**Splunk** is a software technology that uses the data generated by the computer to track, scan, analyze, and visualize it in real-time. It tracks and read store data as indexer events and various types of log files. It enables us to view data in different Dashboard formats.

Splunk is a program that enables the search and analysis of computer data. It analyzes semi-structured data and logs generated by various processes with proper data modeling as per the need of the IT companies. The user produces the data by means of any device like- web apps, sensors, or computers. It has built-in functionality for defining data types, field separators, and search process optimization. For the searched result, it also provides visualization of data.

Students, IT developers, and experts in IT infrastructure management who want to develop a strong understanding of basic Splunk concepts must-read study this tutorial. You can attain intermediate expertise in Splunk after completing this tutorial, and quickly draw on your skills to solve more difficult problems.

The reader should be familiar with the language of querying, like SQL. General awareness of standard operations would be particularly useful when using computer applications such as data storage and retrieval and reading computer programs generated logs.

What is splunk

* Big data platform for machine generated data
* convert raw unstructured data into searchable events
* organize data into indexes
* users can create dashboard, alerts and reports

Buckets in splunk

**How data ages**

A bucket moves through several states as it ages:

* hot
* warm
* cold
* frozen
* thawed

As buckets age, they "roll" from one state to the next. When data is first indexed, it gets written to a hot bucket. Hot buckets are buckets that are actively being written to. An index can have several hot buckets open at a time. Hot buckets are also searchable.

When certain conditions are met (for example, the hot bucket reaches a certain size or the indexer gets restarted), the hot bucket becomes a warm bucket ("rolls to warm"), and a new hot bucket is created in its place. The warm bucket is renamed but it remains in the same location as when it was a hot bucket. Warm buckets are searchable, but they are not actively written to. There can be a large number of warm buckets.

Once further conditions are met (for example, the index reaches some maximum number of warm buckets), the indexer begins to roll the warm buckets to cold, based on their age. It always selects the oldest warm bucket to roll to cold. Buckets continue to roll to cold as they age in this manner. Cold buckets reside in a different location from hot and warm buckets. You can configure the location so that cold buckets reside on cheaper storage.

Finally, after certain other time-based or size-based conditions are met, cold buckets roll to the frozen state, at which point they are deleted from the index, after being optionally archived.

If the frozen data has been archived, it can later be thawed. Data in thawed buckets is available for searches.

Settings in [indexes.conf](http://docs.splunk.com/Documentation/Splunk/9.3.0/admin/Indexesconf) determine when a bucket moves from one state to the next.

Here are the states that buckets age through:

|  |  |  |
| --- | --- | --- |
| **Bucket state** | **Description** | **Searchable?** |
| **Hot** | New data is written to hot buckets. Each index has one or more hot buckets. | Yes |
| **Warm** | Buckets rolled from hot. New data is not written to warm buckets. An index has many warm buckets. | Yes |
| **Cold** | Buckets rolled from warm and moved to a different location. An index has many cold buckets. | Yes |
| **Frozen** | Buckets rolled from cold. The indexer deletes frozen buckets, but you can choose to archive them first. Archived buckets can later be thawed. | No |
| **Thawed** | Buckets restored from an archive. If you archive frozen buckets, you can later return them to the index by thawing them. | Yes |

**Note:** For indexes enabled with the **[SmartStore](https://docs.splunk.com/Splexicon:SmartStore" \o "Splexicon:SmartStore" \t "_blank)** feature, which places data on a remote store such as S3, the cold state does not ordinarily exist. See [Bucket states and SmartStore](http://docs.splunk.com/Documentation/Splunk/9.3.0/Indexer/SmartStoreindexing#Bucket_states_and_SmartStore).

**What the index directories look like**

Each index occupies its own directory under $SPLUNK\_HOME/var/lib/splunk. The name of the directory is the same as the index name. Under the index directory are a series of subdirectories that categorize the buckets by state (hot/warm, cold, or thawed).

Each bucket is a subdirectory within those directories. The bucket names indicate the age of the data they contain.

Here is the directory structure for the default index (defaultdb):

|  |  |  |
| --- | --- | --- |
| **Bucket state** | **Default location** | **Notes** |
| **Hot** | $SPLUNK\_HOME/var/lib/splunk/defaultdb/db/\* | Each hot bucket occupies its own subdirectory. |
| **Warm** | $SPLUNK\_HOME/var/lib/splunk/defaultdb/db/\* | Each warm bucket occupies its own subdirectory. |
| **Cold** | $SPLUNK\_HOME/var/lib/splunk/defaultdb/colddb/\* | Each cold bucket occupies its own subdirectory. When warm buckets roll to cold, they get moved to this directory. |
| **Frozen** | When buckets freeze, they get deleted or archived into a location that you specify. | Deletion is the default. See [Archive indexed data](http://docs.splunk.com/Documentation/Splunk/9.3.0/Indexer/Automatearchiving) for information on how to archive the data instead. |
| **Thawed** | $SPLUNK\_HOME/var/lib/splunk/defaultdb/thaweddb/\* | Buckets that are archived and later thawed reside in this directory. See [Restore archived data](http://docs.splunk.com/Documentation/Splunk/9.3.0/Indexer/Restorearchiveddata) for information on restoring archived data to a thawed state. |

There is a variety of benefits that are offered by the Splunk, as follows:

* Real-time screen visibility.
* Splunk offers Better Interface.
* By offering instant results, it reduces troubleshooting and time-solving.
* It is the most effective method for the study of root causes.
* Splunk permits the generation of graphs, warnings, and dashboards.
* Similar findings can be quickly checked and analyzed using Splunk.
* It enables us to troubleshoot any failure state to improve performance.
* Helps you to track and make educated decisions on every company measure.
* Splunk allows Artificial Intelligence to be incorporated into the data strategy.
* Helps you to gather useful Operational Intelligence from your system data
* Splunk allows us to recognize any data type such as **.csv, json, log** formats, etc.
* Provides the most powerful search and visualization tools to enable all types of users.
* Allows us to establish a central server, where Splunk data can be searched from various sources.

Splunk has some essential features:

* It accelerates the Development & Testing.
* The building of Real-time Data Applications.
* Generate ROI faster
* Agile figures and Real-time architecture documentation.
* Splunk also provides search, analysis, and visualization capabilities to empower users.

Splunk Versions

There are three different versions of Splunk

* Splunk Enterprise
* Splunk Light
* Splunk Cloud

**Splunk Enterprise**

Big IT enterprise uses the Splunk Enterprise Version. With the help of the Splunk tool, we can collect and analyze the data from mobile phones, websites, and applications, etc.

**Splunk Cloud**

Splunk Cloud is a website that is the host. It possesses the same features as the company version. It can be used from Splunk or the cloud platform AWS.

**Splunk Light**

The free version of Splunk Illumination. It enables scanning, recording, and editing of your log data. Compared with other versions, it has limited functionalities and features.

Features of Splunk

We are going to tell you all the features of the Business version of the Spunk.

**Data Ingestion**

In Splunk, we can import or insert the date from different data formats like - JSON, XML, and weblogs and application logs that have unstructured system data. The unstructured data can be modeled as the consumer wants in a data structure.

**Data Indexing**

Splunk indexes the ingested data for speedier search and query on different conditions.

**Data Searching**

Splunk analysis involves using the indexed data to establish graphs, to forecast future trends, and to find patterns in the data.

**Using Alerts**

Used to trigger emails or RSS feeds when a certain requirement is identified in the data that is being analyzed.

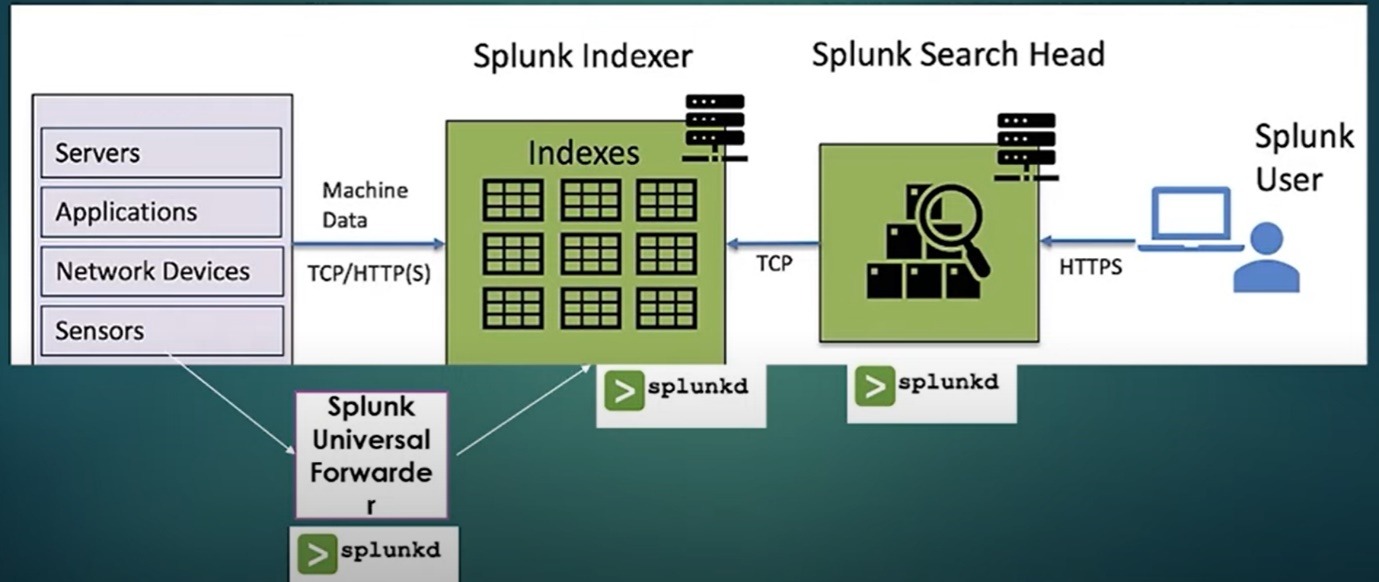
**Dashboards**

When we searched anything, the search result is displayed in the dashboard in the form of maps, reports, pivots, etc.

**Data Model**

The indexed data may be modeled into one or more data sets based on domain expertise. It leads to more straightforward navigation by end-users who evaluate the business cases without understanding the language techniques used by Splunk

Splunk architecture



To download splunk

We can download the Splunk from the link given below: <https://www.splunk.com/en_us/download/splunk-enterprise.html>

Dowload Tutorialdata.zip, prices.csv.zip

SPL(Search Processing language)

source="tutorialdata.zip:\*" host="LAPTOP-CLRL2B1O" index="main"

index=main sourcetype = vendor\_sales/vendor\_sales

1. generate report of number of events grouped by categoryId.

index=main

sourcetype=access\_combined\_wcookie

status!=200

| timechart span=6h count by categoryId

index=main

sourcetype=access\_combined\_wcookie

|stats count by categoryId

to display topmost 2 categories that appears frequently

index=main sourcetype="access\_combined\_wcookie"

| stats count by categoryId

| sort -count limit=2

1. generate a report of top 10 vendorIds for vender codes D or E

index=main sourcetype = vendor\_sales/vendor\_sales Code=D or Code=E

|top VendorID

index=main sourcetype = vendor\_sales/vendor\_sales Code=D or Code=E

|top VendorID limit=5 showperc=false

Note : showperc=false will hide percentage information

index=main

sourcetype=access\_combined\_wcookie

|stats max(bytes) as “Biggest Response”, min(bytes) as “smallest response” median(bytes) count as “total number of request”

host=”web\_application”| highlight safari , butter

extra commands

buttercupgames (error OR fail\* OR severe)

sourcetype=access\_combined error | top 5 uri

source="Prices.csv.zip:.\\prices.csv" | sort price

source="Prices.csv.zip:.\\prices.csv" | sort product\_name

source="tutorialdata.zip:.\\vendor\_sales/vendor\_sales.log" | sort VendorID, -Code

source=job\_listings | where salary > industry\_average

source="Prices.csv.zip:.\\prices.csv" | where price > 25

source="Prices.csv.zip:.\\prices.csv" | head 8

source="Prices.csv.zip:.\\prices.csv" | search "Pony Run"

sourcetype=access\_\* | top action by referer\_domain

sourcetype=access\_\* status=200 action=purchase | top itemId

sourcetype=access\_\* | stats count by host

buttercupgames | stats count by status

source="Prices.csv.zip:.\\prices.csv" | table productId, product\_name, price