
# Splunk Cloud Quickstart Deployment

August 2020

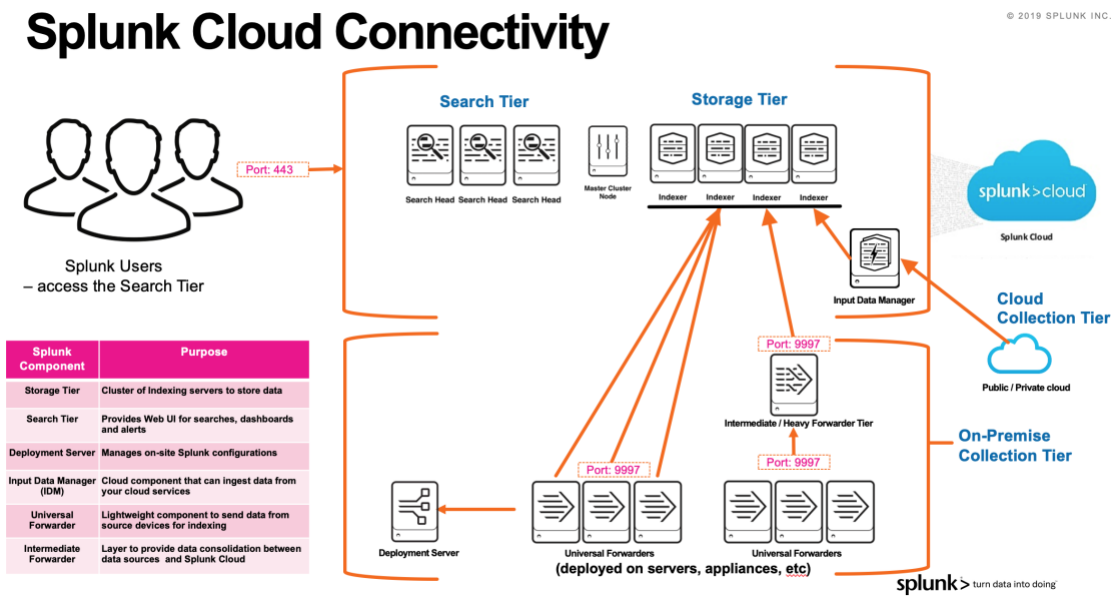
# Introduction

This document provides a high-level introduction to the components and concepts of the Splunk Cloud architecture and the steps needed for a typical Splunk Cloud quick-start deployment.

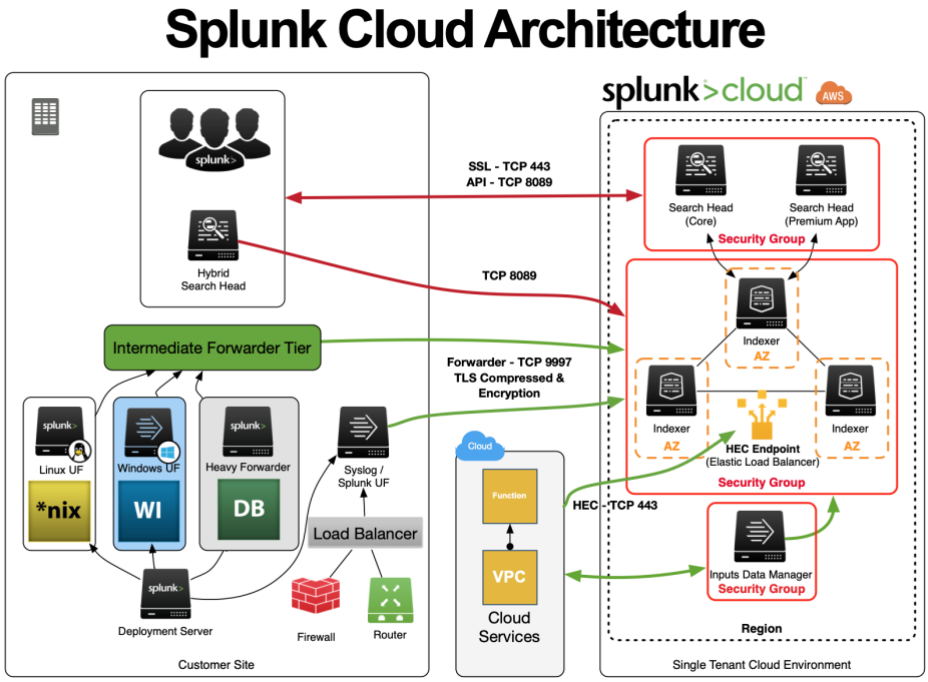
## Splunk Cloud Architecture

Splunk Cloud provides a high availability, scalable platform for Splunk deployment. All that is needed in addition to this, are the components to securely, reliably and efficiently deliver your corporate data into the Splunk cloud. The best delivery approach for your situation is governed by the nature, location and sources of the data, and also by the rules and requirements you may have within your infrastructure.

The following diagram shows the typical components that can be found in a Splunk Cloud deployment, and the connectivity that is needed between your organisation and the cloud.




| Splunk Component | Purpose |
|---|---|
| Storage Tier | Cluster of Indexing servers to store data |
| Search Tier | Provides Web UI for searches, dashboards and alerts |
| Deployment Server | Manages on-site Splunk configurations |
| Input Data Manager (IDM) | Cloud component that can ingest data from your cloud services |
| Universal Forwarder | Lightweight component to send data from source devices for indexing |
| Intermediate Forwarder | Layer to provide data consolidation between data sources and Splunk Cloud |




*Data Collection Considerations*

The data collection tier takes data from your on-premise and cloud systems and devices, and sends it to the Splunk Storage tier, where it is then available for Searching.

The key component for data collection is the specialised and lightweight Universal Forwarder application which needs to be deployed to each device within your organisation. Whilst the initial deployment of this application is typically managed using your own SCCM tools, the configuration of the Splunk estate is best managed and controlled using one or more Splunk Deployment servers

The simplest approach is for each Universal Forwarder to send its data directly to the Splunk Cloud storage tier – the indexers. In some cases, there may be security or network considerations which require the number of network egress points to be minimised, so in such cases a further layer of on-premise Intermediate Forwarders may be used. These can all be managed using the Splunk Deployment Server.

For some specific data sources it is not possible or sensible to deploy a Universal Forwarder onto it to collect the data; instead another Splunk instance may be configured to either pull or receive the broadcast data. This type of instance is termed a Heavy Forwarder and can receive and then forward the data to the indexing tier on behalf of the source system.

Where data from Cloud service providers such as AWS, Azure or GCP is to be ingested, this can be done either via an on-premise route using a Heavy Forwarder, or more efficiently it can be ingested directly into Splunk Cloud using the Input Data Manager component.

## Splunk Cloud Quickstart Approach

To implement a successful Splunk Cloud deployment requires a standard set of discovery, deployment and configuration activities to be undertaken as part of an agreed implementation plan.

The first activity is to create and agree the detailed plan based on the specific use cases, data sources and infrastructure for your organisation. The key steps to achieving this are:

- Identify and prioritise the Splunk Use Cases that you would like to have
    - Identify the use cases that you require from Splunk, and the priority order for delivery of these
    - Review Splunk Apps such as Splunk Security Essentials, Infosec App that provide common out-of-the box use-cases
    - Identify user groups and roles that will access data and use cases
    - Identify actions and operations that may arise from use cases within Splunk:
        - Dashboards to show current status
        - Reports that can be run on-demand
        - Alerts – action resulting from detection of scenarios. sms, email or integration with ticketing and incident management systems
- Identify and document the data sources within your organisation
    - Create a Data Source Register or undertake a Data Source Assessment
    - Classify data sources by Type
    - Identify security requirements for each data source
    - Identify daily storage volumes and retention periods
    - Prioritise the data sources in terms of implementation sequence
- Design the data collection tier
    - Agree a high-level architecture and create a High Level Design (HLD) document to capture the overall solution and the requirements it is meeting
    - Incorporate the required end-to-end availability, reliability and security needs within the HLD
    - Identify infrastructure changes needed to support the required data collection and routing
- Define mutual implementation plan including:
    - Provision of on-premise environments for Splunk servers
    - Installation and configuration of on-premise Splunk components
    - Deployment of Splunk Universal Forwarders onto infrastructure
    - Provision of Splunk cloud environment and the agreed Splunk apps and configurations
    - Customer change control processes for e.g. network/firewall routes to allow UF delivery of data to Splunk cloud
    - Data onboarding for each data source in priority order
    - Testing of each use case and data source

- Training and Knowledge transfer

# Additional Considerations

## Syslog Ingestion

A common requirement for a Splunk Enterprise deployment is to ingest *syslog* format log data from an open-source syslog consolidation server such as *syslog-ng* or *rsyslog*.

In these cases, a Splunk Universal Forwarder will be installed on the *syslog-ng* server(s) instead of being installed on the source servers or appliances that generated the log information. It will forward all of the log data to Splunk.

Another scenario where a syslog server is needed is where devices distribute their log information using UDP data, which is a potentially unreliable fire-and-forget protocol. If this data were sent directly to Splunk and the Splunk server was unavailable for any reason, the record could potentially be lost. To avoid this, the syslog server is implemented with high availability, to receive and store syslog data. The Splunk UF then forwards data to Splunk, using a guaranteed delivery protocol.

Splunk provide an out-of-the-box *syslog-ng* Docker configuration named *Splunk Connect For Syslog (SCFS)* that can be deployed in a Docker container. If Docker is not used by the customer, the Splunk configuration files from *SCFS* can be used with the on-premise *syslog-ng* to implement a best-practice configuration.

Commercial versions of syslog-ng are also available, if a fully 3rd-party supported software stack is required.