splunk>

TECH BRIEF

DEPLOYING SPLUNK® ENTERPRISE ON AMAZON WEB SERVICES

Splunk Enterprise is the leading platform for real-time Operational Intelligence. It takes the machine data generated by IT systems and technology infrastructure—whether it's physical, virtual or in the cloud—and turns it into valuable insights.

Splunk software indexes machine-generated data in real time, enabling deep data drilldown when needed, powerful statistical analytics, and real-time dashboards and views for users at any level of an organization—from server teams to business users. It scales linearly across commodity hardware, allowing IT to rapidly draw correlations between massive amounts of data from various sources.

Cloud adoption is a commonplace strategy for IT organizations seeking to cut costs, increase agility and decrease time to market. In fact, many organizations have deployment policies that require cloud usage. That's why Splunk Enterprise is geared for deploying within the cloud, as well as across hybrid environments—with a mixture of on-prem and cloud infrastructure. Organizations with these mixed environments can gain visibility that was previously unobtainable. This document covers guidelines for deploying Splunk Enterprise on Amazon Web Services (AWS).

Splunk Deployment Components

The typical components that make up a Splunk deployment include Splunk forwarders, indexers and search heads. Splunk Enterprise is a single package that can perform one or many of the roles that each component would normally deliver, in addition to others. Users can install the software within minutes to their choice of hardware (physical, cloud or virtual) and operating system. The package is available via a public AMI (Amazon Machine Image) in addition to downloadable packaged forms for most operating systems. While all major Splunk components can be run from a single installation on a single cloud instance, they can also run independently from within different cloud instances. Depending on the deployment infrastructure, considerations must also be taken to allocate the proper amount of resources per component type.

Forwarders perform data collection, data forwarding and data load balancing. Low amounts of resources are required to run a forwarder as they typically read and send data with minimal overhead. A Universal Forwarder is a lightweight package of the Splunk software that can perform most, if not all, of the forwarder functionality.

Indexers write the data to a storage device and perform searching on the data. These can be resource intense and require I/O and CPU allotment.

Search heads search for information across indexers and require CPU and memory allotment.

Budgeting system resources and bandwidth to enable search and index performance 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: 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. Deploying additional indexers will distribute the load of increased data volumes, resulting in reduced contention for resources during searches and accelerated 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, 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.

Other Splunk components include the Deployment Server (configuration publishing), License Master (license management) and Master Node (manages index replication).

Performance Considerations Within AWS

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

AWS Instance: 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; thus, reserved instances are preferred. The following are recommended minimum EC2 instance requirements:

- **4 vCPU**
- **8GB of RAM**

Splunk software is well suited for AWS, as it scales horizontally. Adding Splunk instances offers more performance and capacity depending on data volume requirements. See tables 1-3 for more detail on recommended sizes.

AWS Storage: Splunk recommends using EBS volumes as a root partition to store Splunk configurations and the OS. EBS volumes are also highly recommended for storing data. For clustered deployments, instance store can be considered as an alternative. AWS's Elastic Block Storage (EBS) is preferred for the following reasons:

- **EBS volumes are highly available, reliable and can grow up to 16TB in size**
- **EBS General Purpose SSD (gp2) types provide balanced price and performance**
- **EBS Throughput Optimised HDD (st1) volumes types are optimal for throughput intensive workloads targeting cold data**
- **EBS Provisioned IOPS SSD (io1) volume types provide the highest performance and are ideal for special use cases**
- **EBS volumes can be deployed in a RAID architecture, if necessary**

For backing up Splunk data, S3 snapshots can be considered. When planning 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 indexing 10GB/day, users should expect to utilize approximately 5GB of storage per day. The number and size of EBS volumes should be based on retention requirements and expected daily indexing volume.

AWS AMI: Splunk Enterprise runs on most widely available operating systems including Windows and *NIX platforms. When choosing the OS for the search head and indexers, a 64-bit architecture is highly recommended. Splunk offers a public AMI containing Splunk Enterprise on top of a 64-bit Linux Amazon OS, via the AWS Marketplace.

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	c4.4xlarge	Up to 100
1	c4.8xlarge	100–250

Table 2: Search Heads

Qty.	Instance Size (Type)	Concurrent Users	Performance
1	c4.4xlarge	Up to 8	Good
1	c4.8xlarge	Up to 16	Better

Table 3: Deployment Server, License or Cluster Master

Qty.	Instance Size (Type)	Performance
1	c3.2xlarge	Good
1	c3.4xlarge	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 not uncommon for this type of instance to be deployed for indexing volumes in the single digit GB/day range.

- **1 – c4.4xlarge with EBS-backed storage**
- **N – Universal Forwarders (data sources)**

Architecturally, this is a single Splunk instance performing indexing and searching. Data can be sent to this system via Splunk forwarders, local files, NFS mounted files and scripted calls or modular inputs. The number and size of EBS volume(s) should be based on 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 – c4.8xlarge with EBS-backed storage (Indexers)**
- **1 – c4.8xlarge with EBS-backed storage (Search Head)**
- **1 - c3.xlarge (License Master)**
- **N – Universal Forwarders (data sources)**

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 the license master. The number and size of EBS volumes on indexers should be based on 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 – c4.8xlarge with EBS-backed storage (Indexers)**
- **1 – c4.8xlarge with EBS-backed storage (Search Head)**
- **1 - c3.xlarge (License Master)**
- **N – Universal Forwarders (data sources)**

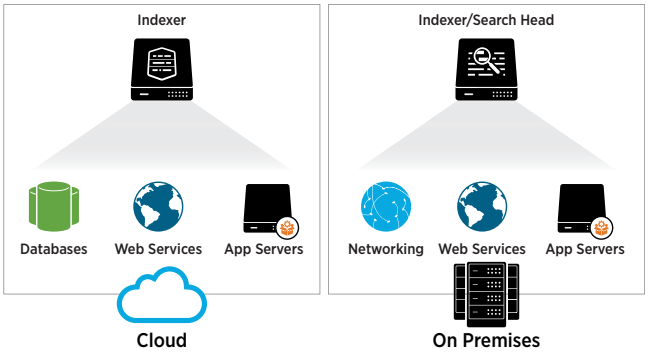
Architecturally, there is a single search head and five indexers. Any N number of Splunk forwarders can distribute data to these indexers. The number and size of EBS volumes should be based on retention requirements and expected daily indexing volume.

Clustered Deployment

The following specifications are an example of a large-scale deployment leveraging the index replication feature. Index replication creates and manages multiple copies of indexes' buckets so they are readily available in the rare event of a Splunk indexer outage. One benefit to this feature is the ability to leverage the ephemeral storage on each instance and allow Splunk to manage replicated data between indexes. This is an alternative to using EBS backed storage for the indexes. 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.

- **5 – d2.8xlarge with local storage (Indexers)**
- **1 – c4.8xlarge with EBS-backed storage (Search Head)**
- **1 – c3.xlarge (License Master and Master Node)**
- **N – Universal Forwarders (data sources)**

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 consider larger instance sizes.



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 enables insight into both environments from a single interface.

Additional Considerations

- **Leverage Splunk Universal Forwarders to gather data from existing systems.**
- **Use the Splunk deployment server to manage and propagate 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 (like EBS with RAID).**
- **To automate delivery of non-clustered deployments, Splunk has made available AWS CloudFormation templates for download.**

Summary

For best performance when deploying Splunk Enterprise on Amazon Web Services, use the recommended instance sizes/types and plan according to expected daily volume requirements. As AWS EC2 is friendly to horizontal scaling, deploy additional Splunk instances to gain capacity and performance.

Use Splunk Solutions on AWS

Splunk Enterprise. Download Splunk Enterprise for free or find the Splunk Enterprise AMI in the AWS Marketplace. 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.

Splunk Cloud. Sign up for Splunk Cloud, which delivers Splunk Enterprise as a service.

Splunk App for AWS. Get started with the Splunk App for AWS to gain operational visibility and security for your AWS environment.