Deploying Splunk Enterprise on Amazon Web Services

Splunk Enterprise is an integral part of Splunk's Data-to-Everything Platform, providing collection, indexing, visualization and deep analysis capabilities of your data. Building actionable insights with an organization's data is the key to turning data into action. While Splunk Cloud is the optimal way to realize the value of the Splunk platform, some organizations may have requirements to deploy and administer Splunk Enterprise themselves. This tech brief applies to those customers choosing to deploy and operate Splunk Enterprise on Amazon Web Services (AWS). It is suited for Splunk administrators familiar with the Splunk Validated Architectures (SVAs), SmartStore capability and AWS. Administrators should also explore the Splunk Quickstart for AWS to automate deploying Splunk products in AWS.

Note: Organizations that are in the process of cloud adoption should look first to Splunk Cloud to get all the benefits of the Splunk platform with the advantages of cloud delivery. In addition, for real-time performance monitoring of cloud infrastructure, services and applications, customers should use Splunk Infrastructure Monitoring alongside Splunk Enterprise. A seamless workflow integration between Splunk Infrastructure Monitoring and Splunk Enterprise eliminates context switching, provides granular insights and accelerates root-cause analysis.

SmartStore Using AWS S3

Splunk supports data storage on compatible object storage, including AWS S3. This Splunk capability, called SmartStore, can significantly reduce the overall cost of running Splunk in AWS. SmartStore using AWS S3 also makes long-term data retention cost-effective and simplifies the implementation of data durability and security.

SmartStore cache sizing

SmartStore's caching capability will automatically move data between local instance storage and S3. The cache is stored on the instance volume, and keeps frequently indexed and searched data. Sizing the volume therefore depends on the expected data ingestion rate and the number of days of data typically being searched.

Cache size can be anticipated from the daily data-ingestion rate of the Splunk landscape, assuming a 2:1 compression ratio for stored data on indexers in the cluster, and adjusted for search and replication factors. For a discussion on setting search and replication factors and associated storage requirements, refer to Splunk documentation.

Typical Splunk software installations work well with 30 days of cached data. It is recommended when running Splunk Enterprise Security to increase the cache size for 90 days of retained data.

To provide maximum flexibility, SmartStore can be configured on a per-index basis, allowing administrators to adjust performance and cost to their specific use cases. Because of the significant advantages of using SmartStore on AWS, the guidance provided in this brief only discusses SmartStore used for all indexes. For other scenarios, follow Splunk documentation for SmartStore and system requirements for clustered deployments.

SmartStore access and security configuration

To properly configure S3 and SmartStore, refer to Splunk documentation for configuring the remote store for SmartStore. You will need to consider the available options for S3 bucket access, bucket settings and security.

A minimal, easy-to-manage and secure configuration is presented on the next page, using roles to grant access to a single bucket that stores all of the indexed data. SmartStore has the flexibility to accommodate different bucket settings and access schemes if your organization

has specific requirements. In addition, extending the pattern below is straightforward to enable multiple remote stores.

In this basic policy the S3 bucket acting as the remote store is configured with no public access or access points, and requires no specific access policies. In addition, all properties are left to their default creation states (disabled), except for AES-256, which should be enabled. Bucket versioning, event logging and tags can be set according to your organization's policies.

When configuring the remote store in Splunk, the encryption scheme matching this bucket configuration should be `remote.s3.encryption = sse-s3`. This matches the AES-256 option on the S3 bucket, encrypting everything in the remote store with S3 providing the key management.

Indexers are then granted access to the S3 bucket by assigning an AWS IAM role to each indexer in the cluster. The role should include a policy with the following capabilities:

```
{
	"Version": "2012-10-17",
	"Statement": [
	{
		"Effect": "Allow",
		"Action": [
		"s3:ListBucket"
		],
		"Resource": [
		"arn:aws:s3:::<my-smartstore-
bucket-name>"
		]
	},
	{
		"Effect": "Allow",
		"Action": [
		"s3:ListBucket",
"s3:GetObject",
"s3:ListBucketVersions"
"s3:DeleteObject"
"s3:PutObject",
"s3:PutObjectAcl"
		],
		"Resource": [
		"arn:aws:s3:::<my-smartstore-
bucket-name>/*"
		]
	}
	]
}
```

Note that `s3:ListBucketVersion` should be specified when enabling versioning on the bucket.

Finally, configure indexes.conf according to Splunk documentation to define the remote index volume and which indexes to store on it, using the name of the S3 bucket you created.

SmartStore and clustering

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 S3 when using SmartStore on AWS, replication is still required for the most recent indexed data. You must set search factor and index replication factor to the same value when using SmartStore.

Performance Considerations Within AWS

There are several performance factors to consider when deploying Splunk software on AWS, including which AWS services to leverage, which instance types to use and how to optimize the storage cost and performance.

While spot instances can save money, we strongly recommend against using them to run Splunk Enterprise, as Splunk Enterprise is always-on software that is intended to gather and index data at all times.

Instances that require EBS should use gp2 volumes to meet the minimum IOPS outlined in the Splunk reference hardware documentation. To meet this requirement, the gp2 volume size should be at least 300GB.

Though it is possible to run Splunk software on a wide variety of operating systems, using SmartStore on AWS requires the use of Linux operating systems. Follow guidance in Splunk documentation for system requirements for system-wide limits and THP settings on *.nix operating systems. Splunk also provides a free public AMI containing Splunk Enterprise on top of a 64-bit Linux Amazon OS via the AWS Marketplace.

Choosing AWS Instance Types

With SmartStore as the preferred storage in AWS, compute resources can be scaled independently from total storage. This provides additional options to save cost and optimize performance, including making it possible to use instances with ephemeral storage to improve search speeds.

Choosing from the variety of instance types AWS offers can be a challenge. The guidance here is intended to provide excellent indexing and search performance while minimizing per instance cost and assumes typical ingestion and search load. Splunk documentation for capacity planning provides recommendations for your specific environment.

Single-Instance Splunk Enterprise Deployments

c5.4xlarge	Up to 200GB/day
c5.9xlarge	Up to 300GB/day

Search Heads

c4.4xlarge	Up to 8 concurrent users
c5.9xlarge	Up to 16 concurrent users

Indexers

i3.4xl	Up to 200GB/day
i3en.6xl	Up to 300GB/day

Utility (Deployment Server, Cluster Master, Monitoring Console, Deployer, License Master)

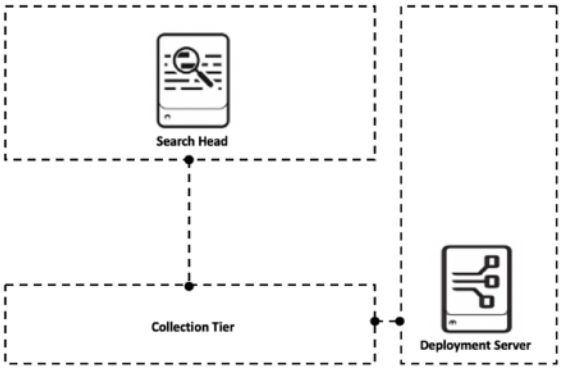
c5.2xlarge	*small/medium deployments/clusters
c5.4xlarge	*large deployments/clusters

* The Splunk Validated Architectures describe the situations where it is possible to combine some utility roles on a single instance.

Deployment Examples

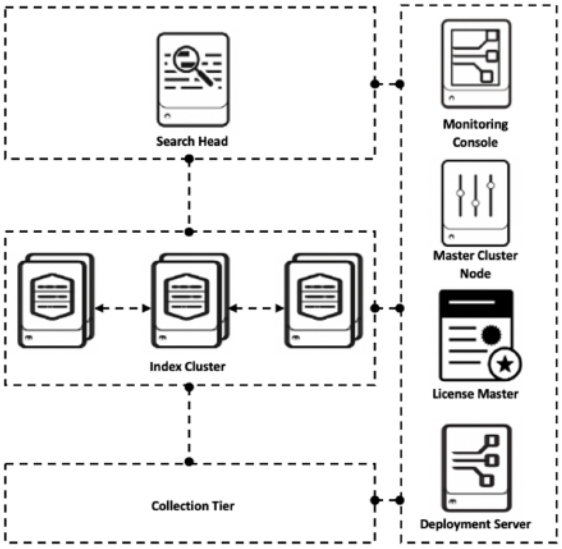
Below are a few examples of Splunk Validated Architectures (SVA) deployed on AWS. Please refer to the Splunk SVA prior to designing your deployment. In these examples, "N" is used to denote one or more instances of servers of that role. Splunk documentation for capacity planning, along with the SVA, will help you determine N in each scenario.

Single Server Deployment (SVA type S1)



1 - c5.9xl acting as Search Head and Indexer
1 - c5.2xl - Deployment Server

Distributed Clustered Deployment (SVA type C1/C11)

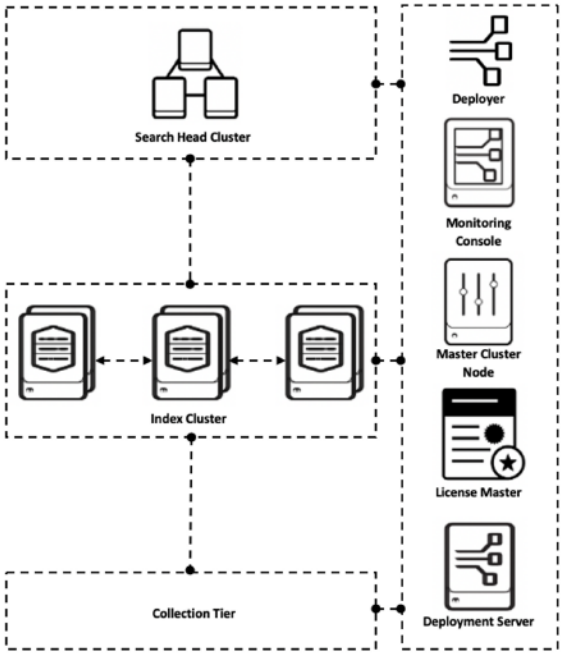


1 - c5.9xl - Search Head

N - i3en.6xl - Indexers

1 - c5.2xl - License Master and Monitoring Console combined
1 - c5.2xl - Deployment Server. Can use c5.4xl for very large numbers of deployment clients.
1 - c5.2xl - Cluster Master Node

Distributed Clustered Deployment (SVA type C1/C11)



3 or more - c5.9xl - Search Head

N - i3en.6xl - Indexers

1 - c5.2xl - License Master and Monitoring Console combined
1 - c5.2x - Search Head Cluster Deployer
1 - c5.2xl - Deployment Server. Can use c5.4xl for very large numbers of deployment clients.
1 - c5.4xl - Master Cluster Node

Summary

For best performance when deploying Splunk Enterprise on Amazon Web Services, leverage Splunk SmartStore with AWS S3, and use ephemeral instances such as i3en. Also, refer to the Splunk SVA guide to design a simple and robust Splunk deployment for the use case and topology your installation requires.

Other Resources for Splunk Software on AWS

Below is a list of assets to help Splunk customers get Splunk running on AWS and using Splunk for operational visibility into their other AWS workloads. Consult Splunk technical sales, customer success and support for more information about Splunk and AWS.

Whitepapers

- Getting Data Into (GDI) Splunk From AWS — a discussion about the AWS data sources available to Splunk users, and how to get that data into Splunk

Product Information

- Splunk Cloud — the flexible, secure and cost-effective data platform service
- Cloud monitoring and Observability
- AWS Quickstart and corresponding cloudformation

Getting AWS data into Splunk

- Cloudformation to automate AWS data ingestion (Project Trumpet)
- Splunk Add-On for AWS
- AWS Kinesis Data Firehose and Splunk Add-On for AWS Firehose

Learn more about Splunk Enterprise.

21-16646-Deploying Splunk ES on AWS-112-TB