Deploying Splunk Enterprise on Microsoft Azure

Splunk® 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 11,000 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 Microsoft Azure, an open and flexible cloud platform with a growing collection of integrated cloud services, including analytics, computing, database, mobile, networking, storage and web.

Splunk Deployment Components

A typical Splunk deployment includes 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. The software can be installed within minutes on your choice of hardware (physical, cloud or virtual) and operating system. 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. 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 volume, resulting in reduced contention for resources and improved search performance.

Common Azure deployments leverage a combination of forwarders and network streams to send data to the Splunk indexer(s). While forwarders are 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 method 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, a highly efficient and secure mechanism to send high volumes of data directly to Splunk, can be used to collect data from various API sources.

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

Performance Considerations Within Microsoft Azure

There are several performance factors to consider when deploying Splunk software on Microsoft Azure. These considerations are Azure Virtual Machine (VM) image and size, and underlying Azure Storage.

Azure VM image

Splunk Enterprise runs on most widely available operating systems including Windows and *nix platforms. Splunk is persistent software that is intended to gather and index data at all times; thus, reserved instances are preferred.

Azure VM size

The size of an Azure VM is defined by the number of CPU cores, the generation of CPU, the amount of available memory, the maximum network bandwidth, and number of data disks that can be attached. The following are recommended minimum Azure VM requirements:

- 8 CPU cores (compute optimized series)
- 14GB of RAM

Splunk Enterprise scales horizontally, making it well suited for Microsoft Azure. Adding Splunk instances can give you more performance and capacity depending on usage and data volume requirements. See below for more detail on recommended sizes.

Azure Storage

Azure VM has two types of disks: a local temporary disk and a network-attached persistent disk or virtual hard disk (VHD). Each VM comes with a local disk, one OS disk as VHD, and can have one or more data disks as VHDs.

A local disk is generally not suitable to store Splunk indexes since it's intended for temporary data only: data on local disk may be lost in case of a hardware failure or upon VM resize or reboot.

A VHD is stored in a standard or premium storage account in Azure. VHDs can be managed or unmanaged. Splunk recommends using managed VHDs for Splunk storage. More specifically, you can store Splunk application and configurations in the persistent OS disk and store Splunk indexes across multiple persistent data disks.

Managed Disks are preferred for various reasons:

- Managed Disks transparently handle storage accounts. With Unmanaged Disks, if IOPS across all disks in a storage account approach storage account limits, you must create additional storage accounts as well as rebalance your virtual machine disks across the storage accounts to insure they stay within the IOPS limit. Managed Disks remove the need to provision additional storage accounts, effectively removing these IOPS limits.
- Managed Disks provide greater reliability for Availability Sets by ensuring disks are sufficiently isolated to avoid single points of failure. This ensures that VMs in an Availability Set will not be stored in the same storage scale unit. Therefore if one VM in an Availability Set goes down due to hardware of software failure, the other VMs will not be impacted.
- VHDs configured as Managed Disks are highly available and are designed for 99.999% availability.

Refer to **Microsoft's documentation on specific storage limits**.

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.

Small-Scale Deployment

Table 1: Indexers

Instance Size (Type)	Daily Indexing Volume (GB)	Performance
Standard_DS4_v2	Up to 100	Good
Standard_DS5_v2	100-500	Better
Standard_DS15_v2	150-250	Best

Table 2: Search Heads

Instance Size (Type)	Concurrent Users	Performance
Standard_DS5_v2	Up to 8	Good
Standard_DS15_v2	Up to 16	Better

Table 3: Deployment Server, License or Cluster Master

Instance Size (Type)	Performance
Standard_DS3_v2	Good
Standard_DS4_v2	Better

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 – Standard_DS4_v2 with VHDs-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, HTTP event collector, local files, NFS mounted files, SMB file shares, and scripted calls or modular inputs. The number and size of your VHD volume(s) should be based on your retention requirements and expected daily indexing volume.

Distributed Deployments and Azure Availability Sets

An Availability Set is a logical grouping capability used in Azure to ensure that the VM resources placed within it are isolated from each other when deployed within an Azure datacenter. Azure ensures that the VMs placed within an Availability Set run across multiple physical servers, compute racks, storage units and network switches. When more than one VM is used to fulfill a role, Splunk recommends using an availability set for that role. For example, if more than one indexer is used in a deployment, Splunk recommends placing the indexers in an Availability Set. If search head clustering is used, place the search heads in a separate Availability Set.

Medium-Scale Deployment

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

- 5 – Standard_DS5_v2 with VHDs-backed storage in an Availability Set (Indexers)
- 1 – Standard_DS5_v2 with VHDs-backed storage (Search Head)
- 1 – Standard_D(S)3_v2 (License Master)
- N – Universal Forwarders (data sources)

Architecturally, this deployment consists of six Splunk instances in a traditional distributed configuration. Five of these instances act as indexers and one acts as the search head. This deployment leverages the horizontal scalability of Splunk software. The number and size of your VHD volumes 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 – Standard_DS15_v2 with VHDs-backed storage in an Availability Set (Indexers)
- 1 – Standard_DS15_v2 with VHDs-backed storage (Search Head)
- 1 - Standard_D(S)3_v2 (License Master)
- N – Universal Forwarders (data sources)

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

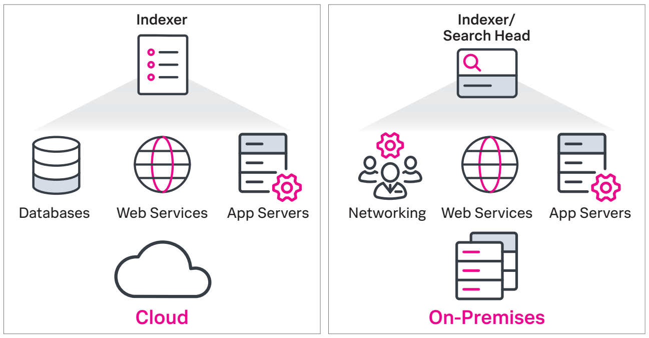
Clustered Deployment

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. This deployment is capable of indexing 500GB/day, with a concurrent search load of up to eight users. Similar to the previous example, adding indexers or search heads will increase performance or capacity when appropriately applied.

- 5 – Standard_DS5_v2 with VHDs-backed storage in an Availability Set (Indexers)
- 1 – Standard_DS5_v2 with VHDs-backed storage (Search Head)
- 1 – Standard_D(S)3_v2 (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 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 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 VHD with RAID).
- Consider provisioning an Azure VM for use as a jump box for SSH or RDP console access. To secure the jump box, add a NSG rule that allows connections only from a safe set of public IP addresses.

Summary

For best performance when deploying Splunk Enterprise on Azure, use the recommended Azure VM sizes and Azure Storage volumes, and plan according to your expected daily volume requirements. As Azure VM and Azure Storage are friendly to horizontal scaling, deploy additional Splunk instances and data disk volumes to gain capacity and performance.

Use Splunk Solutions on Microsoft Azure

Download Splunk Enterprise for free to quickly deploy Splunk Enterprise as either a single instance or a distributed cluster on Azure. 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 any time before then, you can convert to a perpetual free license or purchase an enterprise license by contacting Splunk at **www.splunk.com/asksales**.

Splunk Add-on for Microsoft Cloud Services. Get started with the **Splunk Add-on for Microsoft Cloud Services** to gain operational visibility and security from a variety of Office 365 and Azure services.

splunk>

Learn more: www.splunk.com/asksales

www.splunk.com