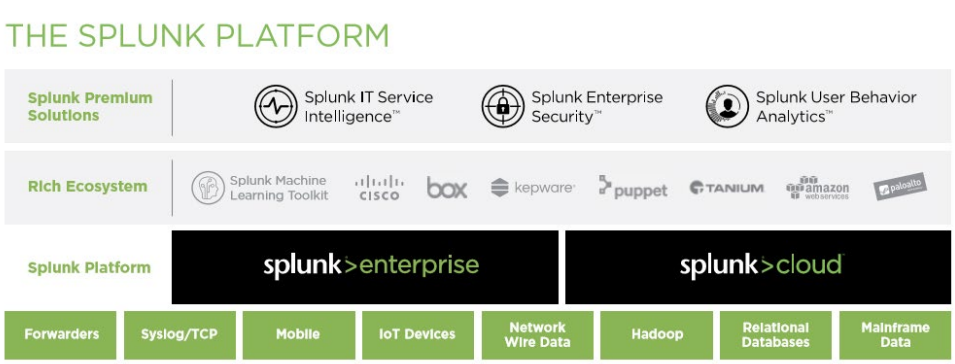
SOLUTION BRIEF

Infinidat Federal InfiniBox® and Splunk® Enterprise

Splunk turns machine data into answers, regardless of your organization's size or industry. Splunk helps you with the answers you need to solve your toughest IT, security, and business challenges. Splunk solutions are deployable on-premises, in the cloud, or via a hybrid implementation.

splunk>

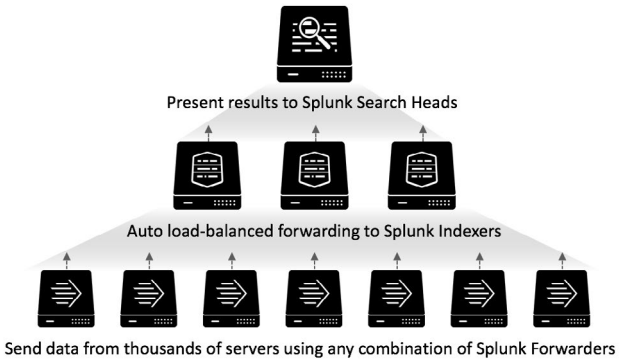
Splunk helps organizations address challenges across infrastructure and IT operations, application delivery, security and compliance, business analytics, IoT and industrial data, and more.



The data that flows through and is analyzed by Splunk can provide answers to questions like: How can I prevent downtime for my critical IT services? Why is my application performing poorly? How can I speed up security investigations and reduce the impact of insider threats? Can I drive more revenue through my website or mobile application? How can I monitor and analyze data from tens of thousands of sensors in real time?

Splunk data storage is often an afterthought that can stop a Splunk project in its tracks–particularly when the scale gets beyond proof-of-concept experiments. Splunk was originally designed with direct-attached storage (DAS) in mind, relying on storage that resides within the scale-out server nodes. For many small- or medium-sized businesses, this approach delivers the necessary performance and reliability. At scale, the DAS approach becomes problematic and inefficient.

For large-scale implementations, Splunk often requires multiple TB-per-day ingest rates, producing storage buckets that may be in the multi-petabyte range. I/O is usually a bottleneck with highly variable performance requirements. DAS cannot accommodate these requirements efficiently–in many cases, DAS approaches require more storage than necessary–typically pushing Splunk administrators and executives to consider SAN storage.



At that point, the Splunk administrator has three options:

1. **DIY using Splunk's sizing calculator.**
2. **Follow overly prescriptive reference architectures that may not be tailored to the exact requirements of the applicable use case.**
3. **Do it the easy way, leveraging Infinidat Federal's unified storage platform.**

SPLUNK STORAGE OPTIONS

OPTION 1

DIY using Splunk's sizing calculator

Dating back to 2013 and earlier, Splunk has been writing blogs to help administrators estimate the storage requirements for Splunk.[1,2] It began with relatively simple calculations, focused on the biggest data consumer: the indexer. To calculate storage requirements for the indexers, the administrator only needed to determine the replication policy and the retention period. After applying a couple of best practice assumptions, the storage requirements could be easily determined.

However, as the system scales, in terms of data ingested, number of nodes required, and number of Splunk components/applications required, the sizing calculations get more complicated. The additional layers of complexity encouraged Splunk to build a sizing calculator, which is available at **http://splunk-sizing.appspot.com**.

Within the Splunk sizing calculator, administrators can plug in the daily data volume, the raw compression factor, the metadata size factor, the hot/warm data retention duration, the cold data retention duration, the archived data retention, the architecture, the use case or application, the maximum volume per indexer, the number of nodes, the storage configuration (which bucket of data resides on which volume), the RAID level per volume, the size of individual disks and the contingency required for those volumes.

The sizing calculator also contains a number of basic recommendations:

- *"Specify the location of the storage configuration. If possible, spread each type of data across separate volumes to improve performance. Hot/Warm data should be on the fastest disk, cold data on slower disk and archived data on the slowest disk."*
- *"Specify the RAID level, size of individual disks and contingency required for this volume. RAID configurations that stripe will yield significantly superior performance to parity-based RAID. That is, RAID 0, 10 and 0+1 will give the best performance, while RAID 5 will offer the worst performance."*
- *"The selected storage configuration would typically be expected to achieve about 800 IOPS when doing 100% read operation, and about 800 IOPS for 100% write operation. These numbers assume that array is dedicated to Splunk and consists of a single volume with 4 disk(s) (typically 200 IOPS per disk)."*

There are some additional recommendations[3] , but they are equally complicated and hardware-dependent.

[1] **https://www.splunk.com/blog/2013/01/31/disk-space-estimator-for-index-replication.html**
[2] **https://www.splunk.com/blog/2015/02/18/splunk-sizing-made-easy.html**
[3] **http://docs.splunk.com/Documentation/Splunk/7.2.1/Capacity/Referencehardware**

OPTION 2

Follow complex reference architectures that may not be tailored to individual requirements

Many storage vendors recommend the traditional tiered storage approach. Tiered storage has been around for decades and provides the ability to move (manually or automatically) data between different tiers of storage, depending on the "temperature" of the data–typically hot, warm, cold or archive.

The tiers of storage could be within a single storage array, distributed among multiple storage arrays.

Historically, the benefit of tiered storage was to optimize the cost/benefit of the various storage platforms. Some storage arrays were designed for high-performance workloads but were very expensive ($/TB, $/IOPS). Conversely, some storage arrays were designed for high-capacity but were very slow.

The downside to tiered storage is complexity. The more tiers and the more platforms exponentially increase the complexity for storage administrators. Tiered storage also typically increases total cost of ownership due to increased operational expenses.

Instead of buying and managing one platform, administrators now need to:

- Negotiate multiple contracts
- Navigate multiple support processes
- Manage multiple architectures
- Manage many more network connections/ports/cables
- Map/Zone more hosts
- Learn to use more management screens
- Work with and troubleshoot more automation scripts

It quickly turns into a large headache and it is no different for a Splunk implementation.

Moreover, most competitive storage solutions either explicitly or implicitly recommend deduplication. Most solutions highlight the advantages of their deduplication to reduce the effective price per TB. However, what if the application, like Splunk, already pre-compresses its data? What if the deduplication rates do not yield the appropriate cost per usable TB? These complications and unknowns make this a less desireable option.

OPTION 3

Leverage Infinidat Federal's InfiniBox® unified storage platform

Splunk requires high-performance, high-resiliency storage at scale. That storage must be easy to manage so that administrators can focus on the needs of the business instead of the needs of the infrastructure, and it excels at delivering a low total cost of ownership (TCO) with a rapid return on investment (ROI).

InfiniBox is built on five fundamental principles that are a perfect match for Splunk solutions:

1. High Performance

A truly innovative cache management algorithm combined with an ultra-efficient data layout delivers maximum performance at a fraction of the cost of All-Flash-Arrays. High throughput, at sub-millisecond latency, is the key to high performance operation as well as powering synchronous and asynchronous replication for block and file.

2. High Availability and Reliability

InfiniBox's self-healing architecture combined with our patented InfiniRaid® and predictive analytics delivers unmatched availability. The InfiniBox "component group redundancy" design enables rapid recovery from any component failure without impacting performance.

3. Multi-Petabyte Scale

Maximum system capacity utilization is possible due to extremely efficient thin provisioning, continuous space reclamation, and inline data compression. Packaged in a single 42U rack and scaling to well over 8PB or more effective capacity, multiple system consolidations are easy to accomplish and remarkably cost-effective.

4. Simple and Powerful Management

An intuitive HTML5 GUI simplifies the most complex storage management operations. A comprehensive RESTful API and a powerful CLI help automate complex tasks, including policy management for quality of service. Easily facilitate service level coordination across tenants, workloads, and volumes. Monitor and measure all feature performance elements using InfiniMetrics®.

5. Low Total Cost of Ownership (TCO) and High Return on Investment (ROI)

Infinidat Federal offers a TCO of 30-50% less on a price per TB basis compared the competition. According to Forrester's Total Economic Impact (TEI) assessment[4], InfiniBox delivers an average ROI of 125% on a payback period of less than six months.

[4] https://www.infinidat.com/news/press-releases/independent-study-finds-infinidat-storage-can-provide-customers-125-roi-over-three-years-and-payback-in-less-than-six-months/

SOLUTION:

Infinidat Federal InfiniBox for Splunk

The InfiniBox enterprise storage solution delivers faster-than-all-flash performance, high availability, and capacity density at multi-petabyte scale for a multitude of mixed workloads. Zero-impact snapshots, synchronous and asynchronous replication, and data-at-rest encryption assure maximum data security and reliability.

With InfiniBox, Splunk implementations are simpler to implement, operate, and protect while reducing costs and delivering the results that help companies discover, innovate, and drive their success.

Key Features

- High Availability–seven nines uptime
- High Reliability–continuous data integrity checking
- High Performance–over 1.3M IOPS & 15.2GB/s throughput
- Multi-protocol–FC, iSCSI, and NFSv3 all supported
- Multi-petabyte Scale–up 8PB in one rack
- Simple management–HTML5 GUI, RESTful API
- Cost effective–all features included at no additional cost

SOLUTION VALUE:

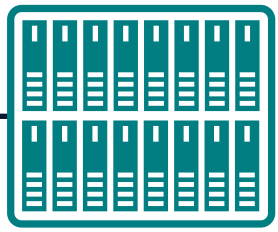
InfiniBox for Splunk in Action

One Fortune 500 insurance company has had great success running Splunk on InfiniBox[5]. The original requirements for the storage system supporting their Splunk environment were that it must be capable of ingesting 100GB per day per indexer and have the capacity to hold that data in the hot/warm bucket for 60 days, totaling hundreds of TB per location.

The customer's Splunk environment currently stores over 800TB of data across two data centers.

InfiniBox F6240
1.9PB Raw / 1.38PB Usable

Search heads
Indexers



Virtual Machines

- Forwarders
- Deployment Servers
- ES SHC Deployer
- License Manager
- DX Cluster Server
- Dist Mgmt Console
- Cluster Master

[5] https://www.infinidat.com/blog/infinidat-turns-digital-exhaust-digital-fuel-big-data-analytics/

CONCLUSION

By running Splunk on InfiniBox, customers can focus on the needs of the business instead of worrying about the cost and complexity of the storage infrastructure. InfiniBox provides the best reliability, the fastest performance, and the lowest TCO at multi-petabyte scale.

- **FASTER INGEST**–increase the ingest speed and overall amount of data ingested WHILE supporting multiple user reports
- **HIGHEST AVAILABILITY**–capture all critical data with ultra-low latency and near instant write acknowledgement. Seven nines availability means your Splunk/InfiniBox solution is always open for business.
- **SPEED/CAPACITY/LOWEST TCO**–with faster-than-flash speed AND petabyte-scale capacity, InfiniBox addresses the speed and capacity in a single solution, at the lowest total cost of ownership.

INFINIDAT FEDERAL