Solution Brief

NetApp ONTAP 9 Solution for Splunk Enterprise

Lower costs, increase performance, and get an enterprise-grade data platform

Key Features

Reduce Infrastructure Costs by Up to 50%

- Reduce server and OS license costs with independent scaling of storage and compute.
- Dramatically shrink dataset space requirements with patented storage efficiency technologies.

Increase Business Agility with Seamless Hybrid Cloud Deployments

- Move data efficiently between on-premises, near cloud, and public cloud environments.
- Lower expenses by automatically moving aging data to more cost-effective storage.

Increase Splunk Performance by Up to 300%

- Complete indexing operations 2 times faster than with DAS.
- Perform searches up to 4 times faster than with DAS.

Maximize Splunk Availability for Greater Insight and ROI

- Maintain performance and data availability during controller failures.
- Sustain performance and quickly recover from failed media.

Simplify Enterprise Data Management

- Back up, restore, and replicate terabytes of Splunk data in seconds.
- Reduce IT support costs with a single-interface management view of all global data repositories.

NetApp®

The Challenge

Big data analytics is quickly becoming critical to the success of enterprises, and Splunk is a key component of that success. Enterprises use Splunk to monitor, to report on, and to analyze massive amounts of machine data, both in real time with incoming live streams and in batches by using vast historical datasets. Splunk provides operational insight from big data that helps enterprises like yours mitigate security risks, improve service levels, reduce IT operational costs, and develop improved products and service offerings.

Managing these rapidly expanding datasets requires a platform that can easily and economically grow storage capacity and performance without sacrificing security, resiliency, and simplicity. However, many big data applications, including Splunk, were initially deployed on clusters of commodity servers with internal disks (direct-attached storage [DAS]), a platform that does not scale well. As Splunk data inevitably grows, the server-based DAS model requires additional servers to house the storage, even if there are no additional computational requirements. If your enterprise cannot scale your storage and compute needs independently, you face increased costs for extra servers, application licenses, rack space, power, and cooling.

The DAS platform does not provide the performance and resiliency that you need as the applications become mission-critical. Drive and controller failures heavily degrade performance, lead to unpredictable availability, and can result in lost revenue and lower customer satisfaction. Your enterprise needs the flexibility to move aging data to lower-cost storage and to leverage the cloud for additional compute and archive capabilities. Manageability, data protection, and data governance become critical as the datasets grow, but server-based storage struggles to meet these requirements.

The Solution

The optimal solution for effectively leveraging Splunk is NetApp® ONTAP® 9 software, the enterprise data management software that powers the NetApp AFF and FAS engineered systems, as well as software-only ONTAP Cloud. ONTAP supports independent scaling of compute and storage resources, reducing costs and facilitating data tiering. ONTAP provides NetApp Integrated Data Protection to safeguard operations with near-instant backup and recovery by using the highly efficient NetApp Snapshot™, SnapRestore®, and SnapCenter® technologies.

The NetApp Data Fabric weaves together data storage regardless of location—on the premises, near the cloud, in private or public clouds—into a unified data system, enabling tremendous flexibility. Data moves seamlessly wherever you need it within the hybrid cloud, supporting data security and governance, resiliency, and data management efficiency. A unified management console provides a single, global view of all data repositories. ONTAP supports multitenancy and adaptive quality of service (QoS). All of these data management features are available with both NFS and SAN and can be accessed via FC, NFS, and iSCSI. This enables Splunk and other big data applications to run on the same enterprise-grade IT storage infrastructure that you already use for traditional core applications, including Oracle and SAP. When big data applications and traditional enterprise applications are consolidated on the same ONTAP platform, hardware, software, and management costs are reduced.

AFF systems that run ONTAP 9 are optimized specifically for flash, delivering superb performance for Splunk with consistent submillisecond latency. Hybrid FAS systems combine flash with HDDs to provide excellent performance with additional cost savings as data ages and transitions from hot to warm to cold. FabricPool extends tiering choices for cold data to the cloud.

As your Splunk deployment grows, your enterprise can maintain continuous operations without interruption while you add or update controllers and disk shelves or move data between on-premises and cloud environments. ONTAP supports a mix of drive types, models, and hardware generations, so older systems can transition from storing hot data to storing warm or cold data. With the shared storage topology of ONTAP, you can handle your data as one large pool, rather than many distributed smaller pools, thereby speeding up access times. You get maximized performance, superior flexibility, and lower TCO.

Splunk Performance Is 2 to 4 Times Faster with ONTAP

ONTAP software delivers the data management features that your enterprise Splunk production deployment needs, but performance is also a critical element. Tests were conducted to compare the performance of the traditional clusters of commodity servers with internal storage (DAS) against a NetApp ONTAP solution with a NetApp AFF storage array. Figure 1 shows the test configuration.

The tests were conducted with equivalent compute structures, and the only difference was in data storage: internal drives in the servers compared with a NetApp AFF system. You can view the complete details and test results in TR-4650: NetApp ONTAP and Splunk Enterprise.

Splunk has two primary processes:

- Indexing, where incoming data is broken down into individual events, prepared for analysis by indexing key fields, and stored
- Queries, where rich search functionality allows analysis of the data and provides multiple views of the results, including events, patterns, statistics, and visualization

The first test measured the performance of the system while new data streamed in and was indexed and stored. The observed average indexing rate was more than twice as fast on NetApp AFF than on the traditional server-based DAS system. Table 1 summarizes the results.

The ability to perform queries and analysis was tested in multiple scenarios, differentiated by the amount of data found meeting the search criteria. The testing revealed that NetApp AFF was

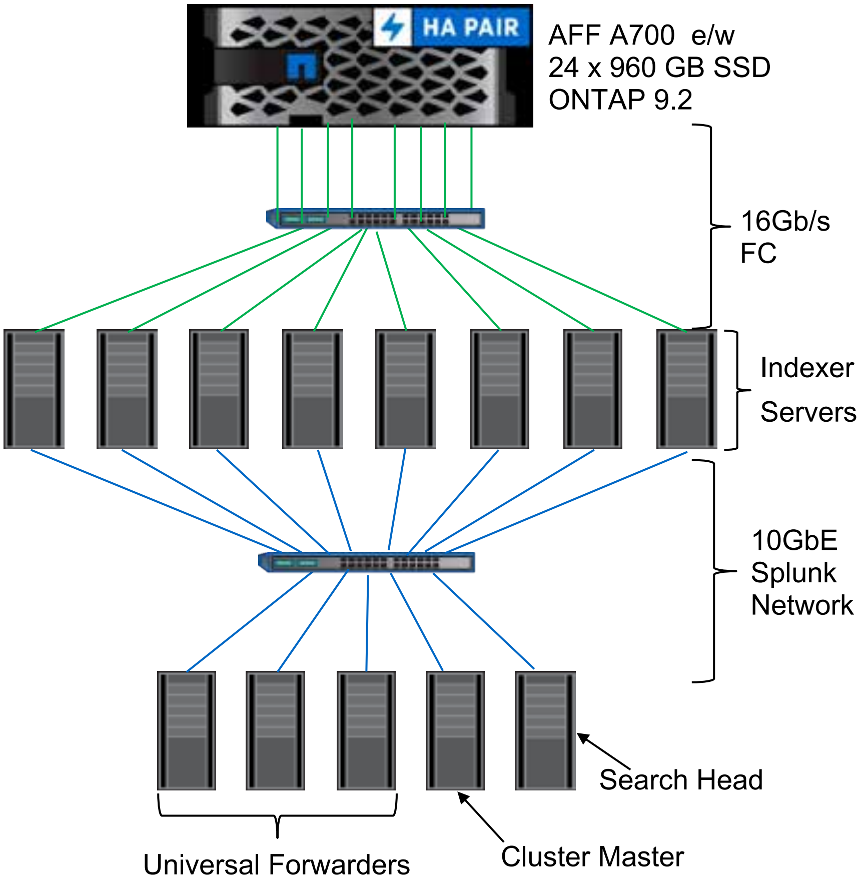

Figure 1) Test configuration

Data Ingest Volume	Average Indexing Rate Observed		
	AFF A700	Servers with DAS	AFF A700 Rate Increase
~1,000GB/day	31.114MBps	14.872MBps	+109%

Table 1) Testing showed that Splunk indexing is more than 2 times faster with ONTAP

Forwarder to Indexer Search Performed	Streaming Search Time (Seconds)			Static Search Time (Seconds)		
	AFF A700	Commodity Servers with DAS	% Faster AFF A700 Compared with DAS	AFF A700	Commodity Servers with DAS	% Faster AFF A700 Compared with DAS
Dense	8.128	19.79	143%	8.136	32.7	302%
Very dense	24.147	43.35	80%	20.226	49.15	143%
Sparse	2.065	2.07	0%	2.061	5.02	144%
Very sparse	2.065	2.07	0%	2.061	2.07	0%

Table 2) Testing showed that Splunk searches are up to 4 times faster with ONTAP

generally 2.5 to 4 times faster at performing queries than the traditional DAS approach. Table 2 summarizes the results.

ONTAP Storage Efficiency Reduces Your Data Footprint

After the performance tests, storage efficiency was compared between NetApp AFF and the traditional server-based DAS. The Splunk dataset that was generated for the test was approximately 1.9TB. ONTAP provides multiple storage efficiency technologies, including in-line deduplication, in-line compression, and in-line compaction. With these data reduction technologies applied, the NetApp AFF consumed roughly 50% of the physical storage space that the traditional servers with internal storage (DAS) required, reducing storage and electrical costs while providing additional data management features.

ONTAP Enhances System Resilience and Data Protection

Backup and restore functionality and performance were tested on the AFF system. NetApp OnCommand® System Manager was used to create Snapshot copy backups of all AFF Splunk data volumes. The Snapshot copies were created while a query was being run. Query performance was monitored during the Snapshot creation, and there was no visible impact. All Splunk data was then deleted to simulate total data loss. The entire restore process, from start to finish, took less than 5 minutes. Performing a total restore of data in a DAS deployment would require much more time because data would have to be copied from the backup location back to the production system.

A test of the ONTAP system resiliency was conducted to observe performance during a storage controller failure, which is typically a major issue for the traditional DAS platform. A Splunk workflow was run on a healthy system, and then a controller panic was induced to observe the response from Splunk. There was no visible impact on Splunk performance during the entire failure sequence, and the workflow completed successfully.

Storage media failure is a given in any system, so the performance of the NetApp AFF system with a drive failure that required RAID rebuild was tested. An SSD was failed during a Splunk index operation, and the rebuild automatically began. Reconstruction of the NetApp AFF system completed in approximately 20 minutes, while the Splunk workflow continued to run with no visible impact on performance. With DAS, this resilience is not the case. The loss of a single disk, part of a mirrored pair, means that bandwidth of the mirrored pair is cut in half, and query performance is significantly affected.

Get Enterprise Data Management, Lower Costs, and Improved Performance for Splunk

NetApp AFF systems for Splunk outperform traditional server-based storage (DAS) by wide margins while providing greater flexibility, resiliency, security, ease of management, and a significantly lower TCO. With ONTAP support for multitenancy, adaptive QoS, and both SAN and NAS, Splunk can be hosted alongside other critical enterprise applications on common IT infrastructure, further reducing costs. The NetApp Data Fabric weaves together data storage regardless of location into a unified data system, providing maximum flexibility. The shared storage topology with ONTAP enables data to be handled as one large pool, speeding up access times. You get maximized performance for Splunk and for other core enterprise applications regardless of what other applications are running on the same system.

About Splunk

Splunk offers the leading platform for Operational Intelligence. It enables the curious to look closely at what others ignore—machine data—and find what others never see: insights that can help make your company more productive, profitable, competitive, and secure. What can you do with Splunk? Just ask.

About NetApp

NetApp is the data authority for hybrid cloud. We provide a full range of hybrid cloud data services that simplify management of applications and data across cloud and on-premises environments to accelerate digital transformation. Together with our partners, we empower global organizations to unleash the full potential of their data to expand customer touchpoints, foster greater innovation and optimize their operations. For more information, visit www.netapp.com. #DataDriven