Accelerating Splunk Deployments

Solution Overview

Key Benefits

- Increase Search times up to 100x for faster analysis and real-time decision making
- Leverage standard ethernet networking for both servers and storage
- Reduce infrastructure footprint, and associated power and cooling costs, by a factor of 4

Splunk Benefits

- Identify and resolve issues and reduce escalations by up to 90%
- Monitor systems, infrastructure, and key performance indicators (KPI) in real time
- Proactively detect and investigate security incidents

Pavilion Benefits

- Latency of local DAS
- High Capacity - up to 460 PB in 4U
- Frictionless Deployment
- Data Resiliency & High Availability
- Space-Efficient, Instant Snapshots and Clones
- Thin Provisioning
- Standard Ethernet
- Pay As You Grow Scalability and Modularity

Move analytics from batch to real-time. Make better decisions faster with the Pavilion Memory Array and Splunk.

Splunk

Splunk is a leading event and log data analytics vendor, and application provider, which allows IT organizations to search, analyze, and visualize data gathered from different sources. Machine data can be collected so that it can be queried and analyzed to gain important information and insight, allowing organizations to proactively predict problems before they occur.

The data collected by Splunk can come from a variety of sources, including applications, networking devices, host and server logs, mobile devices, and more. As the amount of data grows, challenges arise related to analyzing larger data sets in a timely manner.

Splunk architecture consists of the following three key components: indexers, forwarders, and search heads, that can be deployed on a single server or distributed across multiple servers. Indexers store the collected data and also index it to be used for searches. Search heads distribute searches to indexers, and forwarders forward search requests to remote indexers.

The Pavilion Memory Array

The Pavilion Memory Array delivers 25X the performance and 10X better latency than typical networked All-Flash-Arrays. As a result, Big data analytics applications like Splunk can now analyze much larger data sets, and therefore deliver more accurate and timely answers to critical queries. Decisions can be made in real-time at the speed of digital business by analyzing more data quickly.

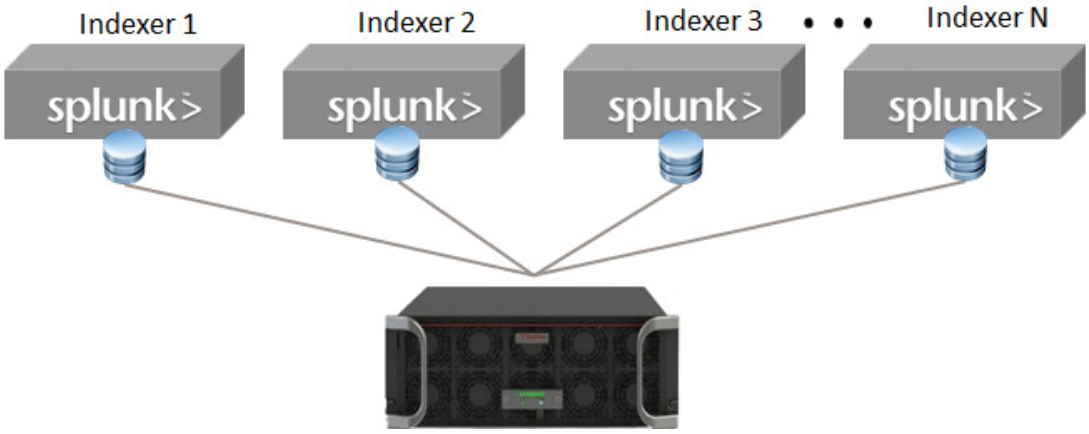
Ultra-low latency (10s of µs) allows customers to make provisioning decisions at application run time, as opposed to procurement time, by replacing direct-attached SSDs in Splunk Indexer nodes, leading to dramatically increased storage utilization and lower costs in these environments.

Splunk Storage Challenges

The traditional approach is to store the data in the Indexer nodes on direct-attached storage captive inside that node. The first problem with this approach is that storage capacity is stranded in individual nodes, leading to under-utilization of storage assets and inflexibility. Another issue occurs when the size of the data outgrows the storage in the nodes, leading many Splunk users to use an external storage array to provide maximum capacity for the indexer nodes. However, traditional storage arrays add latency, making it harder to analyze large data sets in real time and get timely answers to problems. This leads to organizations reacting to problems, as opposed to predicting them and taking corrective action ahead of time.

Real-Time, High-Speed scalable analytics Solution with Pavilion and Splunk

The Pavilion Memory Array is an ideal solution for Splunk storage for organizations requiring the most aggressive analytics solution that can analyze massive amounts of data in real time. By deploying a Pavilion Memory array as centralized storage for Splunk Indexers, a Splunk deployment can grow to much larger scale by deploying hundreds of indexers if needed using a single shared storage pool. Administrators can deploy additional Indexers at will to handle more search traffic, without having to worry about scaling storage at the same time if needed. A single high speed 10 to 40 GB Ethernet network can be leveraged for the Splunk servers as well, alleviating the need for multiple networks or protocols to be deployed.



In addition, concerns about growing data footprint can be alleviated due to the high performance and high capacity of the Pavilion System. A single Pavilion Memory Array can provide up to 460 TB of high-speed storage, as well as support up to 120 GB/s of bandwidth and 20 million read IOPS, which can easily accommodate a large pool of several hundred Splunk Indexers. In addition, the low latency shared storage array will deliver the same performance as direct-attached flash SSDs, allowing for quick turnaround on search requests. Search times can be improved up to 100X by using Pavilion, depending upon the type of search.

Pavilion supports thin provisioning as well, so a large amount of storage can be presented to each indexer, while the array will only allocate the amount of physical storage needed at a specific time. This will deliver savings on the raw storage footprint required when compared to direct-attached SSDs. The solution is also highly-available so that users can be assured that there will be minimal downtime, as all components of the array are hot swappable and redundant, including individual SSDs in the array.

Given the high capacity and throughout of the array, multiple tiers (Splunk buckets) can all be consolidated on one appliance, alleviating the need for complex storage tiering within Splunk. This will also allow for reduced infrastructure footprint, given the performance and capacity density of the Pavilion Memory Array. Given that the array can serve as multiple storage tiers, and the high performance will allow for less Indexers to be used, infrastructure footprint, as well as associated power and cooling costs will be reduced dramatically.

By leveraging Pavilion as high-speed, low-latency, high capacity networked storage solution for Splunk, real time, high speed data anlytics can become a reality for today's IT Organizations, allowing them to make better decisions, faster.

Pavilion DATA SYSTEMS

Pavilion Data Systems, Inc.
2560 N First St., Suite 220, San Jose, CA 95131
E-mail: sales@paviliondata.com