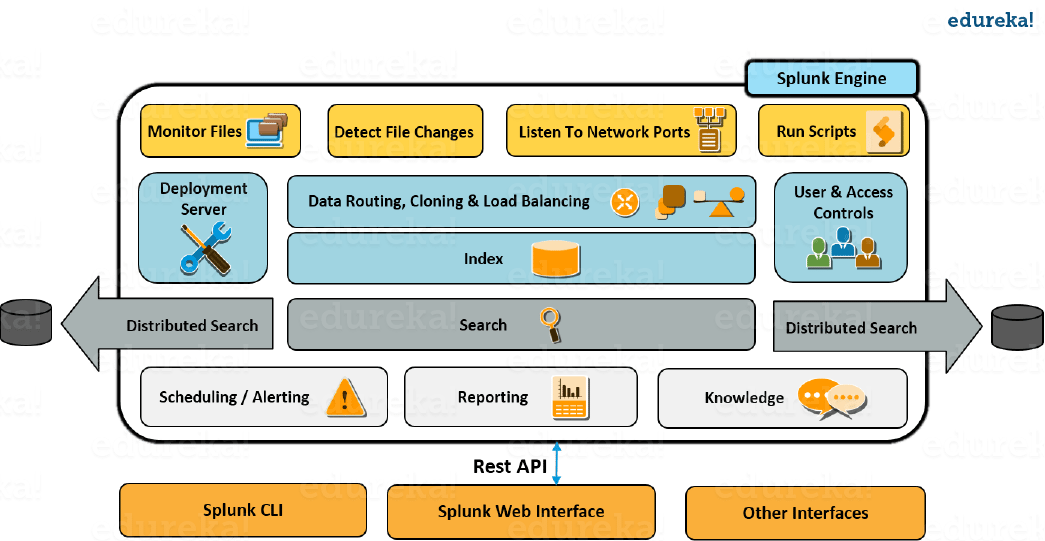
**Licence Master-** A license master controls one or more [**license slaves**](https://docs.splunk.com/Splexicon:Licenseslave). From the license master, you can define [**stacks**](https://docs.splunk.com/Splexicon:Stack), [**pools**](https://docs.splunk.com/Splexicon:Pool), add licensing capacity, and manage license slaves.

**Splunk Port-**

****

**Splunk Architecture-**





**Question: what is Search head?** Answer: it's a default first page where we can write query, make report, create dashboard. it's searching unit of Splunk.

Note- Inbound port for incoming traffic, and outbound port for outbound traffic

Select the main background process in Splunk? -> Splunkd and splunkWeb

Enabling receiving port-setting-> forwarding and receiving -> receiving data -> Add-new

HF vs UF - UF forward send data as it is, whereas HF do index, and send parsed data.

Replication port- it’s for indexer-to-indexer communication. Port number is 8080.

Licence Master- A license master controls one or more [license slaves](https://docs.splunk.com/Splexicon:Licenseslave). From the license master, you can define [stacks](https://docs.splunk.com/Splexicon:Stack), [pools](https://docs.splunk.com/Splexicon:Pool), add licensing capacity, and manage license slaves.

3 Tier architecture of Splunk- Data layer -> Processing layer and searching layer.

UF installation- There is a separate installation software for UF, during the installation it will ask for SSL port. Here you need to share details of your additional SSL certificate if you have, however UF is secured by Splunk itself.

During the installation it may ask which data you want to pick, We are advised to leave that page as it is.

How to start Splunk services?

Navigate to **System > Server controls**. Click **Restart Splunk**.

Start, stop, or restart both processes at once by going to %SPLUNK\_HOME%\bin and typing

> splunk [start|stop|restart]

Question: How many ways we can access the splunk Console? Answer: CLI, splunk web and API.

Enterprise licence- its’s for Business application.

Enterprise trial licence- for trial purpose, later you can upgrade.

Free Licence- if you will not upgrade your Enterprise trial licence, so after 60 days. It will convert to free licence.

Forwarder licence- This licence applicable for HF(Who dos indexing), not to UF.

Master slave concept to licence master- All the component of splunk architecture such as deployment server, cluster master, indexer report to licence master as a slave, whereas licence master is utilization meter for them.