

# DEPLOYING SPLUNK® ENTERPRISE ON GOOGLE CLOUD PLATFORM

Splunk Enterprise provides the leading platform for Operational Intelligence. Splunk software searches, monitors, analyzes and visualizes machine-generated big data from websites, applications, servers, networks, sensors and mobile devices. More than 13,000 organizations use Splunk software 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 enables organizations to increase agility, reduce costs, decrease time to market and empower 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 (GCP).

## 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 by adding additional indexers and search heads to 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, considerations must also be taken to allocate the proper amount of resources per component type.

**Forwarders** forward data between Splunk Enterprise instances, either an indexer, another forwarder or to 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, budgeting system resources and bandwidth depend 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 disk, decompressing it, extracting knowledge and reporting. Since indexers incur most of the workload, increases in indexing volume should be tied to an increase 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 GCP 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, both modular inputs, which are extensions to Splunk Enterprise that define a custom data input, and HTTP Event Collector, 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 Within GCP

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.

**Compute:** While one can save money with preemptible instances, Splunk is persistent software intended to gather and index data at all times; thus, **regular instances** are preferred. The following are recommended minimum VM instance requirements (corresponding to **n1-standard-4** VM Class):

- **4vCPU**
- **15GB of RAM**

Splunk software is well suited for GCP as it scales horizontally. Adding Splunk instances can give you better performance and capacity depending on data volume requirements. See tables below for more detail on recommended sizes.

**Storage:** Splunk recommends using root persistent disks to store Splunk configurations, the OS, and indexed data. For clustered deployments, Local SSDs can be considered as an alternative option. Below are the key characteristics of **persistent disks** and **local SSDs**:

- **Persistent disks are highly available, reliable and durable, and can grow to 64TB in size**
- **Persistent disks come in two forms; standard hard disk (PD-HDD) and solid-state (PD-SDD)**
- **PD-SDD are a better fit for hot/warm data and PD-HDD are more suitable for cold**
- **Local SSDs are attached to the server that hosts the VM instance, limited to 3TB/instance**
- **Local SSDs have a much higher throughput and lower latency than PD-HDD and PD-SDD**
- **Local SSDs are available via two interfaces; SCSI or NVMe**

For backing up Splunk data, **Persistent Disk Snapshots** can be considered when using PD-SSD or PD-HDD. 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/day, you should expect a daily utilization of approximately 50GB. The number and size of your PD-HDD/SSD or Local SSD volumes should be based on your retention requirements and expected daily indexing volume.

### 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.

**Table 1: Indexers**

Qty.	Instance Type	Daily Volume (GB)
1	n1-standard-16	Up to 100
1	n1-standard-32	100 - 250

**Table 2: Search Heads**

Qty.	Instance Type	Concurrent Users	Performance
1	n1-standard-16	Up to 100	Good
1	n1-standard-32	100 - 250	Better

**Table 3: Deployment Server, License or Cluster Master**

Qty.	Instance Size (Type)	Performance
1	n1-standard-8 n1-highcpu-8	Good
1	n1-highcpu-16 n1-highcpu-16	Better

### Small-Scale Deployment

The following specifications outline an example of a small-scale deployment that is capable of indexing up to 100GB/day, with a maximum of six concurrent searches running at all times. It is common for this type of instance to be deployed for indexing volumes in the single digit GB/day range.

- **1 x Search Head/Indexer (Instance Type: n1-standard-16)**

- **Storage Option 1\*:**

- PD-SSD for Hot/Warm and Cold

- **Storage Option 2\*:**

- PD-SSD for Hot/Warm, PD-HDD for Cold

- **N - Universal Forwarders (data sources)**

*\* Storage selection will depend on your cost considerations.*

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, etc.

The total size of available storage should be based on your retention requirements and expected daily indexing volume.

### Medium-Scale Deployment

The following specifications outline an example of a medium-scale deployment that is capable of indexing 500GB/day, with a search load of eight to 16 users.

- **3 x Indexers (Instance Type: n1-standard-32)**

- **Storage Option 1\*:**

- PD-SSD for Hot/Warm and Cold

- **Storage Option 2\*:**

- PD-SSD for Hot/Warm, PD-HDD for Cold

- **1 x Search Head (Instance Type: n1-standard-32)**
  - **Storage Option:** PD-SSD
- **1 x Admin Node (Instance Type: n1-highcpu-8)**
  - **License Master, DMC, etc.**
- **N - Universal Forwarders (data sources)**

*\* Storage selection will depend on your cost considerations.*

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 the other as an admin node. The total size of available storage should be based on your retention requirements and expected daily indexing volume.

### Large-Scale Deployment

The following specifications outline an example of a large-scale deployment that is capable of indexing 1TB/day, with a concurrent search load of 16 users. As noted earlier, Splunk software scales horizontally. To increase the capacity or performance of this installation, simply add indexers or search heads when appropriate.

- **5 x Indexers (Instance Type: n1-standard-32)**
  - **Storage Option 1\*:**  
PD-SSD for Hot/Warm and Cold
  - **Storage Option 2\*:**  
PD-SSD for Hot/Warm, PD-HDD for Cold
- **1 x Search Head (Instance Type: n1-standard-32)**
  - **Storage Option:** PD-SSD
- **1 x Admin Node (Instance Type: n1-highcpu-8)**
  - **License Master, DMC, etc.**
- **N - Universal Forwarders (data sources)**

*\* Storage selection will depend on your cost considerations.*

Architecturally, there is a single search head and five indexers. Any number of Splunk forwarders can distribute data to these indexers. The total size of available storage should be based on your retention requirements and expected daily indexing volume.

### Clustered Deployment

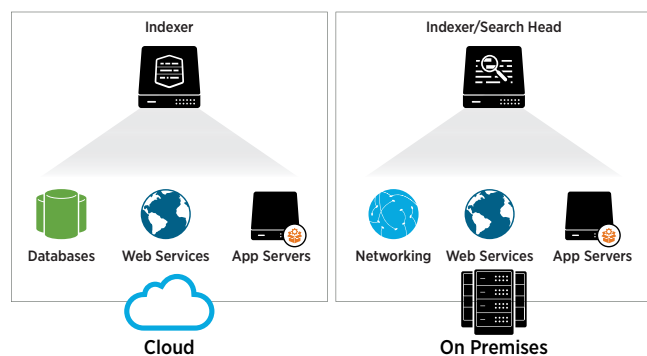
The following specifications are an example of large-scale deployment leveraging the index replication feature. Index replication creates and manages multiple copies of indexers' data so they are readily available in the event of a Splunk indexer outage. This deployment is capable of indexing 1TB/day, with a concurrent search load of up to 16 users. Similar to the previous example, adding indexers or search heads will increase performance or capacity when appropriately applied.

- **8-10 x Indexers (Instance Type: n1-standard-32)**
  - **Storage Option 1\*:**  
Local SSD for Hot/Warm, PD-HDD for Cold
  - **Storage Option 2\*:**  
PD-SSD for Hot/Warm and Cold
  - Storage Option 3\*:**  
PD-SSD for Hot/Warm, PD-HDD for Cold
- **1 x Search Head (Instance Type: n1-standard-32)**
- **1 x Admin Node (Instance Type: n1-highcpu-8)**
  - **Master Node, License Master, DMC, etc.**
- **N - Universal Forwarders (data sources)**

*\* Storage selection will depend on your cost considerations.*

Architecturally, there are five Splunk indexers and a single Splunk search head. All of these components communicate with the cluster and license manager instance for replication and licensing purposes. Like the previous example, the search head distributes search to

all five indexers, although it does so based on information from the cluster master. To increase retention, capacity or both, simply add more indexers and/or more storage.



### Hybrid Environment

The graphic above represents a hybrid environment where Splunk Enterprise is installed on-premises and in the cloud. Splunk software's distributed search capability allows you to peer into both environments from a single interface.

### Additional Considerations

- **Leverage Splunk Universal Forwarders to gather data from existing systems.**
- **Use the deployment server to manage Splunk apps and configurations from a central Splunk instance.**
- **The index replication feature allows for high availability of the indexed data across multiple Splunk systems. Availability is managed at the Splunk software layer versus traditional storage availability methods.**

### Summary

For best performance when deploying Splunk Enterprise on Google Cloud Platform, use the recommended instance sizes and types and plan according to your expected daily volume requirements. As GCP is friendly to horizontal scaling, deploy additional Splunk instances to gain capacity and performance.

## Use Splunk Solutions on GCP

**Splunk Enterprise.** [Download Splunk Enterprise](#) for free. You'll get a Splunk Enterprise license for 60 days and you can index up to 500 megabytes of data per day. After 60 days, or anytime before then, you can convert to a perpetual free license or purchase an enterprise license by contacting Splunk at [www.splunk.com/asksales](http://www.splunk.com/asksales).

**Splunk Add-on for Google Cloud Platform.** Get started with the [Splunk Add-on for Google Cloud Platform](#) to collect platform events, logs, metrics and billing data using Google Cloud Platform API.