

Deploying Splunk Inside Virtual Environments

Configuring VMware Virtual Machines to Run Splunk

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.

VMware technology adoption is a standard in datacenters around the world, being used to cut costs, reduce downtime and achieve operational efficiencies. In fact, many organizations have a virtualization first or virtualization only policy.

To gain visibility across their virtual and physical infrastructures organizations frequently deploy Splunk inside VMware virtual machines. Given virtualization overhead, these Splunk deployments should be carefully planned and configured to achieve the best performance. This tech brief describes performance considerations and guidelines for deploying Splunk inside VMware virtual machines.

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 virtual machines. However, the system resources required to run these components vary and should be planned beforehand:

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

Indexers write data to a disk in key value pairs and 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 other) at any time.

Indexers, in addition to rapidly writing data to disk, do 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.

Performance within VMware Environments

There are several performance factors to consider when deploying Splunk inside VMware virtual machines. These considerations are disk, storage, CPU and memory resources.

- **Disk:** Splunk indexers are usually CPU-and disk I/O-intensive, so disk exposed to these indexers within virtual machines should be capable of 800-1000 I/O operations / second (IOPS). In virtual environments, with virtual machines moving from one type of storage to another, there is less control or guarantee over the type of disk that Splunk virtual machines can access. In VMware environments specifically, VMFS introduces another layer of latency in disk I/O and can impact Splunk performance.
- **CPU:** Since Splunk search heads and indexers are CPU intensive, sharing CPU with other virtual machines running on the same host can result in high wait times which might impact Splunk performance. CPU sharing should be restricted or Splunk indexer/search virtual machines should be given higher priority to ensure good performance.
- **Memory:** is critical for Splunk search heads and indexers. Splunk virtual machines should have reserved memory available to them. VMware hosts running Splunk should not be configured with memory overcommit, as overcommitting memory might result in poor performance due to page swapping to the hard disk.

Deployment Best Practices

No matter how well you provision your virtual machine, Splunk performance may still be affected by the quality of the underlying hardware. All performance guidelines use the following reference server configuration.

Your physical hardware should reflect these minimum requirements:

- Intel x86-64-bit chip architecture
- 2 CPU, 4 core per CPU, 2.5–3Ghz per core
- 8GB RAM
- 4x300GB SAS hard disks at 10,000 rpm each in RAID 10
- 800 IOPS disk performance
- Standard 1Gb Ethernet NICs , management, VMotion traffic is over separate networks

Splunk is often constrained by disk I/O first, so always consider that first when selecting your hardware.

Virtual Machine System Requirements

Here are the recommended minimum system requirements for VMware virtual machines running Splunk:

- 2 virtual CPUs at 2.5 GHz each
- 4 GB of RAM
- Highest priority for CPU and memory shares
- No memory overcommit
- Raw volume (if you're using VMFS, assume a 30% lower indexing performance)
- 800 IOPS disk performance

For Data Volumes up to 20 GB of Data Per Day

If you're indexing less than 20 GB of data per day, you can deploy Splunk as a standard 2-way virtual machine described above. However, depending on the number of concurrent users, you can deploy additional Splunk instances.

- For 2 or fewer users: the single Splunk instance should suffice as both index and search head
- For 2-4 users: Add 2 additional vCPUs to the Splunk virtual machine, to allow for concurrent searches

Greater than 20GB/day or 4+ Concurrent Active Searches

- Splunk's scale-out capabilities include [auto load balancing](#) and [distributed search](#) to run searches in parallel across indexers.

For greater volumes of data or to accommodate many concurrent searches, split Splunk search and indexing functions and parallelize indexing. A good rule of thumb is to add 1 additional virtual indexer per additional 20GB of data; or per additional 2 users.

You have the option to use multiple dual-purpose Splunk virtual machines that search across each other. In this type of distributed search, each Splunk VM functions as both a search head and a search peer.

You can also scale the number of concurrent searches by adding multiple dedicated search head virtual machines. Dedicated search heads are far more CPU bound than indexers—and for a constant number of indexers, they scale by adding additional vCPUs. However, we recommend that you use multiple indexers to scale vs search heads, since indexers do the bulk of the processing work.

Technically, there is no practical limitation on the number of Splunk search heads an indexer can support, or the number of indexers a search head can search against. However best practices suggest a ratio of approximately 8 search to 1 indexer for most deployments. That is a rough guideline; if you have many search heads compared to your total data volume, more search heads will likely help with performance.

In general, the best use of a separate search head is to populate summary indexes. With summary indexing, you set up a search that extracts the precise information you want, on a scheduled basis. This search represents a new, unique index — a summary index. The separate search head will perform like an indexer to the primary search head that users log into.

Additional Considerations

- The recommended best practice for Splunk forwarder management is to use auto load balancing
- You can use the Splunk deployment server to propagate Splunk apps and user preferences between instances
- Fiber Channel direct access to a dedicated LUN is strongly encouraged
- NAS or virtualized storage is strongly discouraged, as is anything with higher than local disk latencies

For high availability and redundancy requirements, refer to the relevant section in our documentation for physical hardware (<http://www.splunk.com/base/Documentation/latest/Installation/apacityplanningforalargerSplunkdeployment>)

Note, while there are lower limits on data volumes within virtual environments, virtualization offers better resource sharing and provisioning.

VMware vMotion, Storage vMotion, DRS, HA

While not formally tested by Splunk, the resources needed for various vSphere migrations will have a performance impact on Splunk deployments within VMware environments. Giving these virtual machines “reserved” memory and CPU is probably safest, but while they are being migrated, they do get quiesced by the hypervisor, so it is recommended that you do these migrations during off peak hours.

Note, if there is a hardware failure and HA restarts virtual machines running Splunk, during the restart some data may not get collected or indexed.

Summary

As is expected with most applications, you should expect less performance when running Splunk within virtual environments. However, there are many additional benefits to consider. Virtualization offers better resource sharing and utilization, might support a corporate mandate and makes provisioning an easier exercise.

For best performance put a high priority on CPU and memory shares for virtual machines running Splunk. Disk quality is also critical to Splunk performance—make sure you are using the best disk available. And to keep up with increasing data volumes, scale your deployment by adding additional Splunk indexers.

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.