Technical Report

NetApp E-Series E5700 and Splunk Enterprise

Mitch Blackburn, NetApp

September 2017 | TR-4623

Abstract

This technical report describes the integrated architecture of the NetApp® E-Series and Splunk design. Optimized for node storage balance, reliability, performance, storage capacity, and density, this design employs the Splunk clustered index node model, with higher scalability and lower TCO. Decoupling storage from compute provides the ability to scale each separately, saving the cost of overprovisioning one or the other. In addition, this document summarizes the performance test results obtained from a Splunk machine log event simulation tool.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Introduction

NetApp E-Series enables Splunk environments to maintain the highest levels of performance and uptime for workloads by providing advanced fault recovery features and easy in-service growth capabilities to meet ever-changing business requirements. By decoupling storage from compute, you gain the ability to scale capacity and compute separately, saving the cost of overprovisioning one or the other.

The E-Series is designed to handle the most extreme application workloads with very low latency. Typical use cases include application acceleration; improving the response time of latency-sensitive applications; and improving the power, environmental, and capacity efficiency of overprovisioned environments. E-Series storage systems leverage the latest solid-state disk (SSD) and SAS drive technologies and are built on a long heritage of serving diverse workloads to provide superior business value and enterprise-class reliability.

Splunk is the leading operational intelligence software. It enables you to monitor, report, and analyze live streaming and historical machine-generated data, whether it is located on the premises or in the cloud. An organization's IT data is a definitive source of intelligence because it is a categorical record of activity and behavior, including user transactions, customer behavior, machine behavior, security threats, and fraudulent activity. Splunk helps users gain visibility into this machine data to improve service levels, reduce IT operations costs, mitigate security risks, enable compliance, and create new product and service offerings. Splunk offers solutions for IT operations, applications management, security and compliance, business analytics, and industrial data.

This technical report describes the integrated architecture of the NetApp E-Series E5700 and Splunk Enterprise 6.6. The E5700 used in these tests used SANtricity® release 11.40.

All testing was carried out using Dynamic Disk Pools (DDP). Primary areas of E-Series E5700 performance test include:

- All SSDs for hot tier (equivalent to a NetApp E-Series EF570 all-flash array)
- All 10K SAS hard disk drives (HDDs) for hot tier
- All NL-SAS 7.2K HDD cold tier and SSD or 10K SAS hot tier
- E5700 configured with DDP for enhanced recovery times when a drive fails (performance on failure) for both tiers

2 Splunk Overview

All of your IT applications, systems, and technology infrastructure generate data every millisecond of every day. This machine data is one of the fastest growing and most complex areas of big data. Splunk collects all of your data sources—streaming and historical—by using a technology called universal indexing. Splunk is scalable enough to work across all of your data centers, and it is powerful enough to deliver real-time dashboard views to any level of the organization. However, using this data can be a challenge for traditional data analysis, monitoring, and management solutions that are not designed for large-volume, high-velocity diverse data.

Splunk offers a unique way to sift, distill, and understand these immense amounts of machine data, which can change how IT organizations manage, analyze, secure, and audit IT. Splunk enables users to develop valuable insights into how to innovate and offer new services as well as into trends and customer behaviors.

2.1 Primary Use Cases

Splunk can be deployed for use in a wide variety of use cases, and it provides creative ways for users to gain intelligence from data.

Application Delivery

Gain end-to-end visibility across distributed infrastructures, troubleshoot application environments, monitor performance for degradation, and monitor transactions across distributed systems and infrastructure.

Security, Compliance, and Fraud

Enable rapid incident response, real-time correlation, and in-depth monitoring across data sources. Conduct statistical analysis for advanced pattern detection and threat defense.

Infrastructure and Operations Management

Proactively monitor across IT silos to enable uptime, rapidly pinpoint and resolve problems, identify infrastructure service relationships, establish baselines, and create analytics to report on SLAs or track service provider SLAs.

Business Analytics

Provide visibility and intelligence related to customers, services, and transactions. Recognize trends and patterns in real time and provide valuable understanding of new product features' impact on back-end services. Gain valuable understanding of the user experience for greater user satisfaction and prevent drop-offs, improve conversions, and boost online revenues.

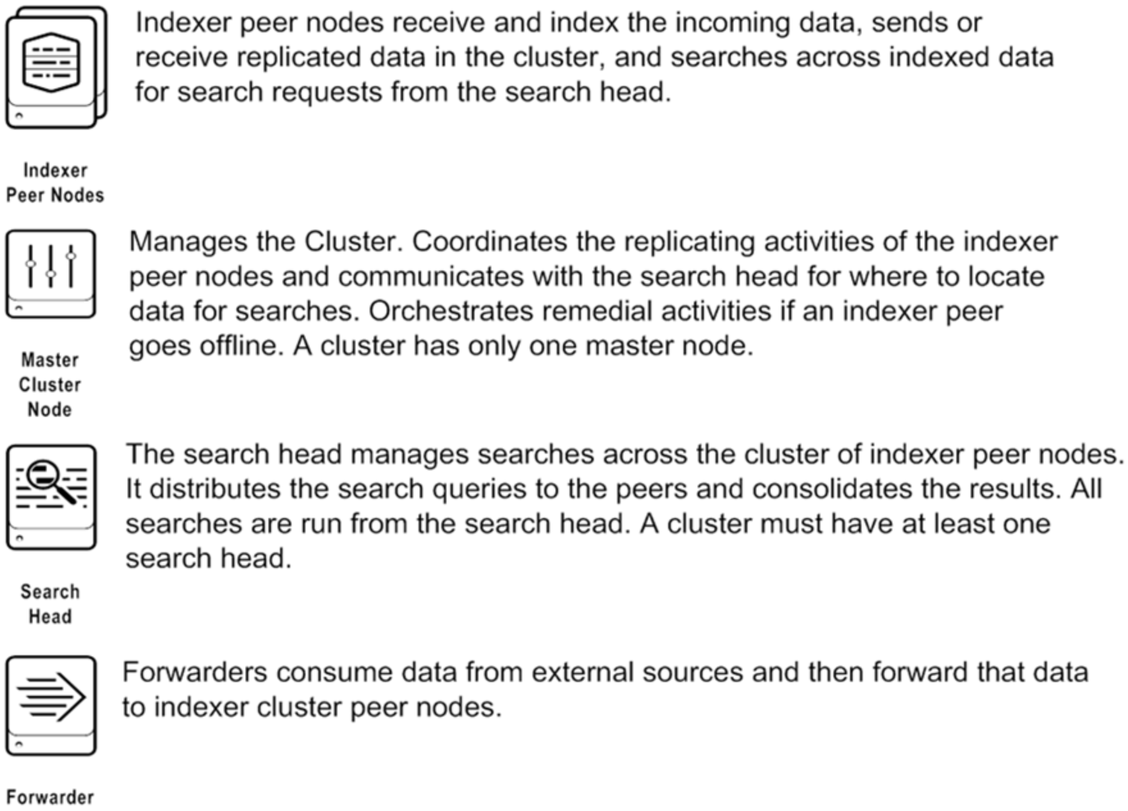
2.2 Splunk Architecture Overview

Splunk's architecture provides linear scalability for indexing and distributed search. Splunk's implementation of MapReduce allows large-scale search, reporting, and alerting. Splunk takes a single search and enables you to query many indexers in massive parallel clusters. With the addition of index replication, you can specify how many copies of the data you want to make available to meet your availability requirements.

The Splunk platform is open and has SDKs and APIs, including a REST API and SDKs for Python, Java, JavaScript, PHP, Ruby, and C#. This capability enables developers to programmatically interface with the Splunk platform. With Splunk you can develop your own applications or templates to deploy on your infrastructure.

Splunk can write the indexed data to clustered servers to add additional copies of the raw file and metadata using replication so that the data is available even during indexer node failures. Common Splunk server components are shown in Figure 1.

Figure 1) Splunk cluster server components.



Indexer peer nodes receive and index the incoming data, sends or receive replicated data in the cluster, and searches across indexed data for search requests from the search head.

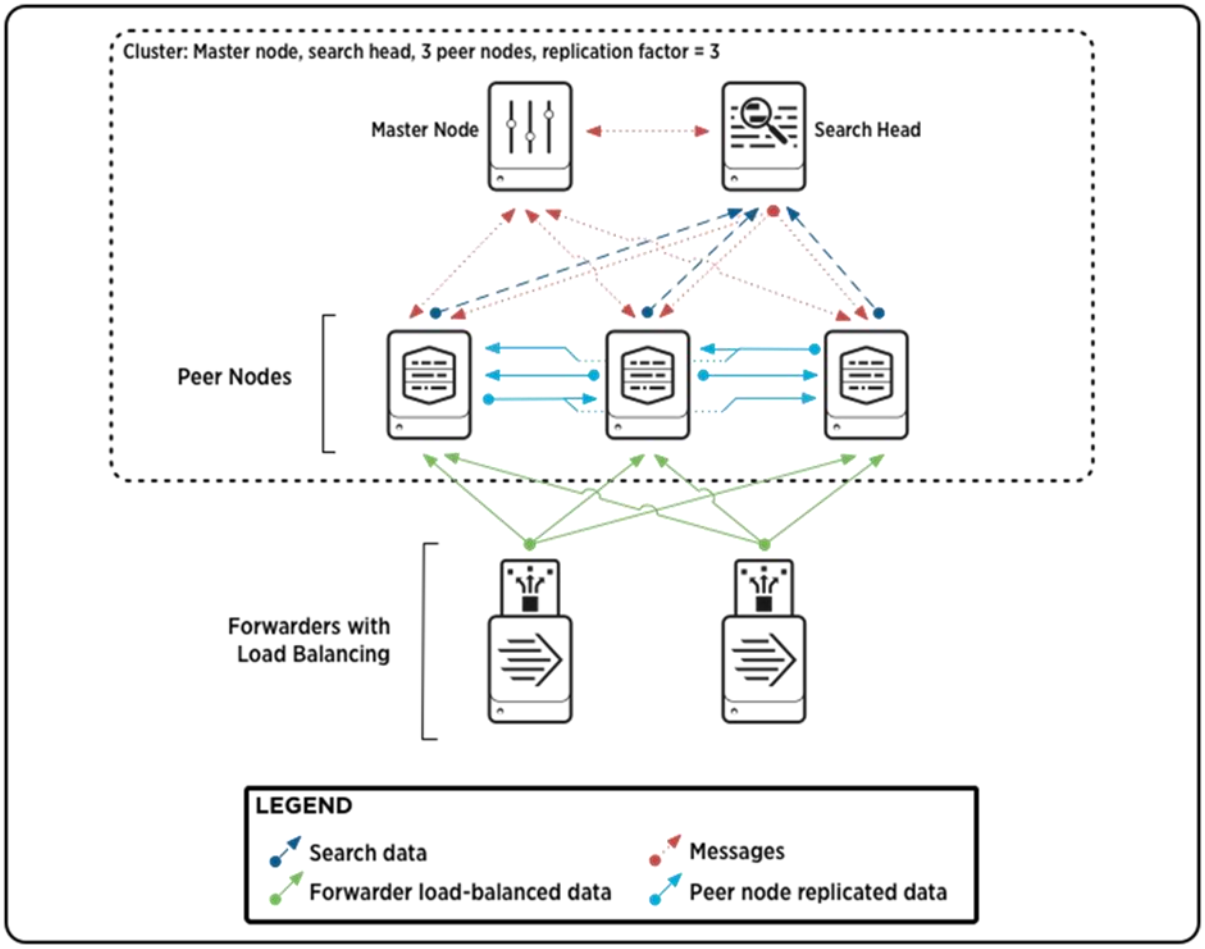
Manages the Cluster. Coordinates the replicating activities of the indexer peer nodes and communicates with the search head for where to locate data for searches. Orchestrates remedial activities if an indexer peer goes offline. A cluster has only one master node.

The search head manages searches across the cluster of indexer peer nodes. It distributes the search queries to the peers and consolidates the results. All searches are run from the search head. A cluster must have at least one search head.

Forwarders consume data from external sources and then forward that data to indexer cluster peer nodes.

The common and recommended replication factor for Splunk running internal direct-attached storage (DAS) is three. In this scenario, the minimum number of needed Splunk index servers is also three. Figure 2 shows the basic Splunk cluster configuration.

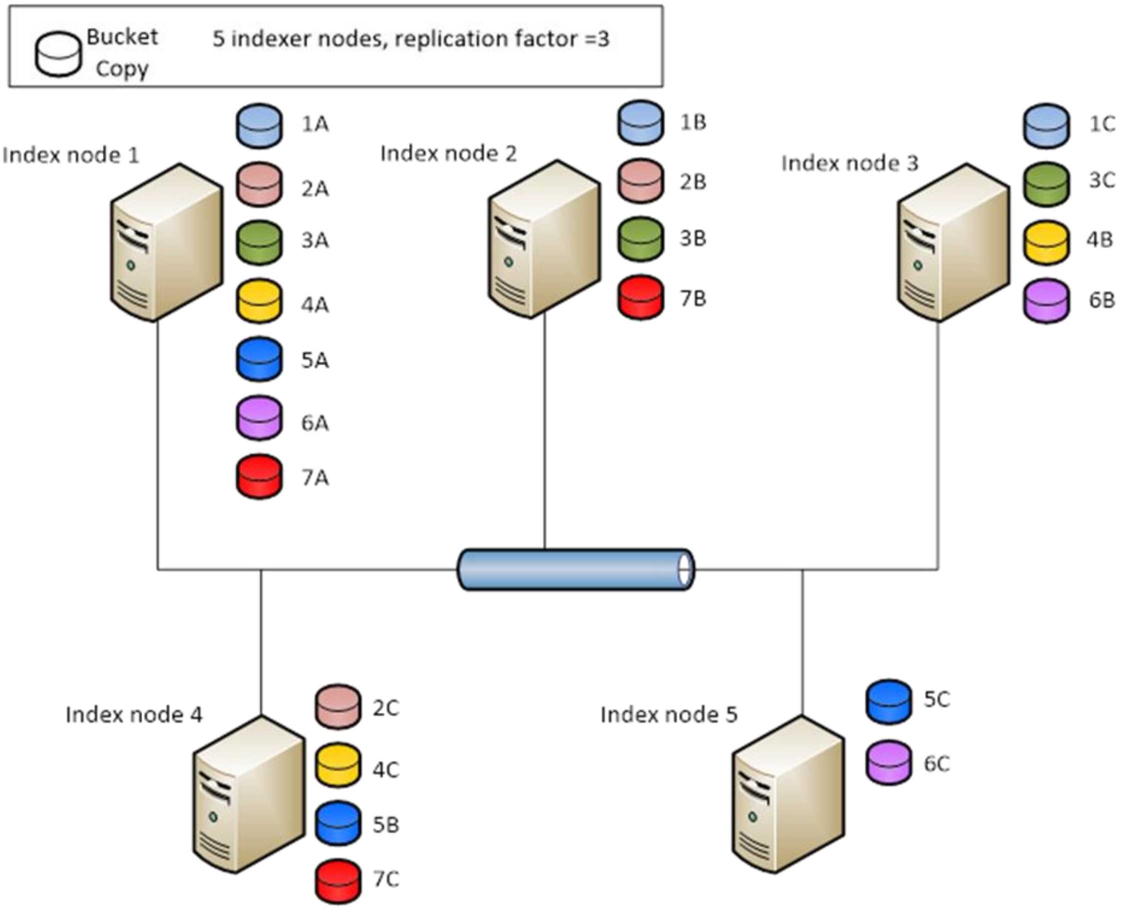
Figure 2) Basic Splunk cluster configuration.


Cluster: Master node, search head, 3 peer nodes, replication factor = 3
Master Node
Search Head
Peer Nodes
Forwarders with Load Balancing
LEGEND
Search data
Forwarder load-balanced data
Messages
Peer node replicated data

The machine log data from the Splunk forwarders sent to the indexer peer nodes uses the recommended data replication factor of three, which makes available three copies of data. The ingested data is compressed and indexed as raw data files and metadata, which are then distributed among the indexer peer nodes for redundancy. Figure 3 depicts the way that Splunk replicates data in a five-indexer cluster.

Splunk places your indexed data in directories, also referred to as "buckets," as it moves through its lifecycle in Splunk. When data is first indexed, it goes in db-hot; then, according to your data policy definitions, it moves into the warm buckets, then cold, and finally "frozen" (which by default means it is deleted). Each time db-hot is rolled to warm, it creates a new directory (known as a warm bucket) named to indicate the time range of the events in that bucket. Rolling to warm occurs automatically when the specified bucket size is reached, so the buckets are all typically the same size unless you have rolled manually at some point. In our configuration, the hot/warm tier uses a volume created on the SSDs. Each time the warm bucket is rolled to the cold bucket, it creates a new directory on the E-Series volume containing the cold tier. In our case this directory is created using the NL-SAS drives. For more information about how buckets work in Splunk, see Understanding how "buckets" work.

Figure 3) Distribution of data in a five-node Splunk cluster.



3 NetApp E-Series Overview

The E-Series E5700 is an industry-leading storage system that delivers high input/output operations per second (IOPS) and bandwidth with consistently low latency to support the demanding performance and capacity needs of science and technology, simulation modeling, and decision support environments. In addition, the E5700 is equally capable of supporting primary transactional databases, general mixed workloads, and dedicated workloads such as video analytics in a highly efficient footprint with extreme simplicity, reliability, and scalability.

The E5700 provides the following benefits:

- Support for wide-ranging workloads and performance requirements
- Fully redundant I/O paths, advanced protection features, and proactive support monitoring and services for high levels of availability, integrity, and security
- Increased IOPS performance by up to 20% compared to the previous high-performance generation of E-Series products
- A level of performance, density, and economics that leads the industry
- Interface protocol flexibility to support FC host and iSCSI host workloads simultaneously
- Support for private and public cloud workloads behind virtualizers such as FlexArray®, Veeam Cloud Connect, and StorageGRID®

3.1 E-Series Hardware Overview

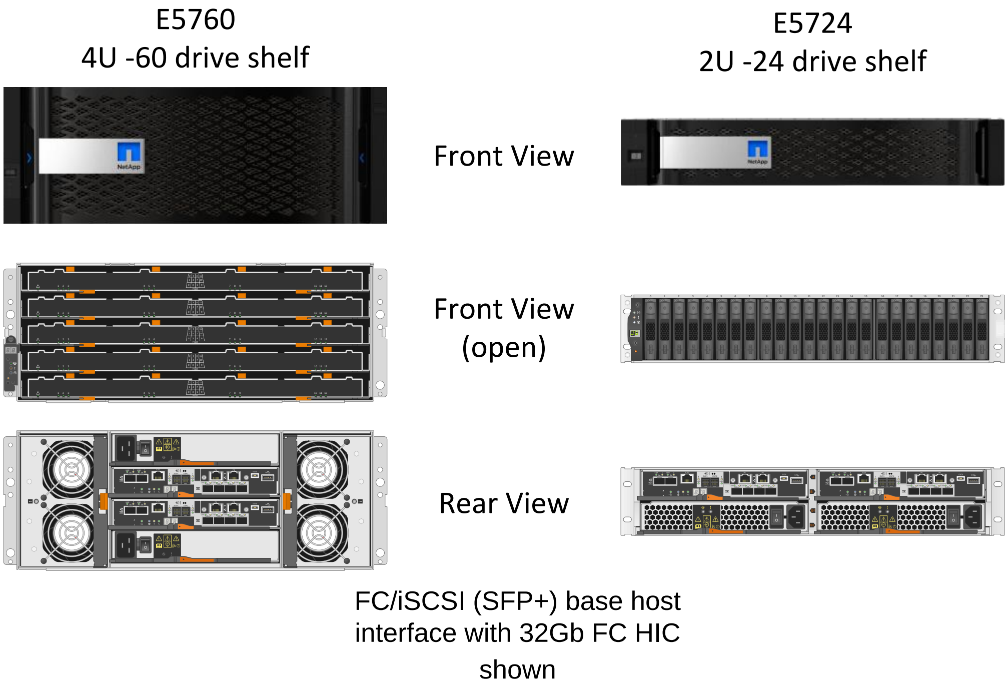
As shown in Table 1, the E5700 is available in two shelf options, which support both HDDs and SSDs to meet a wide range of performance and application requirements.

Table 1) E5700 controller shelf and drive shelf models.

Controller Shelf Model	Drive Shelf Model	Number of Drives	Type of Drives
E5760	DE460C	60	2.5” and 3.5” SAS drives (HDDs and SSDs)
E5724	DE224C	24	2.5” SAS drives (HDDs and SSDs)

Both shelf options include dual-controller modules, dual power supplies, and dual fan units for redundancy. The 24-drive shelf has integrated power and fan modules. The shelves are sized to hold 60 drives or 24 drives, as shown in Figure 4.

Figure 4) E5700 controller drive shelf options.



Each E5700 controller shelf includes two controllers, with each controller providing two Ethernet management ports for out-of-band management. The system has two 12Gbps (x 4 lanes) wide-port SAS drive expansion ports for redundant drive expansion paths. The E5700 controllers also include two built-in host ports, which can be configured as either 16Gb FC or 10Gb iSCSI. The following host interface cards (HICs) can be installed in each controller:

Note: Both controllers in an E5700 array must be identically configured.

- 4-port 12Gb SAS wide port (SAS-3 connector)
- 4-port 32Gb Fibre Channel
- 4-port 25Gb iSCSI
- 2-port 100Gb InfiniBand (IB)

3.2 SANtricity Software

E5700 systems are managed by the SANtricity System Manager browser-based application. The SANtricity System Manager 11.40 is embedded on the controller.

To create volume groups on the array, the first step when configuring SANtricity is to assign a protection level. This assignment is then applied to the disks selected to form the volume group. The E5700 storage systems support DDP as well as RAID levels 0, 1, 5, 6, and 10. DDP was used for all configurations described in this document.

To simplify the storage provisioning, NetApp SANtricity provides an automatic configuration feature. The configuration wizard analyzes the available disk capacity on the array. It then selects disks that maximize array performance and fault tolerance while meeting capacity requirements, hot spares, and any other criteria specified in the wizard.

For further information about the SANtricity Storage Manager and the SANtricity System Manager, see the E-Series Documentation Center.

Dynamic Storage Functionality

From a management perspective, SANtricity offers a number of capabilities to ease the burden of storage management, including the following:

- New volumes can be created and are immediately available for use by connected servers.
- New RAID sets (volume groups) or disk pools can be created at any time from unused disk devices.
- Dynamic volume expansion allows capacity to be added to volumes online as needed.
- Dynamic capacity expansion allows disks to be added to volume groups and disk pools online to meet any new requirements for capacity or performance.
- Dynamic RAID migration allows the RAID level of a particular volume group to be modified online if new requirements dictate a change, for example, from RAID 10 to RAID 5.
- Flexible cache block and dynamic segment sizes enable optimized performance tuning based on a particular workload. Both items can also be modified online.
- Online controller firmware upgrades and drive firmware upgrades are possible.
- Path failover and load balancing (if applicable) between the host and the redundant storage controllers in the E5700 are provided. See the Multipath Drivers Guide for more information.

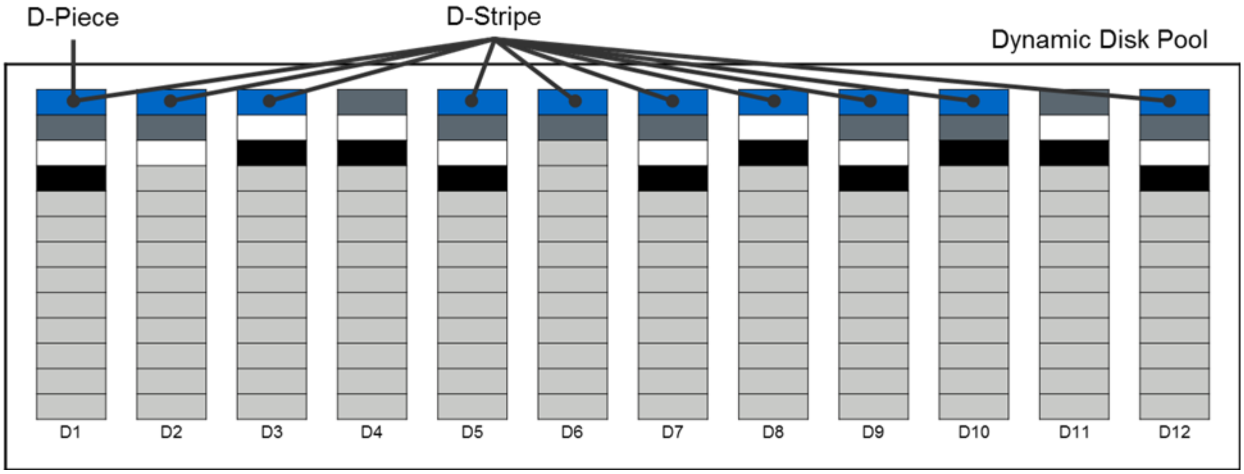
Dynamic Disk Pools

With seven patents pending, the DDP feature dynamically distributes data, spare capacity, and protection information across a pool of disk drives. These pools can range in size from a minimum of 11 drives to all the supported drives in a system. In addition to creating a single DDP, storage administrators can opt to mix traditional volume groups and DDP or even multiple DDPs, offering an unprecedented level of flexibility.

Dynamic Disk Pools are composed of several lower-level elements. The first of these is a D-piece. A D-piece consists of a contiguous 512MB section from a physical disk that contains 4,096 128KB segments. Within a pool, 10 D-pieces are selected using an intelligent optimization algorithm from selected drives within the pool. Together, the 10 associated D-pieces are considered a D-stripe, which is 4GB of usable capacity in size. Within the D-stripe, the contents are similar to a RAID 6 8+2 scenario. There, 8 of the underlying segments potentially contain user data, 1 segment contains parity (P) information calculated from the user data segments, and 1 segment contains the Q value as defined by RAID 6.

Volumes are then created from an aggregation of multiple 4GB D-stripes as required to satisfy the defined volume size up to the maximum allowable volume size within a DDP. Figure 5 shows the relationship between these data structures.

Figure 5) Dynamic Disk Pool components.

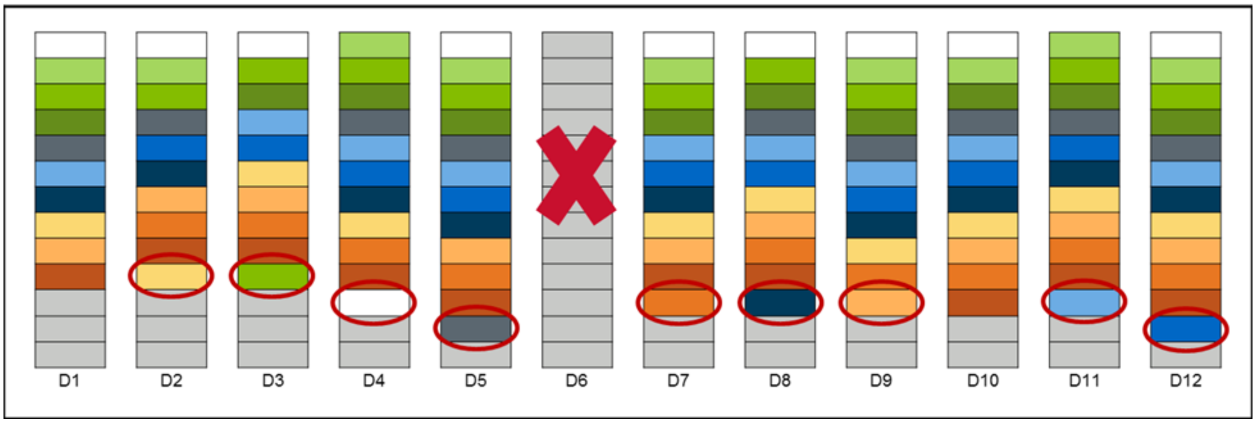


Another major benefit of a DDP is that, rather than using dedicated stranded hot spares, the pool contains integrated preservation capacity to provide rebuild locations for potential drive failures. This approach simplifies management, because individual hot spares no longer need to be planned or managed. The approach also greatly improves the time for rebuilds, if required, and enhances volume performance during a rebuild, as opposed to traditional hot spares.

When a drive in a DDP fails, the D-pieces from the failed drive are reconstructed to potentially all other drives in the pool using the same mechanism normally used by RAID 6. During this process, an algorithm internal to the controller framework verifies that no single drive contains two D-pieces from the same D-stripe. The individual D-pieces are reconstructed at the lowest available logical block address (LBA) range on the selected disk drive.

In Figure 6, disk drive 6 (D6) is shown to have failed. Subsequently, the D-pieces that previously resided on that disk are recreated simultaneously across several other drives in the pool. Because there are multiple disks participating in the effort, the overall performance impact of this situation is lessened, and the length of time needed to complete the operation is dramatically reduced.

Figure 6) Dynamic Disk Pool drive failure.



When multiple disk failures occur within a DDP, priority for reconstruction is given to any D-stripes missing two D-pieces to minimize data availability risk. After those critically affected D-stripes are reconstructed, the remainder of the necessary data is reconstructed.

From a controller resource allocation perspective, there are two user-modifiable reconstruction priorities within DDP:

- Degraded reconstruction priority is assigned to instances in which only a single D-piece must be rebuilt for the affected D-stripes; the default for this value is high.
- Critical reconstruction priority is assigned to instances in which a D-stripe has two missing D-pieces that need to be rebuilt; the default for this value is highest.

For very large disk pools with two simultaneous disk failures, only a relatively small number of D-stripes are likely to encounter the critical situation in which two D-pieces must be reconstructed. As discussed previously, these critical D-pieces are identified and reconstructed initially at the highest priority. Doing so returns the DDP to a degraded state quickly so that further drive failures can be tolerated.

In addition to improving rebuild times and providing superior data protection, DDP can also greatly improve the performance of the base volume when under a failure condition compared to the performance of traditional volume groups.

For more information about DDP, see TR-4115: SANtricity Dynamic Disk Pools BPG.

E-Series Data Protection Features

E-Series has a reputation for reliability and availability. Many of the data protection features in E-Series systems can be beneficial in a Splunk environment.

Encrypted Drive Support

E-Series storage systems provide at-rest data encryption through self-encrypting drives. These drives encrypt data on writes and decrypt data on reads regardless of whether the full disk encryption (FDE) feature is enabled. Without the SANtricity feature enabled, the data is encrypted at rest on the media, but it is automatically decrypted on a read request.

When the FDE feature is enabled on the storage array, the drives protect the data at rest by locking the drive from reads or writes unless the correct security key is provided. This process prevents another array from accessing the data without first importing the appropriate security key file to unlock drives. It also prevents any utility or operating system from accessing the data.

SANtricity 11.40 further enhances the FDE feature by introducing the capability for users to manage the FDE security key using a centralized key management platform such as the Gemalto SafeNet KeySecure Enterprise Encryption Key Management, which adheres to the Key Management Interface Protocol (KMIP) standard. This feature is in addition to the internal security key management solution from earlier than SANtricity 11.40 and is available beginning with the E2800, E5700, and EF570.

The encryption and decryption performed by the hardware in the drive are invisible to the user and do not affect the performance or user workflow. Each drive has its own unique encryption key, which cannot be transferred, copied, or read from the drive. The encryption key is a 256-bit key as specified in the NIST Advanced Encryption Standard (AES). The entire drive, not just a portion, is encrypted.

Security can be enabled at any time by selecting the Secure Drives option in the Volume Group or Disk Pool menu. This selection can be made either at volume group or disk pool creation or afterward. It does not affect existing data on the drives and can be used to secure the data after creation. However, the option cannot be disabled without erasing all the data on the affected drive group or pool.

Figure 7 and Figure 8 show the technical components of NetApp E-Series FDE.

Figure 7) Technical components of NetApp E-Series FDE feature with an internally managed security key.

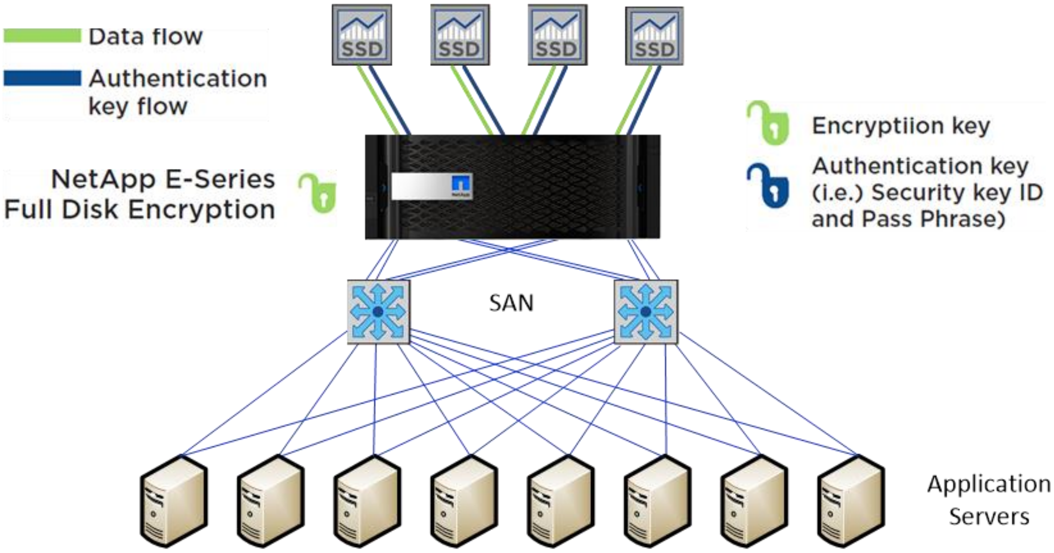
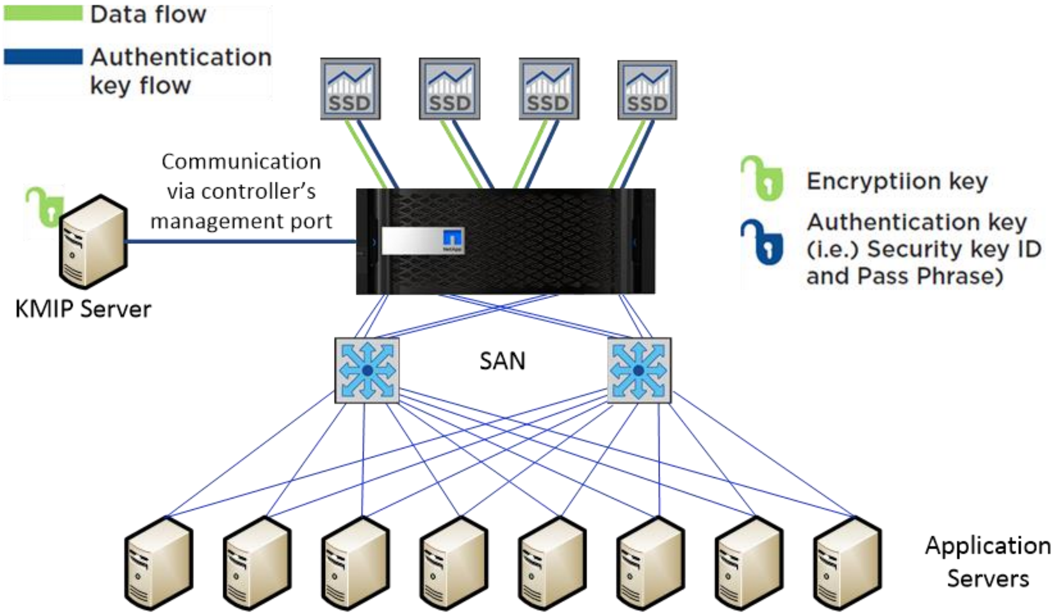

Data flow
Authentication key flow
SSD
SSD
SSD
SSD
NetApp E-Series Full Disk Encryption
Encryptiion key
Authentication key (i.e.) Security key ID and Pass Phrase)
SAN
Application Servers

Figure 8) Technical components of NetApp E-Series FDE feature with an externally managed security key.


Data flow
Authentication key flow
SSD
SSD
SSD
SSD
Communication via controller's management port
KMIP Server
Encryptiion key
Authentication key (i.e.) Security key ID and Pass Phrase)
SAN
Application Servers

For more information about disk encryption, see TR-4474: SANtricity Full Disk Encryption.

Background Media Scan

Media scan is a background process that is performed by the controllers to provide error detection on the drive media. The main purpose of the feature is to detect and repair media errors on disk drives that are infrequently read by user applications and where data loss might occur if other drives in the volume group fail. A secondary purpose is to detect redundancy errors such as data/parity mismatches. A background media scan can find media errors before they disrupt normal drive reads and writes.

Data Assurance (T10 PI)

The data assurance feature provides controller-to-drive data integrity protection through the SCSI direct access block device protection information model. This model protects user data by appending protection information to each block of user data. The protection model is sometimes referred to as data integrity field protection, or T10 PI. This model makes sure that an I/O has completed without any bad blocks written to or read from disk. It protects against displacement errors, data corruption resulting from hardware or software errors, bit flips, and silent drive errors, such as when the drive delivers the wrong data on a read request or writes to the wrong location.

You need both data assurance and media scan. They work complementarily to protect your data.

Unreadable Sector Management

This feature provides a controller-based mechanism for handling unreadable sectors detected both during normal I/O operation of the controller and during long-lived operations such as reconstructions. The feature is transparent to the user and requires no special configuration.

Proactive Drive Health Monitor

Proactive drive health monitoring examines every completed drive I/O and tracks the rate of error and exception conditions returned by the drives. It also tracks drive performance degradation, which is often associated with unreported internal drive issues. Using predictive failure analysis technology, when any error rate or degraded performance threshold is exceeded—indicating that a drive is showing signs of impending failure—SANtricity software issues a critical alert message and takes corrective action necessary to protect the data.

Data Evacuator

With data evacuator, nonresponsive drives are automatically power-cycled to see if the fault condition can be cleared. If the condition cannot be cleared, the drive is flagged as failed. For predictive failure events, the evacuator feature removes data from the affected drive in an effort to move the data before the drive actually fails. If the drive fails, rebuild picks up where the evacuator was disrupted, thus reducing the rebuild time.

Hot Spare Support

The system supports global hot spares that can be automatically used by the controller to reconstruct the data of the failed drive if enough redundancy information is available. The controller selects the best match for the hot spare based on several factors, including capacity and speed.

SSD Wear Life Monitoring and Reporting

If an SSD supports wear life reporting, the GUI provides this information to the user to allow monitoring how much of the useful life of an SSD remains. For SSDs that support wear life monitoring, the percentage of spare blocks remaining in solid-state media is monitored by controller firmware at approximately one-hour intervals. Think of this approach as a fuel gauge for SSDs.

SSD Read Cache

The SANtricity SSD read cache feature uses SSD storage to hold frequently accessed data from user volumes. It is intended to improve the performance of workloads that are performance limited by HDD IOPS. Workloads with the following characteristics can benefit from using the SANtricity SSD read cache feature:

- Read performance is limited by HDD IOPS.

- There is a high percentage of read operations relative to write operations: that is, greater than 80% read.
- A large number of reads are repeat reads to the same or adjacent areas of disk.
- The size of the data that is repeatedly accessed is smaller than the SSD read cache capacity.

For more information about SSD read cache, see TR-4099: NetApp SANtricity SSD Cache for E-Series.

3.3 Performance and Capacity

Performance

An E5700 configured with all SSD, all HDD, or a mixture of both drives is capable of performing at very high levels, both in IOPS and throughput, while still providing extremely low latency. The E5700, through its ease of management, high degree of reliability, and exceptional performance, can be leveraged to meet the extreme performance requirements expected in a Splunk server cluster deployment.

An E5700 with 24 SSDs can provide up to 1,000,000 4K random read IOPS at less than 100µs average response time. This configuration is also capable of delivering 21GBps of read throughput and 9GBps of write throughput.

Many factors can affect the performance of the E5700, including different volume group types, the use of DDP, the average I/O size, and the read versus write percentage provided by the attached servers. Figure 9 and Figure 10 provide further performance statistics across various data protection strategies on the system under generic random I/O workloads.

Note: The system under test used 48 SSDs, 4K and 16K block sizes, and 25% and 75% read workloads.

Figure 9) Write-heavy workload expected system performance E5700.

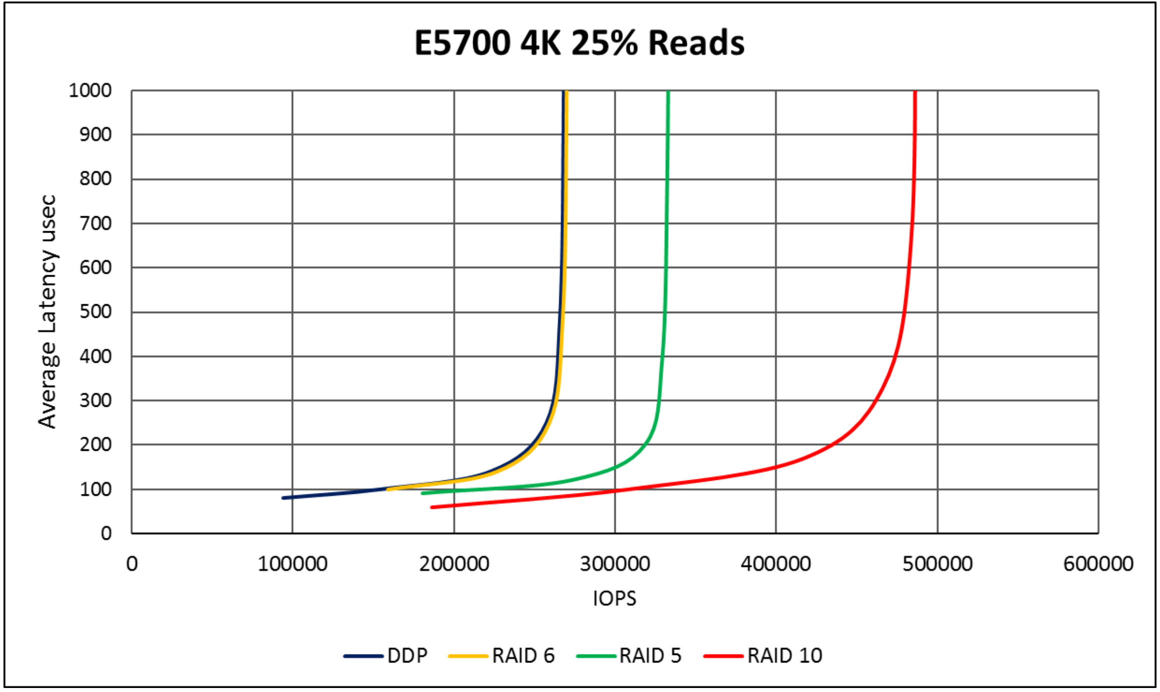
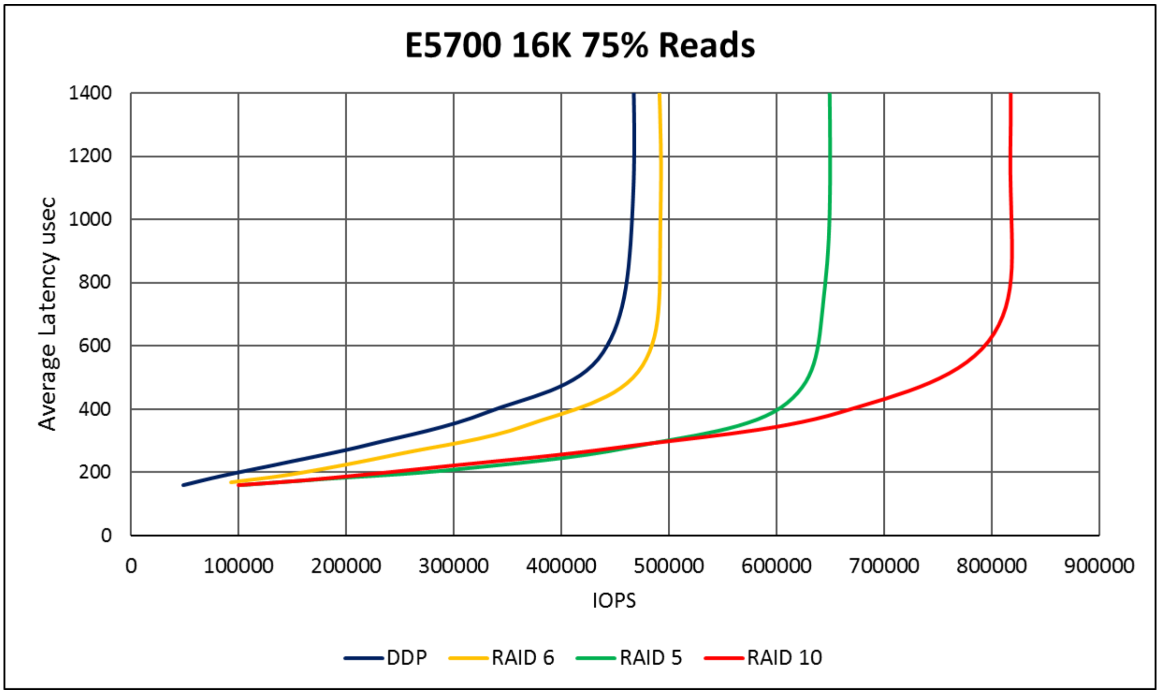


Figure 10) Read-heavy workload expected system performance E5700.



Capacity

The E5760 has a maximum capacity of 4800TB using 480 10TB drives. The E5724 has a maximum capacity of 1800TB using 120 15.3TB drives. See Table 2 for available drive capacities.

Table 2) Available drive capacities for E5700.

Controller Shelf Model	Drive Shelf Model	Number of Drives	NL-SAS HDDs	SAS HDDs	SSDs
E5760	DE460C (4U60)	60	4TB 8TB 10TB	1.2TB 1.8TB	800GB 1.6TB 3.2TB
E5724	DE224C (2U24)	24	NA	900GB 1.2TB 1.8TB	800GB 1.6TB 3.2TB 15.3TB

4 Decoupling Storage from Compute

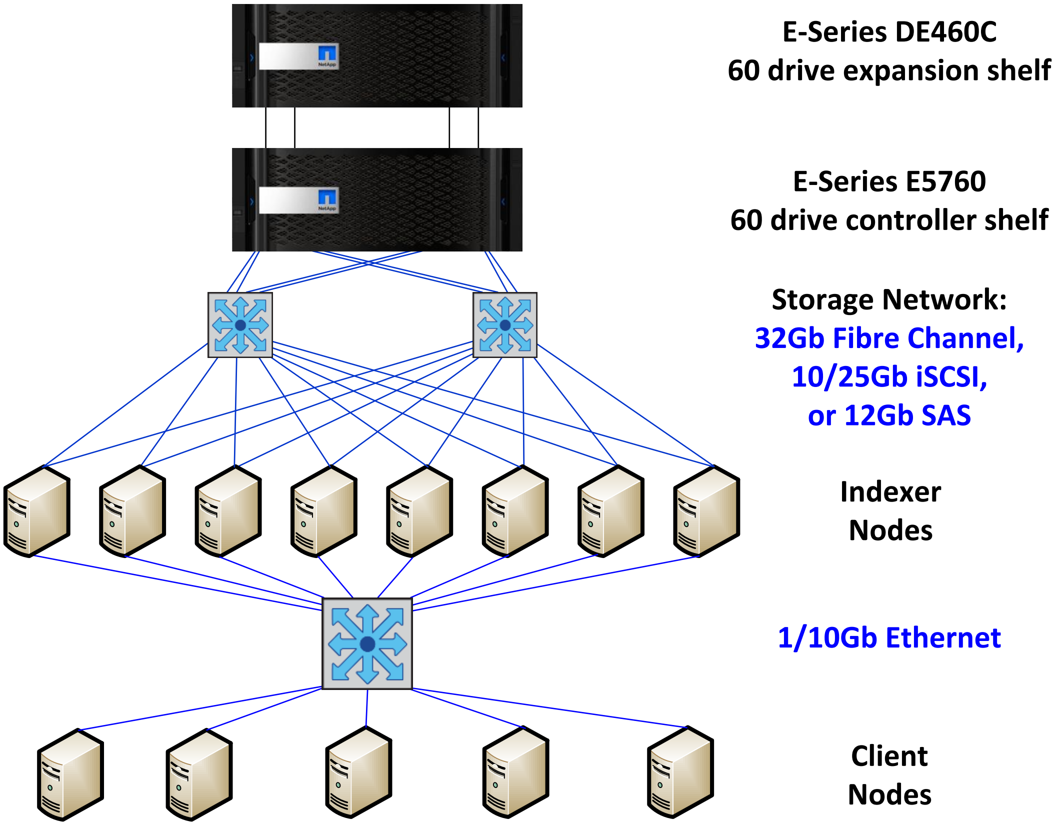
Splunk ingest rates are increasing day over day, retention periods are being extended, SSDs are getting larger, and the use of specialized compute for data analytics is expanding. The ability to decouple storage from compute with Splunk is becoming an economic necessity.

Figure 11 shows a sample Splunk architecture where an E-Series E5760 is connected over iSCSI to eight cluster nodes, and then client nodes are connected over Ethernet to the cluster nodes. With NetApp E-Series it’s possible to begin with 20 SSDs and scale all the way up to 120 SSDs without needing to add more cluster nodes in an all-SSD system. If HDDs are being deployed for the cold tier, it is possible to scale up to 480 drives by adding additional drive shelves to the system shown. With E-Series, 4 to 10 nodes per E-Series array work well depending on storage and performance requirements.

The advantages of decoupling storage from compute include:

- Ability to scale capacity and compute separately, saving the cost of overprovisioning one or the other
- Ability to nondisruptively scale capacity and compute as demanding requirements change
- Ability to refresh compute (which happens more frequently than storage) without a performance-affecting data migration effort
- Flexibility to use excess top-of-rack switch bandwidth for the storage network, use a wholly different storage network such as Fibre Channel, or connect the array as DAS.

Figure 11) Sample of storage and compute separation.



This type of decoupling, for example, can allow a company with 100 nodes to reduce its number of nodes substantially, if 100 nodes of compute aren’t required. This change provides a significant reduction in rack space required and the associated cooling and power requirements. In contrast, if the need is more compute, then less expensive servers can be purchased that don’t require space for additional storage and have a smaller footprint. It also enables the use of blade server technology with their environmental savings.

After the decision has been made to use a separate storage array, sizing and configuring it are straightforward.

4.1 Sizing

Estimate Your Storage Requirements

Before sizing, it is necessary to estimate your daily ingest rate. When Splunk Enterprise indexes your data, it creates two main types of files: the "rawdata" file that contains the original data in compressed form and the index files that point to this data. (It also creates a few metadata files, which don't consume much space.) With a little experimentation, you can estimate how much index disk space you will need for a given amount of incoming data. A rule of thumb used by Splunk for syslog-type data is that after it has been compressed and indexed in Splunk, it occupies approximately 50% of its original size:

- 15% for the rawdata file
- 35% for associated index files

As an example, assume the current environment is split between two logging formats: syslog and Windows. The logs generated by the Windows servers and Active Directory are logged as Windows events. For antivirus software, there is Symantec Endpoint Protection; the Symantec Endpoint Protection Manager runs on a Windows server. The logs ingested into Splunk include the firewall logs from the Cisco ASA devices, Linux logs, webserver logs, Cisco WSA proxy log, Windows logs, Symantec logs, and e-commerce transaction logs (these are our data sources). Retention periods are as shown; assume 30 days of the retention period for the hot/warm tier.

At a high level, Splunk calculates total disk storage for each tier as follows:

(Daily average indexing rate) x (retention policy) x ½

Table 3 shows what the results of our sizing to this point might look like.

Table 3) Sizing example for nonclustered environment.

Data Source	GB per Day	Retention Days	Raw Comp. Rate	Base Size of Raw	Index Comp. Rate	Base Size of Index Files	Estimated Size on Disk	Hot/Warm	Cold
Cisco ASA firewall logs	100	90	0.15	1,350	0.35	3,150	4,500	1,500	3,000
Badge reader log	60	180	0.15	1,620	0.35	3,780	5,400	900	4,500
Cisco WSA proxy	120	30	0.15	540	0.35	1,260	1,800	1,800	0
Linux logs	80	90	0.15	1,080	0.35	2,520	3,600	1,200	2,400
Windows logs	140	90	0.15	1,890	0.35	4,410	6,300	2,100	4,200
Symantec logs	20	90	0.15	270	0.35	630	900	300	600
Web logs	440	90	0.15	5,940	0.35	13,860	19,800	6,600	13,200
E-commerce logs	300	365	0.15	16,425	0.35	38,325	54,750	4,500	50,250
Total (GB)	**1,260**						**97,050**	**18,900**	**78,150**

The best way to get an idea of your space needs is to experiment by indexing a representative sample of your data and then checking the sizes of the resulting directories in `$SPLUNK_HOME/var/lib/splunk/defaultdb`.

For more about estimating storage requirements see the Capacity Planning Module of the online Splunk documentation; in particular, see Estimate your storage requirements. Splunk also provides an online application to aid in sizing: Splunk Storage Sizing.

At this point, we have a good idea of our daily indexing volume, 1260GB/day for our example. We can use Splunk's Summary of performance recommendations in the capacity planning module to estimate the number of reference machines required for indexing and searching. Because the daily indexing volume is just over 1TB/day, let's assume we need 1 search head and 8 indexers.

The last step is to decide if we will cluster any of the indexes and how they will affect our capacity requirements. Let's assume that the e-commerce data is critical and that the Sales Support team needs to be able to always access this data.

So, we need to determine:

- Replication factor (RF), which specifies how many total copies of rawdata the cluster should maintain. This factor sets the total failure tolerance level.
- Search factor (SF), which specifies how many copies are searchable (searchable buckets have both rawdata and index files). This factor determines how quickly you can recover the search capability.

This step is where the power of decoupling storage from compute with an E-Series array comes in. Because high availability is built into the array, it is only necessary to have an RF=2, and because we want to recover the search capability immediately, we set the SF=2. So, it is not necessary to add compute to add storage or install additional unneeded storage in all nodes. For the e-commerce logs, only the amount of capacity required doubles, as shown in Table 4.

Table 4) Increased capacity needs for clustering of e-commerce logs.

Data Source	GB per Day	Retention Days	Raw Comp. Rate	Base Size of Raw	Index Comp. Rate	Base Sixe of Index Files	Estimated Size on Disk	Hot/Warm	Cold
Cisco ASA firewall logs	100	90	0.15	1,350	0.35	3,150	4,500	1,500	3,000
Badge reader log	60	180	0.15	1,620	0.35	3,780	5,400	900	4,500
Cisco WSA proxy	120	30	0.15	540	0.35	1,260	1,800	1,800	0
Linux logs	80	90	0.15	1,080	0.35	2,520	3,600	1,200	2,400
Windows logs	140	90	0.15	1,890	0.35	4,410	6,300	2,100	4,200
Symantec logs	20	90	0.15	270	0.35	630	900	300	600
Web logs	440	90	0.15	5,940	0.35	13,860	19,800	6,600	13,200
E-commerce logs	300	365	0.15	16,425 x RF(2) = 32,850	0.35	38,325 x SF(2) = 76,650	109,500	9,000	100,500

Data Source	GB per Day	Retention Days	Raw Comp. Rate	Base Size of Raw	Index Comp. Rate	Base Sixe of Index Files	Estimated Size on Disk	Hot/Warm	Cold
Total (GB)	**1,260**						**151,800**	**23,400**	**128,400**

Note: Using the traditional architecture, RF=3 and SF=2, would increase the required disk space by an additional16425GB.

Note: This calculation is only for one index, the E-commerce logs, in our example. For all indexes, the increase would be 29115GB.

Now we have an estimate of the number of servers and amount of storage required for our implementation.

Configure Your E-Series Storage Based on Estimate

At this point we just need to complete the configuration of our E-Series storage. We want to use DDP, from which we will create volumes (LUNs). We need one volume per indexer for the hot/warm tier (eight total) and one volume per indexer for the cold tier (eight total).

For the hot/warm tier, the usable storage capacity required is 23400GB. This requirement could be satisfied using either 10k SAS drives (possibly including an SSD read cache) or, for higher performance, all SSDs.

For the cold tier, the usable capacity required is 128400GB. The recommendation in this instance would be to use high-capacity NL-SAS drives, such as 10TB drives. The cold tier might also employ an SSD read cache for improved searching.

We configure our array using the E5760, which is a 4U shelf with five drawers and each drawer containing 12 drives, as shown in Figure 4. We will use an all-SSD drive hot/warm tier. This setup requires approximately 24 1.6TB drives, which provide a usable capacity of 27.5TB, giving us a little room for growth. For the cold tier, we use 20 10TB NL-SAS drives. This configuration provides a usable capacity of 138.9TB, again allowing us a little room for growth.

Now that we have our drives, we can create two DDPs, one for the hot/warm tier using the SSDs and the other for the cold tier using the NL-SAS drives. Each of these DDPs is then cut into eight volumes (LUNs), which are presented to the eight indexers. So each indexer has two volumes presented to it, one for the hot/warm tier and one for the cold tier.

See the NetApp E-Series Splunk Deployment Guide for actual implementation steps.

4.2 Other Considerations

E-Series

To prepare your server for storage access, see:

- Installing and Configuring for Linux Express Guide
- Installing and Configuring for Linux Power Guide for Advanced Users

These documents guide you through:

- Installing SANtricity Storage Manager host-side applications
- Configuring multipath
- Installing NetApp Host Utilities
- Using the iscsiadm open-iscsi utility with E-Series products (if using iSCSI)

The E-Series Interoperability Matrix Tool (IMT) has some 85,000 entries to not only connect to any SAN but also support it. To verify that your configuration is supported and check for any changes that might be required for correct functioning of your E-Series, see the Interoperability Matrix Tool.

Linux Configuration

All servers in the Splunk cluster were tested with RHEL 7.3 with default kernel settings.

For persistent deployments, administrators should consider the following flags when adding a mount into /etc/fstab:

- **`nobarrier`.** Allows data to sit in cache instead of being flushed. There is a large performance gain on particular workloads by allowing `nobarrier`. This option should only be used for E-Series storage, because internal disks might not have battery backup.
- **`noatime`.** Forces file reads to not record their access times to disk, which can increase I/O dramatically on heavy read loads. Setting the `noatime` flag is only recommended for file systems or dependent applications where a record of the last access time of a file for reading is unnecessary.
- **`_netdev`.** Required for configurations using iSCSI and iSER network protocols. The `_netdev` option forces the mount to wait until the network is up before trying to mount. Without this option, the OS attempts to mount the disk prior to the network being completely available, and it could lead to various timeouts or the OS entering recovery mode.
- **`discard`.** If the storage volume is thinly provisioned, providing the `discard` flag allows the file system to reclaim space. This flag can cause performance degradation. Administrators who want to control when discards take place (for example, nightly) should consider using `fstrim` or an equivalent command for the OS.

Note: The use of thin-provisioned volumes is not recommended with Splunk installations.

To increase performance, jumbo frames should be set on the network. Setting jumbo frames for the storage is explained in the E-Series documentation. On the server, they are configured by adding an entry of MTU=9000 to the interface file in the /etc/sysconfig/network-scripts directory and restarting the interface. To validate that jumbo frames have been set, use the `ip link show` command:

```
[root@ictk0103r720-4 ~]# ip link show
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN mode DEFAULT link/loopback
00:00:00:00:00:00 brd 00:00:00:00:00:00
2: eth0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 9000 qdisc mq state UP mode DEFAULT qlen 1000
link/ether b0:83:fe:d5:ae:62 brd ff:ff:ff:ff:ff:ff
3: eth1: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 9000 qdisc mq state UP mode DEFAULT qlen 1000
link/ether b0:83:fe:d5:ae:64 brd ff:ff:ff:ff:ff:ff
```

If you use multipathing, you need to edit the timeout value in the iSCSI configuration file, /etc/iscsi/iscsid.conf. NetApp recommends using a value of 5 seconds.

```
node.session.timeo.replacement_timeout = 5
```

This amount is the length of time to wait for session reestablishment before failing SCSI commands back to the application when running the Linux SCSI layer error handler. The default value is 120 seconds.

Splunk

See the following deployment information for Splunk Enterprise Edition 6.6 to make sure that your Linux environment is set up correctly:

- Splunk Enterprise Installation Manual
- Splunk Enterprise Installation Manual - Install on Linux

5 Splunk Enterprise Edition and NetApp E5700 Testing

NetApp tested a simulated Splunk cluster environment where the index peer nodes were configured to use one E-Series E5700 for hot/warm and cold data buckets. The server hardware was chosen with the following recommendations from the Splunk reference architecture system requirements. The Splunk cluster server hardware that was used is listed in Table 5.

Table 5) Splunk cluster server hardware.

Splunk Cluster	Qty.	Type	CPU	CPUs	Cores/CPUs	Speed	RAM
Indexer peer node	8	Dell 730xd	E2-2670 v3	2	8	2.3Ghz	128GB
Search head	1	Dell 730xd	E2-2670 v3	2	8	2.3Ghz	128GB
Cluster master	1	Dell 730xd	E2-2670 v3	2	8	2.3Ghz	128GB
Forwarder	1	Dell 730	E2-2670 v3	2	8	2.3Ghz	128GB

The ingest machine log data was created using the Splunk workload tool Eventgen. The cluster had eight index peer nodes to handle ingesting ~125GB of simulated machine syslog data per indexer, for a total of ~1TB per day for the entire cluster.

5.1 Overview of Splunk Cluster Testing Used for E-Series

The Splunk cluster configuration components consist of:

- **Forwarder.** Ingest 125GB of machine log data files into the cluster of index node peers.
- **Index peer nodes.** Index the ingested machine syslog data and replicate data copies in the cluster.
- **Search head.** Execute custom searches for dense, very dense, sparse, and very sparse data from the cluster of index peer nodes.
- **Master.** Monitor and push configuration management changes for the cluster. License master of 1TB per day ingest amount for the 8-index peer node cluster.

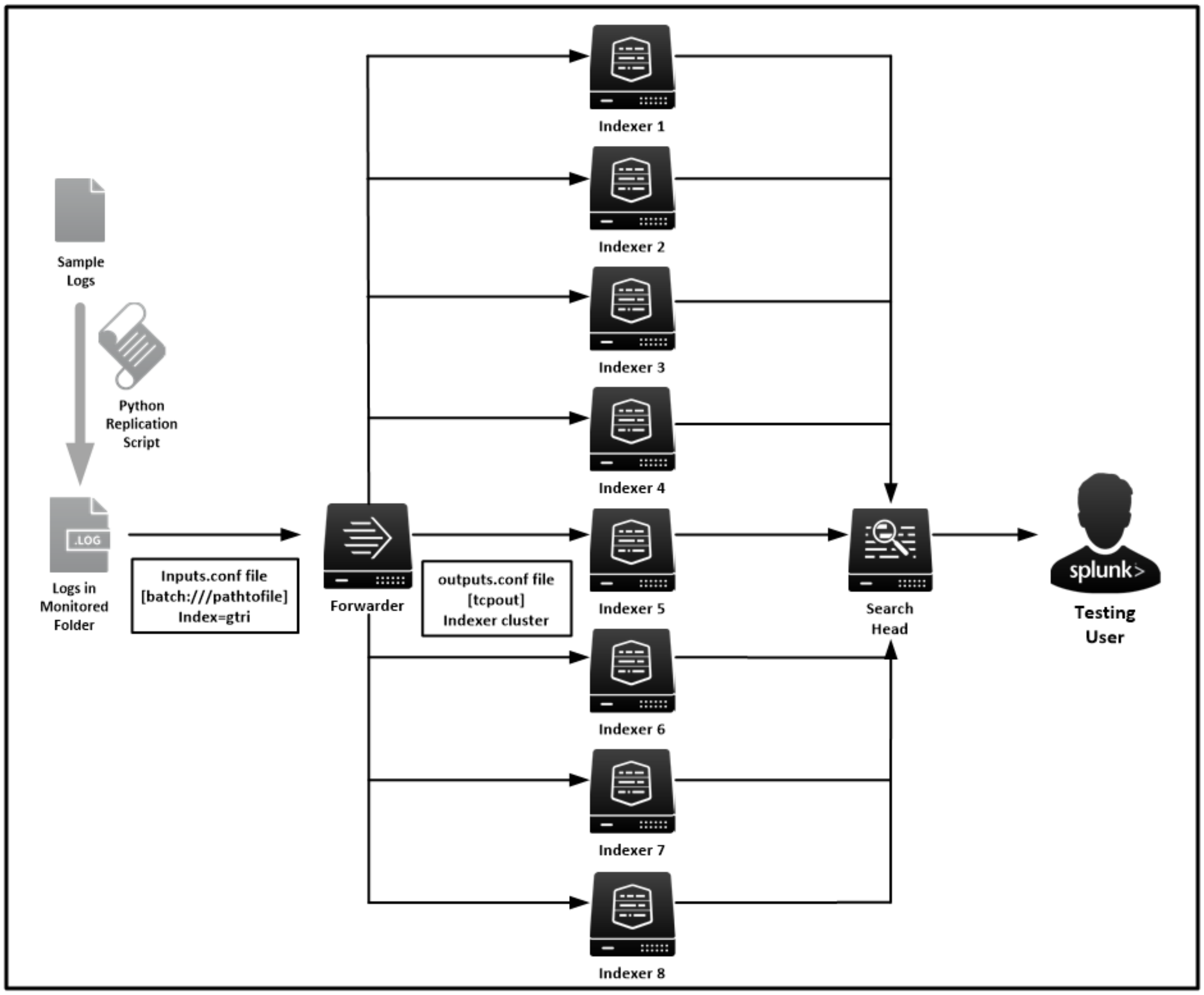
5.2 Eventgen Data

The machine log dataset was created with Splunk’s event generator, Eventgen. The Splunk event generator is a downloadable Splunk app available from the Splunk website. Splunk Eventgen enables users to load samples of log files or exported .csv files as an event template. The templates can then be used to create artificial log events with simulated time stamps. A user can modify the field values and configure the random variance while preserving the structure of the events. The data templates can be looped to provide a continuous stream of real-time data. For more Eventgen information, visit Splunk Eventgen app.

For our testing, Eventgen was loaded into the cluster and was configured to produce a 125GB simulated syslog type file for the Splunk forwarder instance. The file was then split into smaller syslog files on one Splunk heavy forwarder instance, which is forwarding data on a rotating basis to each of the 8 index peer nodes. The total ingested data is ~1TB per day loaded into the cluster. The logical configuration is shown in Figure 12.

Note: The Splunk instance was licensed for 1TB per day. The actual loading of 1TB of data takes considerably less time than one day.

Figure 12) Splunk logical configuration.



Following are the number of rare and dense search terms per 10,000,000 lines:

- **Very dense search.** 1 out of 100 lines; 100,000 occurrences
- **Dense search.** 1 out of 1,000 lines; 10,000 occurrences
- **Sparse search.** 1 out of 1,000,000 lines; 10 occurrences
- **Very sparse search.** 1 out of 10,000,000 lines; 1 occurrence

5.3 Cluster Replication and Searchable Copies Factor

The search factor determines the number of searchable copies of indexed data the indexer cluster maintains for each bucket. The default value for the search factor is 2, meaning that the cluster maintains two searchable copies of all data. The search factor must be less than or equal to the replication factor. The replication factor is the number of copies of data that you want the cluster to maintain. Peer nodes store incoming data in buckets, and the cluster maintains multiple copies of each bucket. The cluster stores each bucket copy on a separate peer node. The number of copies of each bucket that the cluster maintains is the replication factor. The default replication value for a cluster is 3.

The E-Series test was configured with a replication factor of 2 and searchable copies with a factor of 2. The E-Series provides additional redundancy with additional copies of indexed data located in the DDP volumes for each indexer. This additional redundancy enables the replication factor of 2 to seamlessly provide fewer copies of index data in the Splunk cluster for performance and data storage benefits.

5.4 E-Series with DDP Baseline Test Setup

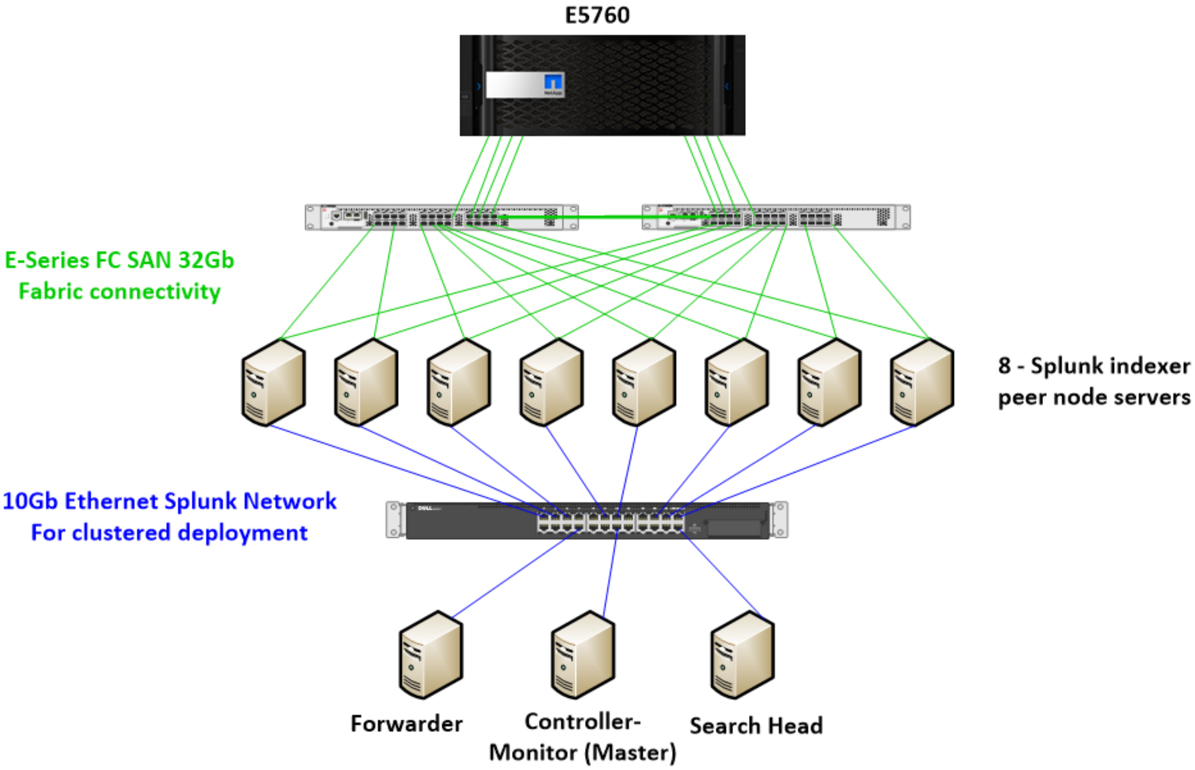
The E5760 configuration for the baseline test was configured with DDP LUNs using:

- 24 800GB SSDs with a pool preservation capacity of 2 drives, offering ~12.8TB of usable capacity for the Splunk cluster hot/warm data buckets
- 22 900GB 10k SAS HDDs with a pool preservation capacity of 2 drives, offering ~13TB of usable capacity for the Splunk cluster hot/warm data buckets (alternate hot/warm tier)
- 12 8TB NL-SAS drives with a pool preservation capacity of 2 drives, offering ~57.7TB of usable capacity for the Splunk cluster cold data buckets
- 10Gb Ethernet private network for index peer nodes
- 32Gb Fibre Channel SAN for E-Series and index peer nodes

The DDP was configured into eight volumes: one each for the eight index peer node hosts. The mounted volumes were configured as ext4 file systems on the RHEL 7.3 OS of each indexer.

See Figure 13 for the E-Series baseline configuration.

Figure 13) Splunk cluster with E-Series.



5.5 Baseline Test Results for E-Series

To make sure of consistency, the same data was loaded using the same scripts and Splunk configurations. The testing pattern was to ingest 1TB of data into the storage configuration using the preceding configurations. After the script used to transfer data and manipulate the data ran, the data began to be copied to the splunk_forward directory, where the forwarder began sending the files to be indexed across the indexers.

At this time, measurements would be taken of the indexing rate for each of the indexers based on the following Splunk query:

- Average and peak indexing rates: index="_internal" source=*metrics.log* series=gtri per_index_thruput | timechart span=5m max(kbps) avg(kbps)

The next step would be to provide “static” searches on data that was indexed and record the times. During static searches, no data was being indexed. For the purpose of this testing, we used a dataset consisting of strings interspersed at given intervals and searched for these in order to determine the performance of a search against this level of density in 125GB worth of data. The levels of frequency and searches are

- 1 match per 100 events: index=gtri “DENSE100” source=*001.log
- 1 match per 1000 events: index=gtri “DENSE128” source=*001.log
- 1 match per 1 million events: index=gtri “RARE1MIL” source=*001.log
- 1 match per 10 million events: index=gtri “RARE10MIL” source=*001.log

Following the static search, a “streaming” search was run. While data was being indexed, the searches just described were run to find the streaming search time.

For all searches, the Splunk Job Inspector was used as a barometer of how quickly the search had run. This barometer can be accessed through the web UI below the search bar by selecting Job > Inspect Job.

This method of testing was repeated for each of the storage configurations and detailed in the following subsections.

Hot/Warm Tier SSD

The first testing scenario used ~1TB of hot/warm storage per indexer using 800GB SSDs. In addition, a 6.5TB cold storage was added using an NL-SAS 7.2K HDD with no read cache on it. The sample logs provided were forwarded to all indexers for improved performance as would be best practice for this architecture. Table 6 shows the indexing results, and Table 7 shows search results.

Table 6) SSD hot/warm tier indexing results.

Data Ingest Volume (GB/Day)	Average Indexing Rate Observed	Peak Indexing Rate Observed
~1000GB/day	28.158MBps	43.776MBps

Note: The Splunk instance was licensed for 1TB per day. The actual loading of 1TB of data takes considerably less time than one day.

Table 7) SSD hot/warm tier searching results.

Search Performed	Streaming Search Time	Static Search Time
Dense100	14.123 seconds	14.175 seconds
Dense1000	4.098 seconds	4.11 seconds
RARE1MIL	2.112 seconds	2.11 seconds
RARE10MIL	2.109 seconds	2.055 seconds

Hot/Warm Tier 10K SAS HDD

The second testing scenario also used ~1TB hot/warm storage per indexer using 900GB 10K SAS HDDs. In addition, a 6.5TB cold storage was added using a near-line SAS drive with no read cache on it. The sample logs provided were forwarded to all indexers for improved performance as would be best practice for this architecture. Table 8 shows the indexing results, and Table 9 shows search results.

Table 8) 10K SAS HDD hot/warm tier indexing results.

Data Ingest Volume (GB/Day)	Average Indexing Rate Observed	Peak Indexing Rate Observed
~1000GB/day	25.133MBps	42.023MBps

Note: The Splunk instance was licensed for 1TB per day. The actual loading of 1TB of data takes considerably less time than one day.

Table 9) 10K SAS HDD hot/warm tier searching results.

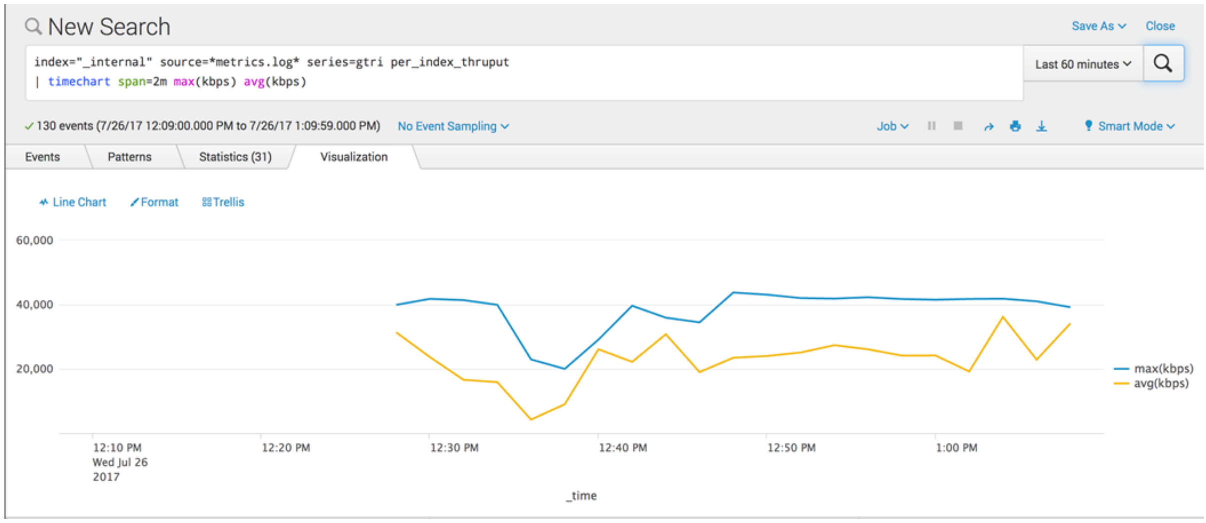
Search Performed	Streaming Search Time	Static Search Time
Dense100	26.397 seconds	22.124 seconds
Dense1000	25.179 seconds	20.184 seconds
RARE1MIL	20.118 seconds	16.117 seconds
RARE10MIL	12.115 seconds	10.114 seconds

To improve search times, an SSD read cache could be used with the all-HDD hot/warm tier. This configuration is dependent on the types of search queries being run against the indexers. If the queries are regularly querying the same data, an SSD read cache would be appropriate for improved performance. This result would also be true if the cold tier is regularly queried.

System Drive Failure Simulation

In order to validate the recommendation of DDP, a drive failure of the all-SSD hot/warm tier was tested to observe the overall impact on Splunk indexing. For the simulation, a drive failure was induced while the overall index rate of the array under heavy ingest rates was monitored. After drive failure, the array recovered and migrated mirrored data to additional SSDs within the pool without noticeably affecting Splunk performance. Figure 14 shows the Splunk ingestion chart with a failed drive and the normal performance fluctuations.

Figure 14) Failed drive impact.



```
index="_internal" source=*metrics.log* series=gtri per_index_thruput
| timechart span=2m max(kbps) avg(kbps)
```

✓ 130 events (7/26/17 12:09:00.000 PM to 7/26/17 1:09:59.000 PM)

Although there was a slight dip in the indexing speed, within approximately eight minutes, the indexing rate was restored to normal.

Summary

The NetApp E-Series 5700 provides a number of significant advantages over internal DAS for Splunk deployments. These advantages include exceptional storage management capabilities, dramatically improved reliability, high availability, and limited performance degradation because of failure conditions such as disk failures. By decoupling storage from compute, you gain the ability to scale capacity and compute separately, saving the cost of overprovisioning one or the other.

The advantages also include excellent performance handling ingest of machine log data and excellent search capabilities at very low latency with an E5700 configuration of all-flash SSDs for hot and warm Splunk buckets or a hybrid configuration employing HDDs, which utilizes SSDs as a read cache. The E-Series E5760 provides excellent performance and reliability for the Splunk cold data bucket tiers as well and can also make use of an SSD read cache to improve query performance as needed.

Organizations that use Splunk often use traditional server-based storage with inefficient, hard-to-scale internal DAS. The NetApp reference design employs the managed DAS model, with higher scalability and performance. The reduction of the Splunk cluster replication factor available when deploying E-Series storage reduces the amount of indexed data stored. This reduction prevents unnecessary purchases of compute nodes for storage-intensive workloads for Splunk environments that need to grow to meet organizational requirements.

The NetApp reference architecture for Splunk is optimized for node storage balance, reliability, performance, storage capacity, and density. From an administrative standpoint, E-Series offers simplified storage management with a browser-based UI. This solution enables new volumes and Dynamic Disk Pools to be created easily and provisioned immediately for use by Splunk cluster servers. In addition, existing volumes and Dynamic Disk Pools can all be increased in size dynamically to provide additional capacity and/or performance as required for the Splunk indexer cluster environment.

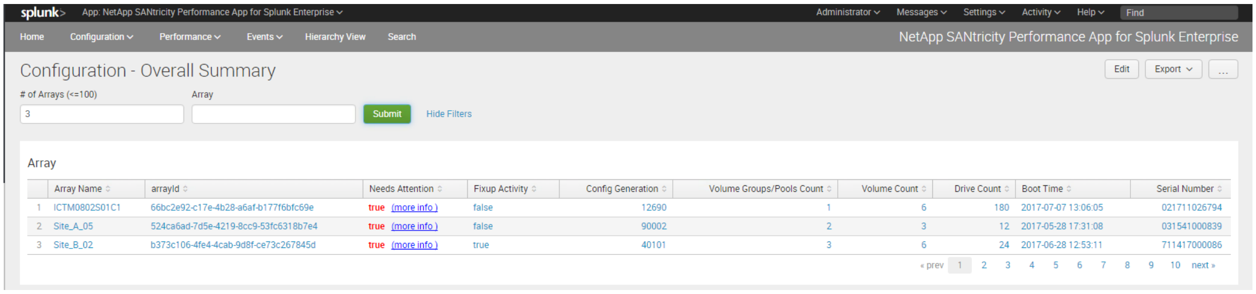
Appendix

Splunk App for NetApp E-Series and EF-Series

The SANtricity performance app for Splunk Enterprise makes it easy to monitor the health and performance of NetApp E-Series and EF-Series storage systems from within the Splunk environment.

The Configuration tab, shown in Figure 15, gives a basic overview of each actively monitored array. A user can drill down to view individual volume groups, volumes, and drives by selecting an individual array. If an array needs attention, a user can click “(more info)” to view major event log (MEL) data specific to the issue, as in Figure 16.

Figure 15) Configuration tab of the SANtricity performance app for Splunk Enterprise.



Configuration - Overall Summary

# of Arrays (<=100): 3

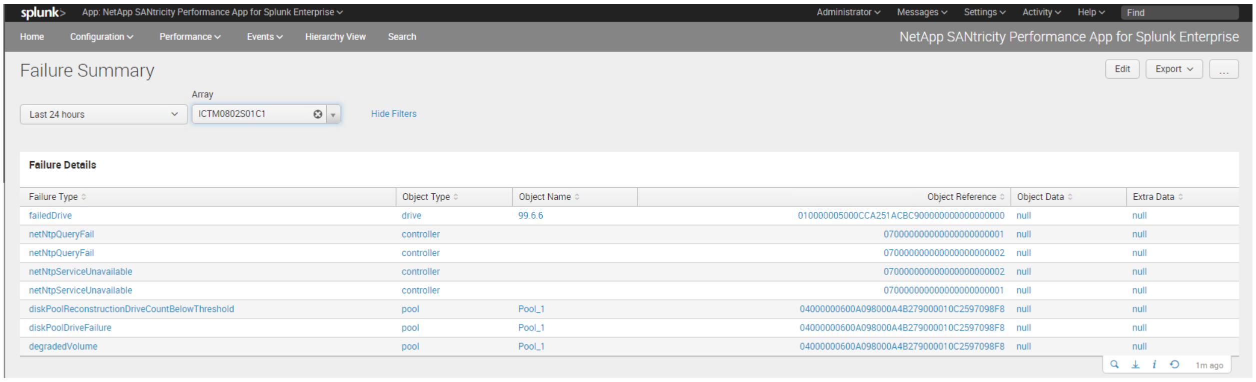
Array:

Array

	Array Name	arrayId	Needs Attention	Fixup Activity	Config Generation	Volume Groups/Pools Count	Volume Count	Drive Count	Boot Time	Serial Number
1	ICTM0802S01C1	66bc2e92-c17e-4b28-a6af-b177f6bfc69e	true (more info)	false	12690	1	6	180	2017-07-07 13:06:05	021711026794
2	Site_A_05	524ca6ad-7d5e-4219-8cc9-53fc6318b7e4	true (more info)	false	90002	2	3	12	2017-05-28 17:31:08	031541000839
3	Site_B_02	b373c106-4fe4-4cab-9d8f-ce73c267845d	true (more info)	true	40101	3	6	24	2017-06-28 12:53:11	711417000086

« prev 1 2 3 4 5 6 7 8 9 10 next »

Figure 16) Major event log (MEL) information for an array needing attention.



Failure Summary

Last 24 hours

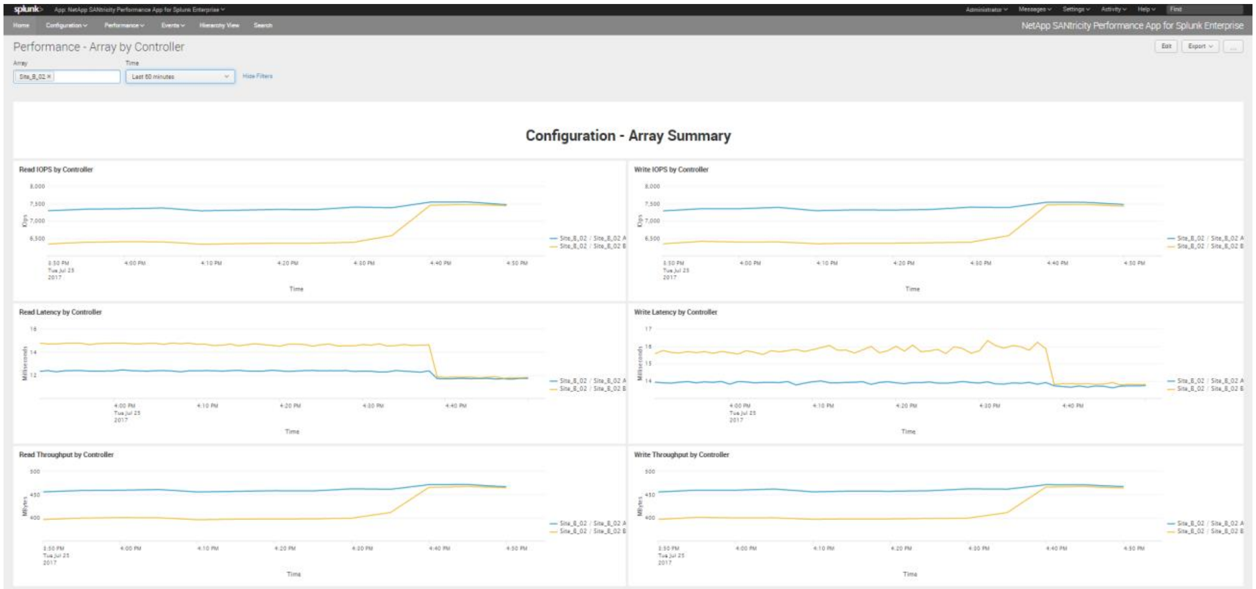
Array: ICTM0802S01C1

Failure Details

Failure Type	Object Type	Object Name	Object Reference	Object Data	Extra Data
failedDrive	drive	99.6.6	010000005000CCA251ACBC900000000000000000	null	null
netNtpQueryFail	controller		070000000000000000000001	null	null
netNtpQueryFail	controller		070000000000000000000002	null	null
netNtpServiceUnavailable	controller		070000000000000000000002	null	null
netNtpServiceUnavailable	controller		070000000000000000000001	null	null
diskPoolReconstructionDriveCountBelowThreshold	pool	Pool_1	04000000600A098000A4B279000010C2597098F8	null	null
diskPoolDriveFailure	pool	Pool_1	04000000600A098000A4B279000010C2597098F8	null	null
degradedVolume	pool	Pool_1	04000000600A098000A4B279000010C2597098F8	null	null

The Performance tab, shown in Figure 17, displays real-time graphical information about read/write operations, read/write latency, and read/write throughput. The view defaults to a single array, but a user can choose to view multiple arrays or to view performance by controller, volume group, volume, or individual drive.

Figure 17) Performance tab of the SANtricity performance app for Splunk Enterprise.



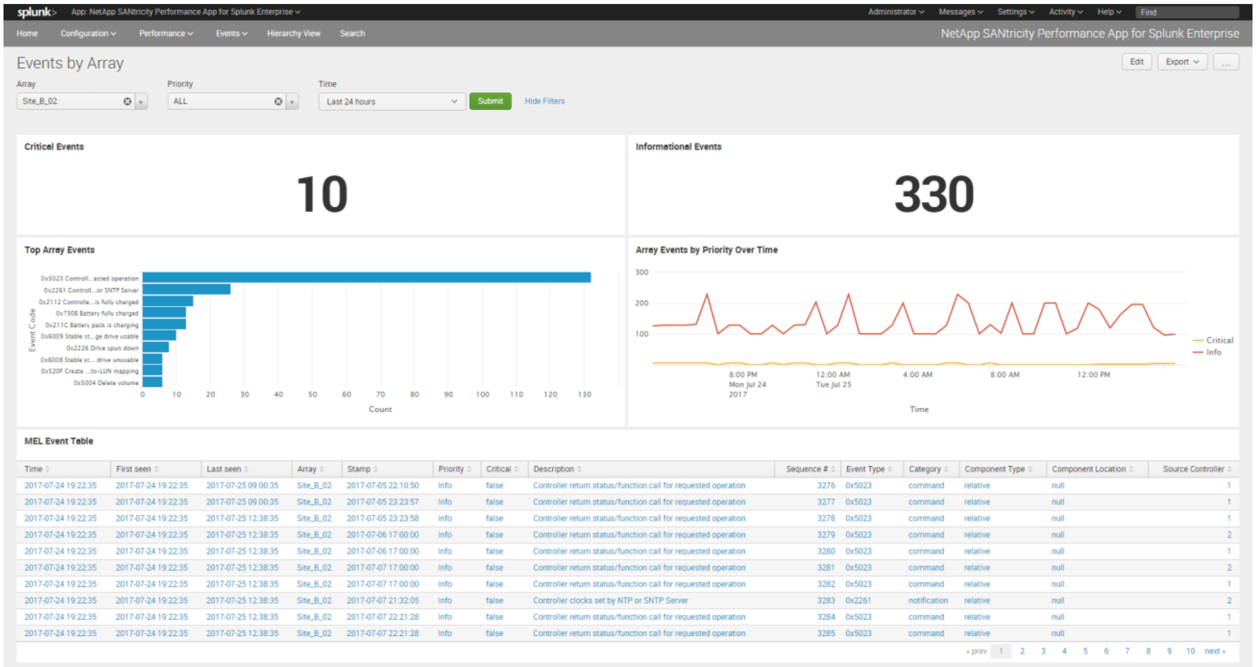
Performance - Array by Controller

Array: Site_B_02

Time: Last 60 minutes

The Events tab, shown in Figure 18, gives information about all MEL events reported by an array. Events can be sorted by a variety of fields and filtered by priority as needed.

Figure 18) Events tab of the SANtricity performance app for Splunk Enterprise.


splunk>
App: NetApp SANtricity Performance App for Splunk Enterprise
Administrator
Messages
Settings
Activity
Help
Find
Home
Configuration
Performance
Events
Hierarchy View
Search
NetApp SANtricity Performance App for Splunk Enterprise
Edit
Export
...
Submit
Hide Filters
Top Array Events
Event Code
0x5023 Controll...ested operation
0x2261 Controll...or SNTP Server
0x2112 Controlle...is fully charged
0x7308 Battery fully charged
0x211C Battery pack is charging
0x6009 Stable st...ge drive usable
0x2226 Drive spun down
0x6008 Stable st... drive unusable
0x520F Create ...to-LUN mapping
0x5004 Delete volume
0
10
20
30
40
50
60
70
80
90
100
110
120
130
Count
Array Events by Priority Over Time
300
200
100
8:00 PM
Mon Jul 24
2017
12:00 AM
Tue Jul 25
4:00 AM
8:00 AM
12:00 PM
Time
Critical
Info
« prev
1
2
3
4
5
6
7
8
9
10
next »

Events by Array

Array: Site_B_02

Priority: ALL

Time: Last 24 hours

Critical Events

10

Informational Events

330

MEL Event Table

Time	First seen	Last seen	Array	Stamp	Priority	Critical	Description	Sequence #	Event Type	Category	Component Type	Component Location	Source Controller
2017-07-24 19:22:35	2017-07-24 19:22:35	2017-07-25 09:00:35	Site_B_02	2017-07-05 22:10:50	Info	false	Controller return status/function call for requested operation	3276	0x5023	command	relative	null	1
2017-07-24 19:22:35	2017-07-24 19:22:35	2017-07-25 09:00:35	Site_B_02	2017-07-05 23:23:57	Info	false	Controller return status/function call for requested operation	3277	0x5023	command	relative	null	1
2017-07-24 19:22:35	2017-07-24 19:22:35	2017-07-25 12:38:35	Site_B_02	2017-07-05 23:23:58	Info	false	Controller return status/function call for requested operation	3278	0x5023	command	relative	null	1
2017-07-24 19:22:35	2017-07-24 19:22:35	2017-07-25 12:38:35	Site_B_02	2017-07-06 17:00:00	Info	false	Controller return status/function call for requested operation	3279	0x5023	command	relative	null	2
2017-07-24 19:22:35	2017-07-24 19:22:35	2017-07-25 12:38:35	Site_B_02	2017-07-06 17:00:00	Info	false	Controller return status/function call for requested operation	3280	0x5023	command	relative	null	1
2017-07-24 19:22:35	2017-07-24 19:22:35	2017-07-25 12:38:35	Site_B_02	2017-07-07 17:00:00	Info	false	Controller return status/function call for requested operation	3281	0x5023	command	relative	null	2
2017-07-24 19:22:35	2017-07-24 19:22:35	2017-07-25 12:38:35	Site_B_02	2017-07-07 17:00:00	Info	false	Controller return status/function call for requested operation	3282	0x5023	command	relative	null	1
2017-07-24 19:22:35	2017-07-24 19:22:35	2017-07-25 12:38:35	Site_B_02	2017-07-07 21:32:05	Info	false	Controller clocks set by NTP or SNTP Server	3283	0x2261	notification	relative	null	2
2017-07-24 19:22:35	2017-07-24 19:22:35	2017-07-25 12:38:35	Site_B_02	2017-07-07 22:21:28	Info	false	Controller return status/function call for requested operation	3284	0x5023	command	relative	null	1
2017-07-24 19:22:35	2017-07-24 19:22:35	2017-07-25 12:38:35	Site_B_02	2017-07-07 22:21:28	Info	false	Controller return status/function call for requested operation	3285	0x5023	command	relative	null	1

The SANtricity performance app for Splunk Enterprise requires:

- Splunk 6.1 or higher running on Linux
- NetApp E-Series/EF-Series storage arrays running firmware 7.84 or higher
- NetApp SANtricity Web Services Proxy 1.3 or higher (2.0 or higher for best feature set) running on Windows or Linux

- NetApp SANtricity Performance App for Splunk Enterprise (available from SplunkBase)
- Technology Add-On for NetApp SANtricity (available from SplunkBase)

To use the app, first install NetApp SANtricity Web Services Proxy on any server with network access to the monitored arrays. Then, upload and install NetApp SANtricity performance app for Splunk Enterprise and technology add-on for NetApp SANtricity from within the Splunk environment. Additional configuration is required to add each array to NetApp SANtricity Web Services Proxy and to Splunk. See the README included with each app or the Details page on SplunkBase for more information.

References

NetApp Documentation

Datasheet: NetApp E-Series SANtricity Software

SANtricity Release 11.30: Documentation Center

TR-4115: SANtricity Dynamic Disk Pools Best Practice Guide

TR-4099: NetApp SANtricity SSD Cache for E-Series

TR-4474: SANtricity Full Disk Encryption

Interoperability Matrix Tool

Installing and Configuring for Linux Express Guide

Installing and Configuring for Linux Power Guide for Advanced Users

NetApp SANtricity Web Services Proxy

Splunk Documentation

Splunk>docs

Splunk>answers

Splunk>wiki

Splunk Enterprise Installation Manual

Splunk Enterprise Capacity Planning Manual

Splunk Enterprise Managing Indexers and Clusters of Indexers

NetApp SANtricity Performance App for Splunk Enterprise

Technology Add-On for NetApp SANtricity

Version History

Version	Date	Document Version History
Version 1.0	September 2017	Original release.

Refer to the Interoperability Matrix Tool (IMT) on the NetApp Support site to validate that the exact product and feature versions described in this document are supported for your specific environment. The NetApp IMT defines the product components and versions that can be used to construct configurations that are supported by NetApp. Specific results depend on each customer's installation in accordance with published specifications.

Copyright Information

Trademark Information