Deploying Splunk Enterprise on Google Cloud Platform

In this document we provide deployment and configuration recommendations, including references to Google Cloud Platform virtual machines, performance considerations which cover public information about Google Cloud Platform capabilities. There are sections in this document on deployment components, performance considerations, deployment guidelines, examples based on Splunk Verified Architectures, and guidelines for using SmartStore. We also provide references to a list of Google-focused Splunk-offerings in order to help your customer better integrate with Google Cloud Platform and Google services. These recommendations are for Splunk customers running Splunk Bring Your Own License (BYOL) on virtual machines as an intermediate step on their cloud journeys and migrating to Splunk Cloud on Google Cloud Platform.

Splunk Enterprise

Splunk® Enterprise provides the leading platform to turn data into doing. Splunk software searches, monitors, analyzes and visualizes machine-generated big data from websites, applications, servers, networks, sensors and mobile devices. More than 18,000 organizations use Splunk to deepen business and customer understanding, mitigate cybersecurity risk, improve service performance and reduce costs. Splunk Enterprise indexes machine data in real time, enabling multiple roles across the organization — from system administrators to business analysts — to rapidly gain insight from the massive amounts of machine data generated by your environment. Adopting a cloud strategy increases agility, reduces costs, decreases time to market and empowers innovation. Splunk Enterprise is perfect for deploying in a cloud environment, offering enterprise-grade availability and scalability to support the collection of hundreds of terabytes of data per day from workloads residing on-premises, in the cloud or across hybrid environments. This document covers guidelines for deploying Splunk Enterprise on Google Cloud Platform.

Deployment components

Splunk Deployment Components

A typical Splunk deployment includes the following components: Splunk forwarders, indexers and search heads. While all major Splunk components can be run from a single installation on a single instance, they can also run independently from within different instances. This enables Splunk Enterprise to scale horizontally through additional indexers and search heads that support extra workloads.

Splunk software can be installed to your choice of hardware (physical, cloud or virtual) and operating system within minutes. The package is available for download for most operating systems. Depending on the deployment infrastructure, the proper amount of resources must be allocated per component type.

Forwarders transfer data between Splunk Enterprise instances and an indexer, another forwarder or a third-party system. Most forwarders are lightweight instances with minimal resource utilization, allowing them to reside easily on the machine generating the data.

Indexers write the data to a storage device and perform searches on the data. Indexers can be resource-intensive and require sufficient I/O and CPU allocation.

Search heads handle search management functions, directing search requests to a set of indexers and then merging the results back to the user. Search heads require sufficient CPU and memory allocation. In order to enable search and index performance, system resources and bandwidth must be budgeted depending on the total volume of data being indexed and the number of active concurrent searches (scheduled or otherwise) at any time. In addition to rapidly writing data to disk, indexers perform much of the work involved in running searches including reading data off a disk, decompressing it, extracting knowledge and

reporting. Since indexers incur most of the workload, increases in indexing volume should occur in tandem with increases in indexer instances. Scaling out to additional indexers will distribute the load of increased data volumes, resulting in reduced contention for resources during searches and accelerated search performance. Common Google Cloud Platform deployments leverage a combination of forwarders and network streams to send data to the Splunk indexer(s). While a forwarder is not required to gather data from the source, they do provide certain benefits such as flexibility, load balancing and reliability. Using a syslog output from a data source or a file mount is also a common form of getting data into the Splunk indexer. Additionally, modular inputs, which are extensions to Splunk Enterprise that define a custom data input, and HTTP Event Collector (HEC), a highly efficient and secure mechanism used to send high volumes of data directly to the Splunk platform, can be used to collect data from various API sources. Other Splunk components include the deployment server (used for configuration publishing), license master (used for license management) and master node (manages index replication) — which we will collectively refer to as Admin Servers below.

Performance considerations

There are several performance factors to consider when deploying Splunk software on Google Cloud Platform. The most important considerations revolve around instance (compute) and storage type. For more information on configurations, please refer to Splunk Capacity Planning Manual.

Compute

For Splunk deployments we recommend using regular compute instances with minimum VM instance requirements (corresponding to n2-standard-8): 8vCPU and 16GB of RAM.

Splunk is well-suited for Google Cloud Platform as it scales horizontally, and adding Splunk instances can give you better performance and capacity depending on data volume requirements.

Storage

Splunk recommends using root persistent disks to store Splunk configurations, OS and indexed data. For clustered deployments, local SSDs are an alternative. Below are the key characteristics of persistent disks (PD) and non-persistent local SSDs. Persistent disks (PD) are highly available, reliable and durable, and can grow to 64TB per disk and you can add up to 257TB overall per VM. There's two types of persistent storage available on Google Cloud Platform: standard hard disk (PD-HDD) and solid-state (PD-SDD). PD-SDD is a better fit for warm and hot data, and PD-HDD is more suitable for cold.

Local SSDs are attached to the server that hosts the VM instance and are limited to 9TB per instance. Local SSDs have a much higher throughput and lower latency than PD-HDD and PD-SSD, but are not persistent. Also, in order to achieve full performance on local SSDs, you need to consider configurations with more vCPU. For example, for 9TB, you need at least 24 vCPU cores.

When planning your storage requirements for the indexes, take into account that Splunk software will compress the data. Typical installations experience an effective 2:1 compression ratio when storing raw data and the associated index and metadata. This means that if you are indexing 100GB per day, you should expect a daily utilization of approximately 50GB. The number and size of your local storage volumes should be based on your retention requirements and expected daily indexing volume. We also recommend using the Splunk SmartStore option in order to optimize your costs — please see deployment guidelines and "Using Splunk SmartStore on Google Cloud Platform" for more details.

Deployment guidelines and examples

The tables below describe general guidelines for mapping instances to Splunk workloads. Best practices for architecting and sizing should still be considered when referencing these guidelines. It is important to remember that overall Splunk load is composed of both indexing and searching. For all configurations we recommend considering n2-standard-8 or n2-standard-16 as baseline VM, but please consider updating it to n2-highcpu-8/16 respectively, in case you observe a lack of CPU resources. Please use these configurations as a starting point for your deployments and consider changing these VMs to different SKUs based on your organization's performance needs and other requirements. This brief provides an introduction to various options of Splunk deployments. Please refer to Splunk Validated Architectures for more information.

Instance Type	Data Volume
n2-standard-16	Up to 100 Gb
n2-standard-32	100 to 250 Gb

Table 1: Indexers

Instance Type	Concurrent Users
n2-standard-16	Up to 100
n2-standard-32	100 to 250

Table 2: Search Heads

Deployment Server

For deployment servers, you may use n2-standard-8 as default and n2-standard-16 for better performance. For storage, you can use PD-HDD as the most cost-effective storage option.

These configurations might be used as baselines. In real-life environments there are many factors which can impact performance and should be considered

when you choose VM SKU. Please refer to this introduction to capacity planning for Splunk Enterprise and this guide on how concurrent users and searches impact performance.

Below you will find three types of configurations with recommendations as baseline configurations. Please refer to Splunk Validated Architectures for more options; we also use reference names for Splunk Validated Architectures (SVA) from this document in our brief.

Small-scale deployment (SVA S1)

This topology is typically used for smaller, non-business critical use-cases (often departmental in nature). Appropriate use cases include data onboarding test environments, small DevOps use cases, application test and integration environments, and other similar scenarios. The primary benefits of this topology include easy manageability, good search performance for smaller data volumes and a fixed total cost of ownership. Multiple, independent single-instance deployments can be managed by a single management tier, as needed. In this configuration we have one VM which is the search head and indexer at the same time.

This configuration provides you with a very cost-effective solution if your environment meets all of the following criteria:

• You do not have any requirements to provide high availability or automatic disaster recovery for your Splunk deployment.

• Your daily data ingest is under ~300GB/day.

• You have a small number of users with non-critical search use cases.
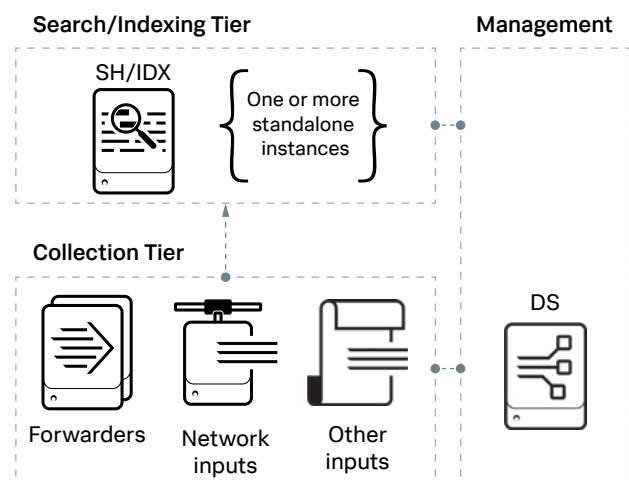
Indexer/Search Head

- VM SKU: 1 x n2-standard-16
- Storage: Local SSD

Management: Deployment Server

- VM SKU: 1 x n2-standard-8
- Storage: PD-SSD

Architecturally, this is a single Splunk instance performing indexing and searching. Data can be sent to this system via Splunk forwarders, local files, remote systems, syslog, HTTP Event Collector, modular inputs and so on. The total size of available storage should be based on your retention requirements and expected daily indexing volume. Using Splunk SmartStore may help you optimize costs for storage options.

## Medium-scale distributed deployment (SVA C1)

This topology introduces indexer clustering in conjunction with an appropriately configured data replication policy. This provides high data availability in case of indexer peer node failure. However, you should be aware that this applies only to the indexing tier and does not protect against search head failure. Multiple independent search heads can be used for availability and capacity reasons, or to run Splunk premium app solutions like ES, although the recommended approach to scale search capacity is to employ search head clustering.

This topology requires an additional Splunk component named the Cluster Manager (CM). The CM is responsible for coordination and enforcement of the configured data replication policy. The CM also serves as the authoritative source for available cluster peers (indexers). Search head configuration is simplified by configuring the CM instead of individual search peers.

You can optionally configure the forwarding tier to discover available indexers via the CM. This simplifies the management of the forwarding tier. Be aware that data is replicated within the cluster in a nondeterministic way. You will not have control over where requested copies of each event are stored. Additionally, while scalability is linear, there are limitations with respect to total cluster size (~50 petabytes of searchable data under ideal conditions). Splunk recommends deploying the Monitoring Console (MC) to monitor the health of your Splunk environment.

Configuration limitations:

- No high availability for the search tier
- Total number of unique buckets in indexer cluster limited to 20MM (V8.x), 40MM total buckets
- No automatic DR capability in case of data center outage
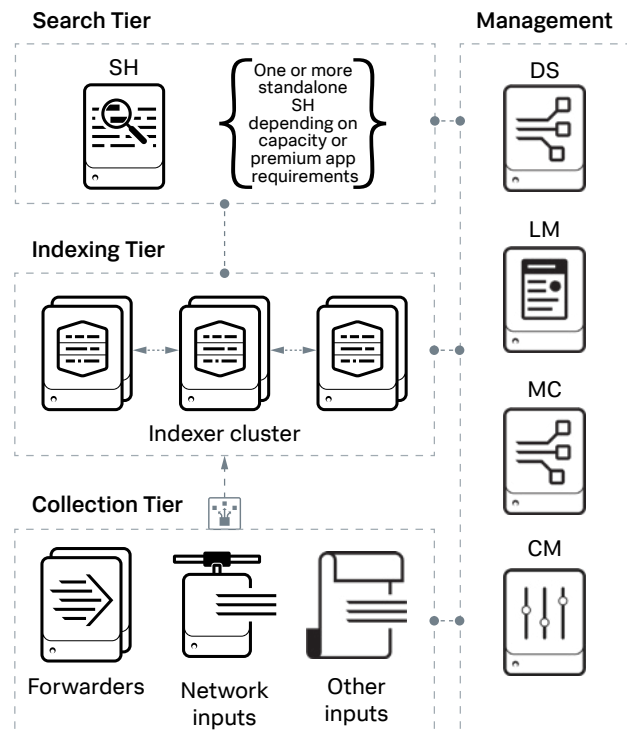
Search Head

- VM SKU: 1 x n2-standard-32
- Storage: Local SSD or PD-SSD

Indexing Tier

- VM SKU: 3 x n2-standard-32
- Storage: Local SSD or PD-SSD

Deployment Server

- VM SKU: n2-standard-8
- Storage: PD-HDD

Monitoring Console/License Master Combined

- VM SKU: n2-standard-8
- Storage: PD-SSD

Cluster Master

- VM SKU: n2-standard-8
- Storage: PD-SSD

Architecturally, this deployment consists of five Splunk instances in a traditional distributed configuration. Three of these instances act as indexers, another acts as the search head, and we also have three nodes in the management layer. The total size of available storage should be based on your retention requirements and expected daily indexing volume. Using Splunk SmartStore may help you optimize costs for storage options.
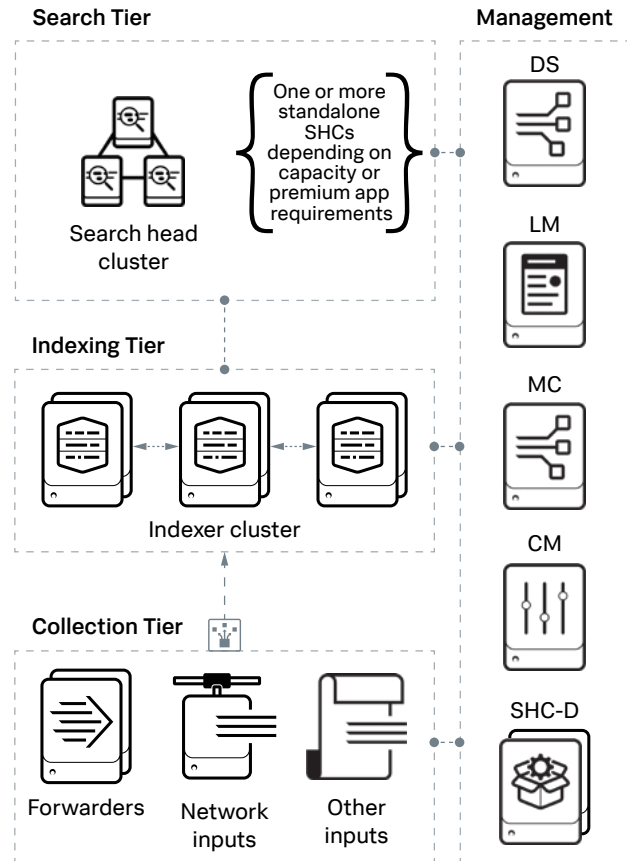
Large-scale distributed deployment (SVA C3)

Search Head Clustering (SHC) adds horizontal scalability and removes the single point of failure from the search tier. A minimum of three search heads are required to implement a SHC.

To manage the SHC configuration, an additional Splunk component called the Search Head Cluster Deployer is required for each SHC. This component is necessary in order to deploy changes to configuration files in the cluster. The Search Head Cluster Deployer has no HA requirements (no runtime role). The SHC provides the mechanism to increase available search capacity beyond what a single search head can provide. Additionally, the SHC allows for scheduled search workload distribution across the cluster. The SHC also provides optimal user failover in case of a search head failure. A network load-balancer that supports sticky sessions is required in front of the SHC members to ensure proper load balancing of users across the cluster.

Configuration limitations:

- There is no disaster recovery capability in case of data center outage.
- Splunk Enterprise Security (ES) requires a dedicated SH/SHC.
- Splunk Professional Services is recommended for implementing ES on a SHC.
- SHC cannot have more than 100 nodes.

Search Tier — Management

DS — LM — MC — CM — SHC-D

Search head cluster — One or more standalone SHCs depending on capacity or premium app requirements

Indexing Tier — Indexer cluster

Collection Tier — Forwarders, Network inputs, Other inputs

Search Head Cluster (SHC)
- VM SKU: 3 or more x n2-standard-32
- Storage: Local SSD or PD-SSD

Indexing Tier
- VM SKU: 3 x n2-standard-32
- Storage: Local SSD or PD-SSD

Deployment Server
- VM SKU: n2-standard-8
- Storage: PD-SSD

Monitoring Console
- VM SKU: n2-standard-8
- Storage: PD-SSD

License Master
- VM SKU: n2-standard-8
- Storage: PD-SSD

Cluster Master
- VM SKU: n2-standard-8
- Storage: PD-SSD

Using Splunk SmartStore on Google Cloud Platform

Splunk supports data storage on compatible object storage for Google Cloud Storage (GCS) as a remote storage device. Using SmartStore may help Splunk customers optimize costs of running Splunk in Google Cloud, including long-term data retention costs, and also improve data security and durability.

SmartStore Cache Sizing

By using SmartStore's caching capability, data will automatically move between VM local storage and Google Cloud Storage. The SmartStore cache is located on VM local volume in order to keep frequently indexed and searched data. The size of VM volume depends on the expected data ingestion rate and number of days included in the searched data. SmartStore cache size can be calculated by assuming a 2:1 compression for stored data on the indexer in the cluster and can be adjusted for search and replication purposes. While most Splunk installations work fine with 30 days of cached data on each indexer, please consider increasing the cache size to 90 days of retained data. For more information, please refer to SmartStore system requirements and information on bucket and indexer clusters.

To provide maximum flexibility, SmartStore also can be configured on a per-index basis, allowing administrators to adjust performance and cost to their specific use cases. Please be sure to follow guidelines on how to deploy multisite indexer clusters with SmartStore. In this brief we will cover Smartstore used for all indexes.

SmartStore Access and Security Configuration

Below we will cover a minimal, easy-to-manage configuration which uses roles to grant access to a single bucket that stores all of the indexed data. For more detailed information on configuring the remote store, please refer to SmartStore on GCS security strategies.

When you configure GCS buckets, the indexer's VM Compute Engine service account must have read, write and delete permissions for the GCS buckets that the indexers are associated with. Provision the buckets to run in the same Google Cloud Platform region as the indexer compute engine instances. Please see Google Cloud documentation on connecting to cloud storage buckets for information on how to create and configure buckets.

SmartStore supports server-side encryption of data-at-rest on GCS. SmartStore supports three encryption schemes: default scheme with Google-managed encryption keys (scheme `gcp-sse-gcp`), customer-supplied encryption keys (scheme `gcp-sse-c`) and customer-manager encryption keys (scheme `gcp-sse-kms`).

Below you may find examples of configuration for different encryption schemes:

Google-managed encryption keys configuration

```
[volume:my_gcs_vol]
storageType = remote
remote.gs.encryption = gcp-sse-gcp
path = gs://your_gcp_test
```

Customer-supplied encryption keys

```
[volume:my_gcs_vol]
storageType = remote
path = gs://your_gcp_test
remote.gs.project_id = your-app-test
remote.gs.encryption = gcp-sse-c
remote.gs.encryption.gcp-sse-c.key_refresh_
interval = 86400
remote.gs.encryption.gcp-sse-c.key_type =
gcp_kms
remote.gs.gcp_kms.locations = us-central1
remote.gs.gcp_kms.key_ring = your_ring
remote.gs.gcp_kms.key = test_key
```

Customer-managed encryption keys

```
[volume:my_gcs_vol]
storageType = remote
path = gs://your_gcp_test
remote.gs.project_id = your-app-test
remote.gs.encryption = gcp-sse-kms
remote.gs.gcp_kms.locations = us-central1
remote.gs.gcp_kms.key_ring = your_ring_us
remote.gs.gcp_kms.key = test_key
```

If you want to use a key associated with Google's cloud storage account, you may skip `gcp_kms` in this configuration

```
[volume:my_gcs_vol_2]
storageType = remote
path = gs://your_gcp_test_2
remote.gs.project_id = your-app-test
remote.gs.encryption = gcp-sse-kms
```

Please note that you configure this attribute on a per-volume basis.

There are several options on how the indexer can obtain its credential. In descending order of precedence, the indexer uses:

- The credential file specified by the `remote.gs.credential_file` setting in `indexes.conf`, if set.
- The credential for the account associated with the `remote.gs.service_account_email` setting in `indexes.conf`, if set.
- The credential for the compute engine's default service account.

The credentials that you use need permission to perform GCS operations. They also need permission to perform GCP Key Management Service (KMS) operations if you are using the `gcp-sse-c` or `gcp-sse-kms` encryption schemes. Choosing an encryption method is a one-time decision and you can't change this later after initial configuration.

For more detailed information on SmartStore configuration please refer to SmartStore on GCS security strategies, the indexes.conf spec file lookup for the `remote.gs` encryption-related settings and Google documentation on configuring cloud storage encryption keys.

Replication and SmartStore

Search and index replication, often referred to as clustering, provide data resilience and redundancy in case of instance or storage failure, and are always recommended for production Splunk deployments.

Though long-term data durability is aided by the use of Google Cloud Storage when using SmartStore on GCP, replication is still required for the most recent indexed data. Search and index replication must be set to the same value when using SmartStore.

References

Splunk Documentation

Splunk Validated Architectures

Splunk Capacity Planning

Configure Google Cloud Storage Remote Store for SmartStore

index.conf Specifications

Buckets and Indexers Clusters

Splunk Offerings for Google Cloud Platform

Splunk Add-On for Google Workspace

Splunk Add-On for Google Chrome

Splunk Observability on Google Cloud

Summary

Splunk's vision and strategy are focused on the cloud journey, but some organizations may need to use a BYOL licensing model from a requirements perspective or as an intermediate step in the migration path to Splunk Cloud. Based on recommendations and considerations mentioned in this document, customers may deploy Splunk Enterprise BYOL on Google Cloud Platform, using the recommended instance sizes and types as a starting point for their configurations. Based on experience and data volume, customers might need to use modified configurations which reflect their needs and cost considerations. Google Cloud Platform provides rich capabilities for optimal combinations of compute, network and storage which meet customer needs as well as provide wide capabilities to easily scale up Splunk deployments.

Access documentation, product updates, hands-on labs and upcoming events on using Splunk Cloud and Google Cloud

splunk>

Learn more: www.splunk.com/asksales **www.splunk.com**