thinkcloudly
Upscale your career in IT

DAY 02

SPLUNK BOOTCAMP

AGENDA

1. SPLUNK ARCHITECTURE AND SETUP
2. KEY SPLUNK COMPONENTS
3. INSTALLATION AND CONFIGURATION

Overview of Splunk Architecture

- **Data Ingestion:** Handles the collection of data from various sources seamlessly.
- **Search Head:** Facilitates user queries and provides search capabilities across data.
- **Deployment Server:** Manages configurations and app distributions across multiple instances.
- **Search Peers:** Enables distributed searching across indexed data efficiently.

Core Components of Splunk

- **Indexing Layer:** Organizes ingested data for efficient storage and fast retrieval.
- **Forwarders:** Transports data from source systems to Splunk for processing.
- **Cluster Master:** Oversees distributed search indexes ensuring data replication.
- **User Interface:** Provides visualization tools for reporting and dashboarding purposes

Core Components of Splunk

		Function	Usage	Importance	Interdependence	Performance
	Splunk Indexer	Indexes and stores data	Processes incoming data	Crucial for data retrieval	Works with Search Head	High
	Search Head	Searches and analyzes data	User queries data	Key for data insights	Depends on Indexer	Medium
	Forwarders	Collects and forwards data	Distributes data to Indexers	Essential for data intake	Supports Indexers	High
	Dashboard	Visualizes data results	User interface for insights	Critical for data interpretation	Requires Search Head	Medium

Key Features of Splunk

- **Real-time Indexing**

Splunk enables users to index data in real-time, allowing immediate access to insights, faster decision-making, and timely responses to security threats or network issues.

- **Advanced Visualization**

With Splunk, create interactive dashboards and dynamic visualizations to present data findings effectively, making it easier for stakeholders to interpret trends and anomalies.

Overview of Splunk

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.

Overview of Splunk

Data Storage Stage

Data storage consists of two phases: Parsing and Indexing.

1. In Parsing phase, Splunk software examines, analyzes, 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:
   a. Breaking the stream of data into individual lines
   b. Identifying, parsing, and setting timestamps
   c. Annotating individual events with metadata copied from the source-wide keys
   d. 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.

Overview of Splunk

Data Searching Stage

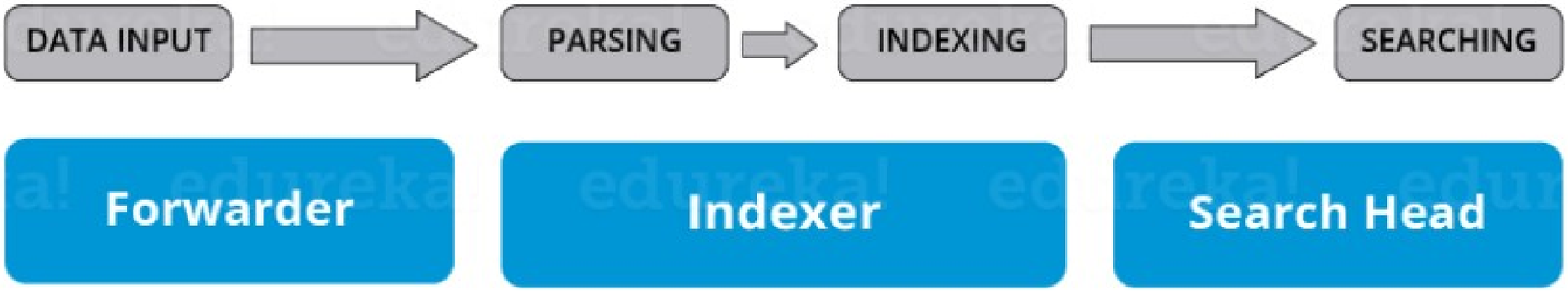
This stage controls how the user accesses, views, and uses the indexed data. As part of the search function, Splunk software stores user-created knowledge objects, such as reports, event types, dashboards, alerts and field extractions. The search function also manages the search process.

Overview of Splunk

Splunk Components

There are 3 main components in Splunk:

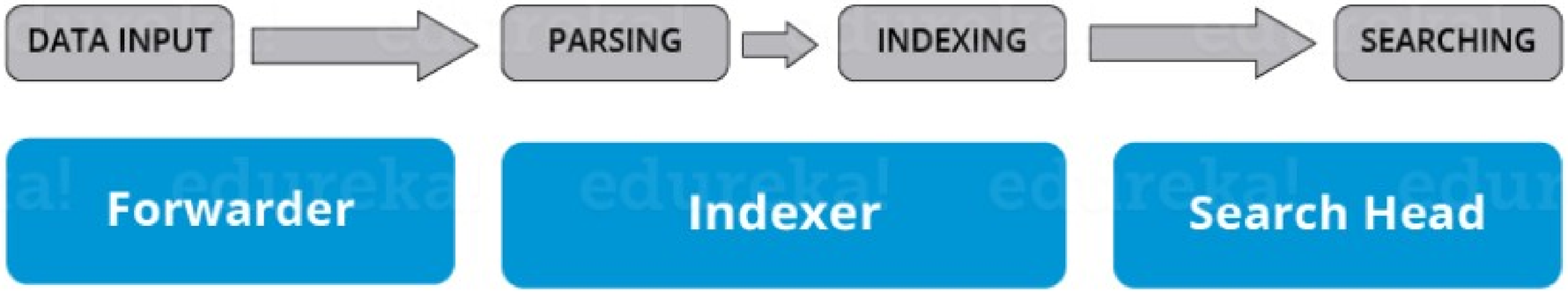
- Splunk Forwarder, used for data forwarding
- Splunk Indexer, used for Parsing and Indexing the data
- Search Head, is a GUI used for searching, analyzing and reporting



Overview of Splunk

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.



Overview of Splunk

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.
- **Heavy forwarder -**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.

Overview of Splunk

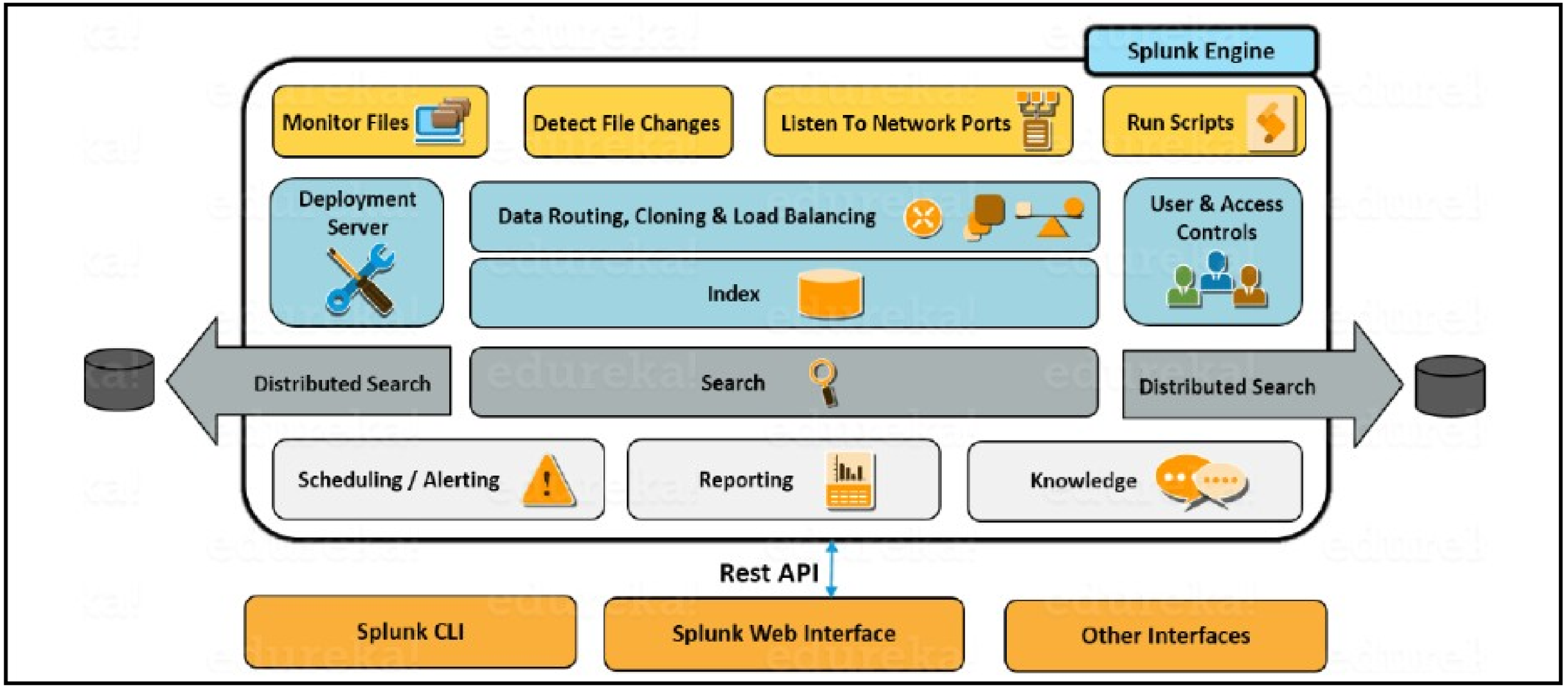
Splunk Indexer

- 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.

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.

Splunk Architecture



Splunk Architecture

- You can receive data from various network ports by running scripts for automating data forwarding
- You can monitor the files coming in and detect the changes in real time
- The forwarder has the capability to intelligently route the data, clone the data and do load balancing on that data before it reaches the indexer. Cloning is done to create multiple copies of an event right at the data source where as load balancing is done so that even if one instance fails, the data can be forwarded to another instance which is hosting the indexer

Continue...

Splunk Architecture

- As I mentioned earlier, the deployment server is used for managing the entire deployment, configurations and policies
- When this data is received, it is stored in an Indexer. The indexer is then broken down into different logical data stores and at each data store you can set permissions which will control what each user views, accesses and uses
- Once the data is in, you can search the indexed data and also distribute searches to other search peers and the results will merged and sent back to the Search head

Continue...

Installation and Configuration (Windows/Linux)

You can download the setup using the below link which is available for both windows and Linux platforms.

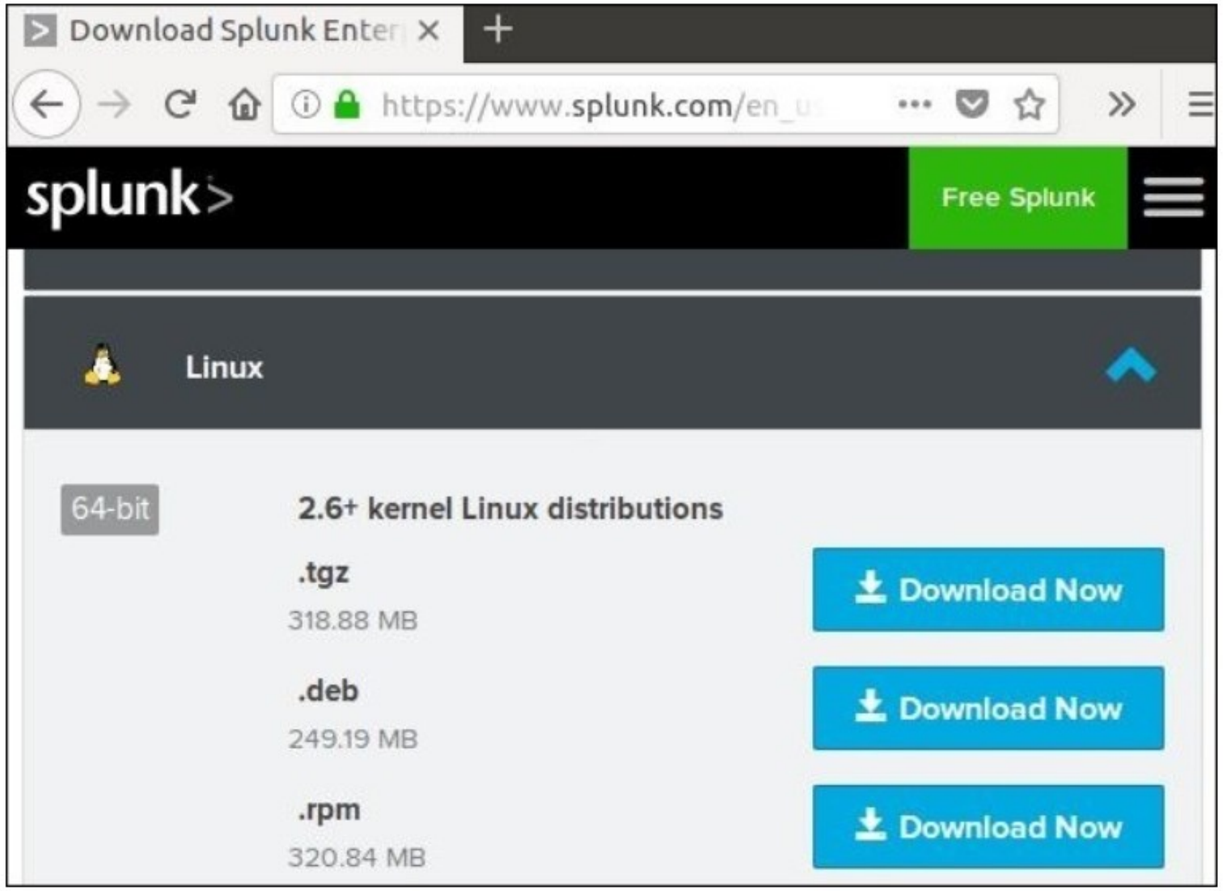
https://www.splunk.com/en_us/download/splunk-enterprise.html.

Installation and Configuration (Windows/Linux)

Linux Versions

Step 1

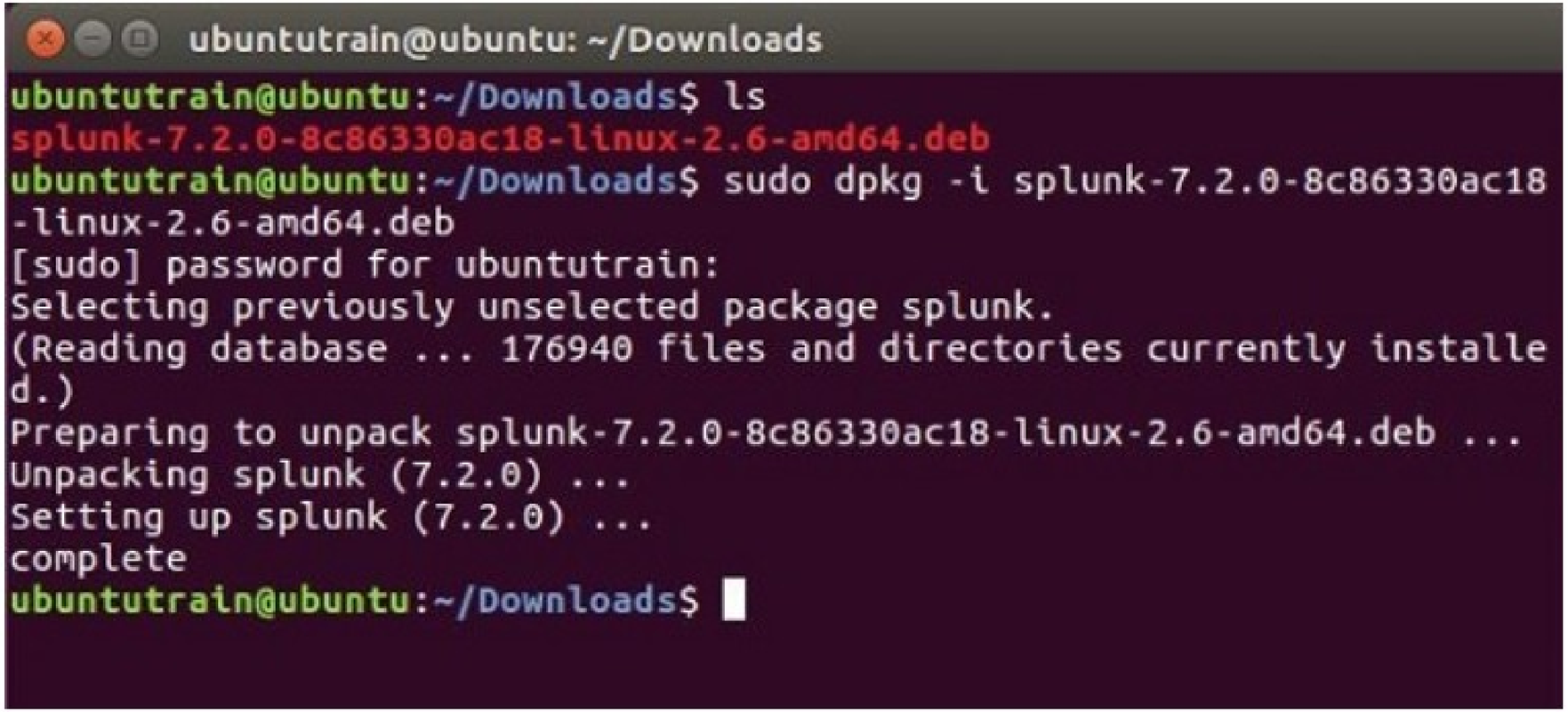
Download the .deb package as shown in the screenshot



Installation and Configuration (Windows/Linux)

Step 2

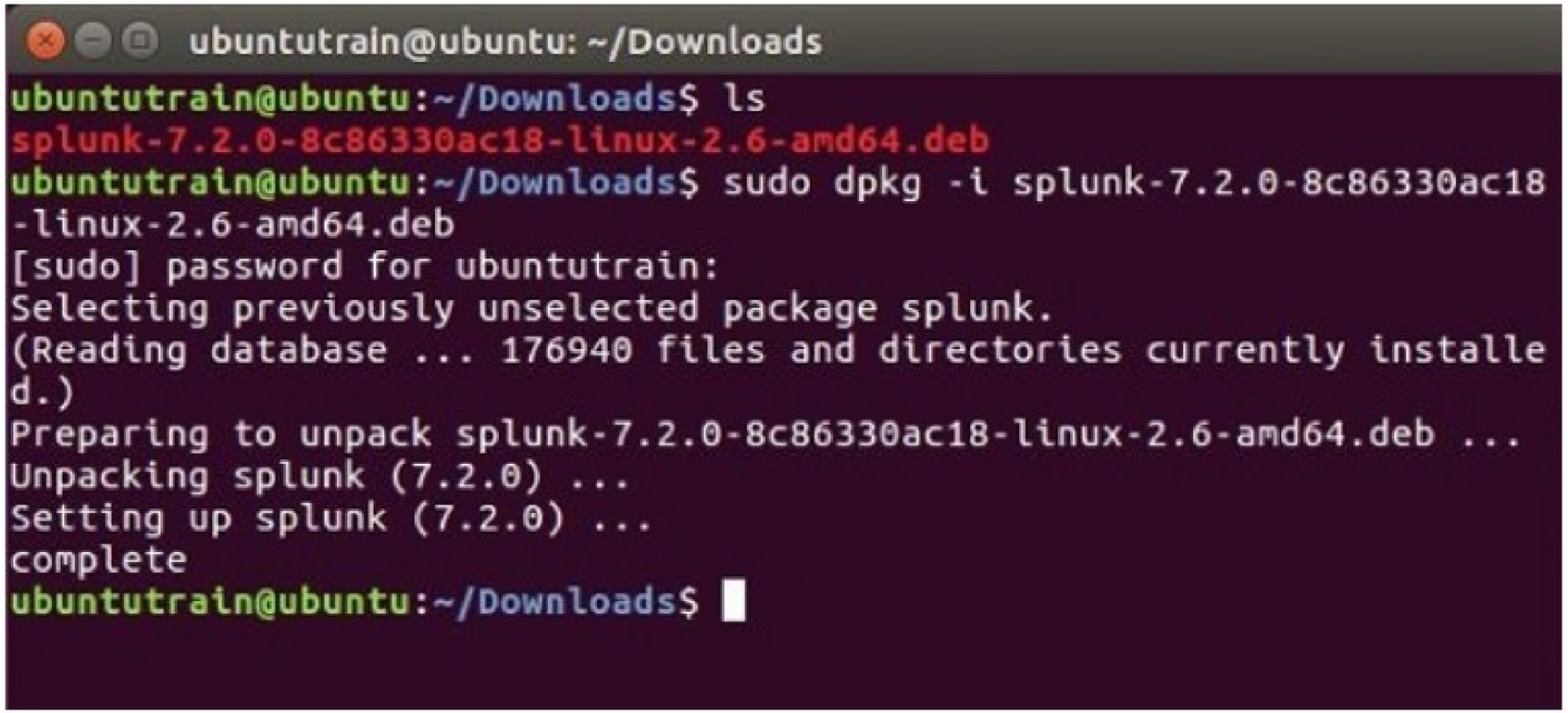
Go to the download directory and install Splunk using the downloaded package.



Installation and Configuration (Windows/Linux)

Step 3

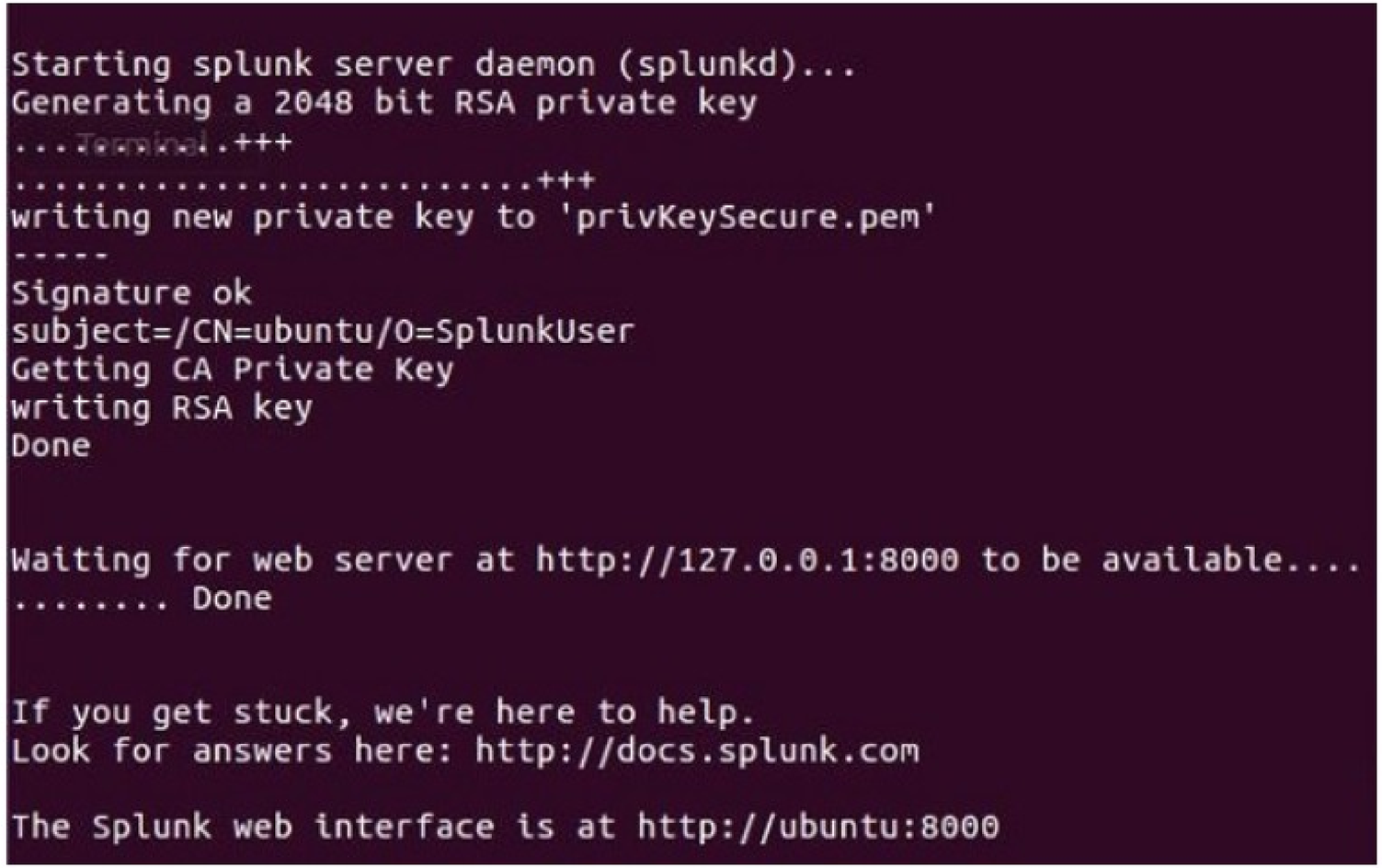
Next you can start Splunk by using the following command with accept license argument. It will ask for administrator user name and password which you should provide and remember.



Installation and Configuration (Windows/Linux)

Step 4

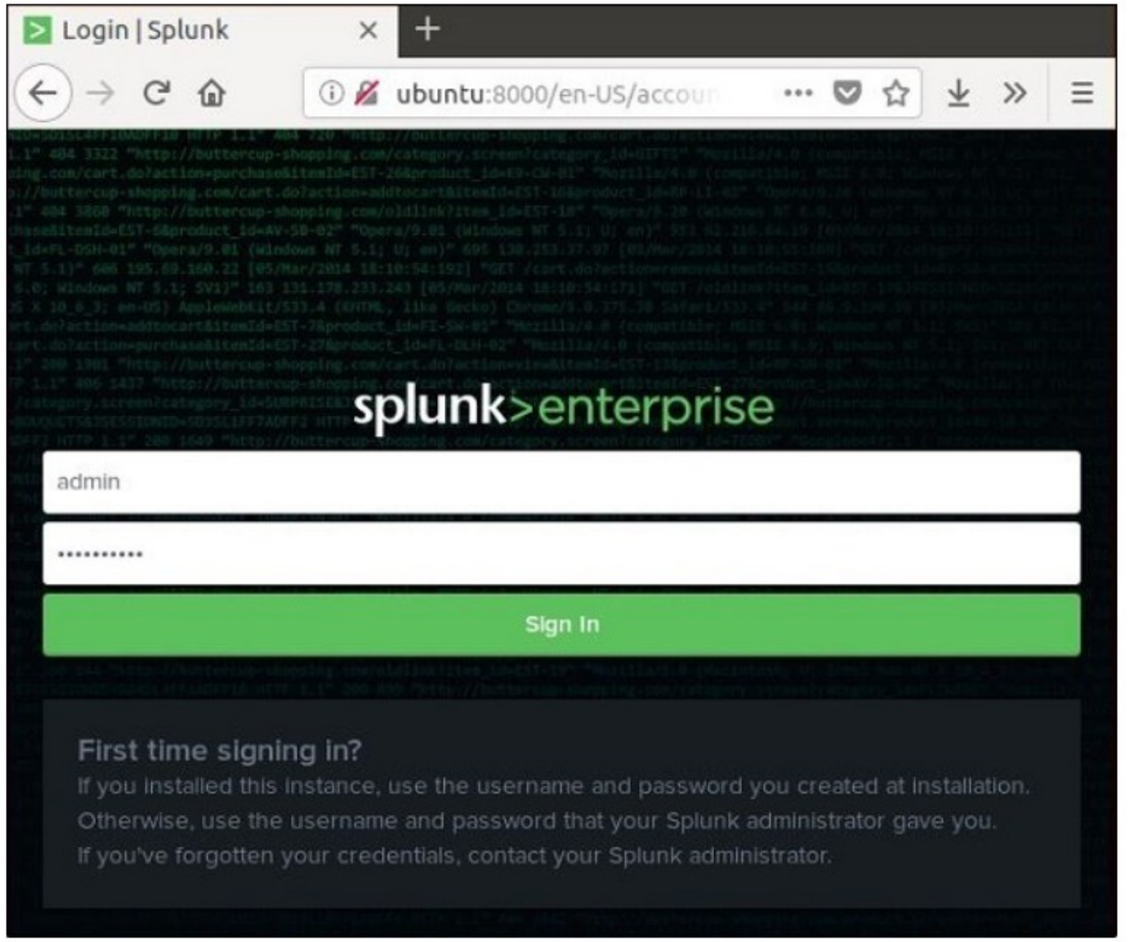
The Splunk server starts and mentions the URL where the Splunk interface can be accessed.



Installation and Configuration (Windows/Linux)

Step 5

Now, you can access the Splunk URL and enter the admin user ID and password created in step 3.

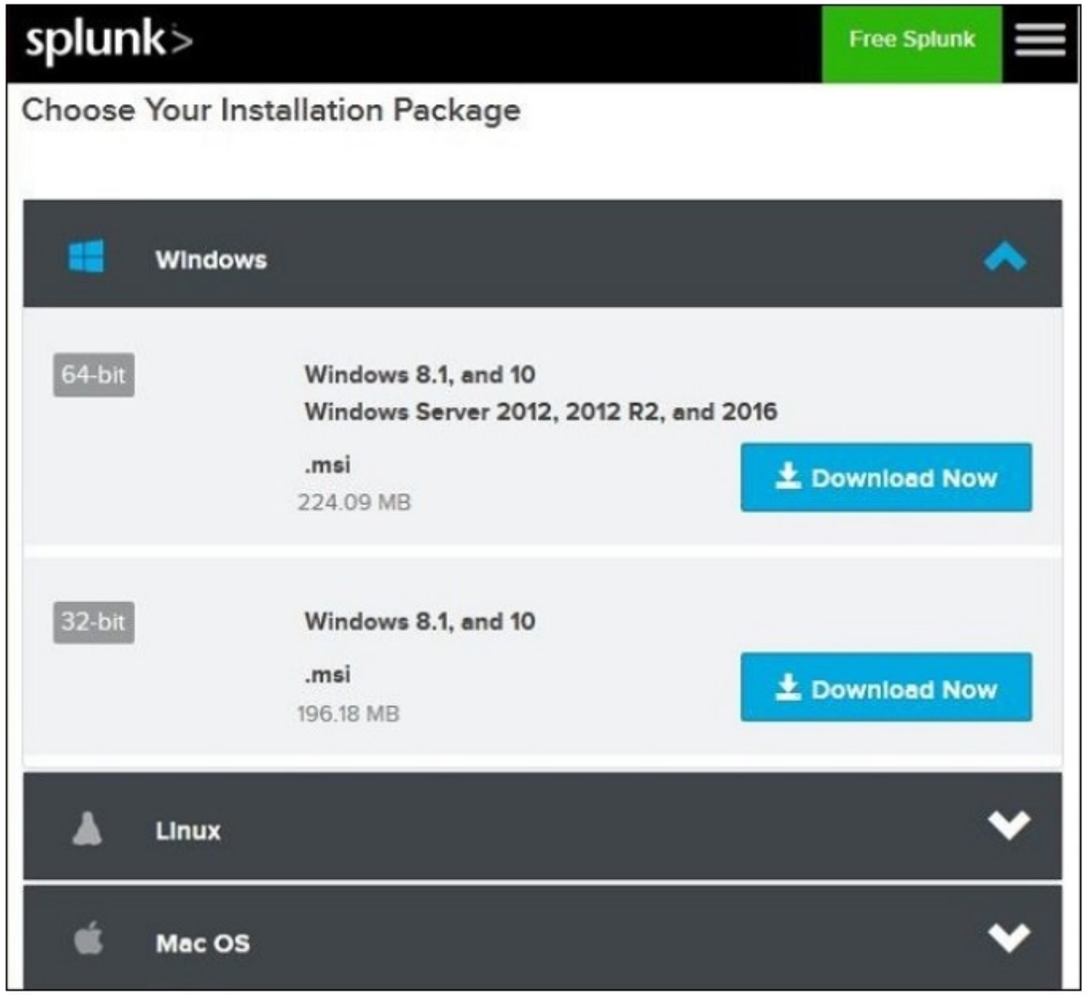


Installation and Configuration (Windows/Linux)

Windows Versions

The windows version is available as a msi installer as shown in the image –

Double clicking on the msi installer installs the Windows version in a straight forward process.



Installation and Configuration (Windows/Linux)

The two important steps where we must make the right choice for successful installation are as follows.

Step 1

As we are installing it on a local system, choose the local system option as given –

Installation and Configuration (Windows/Linux)

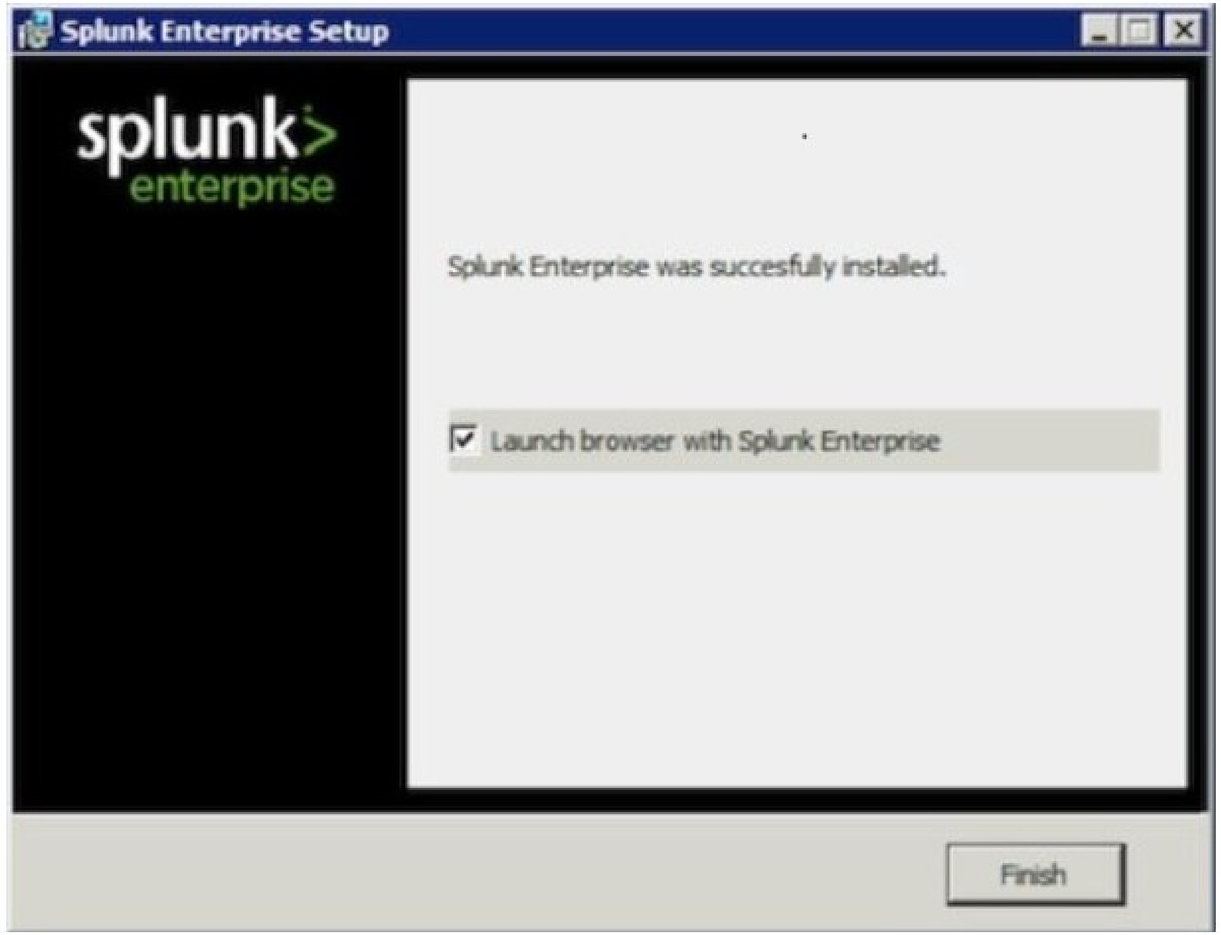
Step 2

Enter the password for the administrator and remember it, as it will be used in the future configurations.

Installation and Configuration (Windows/Linux)

Step 3

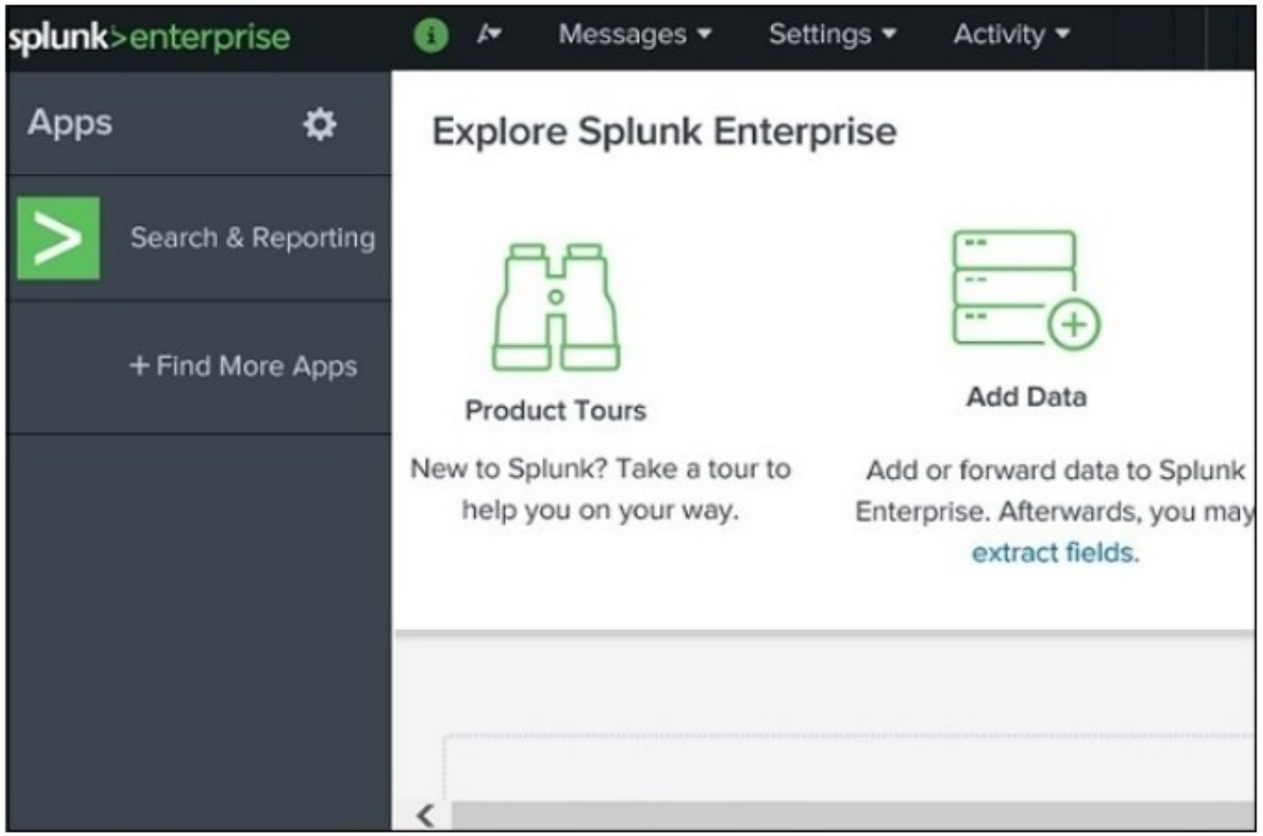
In the final step, we see that Splunk is successfully installed and it can be launched from the web browser.



Installation and Configuration (Windows/Linux)

Step 4

Now, open the browser and enter the given url, http://localhost:8000, and login to the Splunk using the admin user ID and password.



Splunk licensing

- Splunk licensing refers to the system that governs how organizations can use Splunk software and its capabilities based on the volume of data they index or search within a specific time period.
- Splunk's licensing model is primarily designed to ensure compliance with usage limits and provide flexibility for organizations with varying data needs.

Splunk licensing

- Splunk Enterprise takes in data from sources that you designate and processes it so that you can analyze it. This process is called indexing. For information about the indexing process, see How Splunk software handles your data in Getting Data In.
- Splunk Enterprise licenses specify how much data you can index per calendar day (from midnight to midnight by the clock on the license manager).
- Any Splunk Enterprise instance that performs indexing must be licensed to do so. You can either run a standalone indexer with a license installed locally, or you can configure one of your Splunk Enterprise instances as a license manager and set up a license pool from which other indexers, configured as license peer.

Splunk licensing

There are a few types of licenses, such as:

- The Enterprise license enables all Enterprise features, such as authentication and distributed search. As of Splunk Enterprise 6.5.0, new Enterprise licenses are no-enforcement licenses.
- The Free license allows for a limited indexing volume and disables some features, including authentication.
- The Forwarder license allows you to forward data, but not index data, and enables local authentication only.
- The Beta license typically enables Enterprise features, but is restricted to Splunk Beta releases.
- A license for a premium app is used in conjunction with an Enterprise or Cloud license to access the functionality of an app.

Thank You!