Splunk On-Premise Quickstart Deployment Architecture

September 2020

Introduction

This document provides a high-level introduction to Splunk terminology, operation and considerations for on-premise deployment, to allow basic architecture options to be discussed and agreed with a customer.

General Requirements for Splunk

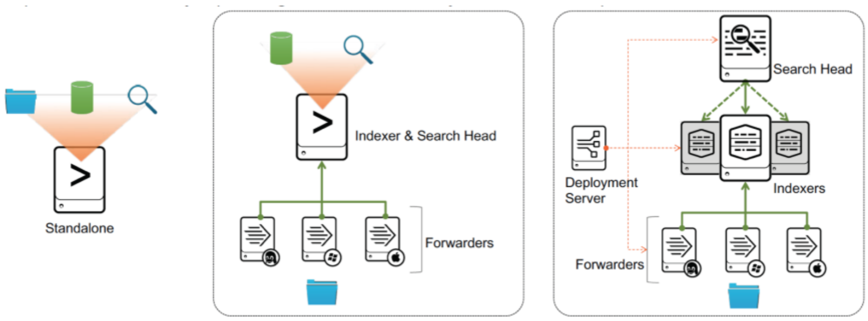
Splunk architecture

Your Splunk system should be implemented in-line with Splunk's recommended architecture patterns to ensure that you get the best performance, scalability and value from your solution from day one.

The range of configuration options and approaches are available online at https://www.splunk.com/pdfs/technical-briefs/splunk-validated-architectures.pdf, however this document provides a distilled set of considerations for a typical new Splunk user with low to moderate data volumes.

Scalability considerations

Splunk installations can be architected and scaled to range from a single standalone server to a multi-clustered, tiered configuration split across global data centres. This is supported by separating the different types of Splunk functionality across different tiers. A single standalone server could be used for small solutions (under 50 GB/day) however it then becomes difficult to scale from there as your Splunk data volumes and use cases increase.



Standalone server – quick and easy configuration for upto 50GB/day and small number of users. However this is difficult to scale, data may be lost if the server is down or too busy, and end-user performance competes with Splunk processing for CPU cycles. Hence the need to consider a multi-server installation.

The three tiers of functionality within Splunk are:

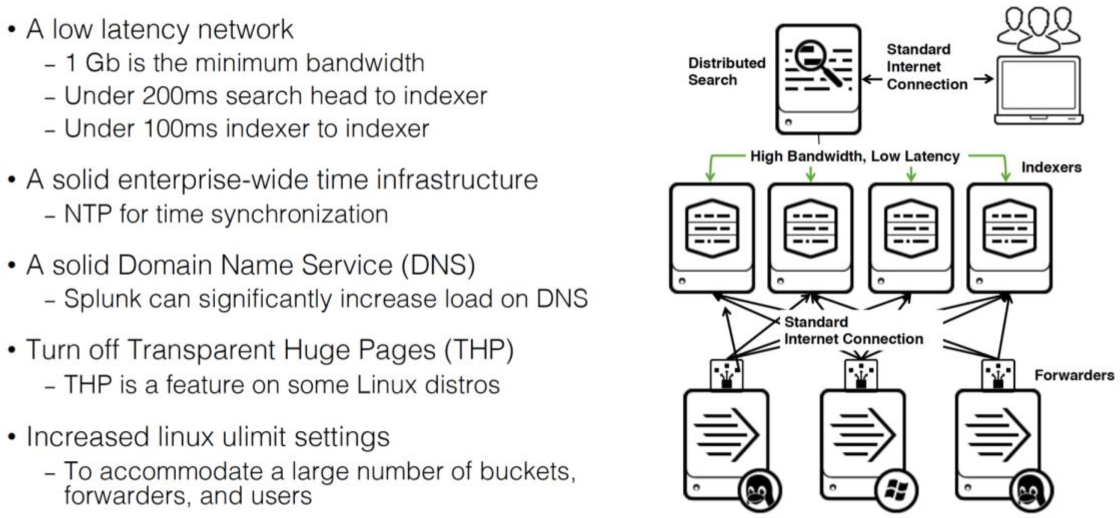
- FORWARDERS. These are Splunk components installed on your infrastructure, that gather and send data from that instance to the INDEXER tier. Typically, a lightweight Universal Forwarder (UF) is installed on each server/desktop/device that is to be monitored. In some cases, additional installations of FORWARDERS may be needed on further Splunk servers, to collate and optimise data delivery across forwarders to the Indexer tier.
- INDEXERS. This is the data storage tier, that receives, indexes and stores data and supports the SEARCH tier. The indexer tier consists of one or more instances, sized to handle the data storage workload. Additional instances can be added to form a cluster that provides redundancy and high availability of the data.
- SEARCH Heads. The SEARCH tier provides the front-end UI – reports, dashboards, alerts. Again, this tier needs to be sized and scaled according to the number of users, number of reports/use cases and taking into account availability/DR requirements for the UI.

In addition to these three tiers, a further Splunk instance known as a Deployment Server is used to simplify the management of the whole Splunk environment. This ensures that the configuration can be easily managed, changed and scaled as needed across all Splunk instances. The DS manages each tier:

- Forwarders – these are configured to 'call home' to the DS periodically to get the latest configuration information.
- Indexers – the index and index cluster information is managed from the DS
- Search Heads – the Search head and cluster configuration is managed from the DS

General environment requirements

A Splunk installation requires some specific infrastructure capabilities to be in place to avoid common potential pitfalls. Linux is the recommended OS, but Windows 64-bit CPU servers can be used.



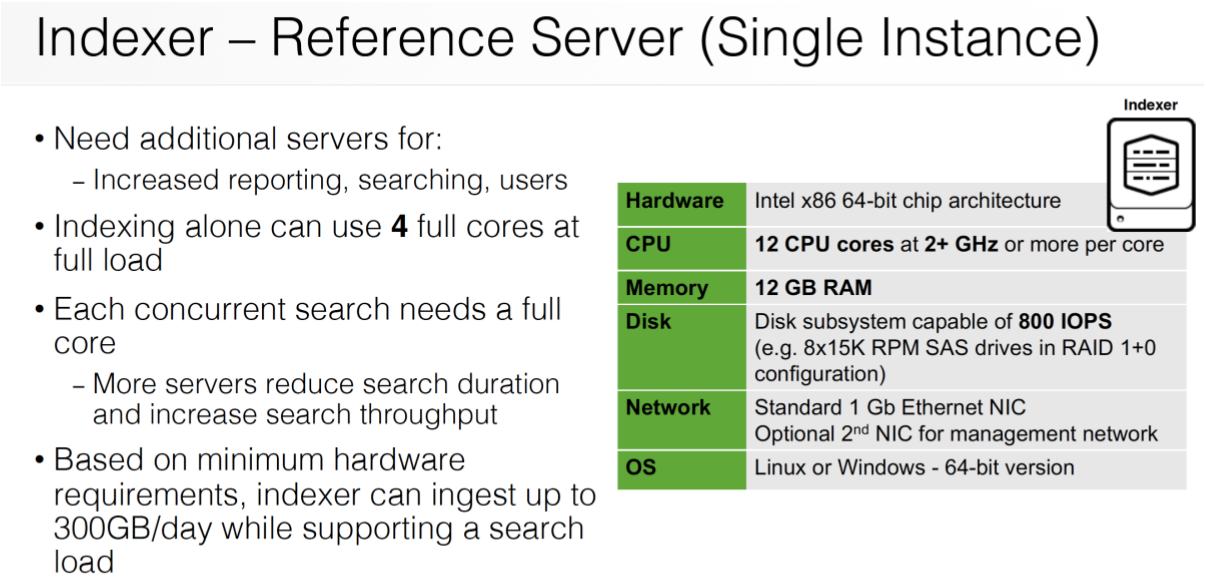
- A low latency network
  - 1 Gb is the minimum bandwidth
  - Under 200ms search head to indexer
  - Under 100ms indexer to indexer
- A solid enterprise-wide time infrastructure
  - NTP for time synchronization
- A solid Domain Name Service (DNS)
  - Splunk can significantly increase load on DNS
- Turn off Transparent Huge Pages (THP)
  - THP is a feature on some Linux distros
- Increased linux ulimit settings
  - To accommodate a large number of buckets, forwarders, and users

Distributed Search
Standard Internet Connection
High Bandwidth, Low Latency
Indexers
Standard Internet Connection
Forwarders

Splunk Server Recommended Specifications

These are the recommended specifications for the individual servers that will be used within your Splunk installation to achieve the best performance throughput as data volumes grow.

Indexer minimum specification



Indexer – Reference Server (Single Instance)

- Need additional servers for:
  - Increased reporting, searching, users
- Indexing alone can use **4** full cores at full load
- Each concurrent search needs a full core
  - More servers reduce search duration and increase search throughput
- Based on minimum hardware requirements, indexer can ingest up to 300GB/day while supporting a search load

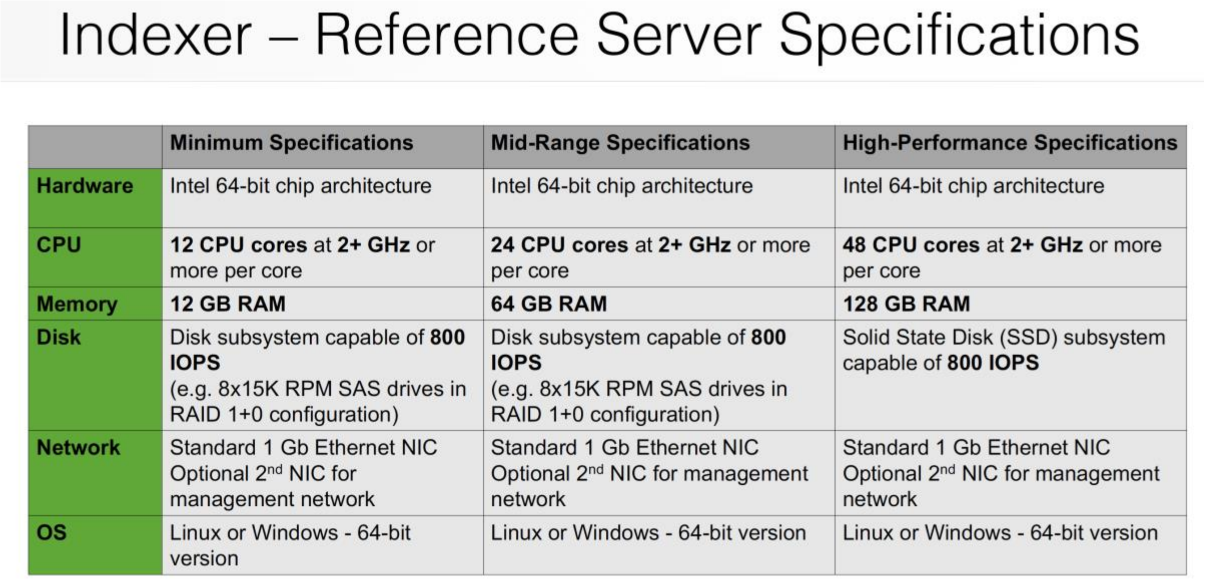
Hardware	Intel x86 64-bit chip architecture
CPU	**12 CPU cores** at **2+ GHz** or more per core
Memory	**12 GB RAM**
Disk	Disk subsystem capable of **800 IOPS** (e.g. 8x15K RPM SAS drives in RAID 1+0 configuration)
Network	Standard 1 Gb Ethernet NIC Optional 2nd NIC for management network
OS	Linux or Windows - 64-bit version

To avoid the loss of any data, INDEXERs need to be clustered for redundancy and availability

- data is replicated across the instances, and also across Data Centres for DR cases
- capacity of each instance depends on the spec of the individual servers: CPU + data storage
- need to size each server based on overall data volumes and number of data sources
- scaled for availability/reliability requirements

Indexer example specifications

The Indexer servers should each be of the same specification so that they can operate equally within the cluster. The actual size and specification of each individual server should be matched to the storage throughput that needs to be supported.



Indexer – Reference Server Specifications

	Minimum Specifications	Mid-Range Specifications	High-Performance Specifications
Hardware	Intel 64-bit chip architecture	Intel 64-bit chip architecture	Intel 64-bit chip architecture
CPU	**12 CPU cores** at **2+ GHz** or more per core	**24 CPU cores** at **2+ GHz** or more per core	**48 CPU cores** at **2+ GHz** or more per core
Memory	**12 GB RAM**	**64 GB RAM**	**128 GB RAM**
Disk	Disk subsystem capable of **800 IOPS** (e.g. 8x15K RPM SAS drives in RAID 1+0 configuration)	Disk subsystem capable of **800 IOPS** (e.g. 8x15K RPM SAS drives in RAID 1+0 configuration)	Solid State Disk (SSD) subsystem capable of **800 IOPS**
Network	Standard 1 Gb Ethernet NIC Optional 2nd NIC for management network	Standard 1 Gb Ethernet NIC Optional 2nd NIC for management network	Standard 1 Gb Ethernet NIC Optional 2nd NIC for management network
OS	Linux or Windows - 64-bit version	Linux or Windows - 64-bit version	Linux or Windows - 64-bit version

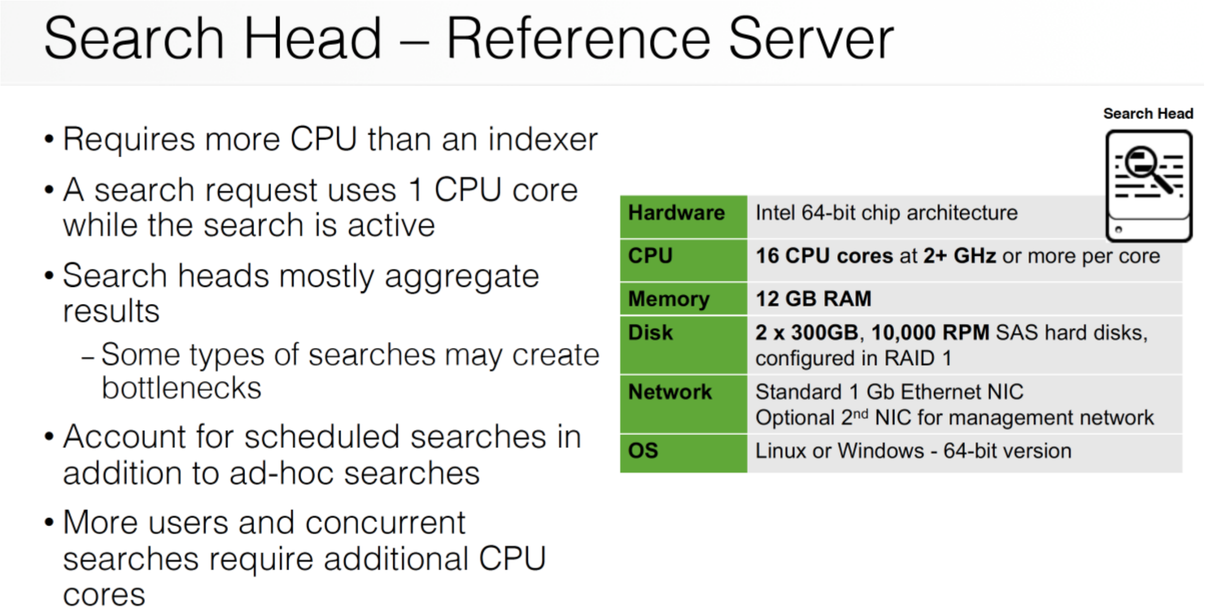
Cluster Master (CM) Specification

When an Index cluster is used, an additional lightweight server is needed to provide co-ordination across the cluster. The specification for a CM server supporting a 3-indexer cluster would be as follows:

- Intel 64-bit chip architecture
- 4 CPU cores at 2+GHz
- 8GB RAM

Search Head Reference Specification

The SEARCH tier provides front-end UI – reports, dashboards, alerts. Again, this tier needs to be sized and scaled according to number of users, number of reports/use cases and taking into account availability and DR requirements.



Search Head – Reference Server

- Requires more CPU than an indexer
- A search request uses 1 CPU core while the search is active
- Search heads mostly aggregate results
  - Some types of searches may create bottlenecks
- Account for scheduled searches in addition to ad-hoc searches
- More users and concurrent searches require additional CPU cores

Search Head

Hardware	Intel 64-bit chip architecture
CPU	**16 CPU cores** at **2+ GHz** or more per core
Memory	**12 GB RAM**
Disk	**2 x 300GB**, **10,000 RPM** SAS hard disks, configured in RAID 1
Network	Standard 1 Gb Ethernet NIC Optional 2nd NIC for management network
OS	Linux or Windows - 64-bit version

Deployment Server (DS) Specification

The specification for a DS server supporting an entry-level Splunk configuration with 1-200 forwarders to be managed would be as follows:

- Intel 64-bit chip architecture
- 4 CPU cores at 2+GHz
- 8GB RAM

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.

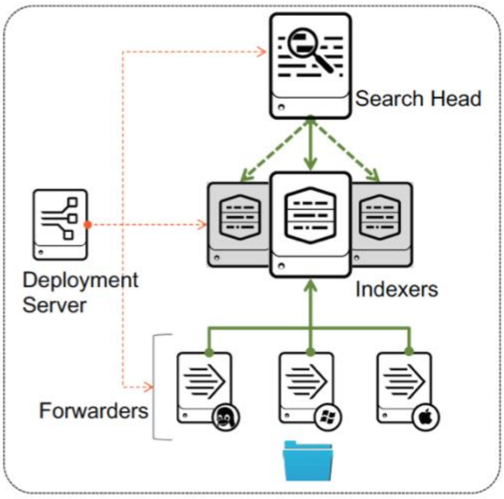
Splunk provide an out-of-the-box *syslog-ng* Docker configuration named *Splunk Connect For Syslog (SCFS)* that can be deployed in a Docker container. If Docker is not used by the customer, the Splunk configuration files from *SCFS* can be used with the on-premise *syslog-ng* to implement a best-practice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

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

Recommended Architecture for High-Availability Low Volume (<100GB / day) Basic Deployment

The recommended architecture for a Splunk RAP deployment is as shown below, consisting of:

- 3 x Indexer instances configured as a cluster to ensure data availability
- 1 x Cluster Master (not shown in diagram)
- 1 x Search Head to support use cases for up-to 10 users
- 1 x Deployment Server to manage overall configuration



The standard specification for an Indexer server can handle up-to 300GB/day. For a situation with an overall daily ingestion of upto 200GB, spread across a cluster of indexers, the specification of these servers can be reduced.

Similarly if there a low number of Search users and/or search use-cases, the Search head can be reduced in specification.

A suitable specification for an entry level deployment would be:

- Search Head: 8 CPU cores, 16GB RAM
- Each Indexer (3 off): 4 CPU cores, 8GB RAM
- Cluster Master: 4 CPU cores, 8GB RAM
- Deployment Server: 4 CPU cores, 8GB RAM

In addition to the Splunk servers, *syslog-ng* syslog server may be needed. A 4 CPU cores, 8GB RAM server is sufficient for supporting an entry level Splunk deployment. This will have open source *syslog-ng* installed to collate syslog data from the agreed data sources (as per agreed project data source register), and a Universal Forwarder to send to the Indexer cluster.

The Splunk data that is indexed is aged over time and the retention period for data depends on the customer's storage capacity and the need for searching on older data.

As a guideline the estimated storage requirement for 25GB/day ingestion stored on three indexers is as follows

- 30 days : 290 GB per Indexer
- 60 days : 580 GB per Indexer
- 90 days : 870 GB per Indexer