Hewlett Packard
Enterprise

HPE Nimble Storage Deployment Considerations for Splunk Enterprise

Published April, 2019

Contents

Notices

The information contained herein is subject to change without notice. The only warranties for Hewlett Packard Enterprise products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. Hewlett Packard Enterprise shall not be liable for technical or editorial errors or omissions contained herein.

Confidential computer software. Valid license from Hewlett Packard Enterprise required for possession, use, or copying. Consistent with FAR 12.211 and 12.212, Commercial Computer Software, Computer Software Documentation, and Technical Data for Commercial Items are licensed to the U.S. Government under vendor's standard commercial license.

Links to third-party websites take you outside the Hewlett Packard Enterprise website. Hewlett Packard Enterprise has no control over and is not responsible for information outside the Hewlett Packard Enterprise website.

Acknowledgments

Intel®, Itanium®, Pentium®, Intel Inside®, and the Intel Inside logo are trademarks of Intel Corporation in the United States and other countries.

Microsoft® and Windows® are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries.

Adobe® and Acrobat® are trademarks of Adobe Systems Incorporated. Java® and Oracle® are registered trademarks of Oracle and/or its affiliates.

UNIX® is a registered trademark of The Open Group.

Publication Date

Wednesday April 17, 2019 07:12:42

Document ID

pej1509553831452

Support

All documentation and knowledge base articles are available on HPE InfoSight at *https://infosight.hpe.com*. To register for HPE InfoSight, click the *Create Account* link on the main page.

Email: *support@nimblestorage.com*

For all other general support contact information, go to *https://www.nimblestorage.com/customer-support/*.

Introduction

Splunk® Enterprise monitors and analyzes machine data from any source to deliver operational intelligence and optimize IT, security, and business performance. With intuitive analysis features, machine learning, packaged applications, and open APIs, Splunk Enterprise is a flexible platform that scales from focused use cases to an enterprise-wide analytics backbone.

Hewlett Packard Enterprise (HPE) offers a predictive flash platform that is ideally suited to Splunk Enterprise deployments. The HPE Nimble Storage portfolio of all-flash arrays and adaptive-flash arrays has a full range of features for management, performance, scalability, and data protection that can easily support the entire lifecycle of Splunk Enterprise data.

Audience

The audience for this guide is Splunk solution architects, storage engineers, system administrators, and IT managers who want to design and maintain a robust Splunk environment on HPE Nimble Storage arrays. The guide focuses on Linux® based deployments and assumes that the reader has a working knowledge of the following technologies:

- Storage network design
- Basic operations on HPE Nimble Storage arrays
- Linux operating systems

The guide presents the overall Splunk Enterprise architecture and discusses considerations for deploying a storage architecture that supports the data lifecycle of Splunk Enterprise workloads. Also covered are general storage sizing guidelines. Customers should consult with their HPE account team or partners to design a storage solution that best meets their individual needs.

Splunk Enterprise Architecture

Splunk Enterprise has a tiered architecture that consists of servers with different roles. These roles might be performed by a single server in small deployments, or they might be distributed across many servers to support larger environments.

The typical components that make up the core of a Splunk Enterprise environment are indexers, forwarders, and search heads. Each component has different storage and performance requirements, with indexers making up the majority of capacity and throughput demands.

Table 1: Splunk Enterprise components

Component	Description
Indexer	• Indexes data • Transforms raw data into events • Searches the indexed data
Forwarder	• Forwards data to another Splunk Enterprise instance (an indexer or another forwarder) or to a third-party system
Search head	• Is used in distributed searches • Handles search management functions, directing search requests to a set of search peers and then merging the results back to the user

Splunk Enterprise data is categorized into buckets that have distinct characteristics. Data is moved between buckets according to criteria that are specified by the Splunk Enterprise architect. An important part of architecting a storage solution for Splunk Enterprise is determining which HPE Nimble Storage solution best meets the needs for each bucket of data.

Figure 1: Types of data bucket

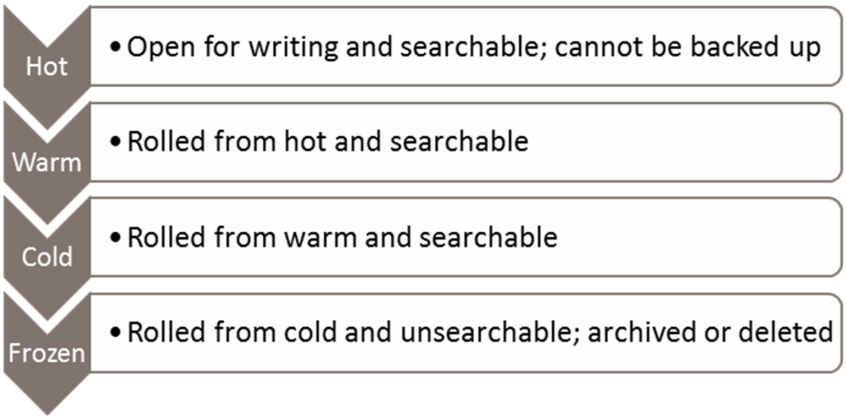
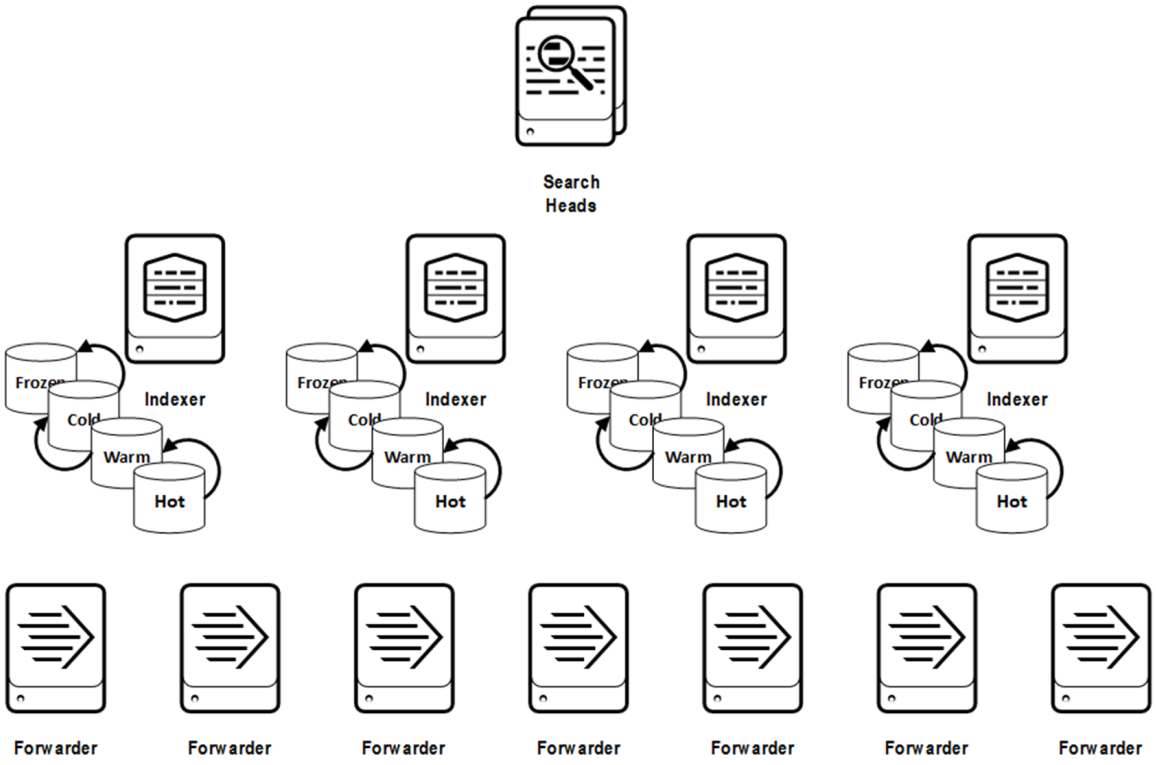


Figure 2: Splunk Enterprise with rolling buckets



Workload Characteristics

Splunk Enterprise workloads consist of two primary operations: sustained writes and periodic reads.

Sustained Writes

Splunk Enterprise categorizes write workloads into a daily ingest rate, which corresponds to the amount of data that is collected in real time from multiple sources. The data is received by the forwarder and sent to the indexer for analysis. The ingest rate can be as low as many gigabytes per day or as high as multiple terabytes per day.

In any case, an ingest rate of one terabyte per day equates to less than 12 megabytes per second. This volume of data can be easily handled by any model of HPE Nimble Storage array.

Periodic Reads

Read workloads in a Splunk Enterprise environment are much less predictable and vary depending on the following conditions:

- The number of users with access to the environment
- The number and type of installed apps
- The restrictions that are placed on the search scope by the Splunk Enterprise administrator

Capacity Requirements

The capacity requirements of Splunk Enterprise are a function of the daily ingest rate and of the data retention policy. For a one-terabyte-per-day ingest rate and a retention policy of 180 days, the total amount of required

data capacity is approximately 180 TB. However, based on data collected from over 150 customers who use HPE Nimble Storage solutions for Splunk Enterprise, HPE InfoSight predictive analytics has been able to determine that the average compression rate for Splunk Enterprise data is 1.5x, which effectively reduces the required capacity to 120 TB.

The use of a *replication factor* might increase the capacity demand of Splunk Enterprise. The replication factor is the number of copies of indexer data that are replicated to peer indexers to improve the performance of searches and reduce the risk of data loss if a server fails. Depending on the criteria for retention and search that are set by the Splunk Enterprise administrator, this demand might be spread across the hot, warm, cold, and frozen buckets.

Server Sizing

The daily ingest rate determines the number and role of the servers that are required for a Splunk Enterprise deployment. Each indexer is capable of processing a certain amount of data per day; therefore, larger deployments need more configured indexers to support the daily ingest rate.

Table 2: Sizing requirements for Splunk Enterprise deployments

Component or Function	Departmental	Small Enterprise	Medium Enterprise	Large Enterprise
Daily indexing volume	0–20 GB	20–100 GB	100–300 GB	300 GB–1 TB+
Number of forwarders	Median < 10; max. 100	Median in the 10s; max. in the 100s	Median in the 10s; max. in the low 100s	Median in the 10s; max. in the 1000s
Number of users	Median < 10	Median in the 10s	Median in the 10s; max. in the low 100s	Median in the 10s; max. 500+
Number of apps	1–10	1–10	1–20+	10–50
Number of indexers	1 indexer	2–3 indexers	4–9 indexers	10+ indexers
Number of search heads	Search head combined with indexer	1 standalone search head	2 search heads	3+ search heads
Configuration of management function	Manual configuration or deployment server	Manual configuration or deployment server	Deployment server or third-party tool	Deployment server or third-party tool

Sizing Tools for Splunk Enterprise

Splunk provides a *sizing tool* for customers and partners to estimate the capacity and the number of servers that are required to support a Splunk Enterprise deployment. A *sizing questionnaire* is also available on HPE InfoSight to help them collect the information for input into the Splunk sizer and decide which HPE Nimble Storage array model is the best fit for the deployment.

The following table shows sample configurations and recommended storage systems for different sizes of Splunk Enterprise deployment. Customers should work with their Splunk architects and their account team to determine the correct solution for their capacity and performance requirements.

Table 3: Sample configurations and storage recommendations for Splunk Enterprise deployments

Configuration	Small Size	Medium Size	Large Size	X-Large Size
Ingest rate (GB/day)	100	250	500	1000

Configuration	Small Size	Medium Size	Large Size	X-Large Size
Retention (days)	90	90	90	90
Number of search heads	1	2	5	10
Search head capacity (GB)	600	1200	3000	6000
Search head IOPS	400	800	2000	4000
Number of indexers	3	6	15	30
Indexer capacity (GB)	4500	11,250	22,500	45,000
Indexer IOPS	3600	7200	18,000	36,000
Total IOPS	4000	8000	20,000	40,000
Total disk capacity (GB)	5100	12,450	25,500	51,000
Flash capacity (GB)	510	1245	2550	5100
Recommended HPE Nimble Storage array model	CS1000H or AF1000	CS1000 or AF1000	CS3000 or AF3000	CS5000 or AF5000

Deploying Splunk Enterprise on HPE Nimble Storage Arrays

The deployment of Splunk Enterprise on HPE Nimble Storage arrays involves a number of considerations. The choice between a distributed configuration or a centralized configuration largely depends on the ingest rate and the performance requirements of the environment. For example, a development or test system will not have the same requirements as a production environment.

Storage Configuration

HPE Nimble Storage systems offer many configuration options to meet the requirements of Splunk Enterprise. In particular, the HPE Nimble Storage predictive cloud platform enables customers to choose the ideal combination of performance, capacity, and flexibility for their workloads in a cloud-ready infrastructure.

Protocol Choice

Although Splunk Enterprise can be deployed on servers that have only internal disk drives, Splunk strongly recommends using shared storage for larger environments. Splunk does not make any recommendation to use either iSCSI or Fibre Channel (FC) for shared-storage environments.

For most customers, the choice of protocol depends on the protocol that they already have in place or the protocol that they are most comfortable implementing. Because HPE Nimble Storage systems perform equally well with iSCSI and FC and support the same features and functionality with either protocol choice, customers are free to use whichever protocol they prefer.

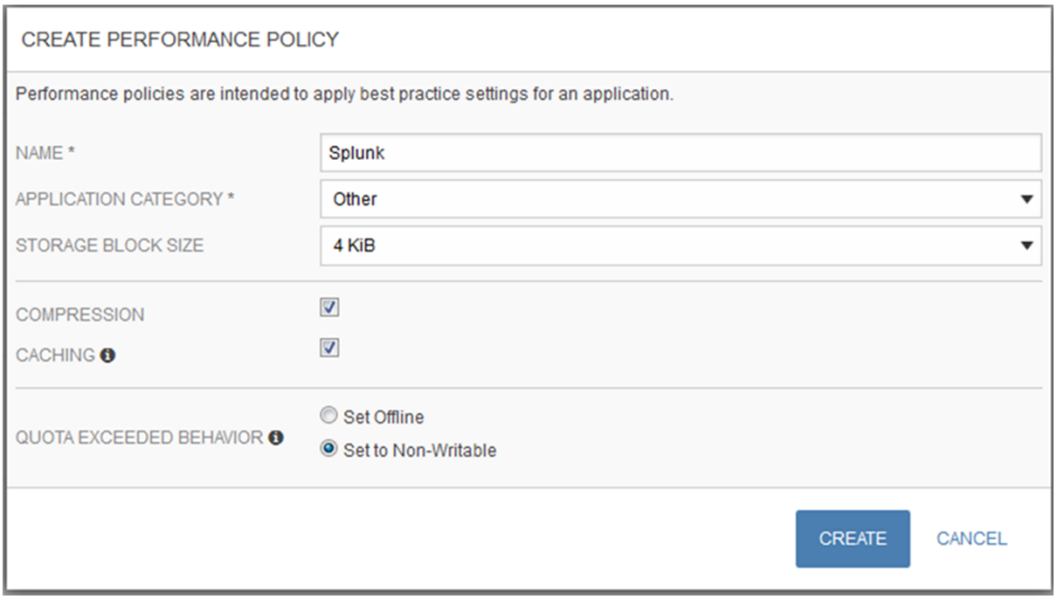
Performance Policy

A performance policy defines the block size, the data reduction strategy, and other characteristics for a storage volume. The NimbleOS operating system that runs on HPE Nimble Storage arrays comes with predefined performance policies for many applications. However, Splunk Enterprise is a unique workload, and NimbleOS does not have a predefined performance policy for Splunk software.

HPE recommends using a 4 KB block size for Splunk Enterprise workloads running on HPE Nimble Storage systems. Customers should create a custom performance policy to help organize and identify Splunk Enterprise volumes.

To create a performance policy, navigate to **Manage > Performance Policies** in the NimbleOS GUI and click **Add**. In the **Create Performance Policy** dialog box, choose a name, the application category, and other settings for the performance policy. HPE recommends enabling compression and caching for the performance policy on all-flash arrays and adaptive-flash arrays.

Figure 3: Example of NimbleOS performance policy for Splunk workloads



Deduplication can provide space saving benefits for large Splunk Enterprise environments with high replication factors. However, with a low replication factor of 2 or a small dataset, deduplication does not yield significant space savings relative to the overhead of processing deduplicated blocks. If the environment has a requirement for highly parallelized search performance and a high replication factor, deduplication can be enabled on adaptive-flash arrays and all-flash arrays to reduce total capacity requirements.

Number of Volumes or LUNs

Using multiple LUNs in a volume group and striping data across those LUNs can be beneficial for achieving a larger logical volume for both capacity and performance. However, HPE does not recommend this configuration for small developments or test environments. In these types of systems, multiple LUNs bring very little improvement in performance at the cost of additional complexity and management effort.

For large production environments, multiple LUNs can improve performance if they are implemented correctly. When creating the logical volume, specify the number of devices in the volume group to ensure that data is written equally across all devices.

Table 4: Recommended number of HPE Nimble Storage volumes for Splunk Enterprise workloads

System Type	Number of Volumes	Volume Usage
Development	1 to 4	Single volume with all buckets or 1 volume per bucket
Test	4 to 16	1 volume per bucket or Up to 4 volumes per bucket with Logical Volume Manager (LVM)

System Type	Number of Volumes	Volume Usage
Production	4 to 32	1 volume per bucket or Up to 8 volumes per bucket with LVM

Splunk Enterprise in the HPE Nimble Storage Predictive Cloud Platform

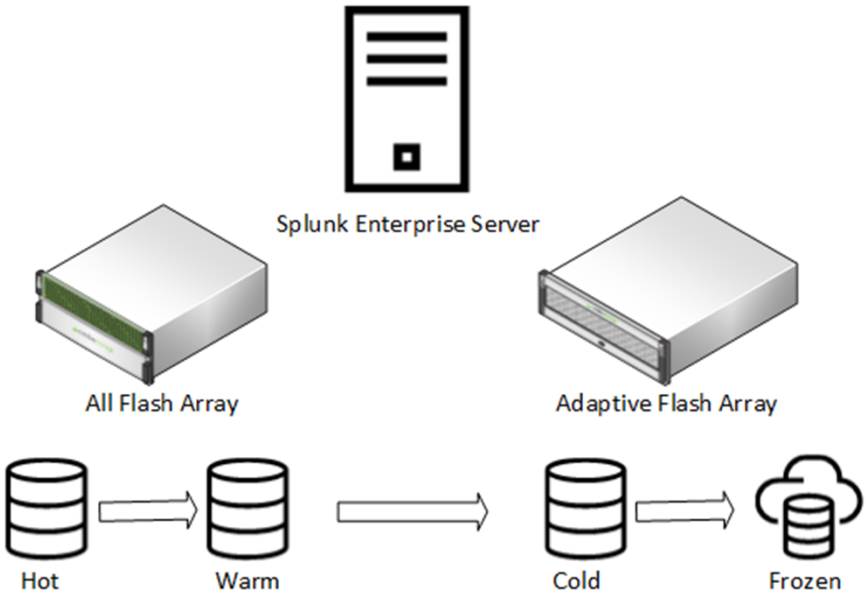
The core components of the HPE Nimble Storage predictive cloud platform are HPE Nimble Storage all-flash arrays, HPE Nimble Storage adaptive-flash arrays, HPE Cloud Volumes, and HPE InfoSight. Together, these components give customers the infrastructure necessary to manage the Splunk Enterprise data lifecycle in the most efficient way by providing a highly available platform that identifies issues before they affect the application.

Splunk Enterprise data is categorized into four buckets: hot, warm, cold, and frozen. Each of these categories has properties that require different performance and capacity levels. With the HPE Nimble Storage predictive cloud platform, customers can choose the best-fit storage that meets their initial needs. Later, they can add arrays and migrate data nondisruptively as those needs change.

The diagram that follows exemplifies a storage configuration for a high-performance environment that uses separate volumes for each bucket of data:

- Hot data and warm data, which are queried more often, reside on an all-flash array for the highest level of performance. Hot data can also be pinned in memory on an adaptive-flash array.
- Cold data, which has a higher capacity requirement and is queried less frequently, resides on an adaptive-flash array.
- Frozen data, which is not searched, is stored on a cloud volume for long-term retention.

Figure 4: Example of storage layout for Splunk Enterprise



This example is only one possible configuration. Customers should work with their HPE account team to determine which HPE Nimble Storage solution works best for their specific requirements.

Operating System Configuration

The storage environment for a Linux-based deployment of Splunk Enterprise can be configured in several ways. Customers should follow the Splunk recommendations for tuning the operating system.

HPE Nimble Storage Connection Manager

The HPE Nimble Storage Connection Manager (NCM) for Linux assists administrators with the task of configuring the operating system to connect to the storage system. This utility can be downloaded from HPE InfoSight along with the guide for installing the tool. For more information about HPE NCM for Linux, visit *HPE InfoSight*.

Although the HPE NCM for Linux is not required for connections to the storage system, it provides an easy way to adjust the operating system settings to optimize the environment.

File System

HPE tested Splunk Enterprise on HPE Nimble Storage arrays with both the XFS and the EXT4 file systems. During testing, there was no significant difference in performance between these two file system types.

Many different file system mount options can be set for XFS and EXT4. Although Splunk does not make any recommendations about file system mount options, HPE recommends setting **noatime** and **nodiratime** at a minimum.

Logical Volume Manager

In a Linux environment, Logical Volume Manager (LVM) simplifies the management of physical devices by combining them into volume groups and logical volumes. LVM gives administrators the ability to combine multiple physical devices or LUNs into a single entity for management and performance reasons.

The following example code creates a volume group, a logical volume, and a file system for Splunk Enterprise data:

```
vgcreate vgSplunkData /dev/mapper/mpathj
lvcreate -n lvSplunkData -L 200G vgSplunkData
mkfs -t ext4 /dev/vgSplunkData/lvSplunkData
```

The following example code creates a volume group and a logical volume with two LUNs, and then creates a file system that uses the logical volume:

```
vgcreate vgSplunkData /dev/mapper/mpathk /dev/mapper/mpathl
lvcreate -i 2 -n lvSplunkData -L 500G vgSplunkData
mkfs -t ext4 /dev/vgSplunkData/lvSplunkData
```

Virtualization

Splunk Enterprise can be installed in virtualized environments. Splunk provides guidance for the configuration of a VMware® environment to support the requirements of Splunk software.

Be sure to review the *Splunk documentation* for running Splunk Enterprise on the VMware virtualization platform.

VMware Integration with HPE Nimble Storage Systems

HPE InfoSight has documentation for the deployment of the VMware integration with HPE Nimble Storage products. Another available resource is the HPE Nimble Storage Connection Manager (NCM) for VMware, which optimizes the hypervisor for HPE Nimble Storage arrays.

Be sure to review the *HPE Nimble Storage documentation* and install HPE NCM for VMware before provisioning storage for Splunk Enterprise or for other workloads on a virtualized environment. For more information about HPE NCM for VMware, visit *HPE InfoSight*.

Performance Policy for Splunk Enterprise on a VMware Environment

The goal of choosing a performance policy for particular workloads in a bare-metal server configuration is to match the common block size of the workload for best performance. Because VMware virtualization adds an abstraction layer to the environment for storage resources, HPE recommends using the preconfigured NimbleOS performance policy for VMware when running Splunk Enterprise on the VMware platform.

VMware Storage Selection

VMware offers a number of options for provisioning and attaching storage to virtual machines. Raw device mapping (RDM), VMDK, Virtual Volumes (VVols), and iSCSI storage directly attached to hosts are all connection options supported by HPE Nimble Storage arrays. Each method of connection provides different benefits and might change the way in which the HPE Nimble Storage features and functionality operate.

Work with your account team to understand which storage connection method best meets the provisioning, performance, and data protection needs of your Splunk Enterprise environment.

Containers

A form of virtualization that is growing rapidly is the use of containers. Containers allow for an extremely dense and flexible resource utilization. Splunk has invested a great deal in support for containerizing Splunk forwarders and Splunk Enterprise security.

Some HPE Nimble Storage tools are built into HPE NCM for VMware to enable customers to take advantage of persistent storage for container workloads. For more information about HPE Nimble Storage container support, visit *HPE InfoSight*.

Data Protection

Splunk Enterprise is often deployed on servers with internal disk drives for storage. This affordable solution enables customers to scale by adding more servers to meet performance requirements. However, because this configuration does not offer the data resiliency features of modern storage arrays, customers need to create copies of their data.

To create those copies, Splunk Enterprise administrators set up a replication factor, which indicates how many copies of the data exist in the environment. Replication improves performance and prevents data loss in the case of a server failure. For example, in a landscape with three indexers, a replication factor of 3 indicates that each of the indexers has a copy of the others' indexes. The trade-off is that this configuration equates to three times the storage requirement for a single indexer.

HPE Nimble Storage arrays have an extremely high level of availability and resiliency, offering a guarantee of 99.9999% uptime. For this reason, it is not necessary to maintain many replicas for data protection. Still, replicating data between indexers for improved search performance is beneficial. With the HPE Nimble Storage data reduction features, the increased storage that is required for the replicas is optimized, so the environment is unlikely to require three times more storage.

To prevent data from becoming unavailable as a consequence of a server failure, HPE recommends a replication and search factor of 2. This setting not only prevents data unavailability in the server failure scenario but also increases search speed.

Snapshot Copies and Replication

One of the most valuable features of HPE Nimble Storage systems is the ability to create snapshot copies of data. In the HPE Nimble Storage predictive cloud platform, a volume collection can be configured to create

a snapshot of volumes on adaptive-flash arrays and all-flash arrays to ensure the consistent backup of all Splunk Enterprise data buckets.

Snapshot copies can also be replicated to HPE Nimble Storage arrays for data protection and long-term retention. This process is very efficient because after the initial copy is moved to the secondary array, only blocks that have been changed are sent to the secondary array.

An added benefit of snapshot copies is the ability to create zero-copy clones for testing and development. Because zero-copy clones are not a physical copy of the data, they are extremely fast and efficient to create. Zero-copy clones can be created on primary or secondary storage without affecting the source volume or interrupting the replication process.

Summary and Useful Resources

Splunk Enterprise has become an important tool for IT professionals. The ability to monitor and analyze huge volumes of data from many disparate data sources is a crucial part of managing IT environments. HPE offers an ideal cloud-ready storage solution—the HPE Nimble Storage predictive cloud platform—that addresses the entire lifecycle of Splunk Enterprise data.

Splunk Resources

- *Splunk Storage Sizing*
- *Splunk Docs*

HPE Nimble Storage Resources

- *HPE Nimble Storage Sizing Questionnaire for Splunk Enterprise Deployments*
- *HPE Nimble Storage Connection Manager (NCM) for Linux*
- *HPE Nimble Storage Connection Manager (NCM) for VMware*
- *HPE Nimble Storage documentation*

About the Author

Craig Sullivan

Master Technical Marketing Engineer

Hewlett Packard Enterprise

Craig Sullivan has over 20 years of experience with SAP Basis and technical architecture. For the first 10 years of his career, Craig was a Basis administrator of global SAP implementations for several companies and consulting firms. In 2010, he brought his skills to the storage industry to develop products and solutions for SAP applications, including SAP HANA. Craig has applied his many years of experience with enterprise applications and storage technologies to helping customers integrate HPE Nimble Storage solutions into their SAP and Splunk environments.

Version History

Version	Release Date	Description
1.0	November 2017	Initial release