



Lesson 200.1 Splunk Basics



Learning Objectives

At the end of this lesson, learners will be able to:

- Explain Machine data.
- Describe Splunk.
- List the components of Splunk.
- Describe the uses of Splunk.
- Define Splunk apps.
- Describe Splunk deployment types.
- Practice navigating in Splunk.
- Practice configuring account and user settings.

Introduction

Splunk is a **platform** that helps organizations search, analyze, and visualize **machine-generated** data. It ingests data from **various sources** and provides real-time insights to quickly identify trends and anomalies.

With its **visualization** tools and scalability, Splunk can be used for a **wide range** of use cases, including IT operations, security and compliance, and business analytics.

As an IT professional, you need to monitor and troubleshoot various systems, applications, and infrastructure components. Splunk can help you **gain visibility** into these systems by ingesting and analyzing the logs and metrics generated by them.

1.1 Machine Data

Machine data refers to any type of data that is generated by **machines**, systems, or devices.

This data is often produced **automatically**, and in **large quantities**, making it **difficult** to manage and analyze without the right tools.

Some examples of machine data include:

- **Server logs:** These logs contain information about **server events**, such as user logins, system errors, and software updates.

```
1513359572.659 53 65.19.167.94 TCP_REFRESH_HIT/200 3681 GET http://www.d11.org/images/MPj04089820000[1].jpg j
1513359622.717 56 125.7.55.180 TCP_REFRESH_HIT/200 4110 GET http://www.d11.org/images/button49.jpg jreistad@b
1513359670.771 53 195.69.252.22 TCP_REFRESH_HIT/200 4210 GET http://www.d11.org/images/MPj03997860000[1].jpg
1513359760.831 58 12.130.60.4 TCP_REFRESH_HIT/200 2777 GET http://www.d11.org/images/button50.jpg jreistad@bu
1513359783.888 55 201.3.120.132 TCP_REFRESH_HIT/200 4204 GET http://www.d11.org/images/MPj04010540000[1].jpg
1513359826.943 53 211.166.11.101 TCP_REFRESH_HIT/200 3727 GET http://www.d11.org/images/button51.jpg jreistad
1513359899.049 104 94.230.166.185 TCP_REFRESH_HIT/200 4524 GET http://www.d11.org/images/MPj04008250000[1].jp
1513359950.107 56 211.166.11.101 TCP_REFRESH_HIT/200 3902 GET http://www.d11.org/images/button1c.jpg jreistad
1513360003.166 58 195.2.240.99 TCP_REFRESH_HIT/200 2995 GET http://www.d11.org/images/CHOICE.jpg jreistad@but
1513360073.230 63 89.167.143.32 TCP_REFRESH_HIT/200 41901 GET http://www.d11.org/images/profile.jpg jreistad@
1513360090.282 50 125.17.14.100 TCP_REFRESH_HIT/200 7185 GET http://www.d11.org/images/index.2.jpg jreistad@b
1513360149.183 899 192.162.19.179 TCP_MISS/200 85168 GET http://www.d11.org/flash/flash.swf?xml_path=flash/f1
1513360201.242 57 2.229.4.58 TCP_REFRESH_HIT/200 1161 GET http://www.d11.org/images/j0431559.png jreistad@but
1513360262.297 53 196.28.38.71 TCP_REFRESH_HIT/200 522 GET http://www.d11.org/images/BD15023_.gif jreistad@bu
1513360326.355 56 210.192.123.204 TCP_REFRESH_HIT/200 1850 GET http://www.d11.org/images/d11.ht3.jpg jreistad
1513360379.409 52 88.191.145.142 TCP_REFRESH_HIT/200 4784 GET http://www.d11.org/images/button40.jpg jreistad
1513360475.478 66 87.194.216.51 TCP_REFRESH_HIT/200 5236 GET http://www.d11.org/images/button41.jpg jreistad@
1513360577.537 58 87.194.216.51 TCP_REFRESH_HIT/200 1212 GET http://www.d11.org/images/rssicon.jpg jreistad@b
1513360668.592 53 220.225.12.171 TCP_REFRESH_HIT/200 4540 GET http://www.d11.org/images/button42.jpg jreistad
1513360718.647 53 64.120.15.156 TCP_REFRESH_HIT/200 692 GET http://www.d11.org/images/MsoPnl_Cnr_b1_21.gif jr
1513360815.707 58 2.229.4.58 TCP_REFRESH_HIT/200 639 GET http://www.d11.org/images/MsoPnl_Cnr_br_23.gif jreis
1513360834.763 54 109.169.32.135 TCP_MISS/404 1955 GET http://www.d11.org/working/images/employees.jpg jreist
1513360930.132 315 64.120.15.156 TCP_MISS/200 2183 GET http://www.crawler.com/ ewarwick@buttercupgames.com DI
1513361012.191 57 222.41.213.238 TCP_MISS/200 1319 GET http://www.crawler.com/xm/style_all.css ewarwick@butte
1513361077.263 69 193.33.170.23 TCP_MISS/200 1313 GET http://www.crawler.com/xm/mlogo_crawler.gif ewarwick@bu
```

image: Screenshot of cisco_ironport_web log file

- [illegible]

1.1 Machine Data

- **Application logs:** These logs contain information about application events, such as user interactions, errors, and performance metrics.
- **Security data:** This data includes information about security events, such as login attempts, system breaches, and virus/malware detections.
- **IoT data:** Data generated by the Internet of Things (IoT) devices, such as smart homes, wearables, and industrial sensors.

1.1 Machine Data - Summary

To summarize, machine data is a **valuable source of information** for organizations, and can provide **insights** into how machines and systems are performing.

This information can be used to **optimize** operations, **improve** security, and **drive** business decisions.

That said, **raw** machine data is often produced in **large** quantities, with a high level of **complexity** and structure that is **not** immediately understandable to humans.

This is where **Splunk** enters the stage.

1.2 Splunk

Splunk is a powerful **platform** that helps organizations **search**, **analyze**, and **visualize** machine-generated data.

It provides a **scalable**, real-time approach to **ingesting** and **processing large volumes of data** from a **wide range of sources**, including servers, applications, network devices, and security systems.

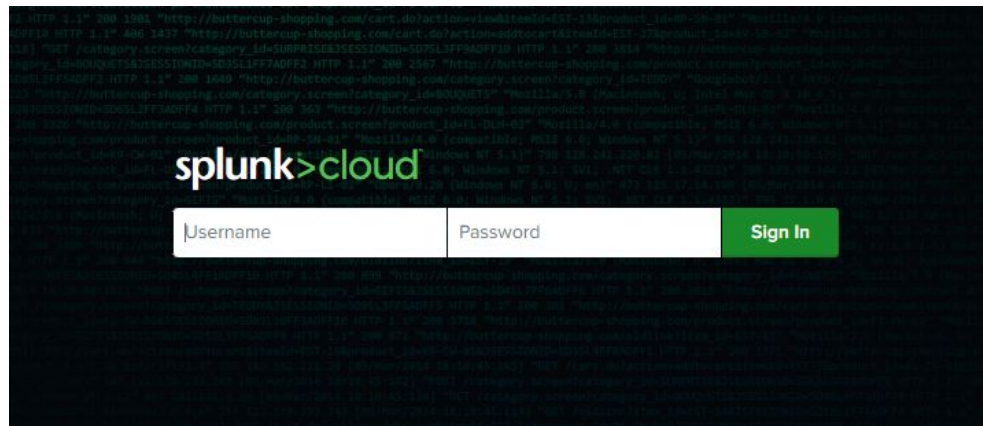


image: screenshot of splunk cloud sign in page

1.2 Splunk (continued)

Splunk's **indexing** engine allows data to be easily searched, correlated, and analyzed, providing IT professionals with a comprehensive view of their environment.

Splunk also provides a variety of **pre-built apps** and **add-ons** that can help IT professionals gain even more value from their machine data, such as monitoring security events, analyzing network traffic, and managing IT operations.

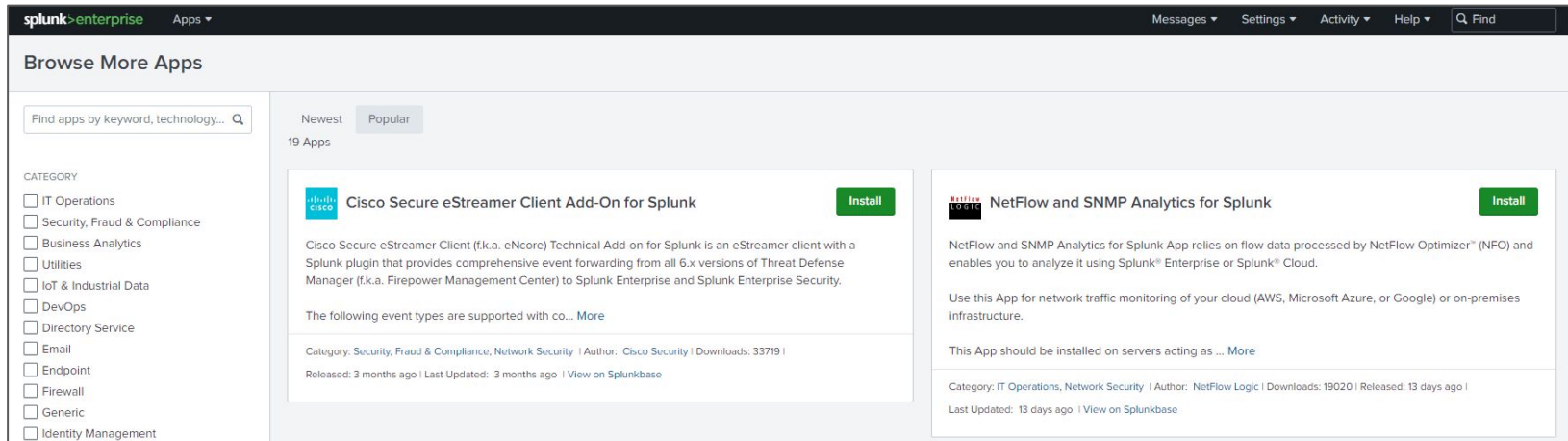


image: screenshot of splunk's App Browser

1.2 Splunk (continued)

With its **visualization** tools, Splunk can also help organizations **quickly identify** trends, anomalies, and potential problems.

Splunk's machine learning capabilities can be used to detect anomalies and predict future issues.

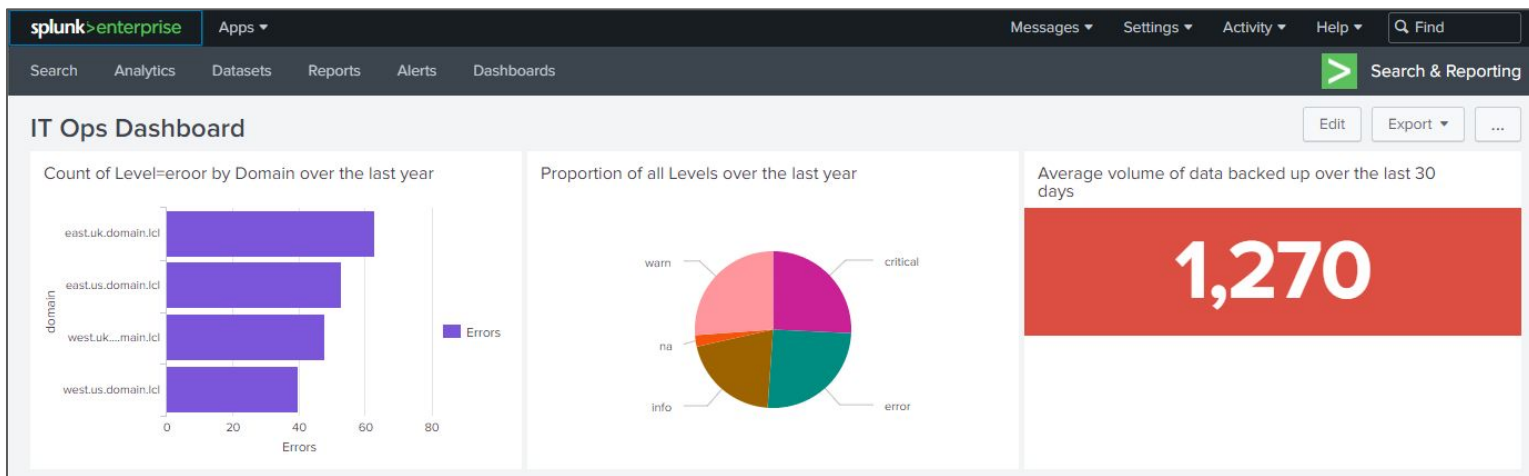


image: screenshot of a splunk dashboard

1.2 Splunk (continued)

Splunk operates through **five** main functions:



PresentationGO.com – The free PowerPoint and Google Slides template library
icons: icons8.com

1.2 Splunk - 5 Functions

Index Data

Splunk **indexing** is the process of **ingesting data** from various sources, such as logs, metrics, and configuration files, and storing it in a format that can be easily **searched**, correlated, and analyzed.

The Splunk indexing engine uses a **multi-step** process to index data, which includes:

- **Breaking** the data into events.
- **Parsing** the events to extract fields.
- **Normalizing** the fields into a common format.
- **Storing** the normalized events in an index.



1.2 Splunk - 5 Functions (continued)

Search and Investigate

Splunk search and investigate allows users to **search**, **filter**, and **investigate** data in real time.

Users can create **complex queries** and **filters** to extract **insights** from their data.

Splunk search and investigate includes powerful capabilities, such as:

- Real-time searching.
- Search Processing Language (SPL), which enables users to create complex search queries.
- Field extraction.
- Visualization.
- Correlation.



1.2 Splunk - 5 Functions (continued)

Add Knowledge

The Add Knowledge function in Splunk refers to the ability to add **custom knowledge** to the data indexed in Splunk.

Users can create **custom** fields, tags, event types, and other data structures that help them better understand and analyze their data.

Here are some of the **key features** of Splunk's Add Knowledge function:

- **Custom fields:** users can create custom fields for raw data extraction.
- **Tags:** users can apply tags to events to group related events together.
- **Lookups:** users can create custom lookups that match values in their data to other data sources; this can provide context to data.



1.2 Splunk - 5 Functions (continued)

Monitor and Alert

The Monitor and Alert function in Splunk is designed to help users **proactively monitor** their data and receive **alerts** when certain conditions are met.

Here are some of the key features of Splunk's Monitor and Alert function:

- **Real-time monitoring:** users can monitor their data in real time.
- **Custom alerts:** users can create custom alerts based on specific conditions.
- **Alert actions:** users can define actions to be taken when alerts are triggered.
- **Alert escalation:** users can define escalation paths for alerts, ensuring that alerts are addressed in a timely manner.



1.2 Splunk - 5 Functions (continued)

Report and Analyze

The Report and Analyze function in Splunk helps users **analyze** and **visualize** their data to gain **insights** into their environment.

Users can create reports and dashboards that provide a **comprehensive view** of their data, allowing them to make informed decisions and take action based on their findings. Users can also **present** and **share** their findings on a **single pane of glass**.

Here are some of the key features of Splunk's Report and Analyze function:

- **Data visualization:** Splunk provides a range of visualization tools, including charts, graphs, and **dashboards**.
- **Custom reports:** users can schedule reports to run at specific intervals.
- **Scheduled reports:** users can define **actions** to be taken when alerts are triggered.
- **Alert escalation:** users can define **escalation** paths for alerts, ensuring that alerts are addressed in a timely manner.



1.2 Splunk - Summary

Splunk is a powerful **software platform** that allows organizations to **collect, index, search, and analyze machine-generated data** from virtually any source.

It enables users to gain **valuable insights** from their data in real time, allowing them to make informed decisions and take action quickly.

With its user-friendly interface and **advanced features**, Splunk has become a popular tool for IT professionals to **monitor, troubleshoot, and optimize** their systems and infrastructure.

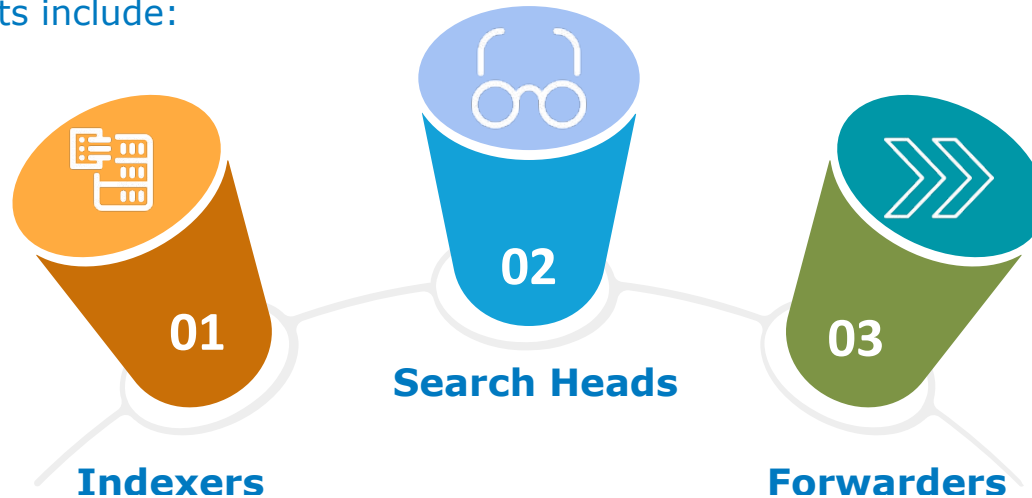
Additional documentations for Splunk can be found at **docs.splunk.com**.

1.3 Splunk Components

Splunk Enterprise is Splunk's **core** product that provides the indexing and search capabilities needed to process machine data.

Splunk is composed of **several components** that work **together** to process, index, and analyze machine data.

These components include:



1.3 Splunk Components (continued)

Indexers are components of Splunk that receive, process, and **index** data from various sources, making it searchable and available for analysis.

Indexers allow data to be stored and retrieved **quickly** and **efficiently**, even across massive datasets.

In larger environments, **multiple indexers** can be clustered together to provide high availability and improved performance.



Indexers

The index is composed of several files, including the **raw data**, **metadata**, and **index** files.

The **raw** data files contain the **original data** that was ingested, while the **metadata** files contain information about each event, such as the **timestamp** and **source**.

Splunk uses a process called "**bucketing**" to manage the **index** files. The index is divided into **time-based buckets**, which **group together** events that occurred within a specific time period, such as one hour or one day.

1.3 Splunk Components (continued)

The **search head** is a component of Splunk that enables users to **search** and **analyze** data in real time. When a user initiates a **search query**, it is sent to the **search head**, which then distributes the search to one or more **indexers**.

- The search head **receives the results** from the indexers and **presents** them to the user in a user-friendly interface. Users can **visualize, analyze, and act** upon the data.
- The search head can also be used to create **alerts** and **reports**.



1.3 Splunk Components (continued)

The **forwarder** is a component of Splunk that is responsible for **collecting** and **forwarding** data from various sources to the Splunk indexers.

The forwarder runs on the **source machine** and collects data in real time from logs, metrics, and other sources, and then forwards the data to one or more indexers for processing.

- **Universal forwarders: lightweight** agents that can be deployed on a **wide range** of machines to collect data and forward it to Splunk.
- **Heavy forwarders: full** Splunk Enterprise instances that can index, search, and change data, as well as forward it.



1.3 Splunk Components - Summary

Indexers, Search Heads, and Forwarders are all **core components** of the Splunk software platform.

Indexers parse and timestamp raw data creating **index** files.

The **Search Head** enable users to interact with Splunk, providing a web-based **interface** for searching and analyzing data.

Forwarders are agents that run on the machines **generating** the data and forward it to the Splunk indexers.

By working **together**, these components enable Splunk to efficiently process and analyze machine data, providing IT professionals with the insights they need to **optimize** their systems and infrastructure.

1.4 Uses of Splunk

Splunk is a powerful platform that can be used for a **wide range of use cases** across industries. Here are some examples of how Splunk can be used:

- **IT Operations:** Splunk can be used to monitor and **troubleshoot** IT infrastructure, applications, and services, enabling IT teams to quickly identify and **resolve** issues that may impact business operations. Splunk can also be used for performance **monitoring**, capacity planning, and log analysis.

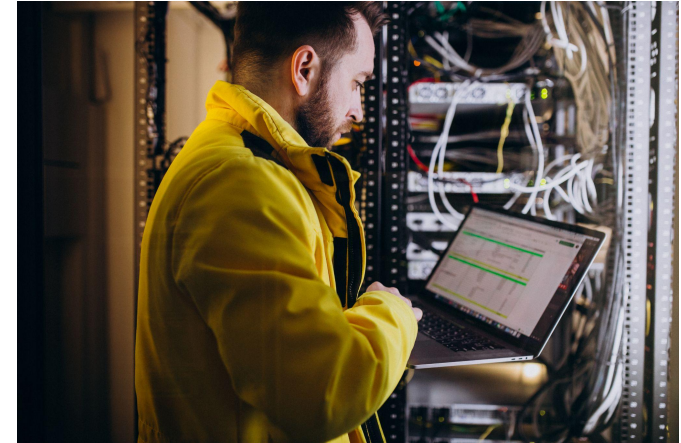


image: Freepik.com

1.4 Uses of Splunk (continued)

- **Security:** Splunk can be used to detect and respond to **security threats** in real time, helping security teams to **identify** and **investigate** suspicious activity, conduct forensic analysis, and manage security incidents. Splunk can also be used for **compliance** reporting and auditing.



image: Freepik.com

1.4 Uses of Splunk (continued)

- **Business Analytics:** Splunk can be used to analyze **customer behavior**, marketing campaigns, and **product performance**, enabling businesses to make data-driven decisions and improve customer engagement.

Splunk can also be used for **financial** analysis, **supply chain** management, and **risk** management.

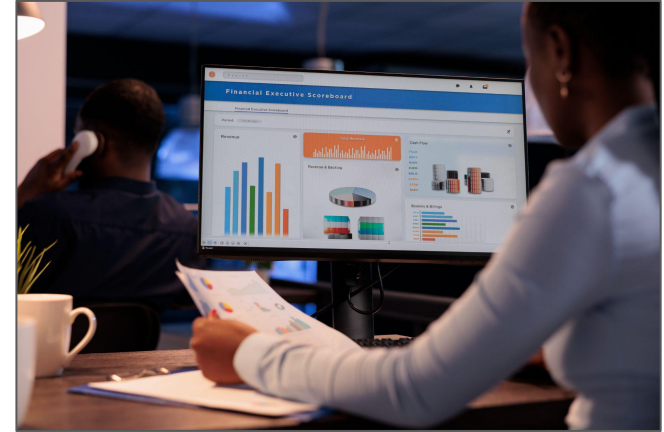


image: Freepik.com

1.4 Uses of Splunk (continued)

- **Internet of Things (IoT):** Splunk can be used to collect and analyze data from **IoT devices** and **sensors**, enabling businesses to monitor and optimize IoT deployments, improve product quality, and enhance customer experiences.



image: Freepik.com

1.4 Uses of Splunk (continued)

Use Case example: Resolving Network Security Breach

A company is alerted to a potential **security breach** on its network. The IT security team needs to quickly identify the **source** of the breach, assess the extent of the damage, and **take action** to **contain** and remediate the issue.

Using Splunk's search and investigation capabilities, the IT security team quickly **searches** for any **suspicious activity**, and correlates events across **multiple systems**. Splunk's dashboards and reports help to **visualize** and **analyze** network traffic, identify **anomalies**, and **patterns**, and gain insights into the attack.



image: Freepik.com

1.4 Uses of Splunk (continued)

Use Case example: Resolving Network Security Breach - continue

The IT security team **identifies** that the breach was caused by a **phishing attack** that led to the **compromise** of a user account. They can see that the attacker used the compromised account to **access sensitive data** on the network, and attempted to exfiltrate the data using a command and control server.

Now, the IT security team can quickly **respond** to the attack by **disabling** the compromised user account and **blocking** the attacker's IP address.



image: Freepik.com

1.4 Uses of Splunk - Summary

Splunk is a powerful **platform** with a wide range of uses. It can be used for **IT operations** and **infrastructure management**, such as monitoring and troubleshooting servers, networks, applications, and databases.

Splunk can also be used for **security** and **compliance** purposes, detecting and responding to security threats, investigating incidents, and ensuring regulatory compliance.

Additionally, Splunk can be used for **business analytics**, analyzing data and providing insights to help drive decision-making and improve business performance.

With its ability to index, search, and analyze machine-generated data from virtually any source, Splunk is a versatile tool that can provide **value** to a variety of industries and use cases.

1.5 Splunk Apps

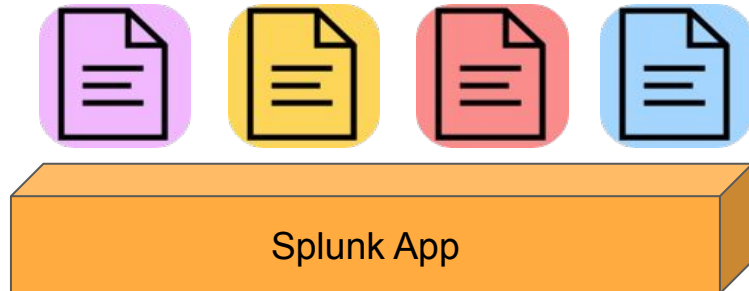
Splunk apps are **pre-built packages** of dashboards, reports, and other content that provide specific functionality or add-ons to the core Splunk platform.

They are designed to help users quickly and easily solve specific use cases or address particular needs, **without requiring extensive knowledge** or expertise in building custom dashboards and reports from scratch.

- Apps extend Splunk's capabilities.
- Apps can be created by **third parties, individuals**, or by **Splunk**.
- Apps marked with the **Splunk Built** logo are made by Splunk.
- Apps marked with the **Splunk Certified** logo are made by Splunk.
- An app can provide plenty of **value** to a user regardless of being certified by Splunk

1.5 Splunk Apps

A Splunk App consists of a collection of **Splunk configuration files** stored under the `C:\Program Files\Splunk\etc\apps\[app name]` folder (on a Windows machine).



icons: icons8.com

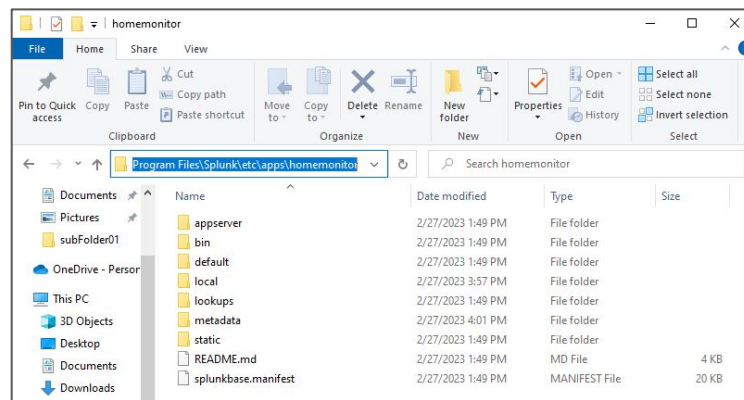
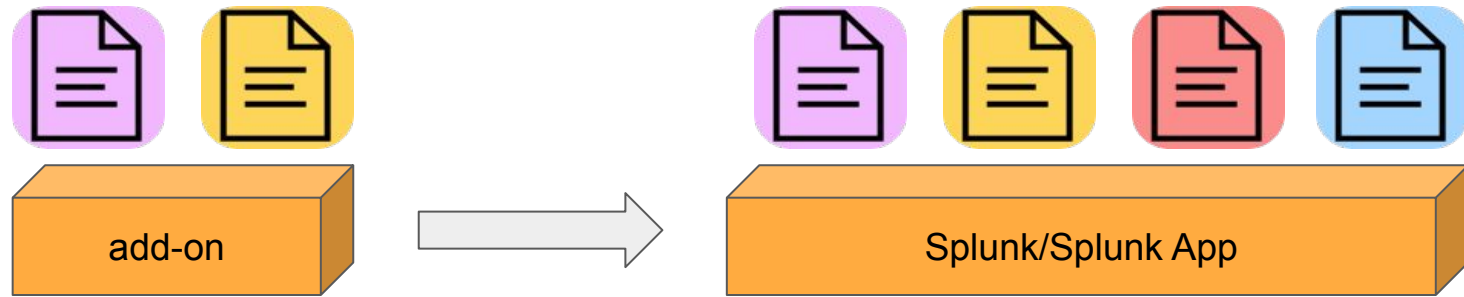


image: Windows screenshot

1.5 Splunk Apps (continued)

A Splunk add-on is a software module that provides **additional functionality** to the Splunk platform or to Splunk apps, usually in the form of **data inputs, parsers**, or other components that allow Splunk to ingest and analyze data from **specific** sources.

Add-ons are often developed by **third-party vendors** and are available for download from the Splunkbase app store, and they can be installed and configured using the Splunk Web Interface.

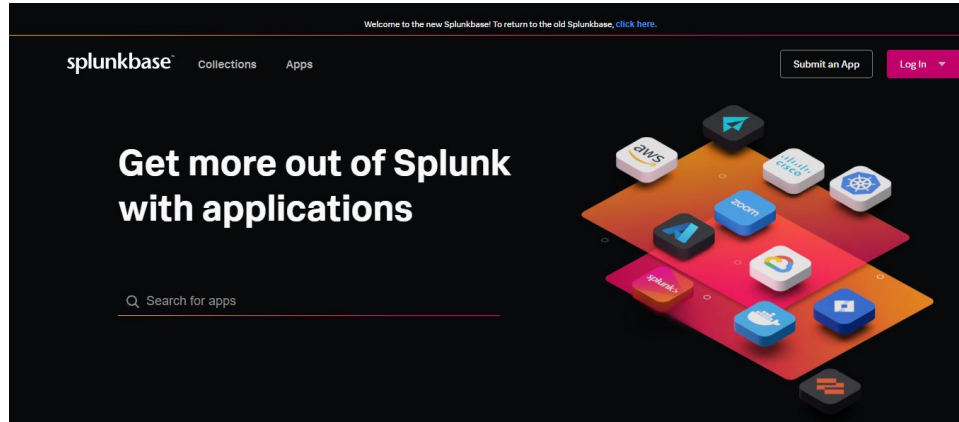


1.5 Splunk Apps (continued)

Many vendors such as Cisco, Amazon, and Microsoft create splunk apps that are tailored to help monitor their products.

Download apps from **within** the Splunk interface or from **Splunkbase.com**

While most apps are available for free, some premium apps require a license.



1.5 Splunk Apps - Summary

Splunk apps are **pre-built packages** of dashboards, reports, and other content that provide specific functionality or add-ons to the core Splunk platform.

They help users **quickly** solve **specific** use cases or address **particular** needs **without** requiring extensive knowledge or expertise in building custom dashboards and reports from **scratch**.

Splunk apps are available for **free** or as **premium** apps that require a license, and they can be downloaded from the **Splunkbase app store**.

Add-ons are configuration files that extend apps.

Additionally, users can develop their **own custom apps** using the Splunk app framework.

1.6 Splunk Deployment Types

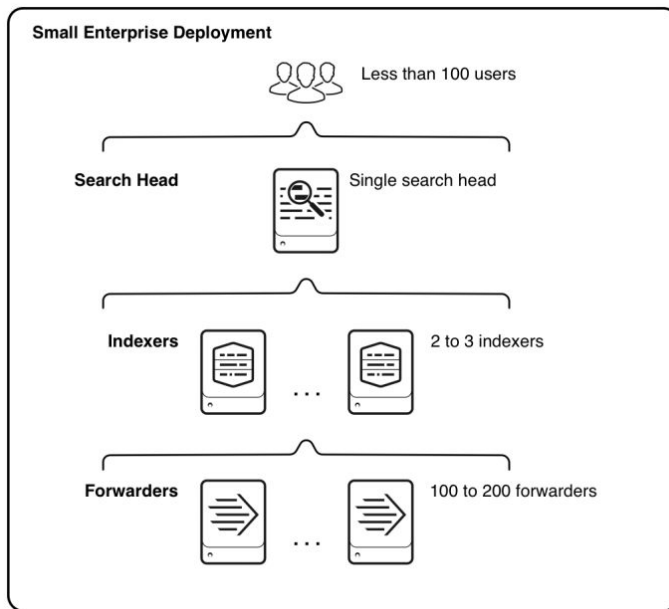
Splunk Enterprise can be deployed in a variety of configurations, from a **single** instance to **distributed** deployments consisting of hundreds or thousands of instances, depending on the specific requirements of the organization.

There are **three** primary deployment types for Splunk Enterprise:

- **Standalone:** A standalone deployment consists of a **single** Splunk instance that performs **all of the necessary functions**, including data ingestion, indexing, search, and visualization.
- **Distributed:** A distributed deployment consists of **multiple** Splunk instances, each of which is responsible for one or more functions. For example, one instance might be responsible for data ingestion, while another might be responsible for indexing and search.
- **Cloud:** A cloud deployment is similar to a **distributed** deployment but is hosted and managed by a **third-party cloud provider**, such as AWS or Microsoft Azure. This type of deployment is often used by organizations that do not want to manage their own infrastructure or need to scale quickly.

1.6 Splunk Deployment Types (continued)

This diagram illustrates the type of deployment that might support the needs of a **small** enterprise.



1.6 Splunk Deployment Types - Summary

Splunk offers **three** primary deployment types: **standalone**, **distributed**, and **cloud**.

A standalone deployment consists of a **single instance** of Splunk that performs all the necessary functions, while a distributed deployment consists of **multiple instances** that are responsible for specific functions.

Cloud deployment is similar to distributed deployment, but it is hosted and managed by a **third-party cloud provider**.

Organizations can choose the deployment type that **best suits** their needs, depending on the size and complexity of their environment.

1.7 Basic Navigation in Splunk

Splunk uses a customizable **web-based interface** that provides users with access to their indexed and stored data.

It includes components such as the **search bar**, search results page, **navigation bar**, and dashboard and visualization editors. It is designed to be intuitive and efficient for users to analyze and understand their data.

Signing into Splunk is done via a **web browser** using port **8000**.

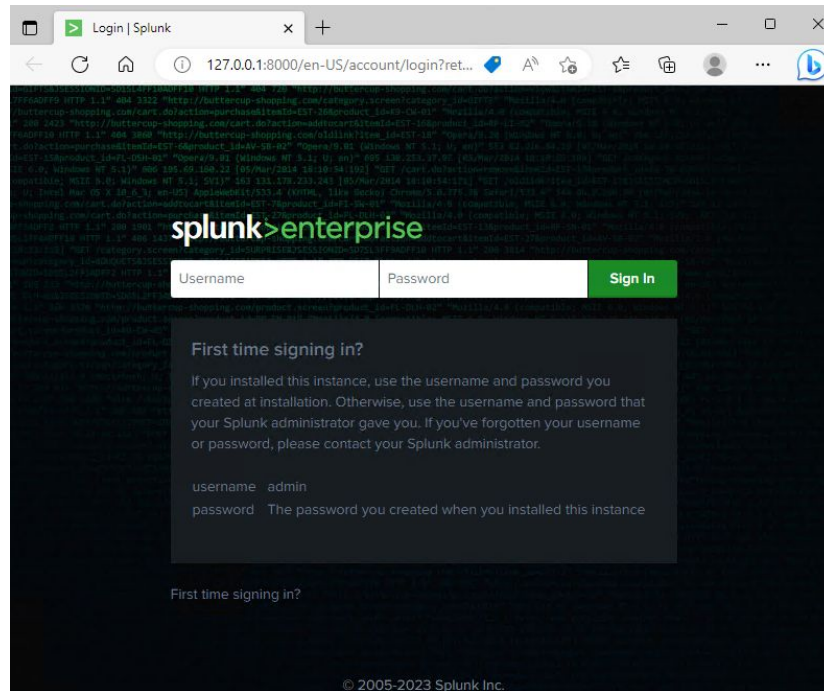


image: screenshot, splunk login page

1.7 Basic Navigation in Splunk (continued)

Instructor demo:

The following is a **partial** list of items within the Splunk platform. If possible, learners should follow along on their Splunk instance.

- Sign into Splunk.
- Home page / launcher.
- App panel / manage applications page.
- Splunk bar - navigation menus.
- Main user roles.

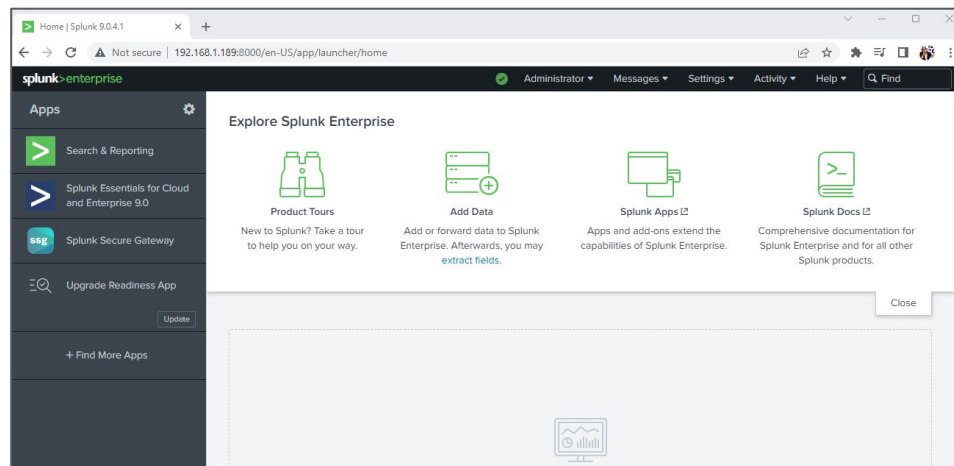


image: screenshot, splunk launcher page

1.8 Customizing User Settings

Instructor demo:

The following is a **partial** list of items within the Splunk user and account settings. If possible, learners should follow along on their Splunk instance.

- Account Settings:
 - Personal (full name, email, and password).
- Preference:
 - Global
 - Time Zone (ensure timestamps are aligned with user).
 - SPL Editor
 - Geneal
 - Themes

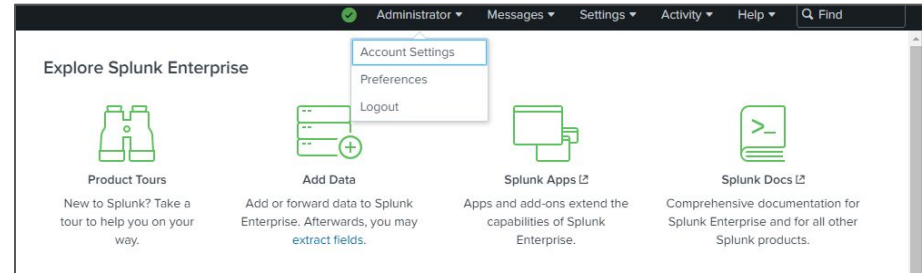


image: screenshot, splunk launcher page

Knowledge check.

- What is machine data? Where does it come from?
- What is the Splunk platform?
- What are Forwarders, Indexers, and Search Heads?
- What is one way an IT professional can use Splunk?
- What are Splunk Apps?
- What is a common Splunk deployment type?
- How do users connect to Splunk?
- Why is it important to set the timezone correctly in Splunk?
- How many main user roles do you have in Splunk?