

# Deploying Splunk Inside Virtual Environments

## Configuring Splunk on Amazon Web Services

### Background

Splunk Enterprise is an engine for machine-generated IT data, giving you real-time visibility and intelligence into what's happening across your IT infrastructure—whether it's physical, virtual or in the cloud.

Splunk indexes machine-generated IT data in real time, enabling deep, data drill down when needed, powerful statistical analytics and real-time dashboards and views for anyone in your organization—from server teams to business users. Splunk scales linearly across commodity hardware, allowing IT to rapidly draw correlations between massive amounts of data from various sources very quickly.

Cloud adoption is a commonplace strategy for IT organizations around the world seeking to cut costs, increase elasticity and decrease time to market. In fact, many organizations have deployment policies that require cloud usage.

Splunk is geared for deploying within the cloud, as well as across hybrid environments where a mixture of physical and cloud devices are deployed. Organizations with these mixed environments can gain visibility that was previously unobtainable. This document covers guidelines for deploying Splunk on Amazon Web Services.

### Splunk Deployment Components

The typical components that make up a Splunk deployment include Splunk forwarders, indexers and search heads. Each of these components can be run independently inside different cloud instances. However, the system resources required to run these components will vary. Before deploying, consider the needs of each component and plan accordingly.

**Forwarders** collect and forward data; these components are usually lightweight and are not resource intensive

**Indexers** write data to a storage device and perform searching on the data. These functions are both CPU- and I/O-intensive

**Search heads** search for information across indexers and are usually both CPU- and memory-intensive

Budgeting system resources and bandwidth to enable search and index performance depends on the total volume of data being indexed and the number of active concurrent searches (scheduled or otherwise) at any time.

Indexers, in addition to rapidly writing data to disk, perform much of the work involved in running searches: reading data off disk, decompressing it, extracting knowledge and reporting. As a result, when scaling up data volumes, additional indexers should be added. These indexers will help handle larger volumes of data, reduce contention for resources during searches and accelerate search performance.

Most EC2 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 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.

### Performance Considerations within Amazon Web Services

There are several performance factors to consider when deploying Splunk on Amazon Web Services. These considerations are AWS EC2 Instance Size, AWS storage type and Amazon Machine Image selection.

**AWS Instance:** Splunk has preferred instance types and sizes. Reserved instances are preferred due to their assured availability. While spot and on-demand instances can save money when not in use, Splunk is persistent software that is intended to gather and index data at all times. Splunk has the following recommended minimum EC2 instance requirements:

- 2 EC2 Compute Units
- 2GB of RAM
- Reserved Instance
- Elastic Block Storage

Splunk is well suited for AWS as it scales horizontally. Adding multiple, smaller instances can give you more performance/capacity, depending on data volume requirements. See table 1 for more detail on recommended sizes.

**AWS Storage:** While each instance includes local storage, Splunk recommends non-local storage to host the Splunk indexes. Amazon Web Services offers two options including Simple Storage Service (S3) and Elastic Block Storage. Elastic Block Storage is preferred for the following reasons:

- EBS behaves similar to raw block devices
- Latency and throughput performance are increased
- EBS volumes can be deployed in a RAID architecture (Splunk recommends RAID 1+0)
- Allows snapshots for backing up current data

While it is possible to run Splunk on a single instance with a single EBS volume, leveraging a RAID1+0 configuration is ideal for performance and redundancy. Provisioning an additional EBS volume for snapshots should also be considered. When planning your storage requirements for the indexes, take into account that Splunk will compress the data. Most Splunk installations experience a 3:1 compression ratio when storing the index data. This means if you are indexing 10 GB/day, you should expect to utilize approximately 3GB of raw storage per day.

**AWS AMI/Platform Image:** Splunk runs on most widely available operating systems including Windows and \*NIX platforms. When choosing the OS for the searching and indexing server, a 64-bit architecture is highly recommended. While Splunk performs well on all platforms, there are slight performance benefits when deploying on \*NIX based OS platforms.

## Deployment Guidelines

| Qty. | Instance Size (Type)   | Daily Indexing Volume (GB) | Performance |
|------|------------------------|----------------------------|-------------|
| 1    | Medium High (CPU)      | 0.5                        | Fair        |
| 3    | Medium High (CPU)      | 10                         | Good        |
| 1    | Large                  | 20                         | Good        |
| 1    | Extra Large            | 100                        | Better      |
| 1    | Extra Large High (CPU) | 100                        | Better      |
| 1    | Quadruple XL           | 100                        | Better      |
| 4    | Large                  | 100                        | Best        |

### Small Scale Deployment

The following specifications are an example of a small-scale deployment. This deployment is capable of indexing 20 GB/day, with a maximum of 2 concurrent searches running at all times.

- 1 – Large EC2 instance Architecturally, this is a single Splunk instance performing indexing and searching. Data can be sent to this system via network inputs (include Splunk Forwarders), local files or NFS mounted files and scripted calls.

### Medium Scale Deployment

The following specifications are an example of a medium-scale deployment. This deployment is capable of indexing 100 GB/day, with a maximum of 3 concurrent searches running at all times.

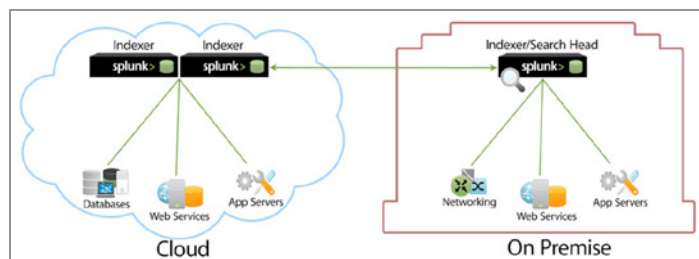
- 4 – Large instances
- 1 Splunk search head
- 3 Splunk indexers

Architecturally, this deployment consists of 4 Splunk instances. Three of these instances act as indexers and another acts as the search head. This deployment leverages the horizontal scalability of Splunk as well as map-reduce technology.

### Large Scale Deployment

The following specifications are an example of a large-scale deployment. This deployment is capable of indexing 500 GB/day, with a maximum of 6 concurrent searches running at all

times. As noted earlier, Splunk scales horizontally. To increase the capacity or performance of this installation, simply adding additional indexers or searchers will suffice.



- 6 – Extra Large instances
- 1 - Splunk search head
- 5 - Splunk indexers
- N – Splunk Forwarders

Architecturally, there are 5 Splunk indexers and a single Splunk search head. The search head distributes search to all 5 indexers, leveraging MapReduce technology. Any N number of Splunk Forwarders can distribute data to these indexers.

### Hybrid Environment

The following graphic represents a hybrid environment where Splunk is installed on premises and in the cloud. Splunk'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.
- You can use the Splunk deployment server to manage and propagate Splunk apps and configurations from a central Splunk instance on data volumes within virtual environments, virtualization offers better resource sharing and provisioning.

### Summary

For best performance when deploying Splunk on Amazon Web Services, use the recommended instance sizes for your expected volume requirements. Additionally, leverage EBS storage as your primary indexing store. As EC2 is friendly to scaling and elasticity, deploy additional Splunk indexers to gain capacity and performance.

### Free Download

[Download Splunk](#) for free. You'll automatically get all of the Enterprise features of Splunk for 60 days and you can index up to 500 megabytes of data per day. Or if you want to get started right away with an Enterprise license contact [sales@splunk.com](mailto:sales@splunk.com).