Apeiron Reference Architecture

for Splunk

Executive Summary

Big Data is all about collecting information from virtually any source, in any format and using it for deriving valuable intelligence and decision making. Splunk® Enterprise is a platform that can provide a central repository for collecting, monitoring and analyzing Big Data in real time.

Modern information technology (IT) advancements including solid-state drives (SSD) and NVM Express® (NVMe) are transforming business. Unfortunately, current enterprise data center architectures are not optimized to efficiently run Splunk and this becomes more apparent as datasets and ingest rates grow. IT solutions reliant on NAS (Network Attached Storage), SAN (Storage Area Networking), virtualization, cloud, and hyper-convergence products create inherent limits that become insurmountable when attempting to introduce the newest technologies such as storage class memory SSDs (Intel® Optane™, for example).

A different and innovative approach is required to fully achieve the potential of Splunk at scale. Apeiron® offers proven solutions available now—the Apeiron Splunk Appliances™ (ASAs)—that support the fastest SSDs without blocking their native performance and support an unlimited number of indexers and search heads while managing petabytes of data. ASAs incorporate NVMe over Ethernet™ technology to create fabric-based direct attached storage (DAS).

The true value of Big Data analytics is in the following use cases:

- Predictive and Proactive Decision Support

  The collection of data and monitoring of key performance metrics enables the early detection and awareness of issues, rapid troubleshooting of problems and the ability to proactively respond with actions that can avoid operational impacts to the business.

- Quality Processes and Customer Satisfaction

  The capturing of historical data and real-time data from transactional systems enables an understanding of trends, a recognition of patterns of activity, capturing of customer buying behaviors for enhanced and personalized customer engagement to achieve a competitive advantage.

Big Data applications are no longer a luxury but have become a necessity to organizations. Whether it is real-time monitoring of security logs, Internet of Things (IoT) events, or other operational events requiring decisions based on rules or machine learning algorithms, the power to proactively perform meaningful search and real-time processing is of

paramount value to the success of any business. Apeiron storage is a perfect fit for addressing Big Data applications like Splunk—based on the following selection criteria:

- Extreme performance resulting from the architecture
- Unique linear scalability to support dynamic growth
- Heightened resiliency of the purpose-built design
- Unprecedented cost-effectiveness of the solution

Why Apeiron for Splunk

Companies and organizations using Splunk, or planning a Splunk deployment, can achieve an order of magnitude (10x) better indexing performance and nearly two orders of magnitude (100x) better search performance using Apeiron Splunk Appliances rather than Splunk reference architecture that assumes traditional controller-based SAN or NAS.

Apeiron storage with Splunk Enterprise provides the integrated platform to ingeniously ask questions about data with the speed required to maximize business decisions, and deliver true customer value. The innovation of Apeiron storage provides extreme performance to meet the demands of large Splunk Enterprise deployments while delivering both capital expense (CAPEX) savings on the storage hardware investment and operating expense (OPEX) savings based on simplicity of operations and management, dramatically lowering the total cost of ownership (TCO).

Apeiron provides an integrated, pre-tested storage solution for deploying Splunk that meets the most demanding service level agreements (SLAs) for performance, scalability and resiliency. Apeiron delivers consistent sub-ms I/O latency, high speed data ingest and indexing, and real-time aggregation for search, analysis and visualization.

Apeiron eliminates the complexity of managing a diverse set of storage silos by providing a single point of storage management and monitoring across Splunk tiers of hot, warm, and cold stages, all from a consolidated and unified storage pool that delivers the highest level of performance. Apeiron's compelling cost effectiveness eliminates the need to ever maintain a frozen tier. It also allows customers to maintain much bigger active datasets enabling much richer queries and detailed studies.

As customers increase their data retention periods, Apeiron can dynamically add higher capacity drives which reduces floor space in the data center and achieves cooling and power reductions for increased cost effectiveness. The use of commercial NVMe SSDs enables customers to make use of the latest in NVM technology (i.e. 15TB/30TB NAND SSDs, Intel Optane SSDs). Apeiron uniquely provides a future-proof storage solution that can address the needs of Splunk's applications today with the ability to organize, store and analyze the rapidly growing datasets and workflows in the future.

1 Products Overview

1.1 Apeiron

Apeiron products offer SSD storage to servers using NVMe over Ethernet, effectively creating a fabric-based direct attached storage (DAS) solution. NVMe over Ethernet adds less than 2 microseconds of latency to native SSD performance, an amount that is invisible to application servers who normally service storage with typical latencies of 100 to 200 microseconds. This results in servers experiencing identical performance whether SSDs are installed in servers, or Apeiron storage systems. Minimizing added latency with NVMe over Ethernet is especially beneficial when using SSDs based on storage class memory (SCM), including Intel Optane and Micron® QuantX™, with typical latencies of 10 to 15 microseconds.

Several years ago, the information technology (IT) industry recognized an opportunity to exploit the low latency, high throughput, and internal parallelism of flash-based solid-state drives (SSDs). This resulted in the creation of Non-Volatile Memory Express (NVM Express or NVMe) technology and standards. NVMe uses the PCI Express (PCIe) bus for connections to offer performance increases over the SATA, SAS or Fibre Channel buses commonly used for storage systems and devices. While PCIe connections are difficult to externalize, they are useful for connecting servers to their internal SSDs and storage systems to their local SSDs. With NVMe over Ethernet these benefits can be extended to external networked SSDs.

Other important NVMe over Ethernet benefits include:

- Production NVMe over Ethernet products are shipping now
- NVMe over Ethernet technology can provide performance greater than SSDs installed in servers by spreading I/O workloads
- NVMe over Ethernet solutions can be used to expand already existing application clusters utilizing internal server storage

1.1.1 Apeiron Storage Systems

The Apeiron storage systems transform standard NVMe SSDs into networked NVMe over Ethernet storage. The ADS1000 storage systems support up to 24 SSDs (2.5-inch U.2) per enclosure and occupy 2 rack units (2RU) of space. Storage capacity and performance (latency, bandwidth, and IOPS) is equivalent to the combined total of installed SSDs. Each Apeiron storage system includes 2 internal, non-blocking NVMe over Ethernet switches providing a total of 32 40Gbit

NVMe over Ethernet external ports. These ports create a 40Gbit NVMe over Ethernet fabric for interconnecting Apeiron storage systems, Apeiron server systems, and x86 standard servers with an installed Apeiron Host Bus Adapter (HBA).

Another key difference of Apeiron storage systems is they use a controller-less architecture. Traditional enterprise storage systems rely on storage controllers—essentially, dedicated servers and software—for host connectivity, device management, data services, and more. Apeiron storage systems connect SSDs directly to servers using a high-speed network instead of storage controllers.

Splunk is storage-aware and designed to directly manage its storage, and includes the powerful storage capabilities of data compression and replication. This means there is no benefit from running similar data services on enterprise storage controllers. The elimination of bottlenecks is a key reason why the same Splunk indexers and search heads achieve much faster performance when using Apeiron storage systems rather than traditional enterprise storage systems.

Figure 1.1.1-1 Apeiron Storage System

1.1.2 Apeiron Server Systems

The Apeiron server systems are ready-to-use NVMe over Ethernet solutions. Each model ADS-N8101 server system occupies 1 rack unit (1RU) of space and includes up to 56 Intel® Xeon® cores with up to 112 threads, 128 GB of DDR4-2133 ECC memory, 1 or 2 Apeiron host bus adapters, dual 10Gbit Ethernet local area network (LAN) ports, , and 2 RAID 0/1 hard drives for booting operating systems and applications. Apeiron server systems are pre-configured to optimize NVMe over Ethernet for specific application roles, so compute and memory configurations may differ between Apeiron servers deployed as Splunk indexers versus search heads.

Figure 1.1.2-1 Apeiron Server System

1.1.3 Apeiron Host Adapters

The Apeiron host adapters model ADS40G add NVMe over Ethernet support to standard x86 servers. Apeiron host adapters feature PCIe Gen3 x8 host interfaces and dual 40Gbit NVMe over Ethernet ports using industry-standard Quad Small Form-Factor Pluggable Plus (QSFP+) transceiver modules. Apeiron host adapters provide the specialized hardware and software x86 servers need to access Apeiron storage systems using NVMe over Ethernet. Depending on the type of QSFP+ module installed, Apeiron host adapters support industry-standard copper cables for short-distance or optical cables for long-distance connections.

Figure 1.1.3-1 Apeiron Host Adapter

1.1.4 Apeiron Splunk Appliances (ASA)

Apeiron recommends the use of a Apeiron Splunk Appliance (ASA) for both distributed and single-instance Splunk environments. ASAs are designed and optimized NVMe over Ethernet solutions that maximize Splunk index and search performance and simplify Splunk capacity planning and deployment. Each Apeiron Splunk Appliance includes one or more Apeiron storage systems and Apeiron server systems interconnected by the Apeiron 40Gbit Ethernet fabric. Several pre-configured ASA models are available, with each supporting a specific combination of Splunk ingest, indexing, searching, and retention capacity. ASAs scale with the addition of Apeiron storage systems and Apeiron server systems to support any Splunk requirements. Apeiron Splunk Appliances include an Apeiron Storage Manager (ASM) which enables a single interface to monitor the appliance performance and allows easy appliance management. The ASM also simplifies the integration into normal data center management environments.

Figure 1.1.3-1 Apeiron Splunk Appliance (ASA)

1.2 Splunk

Splunk (the company) develops software (the products commonly referred to as Splunk) for monitoring, capturing, correlating, searching, and analyzing machine-generated Big Data from sources including technology infrastructure, security systems, business applications, and websites. Splunk software products include the core Splunk Enterprise and extended products including Splunk Enterprise Security (ES), Splunk IT Service Intelligence (ITSI), and Splunk User Behavior Analytics (UBA). You can learn more about Splunk by visiting www.splunk.com.

This *Apeiron Reference Architecture for Splunk* focuses on the core Splunk Enterprise and extended Splunk ES, ITSI, and UBA products. However, Apeiron Splunk Appliances—Apeiron storage systems, server systems, and host adapters—fully support and can benefit other Splunk software products, quick start bundles, apps, and add-ons. Contact Apeiron for assistance with adapting *Apeiron Reference Architecture for Splunk* recommendations for products other than Splunk Enterprise, ES, ITSI, and UBA.

The following descriptions are based upon information published by Splunk and provided for convenience only.

1.2.1 Splunk Enterprise

Splunk Enterprise makes it simple to collect, analyze and act upon the untapped value of the Big Data generated by your technology infrastructure, security systems and business applications—giving you the insights to drive operational performance and business results. Splunk Enterprise monitors and analyzes machine data from any source to deliver Operational Intelligence to optimize your 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.

1.2.2 Splunk Enterprise Security (ES)

Splunk Enterprise Security gives you the answers you need to quickly detect and respond to internal and external attacks. It simplifies threat management while minimizing risk and safeguarding your business. Splunk ES streamlines all aspects of security operations and is suitable for organizations of all sizes and expertise. Splunk ES is a SIEM that provides insight into machine data generated from security technologies such as network, endpoint, access, malware, vulnerability and identity information. Whether deployed for continuous real-time monitoring, rapid incident response, a security operations center (SOC), or for executives who need a view of business risk, Splunk ES delivers the flexibility to customize correlation searches, alerts, reports and dashboards to fit specific needs.

1.2.3 Splunk IT Service Intelligence (ITSI)

Splunk IT Service Intelligence monitors the health and key performance indicators of critical IT services. Splunk IT Service Intelligence is a next-generation monitoring and analytics solution that uses machine learning and event analytics to simplify operations, prioritize problem resolution and align IT with the business. Splunk IT Service Intelligence is a next-generation monitoring and analytics solution that uses machine learning and event analytics to simplify operations, prioritize problem resolution and align IT with the business.

1.2.4 Splunk User Behavior Analytics (UBA)

Detects cyber-attacks and insider threats using data science, machine learning and advanced correlation. Splunk User Behavior Analytics (UBA) is a machine learning-powered solution that delivers the answers you need to find unknown threats and anomalous behavior across users, endpoint devices and applications. It not only focuses on external attacks but also the insider threat. Its machine learning algorithms produce actionable results with risk ratings and supporting evidence that augment security operation center (SOC) analysts' existing techniques for faster action. Additionally, it provides visual pivot points for security analysts and threat hunters to proactively investigate anomalous behavior.

2 Solutions Architecture

Splunk Enterprise deployments can be simple or complex, small or large, and always grow over time as the enterprise adopts the tool for more use-cases. The smallest configuration is a single-instance, combining the roles of indexer and search head on one server. Large configurations use a distributed model with dedicated instances for the various Splunk server roles (i.e. Search heads, Indexers, Cluster Master, etc.). A broad range of configuration options exists between these examples, and Splunk recommends different server specifications for different environments, further complicating Splunk planning and deployment.

One of the biggest challenges faced by new Splunk customers is determining how to best deploy new infrastructure and software to get the required performance. With the standard reference storage solutions, tuning a new Splunk installation can be time consuming and costly.

Some key factors that impact Splunk server guidelines and configurations include:

- Amount of data being ingested
- Amount of indexed data
- Number of concurrent queries
- Number of saved searches
- Types of searches being run
- Whether ES is active

For this reason, the *Apeiron Reference Architecture for Splunk* recommends following the most recent guidelines published by Splunk with certain exceptions detailed below to adapt Splunk's standard guidelines to the unique opportunities created by Apeiron Splunk Appliances with NVMe over Ethernet.

2.1 Apeiron Reference Guidelines

Apeiron Splunk Appliances consist of indexers and search heads connected to direct attached storage made up of NVMe SSDs, with everything interconnected using high-speed, non-blocking, and built-in Ethernet networks. The use of NVMe over Ethernet technology eliminates storage controllers that would limit performance and capacity. ASAs increase Splunk capacity and performance to levels not possible using Splunk reference machines with internal storage, enterprise SAN or enterprise NAS. Ultimately, Apeiron Splunk Appliances create the opportunity to improve on Splunk's guidelines in unique ways.

Splunk reference machines and recommendations assume traditional IT environments, but these include many inefficiencies and duplicate features that are unnecessary for Splunk. These redundant features can reduce the performance of Splunk applications. Splunk Enterprise includes and manages data compression and replication, for

example, but most enterprise storage systems include similar features that are unneeded. The common practice of virtualizing IT infrastructure with traditional virtualized storage systems introduces additional inefficiencies that are not conducive to a Splunk infrastructure. However, these environments can be supported by Apeiron when required.

Perhaps worse of all, enterprise storage systems effectively operate as storage servers running storage software that manages data services for a group of hard disk drives (HDDs) and solid-state drives (SSDs)—and introducing servers between Splunk indexers and search heads and their storage is grossly inefficient. Unlike these traditional storage systems, Apeiron is optimally designed to accelerate Splunk performance by not adding data services that are already built into the Splunk Architecture.

For these reasons, Apeiron offers pre-configured Apeiron Splunk Appliances based on the optimal combination of Splunk and Apeiron recommendations, guidelines, and best practices. ASA capacities begin at sizes designed for Splunk environments with ingest requirements on hundreds of gigabytes per day (GB/day), yet scale linearly to support hundreds of terabytes per day (TB/day) by adding Apeiron storage systems for data retention and server systems as indexers and search heads. In other words, ASA deployments can grow to include tens of servers and hundreds, or thousands of SSDs.

Without Apeiron Splunk Appliances, Splunk deployments using Splunk reference machines could require up to 5x more - servers along with at least one enterprise storage system appropriately configured to support Splunk capacity, performance and retention requirements.

Apeiron Splunk Appliances	ASA 500	ASA 1000	ASA 2000
Splunk Enterprise Ingest	Up to 750 GB/day	Up to 1,500 GB/day	Up to 3,000 GB/day
Splunk ES Ingest	Up to 500 GB/day	Up to 1,000 GB/day	Up to 2,000 GB/day
Splunk ITSI Ingest	Up to 500 GB/day	Up to 1,000 GB/day	Up to 2,000 GB/day
Splunk UBA Ingest	Up to 500 GB/day	Up to 1,000 GB/day	Up to 2,000 GB/day
Splunk Clustering	Enabled	Enabled	Enabled
Splunk Replication Factor	2	2	2
Indexers (Servers)	2	2	4
Search Heads (Servers)	3	3	3
Deployment Servers	1	1	1

Apeiron Splunk Appliances	ASA 500	ASA 1000	ASA 2000
Storage Systems	1	1	3
Storage Capacity	198 TB	352 TB	704 TB

Table 2.1-1 Splunk One Year Data Retention

Apeiron Splunk Appliances	ASA 500	ASA 1000	ASA 2000
Splunk Enterprise Ingest	Up to 750 GB/day	Up to 1,500 GB/day	Up to 3,000 GB/day
Splunk ES Ingest	Up to 500 GB/day	Up to 1,000 GB/day	Up to 2,000 GB/day
Splunk ITSI Ingest	Up to 500 GB/day	Up to 1,000 GB/day	Up to 2,000 GB/day
Splunk UBA Ingest	Up to 500 GB/day	Up to 1,000 GB/day	Up to 2,000 GB/day
Splunk Clustering	Enabled	Enabled	Enabled
Splunk Replication Factor	2	2	2
Indexers (Servers)	2	3	5
Search Heads (Servers)	3	3	3
Deployment Servers	1	1	1
Storage Systems	2	2	5
Storage Capacity	396 TB	704 TB	1,278 TB

Table 2.1-2 Splunk Two Year Data Retention

2.2 Splunk Reference Guidelines

This *Apeiron Reference Architecture for Splunk* assumes environments have grown beyond the capacity of any single-server Splunk instance, so Apeiron Splunk Appliances future-proof distributed deployments with optimal scalability. Table 2.2.2-1 directly compares Splunk Reference Machines and Apeiron Splunk Appliances recommendations to facilitate capacity planning.

For example, the *Splunk Distributed Deployment Manual* includes a Splunk Enterprise performance recommendation of up to 100GB/day per indexer when using Splunk reference machines. In reality, when architecting systems Splunk routinely specifies no more than 70GB/day per indexer if the standard ES queries are active. Apeiron supports up to 500GB/day when using Apeiron Splunk Appliances. The data shown in Table 2.2-1 demonstrate that Apeiron Splunk Appliances can deliver capacity and performance advantages ranging from 5x to 50x.

Splunk Server Type	Splunk Reference Architecture Performance Capability	Apeiron Splunk Appliance Performance Capability
Indexer (Server)	Up to 100 GB/day	Up to 500 GB/day (5x better)
Search Head (Server)	Up to 50,000 events per second for dense searches	Up to 500,000 events per second for dense searches (10x better)
	Up to 5,000 events per second for sparse searches	Up to 50,000 events per second for sparse searches (10x better)
	Up to 1/2 bucket per second for super-sparse searches	Up to 25 buckets per second for super-sparse searches (50x better)
	From 10 to 50 buckets per second for rare searches	From 50 to 250 buckets per second for rare searches (5x better)

Table 2.2-1 Splunk Enterprise Indexer and Search Head Performance Comparison

3 Solution Performance

3.1 Competitive Splunk Deployment Comparison

Enterprise storage systems are commonly selected for Splunk environments, especially distributed deployments, so it makes sense to compare Apeiron Splunk Appliances with traditional SAN and NAS alternatives. Apeiron choose to involve third-parties to facilitate and validate a competitive Splunk deployment comparison:

- Global IT Solution Provider

  This information technology solution provider is an industry leader with billions in annual revenue and thousands of employees worldwide. Their team initially performed more than one month of performance benchmarking activities in their technology center and has been running Splunk testing on Apeiron equipment for over a year.

- **Enterprise Strategy Group® (ESG)**

  ESG is an IT analyst, research, validation, and strategy firm that provides market intelligence and actionable insight to the global IT community. The ESG team audited benchmarking results for three weeks and published their findings as an ESG Labs Spotlight report.

3.1.1 Splunk Reference Configuration

Prior to discovering the Apeiron Splunk Appliances, the traditional enterprise Splunk deployments at the solution provider where implemented based on standard Splunk guidelines, recommendations, and best practices. The purpose of the three Spunk configurations under competitive comparison, and the technology center overall, is to support engineers and customers with building and testing solutions that simulate real world customer workloads.

Apeiron provided an ASA for comparison against two configurations based on Splunk reference architecture including NAS or SAN storage from industry leading providers. All three configurations deployed Splunk using a total of seven (7) servers and one (1) storage system as follows: four (4) indexers, one (1) search head, one (1) cluster master, one (1) ES server, and one (1) storage system. Servers were configured to include the same number of processors, cores, and memory.

The key difference among the three equivalent configurations was storage networks and systems as shown in Figures 3.1.1-1, 3.1.1-2, and 3.1.1-3. This supported a direct Splunk performance comparison.

Figure 3.1.1-1
Splunk with NVMe over Ethernet
(ASA Solution Configuration)

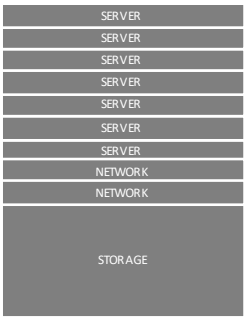

Figure 3.1.1-2
Splunk with Ethernet NAS
(Bare Metal Configuration)

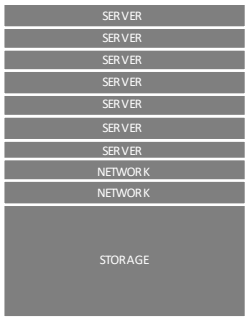

Figure 3.1.1-3
Splunk with Fibre Channel SAN
(Virtualized Configuration)

3.1.2 Apeiron Splunk Appliance (ASA) Configuration

As illustrated by Figure 3.1.1-1, the Apeiron Splunk Appliance provided a complete NVMe over Ethernet solution for Splunk Enterprise. The Apeiron storage system includes 32 dedicated storage networking ports and in this example NVMe SSDs from four distinct suppliers (Intel, Micron, Samsung and Toshiba) as the customer was interested in understanding the interoperability of various NVMe suppliers.

3.1.3 Splunk Performance Comparison

A direct comparison of Splunk dense search performance, using the same data set including approximately 55 million Syslog events, reveals that ASA offers performance advantages approaching one order of magnitude. ASA performance is 7x to 8x higher than "bare metal" environments including physical servers (without hypervisors) with traditional enterprise storage, and 8x to 9x faster than "virtualized" environments including servers running hypervisors with traditional enterprise storage. Details are included in Table 3.1.3-1.

Splunk Dense Search	Bare Metal	Virtualized	ASA Solution	ASA Benefits
Search Time (Seconds)	4,834	5,596	472	10x to 12x Faster
Records per Second	12,388	10,538	124,007	10x Faster

Table 3.1.3-1 Splunk Search Performance Comparison

3.1.4 Splunk Budget Comparison

It is customary for IT equipment suppliers to avoid publishing pricing for their products and this complicates quantitative budgetary comparisons. It seems useful, however, to highlight related factors that impact potential capital expenses and operating expenses. For example, ASA deployments provide a simple server consolidation benefit by reducing the number of indexers to ingest 500GB/day from 5 Splunk indexer reference machines to 1 Apeiron indexer—reducing the number of servers required for indexing by 80%—and consolidation is a proven method for lowering IT related costs and expenses.

Finally, it seems reasonable to assume that reducing IT equipment requirements for Splunk would deliver significant operating expense savings due to reductions in space, power, and cooling requirements. A strong case can be made that replacing disparate brands and models of data center equipment with a unified Apeiron Splunk Appliance including fewer storage systems and server systems—and no external network switches—would deliver powerful budgetary benefits.

3.2 Global IT Solution Provider Validation

3.2.1 Overview of Validation

The solution provider chose to validate Apeiron Splunk Appliance performance using two phases of testing using Splunk Enterprise version 6.5 software and various equipment and methods. This included the use of standard Splunk monitoring and performance tools to report the results. All 60 out of the box Splunk Enterprise correlation searches were run in parallel with ingestion and ad-hoc searches.

During the first phase of testing, they compared the expected performance results using three different Splunk search types: rare, super sparse, and sparse. The solution provider experienced the ASA outperforming the Splunk recommended reference architecture for all searches.

Later, during the second testing phase, the solution provider focused on running the same Splunk dense searches using a data set consisting of 55 million records including 70 billion events across two weeks of time in three different environments: a bare metal environment with traditional storage, a virtualized environment with traditional storage, and an ASA environment. In this context, the ASA environment is best described as a bare metal environment using NVMe over Ethernet instead of SAN or NAS.

3.2.2 Key Findings

The results of the comparative performance testing set new standards for Splunk performance for the solution provider and their technology center. Specifically, they revealed the ASA completed two different super sparse searches 58x and 88x faster than the reference architecture they configured to Splunk recommendations. Splunk sparse searches completed at least 10x faster on ASA than the Splunk reference architecture.

And while Splunk identifies super sparse searches as I/O-bound, which should be expected to complete sooner on the ASA, it is important to note that CPU-bound searches also completed much faster on ASA—Spares searches completed at least 10x faster and rare searches completed 2x to 5x faster. This demonstrates the inherent advantages that NVMe over Ethernet using DAS provides over traditional enterprise SAN and NAS architectures.

Splunk Search Type	Splunk Events Found	Splunk Reference Search Time	Apeiron ASA Search Time	Apeiron ASA Advantage
Rare	23	11.2	2.1	5.3x
	115	11.2	6.1	1.8x

Splunk Search Type	Splunk Events Found	Splunk Reference Search Time	Apeiron ASA Search Time	Apeiron ASA Advantage
Super Sparse	26,802	1,112.2	12.6	88x
	180,850	1,112.2	19.2	58x
Sparse	155,459,317	31,091.9	2842.2	10.9x
	1,126,745,647	225,349.1	14,985.3	15.0x

Table 3.2.2-1 Splunk Reference Architecture vs. Apeiron Splunk Appliance Performance

3.2.3 Summary

The solution provider's validation of Apeiron Splunk Appliances featuring NVMe over Ethernet proved an opportunity to reset and advance reference architectures for Splunk. The limiting factor is not Splunk. Instead, the continued use of traditional enterprise architectures, products, and practices designed for mixed-use environments are holding back the potential of Splunk. The simplest first step is a migration to pre-designed solutions like the Apeiron Splunk Appliances.

3.3 Enterprise Strategy Group (ESG) Audit

3.3.1 Introduction to ESG

Enterprise Strategy Group (ESG), founded by Steve Duplessie, is an integrated IT research, analyst, strategy, and validation firm that is world renowned for providing actionable insight and intelligence to the global IT community. They directly connect research proficiency and operational knowledge, deliver recommendations customized to unique business needs, and blend ongoing market and technology analysis, independent partner research, and best practice know-how from years of experience working with technology growth companies of all types. You can learn more about ESG be visiting www.esg-global.com.

3.3.2 Abstract of ESG Lab Spotlight

ESG audited the performance testing by the solution provider and published the results as an ESG Lab Spotlight titled *Optimizing Splunk Deployments with NVMe Direct Scale-out*. The document, a brief 4-pages in length, independently confirms the performance benchmarks of the Apeiron Splunk Appliance and Splunk reference architecture at the solution provider's technology center.

According to ESG:

- "Organizations want access to all of their data as fast as possible without sacrificing performance, cost, or potential insight. Consequently, it is essential to find a future-proof technology solution that can not only support the entire infrastructure, but improve existing resource utilization to deliver higher levels of ROI [return on investment]."
- "Apeiron delivers extreme performance in a shared infrastructure. In fact, ESG Labs validated the performance delivered by ASA is actually faster than DAS (through the ability to spread workloads across more drives)."
- "The benefits of having Apeiron as the underlying storage infrastructure to support a dynamically growing Splunk deployment are obvious."

3.3.3 Testing Overview: Phase 1

The purpose of the ESG Lab's first phase of audits focused on comparative ingest rates. As shown in Figures 3.3.3-1 and 3.3.3-2, ESG confirmed the Apeiron Splunk appliance under test supported average ingest rates of 10 TB/day with fully healthy status for queues, CPU, and memory when running a high number of concurrent queries. At times, the ASA achieved ingest rates as high as 12 TB/day. When the high number of concurrent queries were removed, the ASA routinely reached ingest rates of 20 TB/day or more. These levels of ingest performance were documented over a period of several weeks and has been repeated for various customers for over a year by the solution provider.

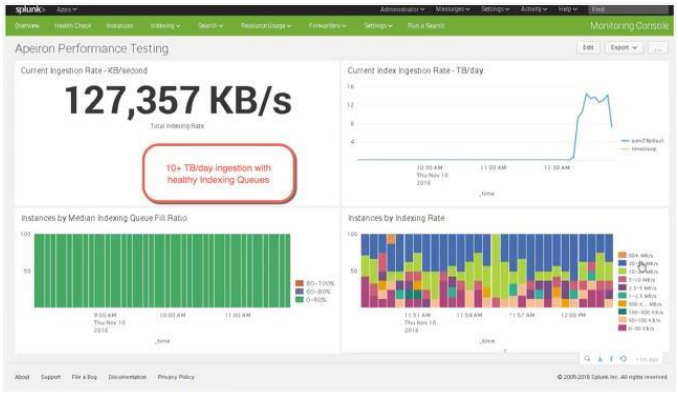

Figure 3.3.3-1 Apeiron Splunk Ingest Performance

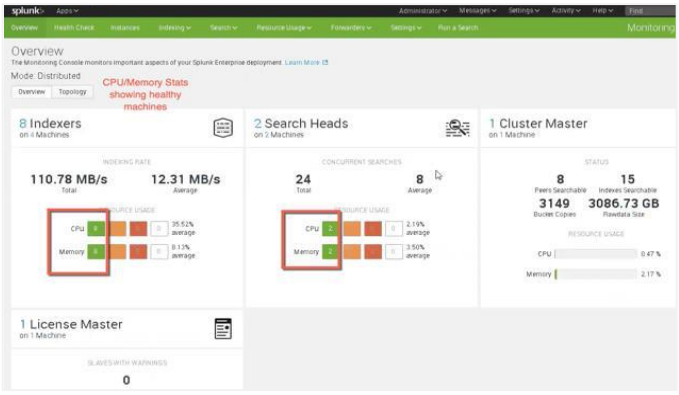

Figure 3.3.3-2 Apeiron Splunk Server Performance

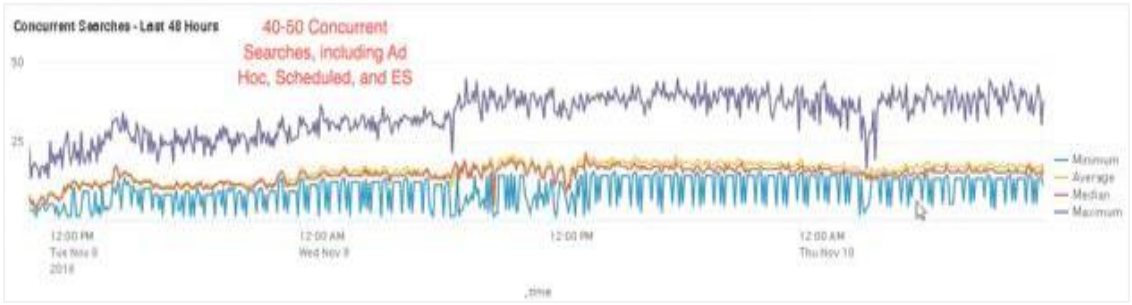

Figure 3.3.3-3 Apeiron Splunk Concurrent Performance

The ASA compared extremely favorably against similarly configured Splunk reference architectures. In the latter case, four indexers and one search head using traditional enterprise storage supported ingest rates of approximately 1 TB/day—a reduction of at least 90% versus ASA. These findings suggest the ingest rates of Splunk reference architectures and reference machines are one order of magnitude slower than Apeiron Splunk Appliances.

3.3.4 Testing Overview: Phase 2

For the second phase of audits, ESG Labs focused on comparative search performance. This effort audited and confirmed the results the solution provider experienced during their performance benchmarking of the Apeiron Splunk Appliance and Splunk reference architecture as shown in Tables 3.3.2-1 and 3.1.3-1. In summary, the ASA outperformed Splunk reference architecture using the same number of indexers and search heads, as follows:

ASA Splunk Search Type Comparisons

- 58x to 88x faster super sparse searches
- 10.9x to 15x faster sparse searches
- 1.8x to 5.3x faster rare searches

ASA Splunk Enterprise Environment Comparisons

- 7x to 8x faster than physical environments
- 8x to 9x faster than virtualized environments

3.3.5 Summary

The audit by ESG Labs verified what the solution provider demonstrated in their technology center—Apeiron Splunk Appliances typically provide one order of magnitude better performance (a 10x increase) than Splunk reference architectures using the same number of similarly configured indexers and search heads. During certain Splunk searches, the performance benefits of Apeiron Splunk Appliances can approach two orders of magnitude (a nearly 100x increase).

Simply adding NVMe and NVMe over Fabric to traditional enterprise storage systems will not be able to achieve the performance of Apeiron due to the innovations of Apeiron. One key enabler of the ASA advantage is the use of Apeiron storage systems supporting NVMe over Ethernet versus traditional enterprise storage systems. Another is pre-designed models of Apeiron Splunk Appliances deliver pre-determined and reliably predictable levels of Splunk ingesting, indexing, searching, and retention capacity.

According to ESG:

- "The benefits of having Apeiron as the underlying storage infrastructure to support a dynamically growing Splunk deployment are obvious."
- "As data analytics tools such as Splunk continue to mature and add value to the business, a shift in focus will be needed. Data sets continue to grow and traditional infrastructures struggle to meet performance, scalability and cost requirements."
- "Traditional storage approaches introduce the potential for this mission-critical workflow to be interrupted. As an example, a traditional storage infrastructure for Splunk deploys three separate networking protocols (FC, IB and Ethernet). This places a tremendous burden on the IT staff and their budget."
- "The Apeiron architecture was designed to provide all the performance NVMe offers, and to leverage a single networking protocol; an Ethernet fabric which drives down both costs and risk.
- "With petabyte-scale NVMe storage, no external switching, and significant server CPU benefits, the consolidation of hardware and IT functions provides a compelling ROI/TCO [total cost of ownership] justification [for Apeiron Splunk Appliances]."

4 Solution Configuration

4.1 Hardware Configuration

Apeiron Splunk Appliances are available in a number of models matched to specific Splunk requirements for indexer and search head performance, data protection, and data retention. ASA models are designed with a capacity to support a specific average ingest rate and are available in options supporting one- or two-years of indexer data using a Splunk replication factor of 2 without multi-site replication. As previously stated, Apeiron Splunk Appliances scale to support any Splunk requirements with the addition of optional Apeiron storage systems and server systems.

4.1.1 Apeiron Splunk Appliance 500

ASA model 500 is designed to support Splunk environments with ingest rates up to 500 GB per day with standard options to support a specific period of indexing data. ASA-500-1 supports one-year of indexing data and includes one Apeiron indexer, three Apeiron search heads, one Apeiron deployment server, and one Apeiron storage system with a capacity of 198 TB. ASA-500-2 supports two-years of indexing data and includes one Apeiron indexer, three Apeiron search heads, one Apeiron deployment server, and one Apeiron storage system with a capacity of 396 TB.

Figure 4.1.1-1 ASA Model 500-1 Configuration

Figure 4.1.1-2 ASA Model 500-2 Configuration

4.1.2 Apeiron Splunk Appliance 1000

ASA model 1000 is designed to support Splunk environments with ingest rates up to 1,000 GB per day with standard options to support a specific period of indexing data. ASA-1000-1 supports one-year of indexing data and includes two Apeiron indexers, three Apeiron search heads, one Apeiron deployment server, and one Apeiron storage system with a capacity of 352 TB. ASA-1000-2 supports two-years of indexing data and includes three Apeiron indexers, three Apeiron search heads, one Apeiron deployment server, and two Apeiron storage systems with a combined capacity of 704 TB.

Figure 4.1.2-1 ASA Model 1000-1 Configuration

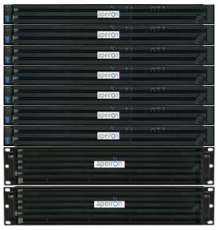

Figure 4.1.2-2 ASA Model 1000-2 Configuration

4.1.3 Apeiron Splunk Appliance 2000

ASA model 2000 is designed to support Splunk environments with ingest rates up to 2,000 GB per day with standard options to support a specific period of indexing data. ASA-2000-1 supports one-year of indexing data and includes four Apeiron indexers, three Apeiron search heads, one Apeiron deployment server, and three Apeiron storage systems with a

capacity of 704 TB. ASA-2000-2 supports two-years of indexing data and includes five Apeiron indexers, three Apeiron search heads, one Apeiron deployment server, and five Apeiron storage systems with a combined capacity of 1,278 TB.

Figure 4.1.3-1 ASA Model 2000-1 Configuration

Figure 4.1.3-2 ASA Model 2000-2 Configuration

5 Conclusions

Splunk Enterprise and related products, solid-state drives (SSD), and NVM Express (NVMe) technology are transforming business. Unfortunately, current IT architectures are not optimized to efficiently support data centers running Splunk and emerging technologies such as storage class memory SSDs (Intel Optane, for example) and NVMe over Fabrics face unsurmountable limits inherent to current IT solutions reliant on NAS, SAN, virtualization, cloud, and hyper-convergence architectures and products. A different and innovative solution is required that matches the potential of Splunk.

Apeiron offers a proven solution for optimizing Splunk that is available now, as documented in this *Apeiron Reference Architecture for Splunk*—the Apeiron Splunk Appliances—that are demonstrated, validated, and audited, as follows:

- Apeiron demonstrated Apeiron Splunk Appliances eliminate the guesswork Splunk customers face during capacity planning and simplify Splunk distributed deployments.
- Apeiron Splunk Appliances eliminated the weeks of work normally involved with tuning new Splunk installations due to storage infrastructure limitations.
- The solution provider validated the Splunk ingest rates, indexing capacity, searching performance, and retention periods of Apeiron Splunk Appliances during many weeks of testing.
- Enterprise Strategy Group audited the findings of the solution provider during a detailed three-week process and independently published the results as an ESG Labs Spotlight.
- The solution provider has been running actual network syslog Splunk data on Apeiron systems deployed at the technology center for over a year.

Companies and other organizations using Splunk, or planning Splunk deployments, can achieve one order of magnitude (10x) better indexer performance and nearly two orders of magnitude (100x) better search head performance using Apeiron Splunk Appliances while dramatically reducing the needed IT infrastructure.