ESG Technical White Paper

Running Splunk on VxRack FLEX

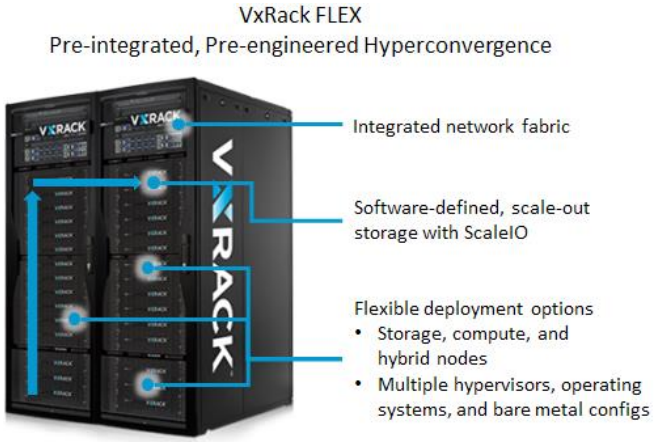
Date: August 2017 **Author:** Mike Leone, Senior Validation Analyst; and Domenic Amato, Associate Validation Analyst

Background

As organizations continue to look for ways to modernize their infrastructures by delivering a cloud-like experience on-premises, hyperconverged offerings are exceeding expectations. In fact, the adoption of hyperconverged infrastructure has more than doubled over the last year and there are no signs of slowing down.[1] Simplified, flexible deployment options that are easily managed and easily scale are just a few of the factors driving the rapid adoption of hyperconverged technology. As consolidation and modernization efforts are well underway across all of IT, hyperconverged offerings continue to meet the needs of the business while satisfying existing and future application SLAs, whether they are related to performance, scalability, reliability, or cost.

With many organizations running enterprise applications and databases of different shapes and sizes, including hypervisors and operating systems, on traditional physical or virtual infrastructure models, all configurations and architectures must be considered when planning and expanding further adoption of hyperconverged infrastructure in the data center.

VxRack System FLEX



VxRack System FLEX (VxRack FLEX) is a rack scale hyperconverged solution that leverages Dell EMC ScaleIO data center grade software-defined storage to deliver flexible, scalable, and performant software-defined storage. The ScaleIO software runs on industry-proven Dell EMC PowerEdge servers to deliver a full hyperconverged stack. Standardizing on VxRack FLEX enables IT organizations to deal with one vendor for all the software, hardware, and support required to modernize their data centers. Key to the solution is the scale-out architecture, offering organizations flexibility to start small and grow based on their needs. Further, this elasticity delivers on the hyperconverged promise of cloud-like scale and flexibility on-premises. While many other hyperconverged offerings disregard networking altogether, VxRack FLEX supports both physical and virtual networking including top of rack switches that control network traffic, management, and redundancy. Put it all together and organizations get a hyperconverged solution that can easily interoperate with all other Dell EMC products and services, on the production side with ScaleIO delivering impressive performance at scale, or with technologies that enhance data protection, availability, and recovery, such as Data Domain with Data Domain Boost and RecoverPoint, for example.

ScaleIO – Data center grade, software-defined storage

The storage power behind the VxRack FLEX solution is ScaleIO, Dell EMC's scale-out software-defined storage solution that abstracts the direct-attached storage found in Dell EMC PowerEdge servers into a pool of shared block storage. By converging the storage and compute on the same physical servers, this single and/or two-layer architecture helps to simplify management and maximize storage efficiency as the infrastructure grows from four to thousands of nodes. Whether using HDDs, SSDs, or even NVMe or PCIe flash, storage is combined into virtual block-storage pools with varying performance tiers. Combined with QoS, snapshots, caching, fault sets and protection domains, and data-at-rest encryption, ScaleIO running within the VxRack FLEX system delivers a data center grade, fully integrated, hyperconverged solution. Leveraging a software-defined storage approach to satisy enterprise application and database block storage requirements enables organizations to potentially break free of large initial investments and high operational costs commonly associated with traditional SANs. Further, fears of technology updates, refreshes, and data migrations impacting costs, risk, and periods of downtime can be all but eliminated.

[1] Source: ESG Research Report, *Hyperconverged Infrastructure Continues to Gain On-premises Momentum*, to be published.

Splunk Enterprise

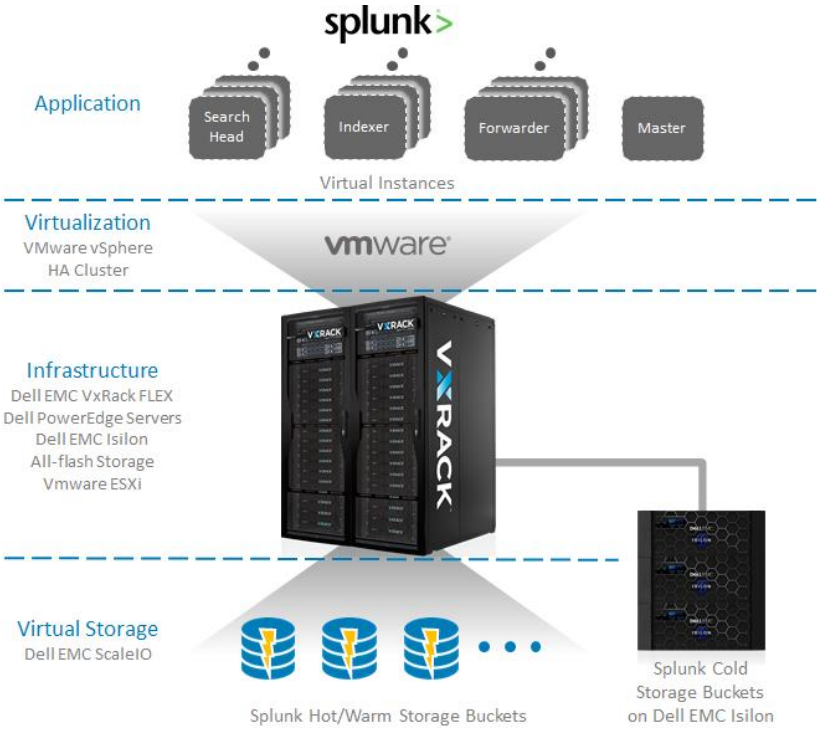
Splunk is a leading data analytics platform capable of providing operational intelligence on machine data from applications, network devices, logs, mobile devices, and more within IT environments. Splunk can gather and index this data in real time natively or through third-party applications, and can visualize insights to better inform organizations about their resources, applications, and security.

Splunk's core architecture is made up of three components: search heads, indexers, and forwarders. Splunk forwarders (agents) are installed on any number of distributed sources to send machine data to a Splunk indexer. A Splunk server running any supported OS platform can forward data to another Splunk instance (as well as others) in real time. Splunk forwarders are not usually resource-intensive. Splunk indexers are the repositories for incoming data which is transformed into events and stored in indexes. The number of indexers determines the amount of raw data a Splunk volume can index. The Spunk indexers also manage policy-based data tiering for hot buckets (newly indexed data, open for writing and search), warm buckets (data rolled from hot buckets, available for search only) and cold buckets (data rolled from warm buckets, available for search only). An indexer cluster is a group of indexers configured to replicate each other's data which prevents data loss in the event of a failure. Splunk indexer's performance is impacted by increases in data streams and requests from search heads. A search head sends queries to a group of indexers, or search peers, which read the indexes and return data to the search head. The search head then merges the results back to the user. Search heads are typically CPU- and memory-intensive.

Solution Architecture

Splunk makes it simple for companies to observe real-time operational metrics about their organization to gain better insights faster. With Dell EMC VxRack FLEX, Splunk deployments gain a cost-effective, scalable, and flexible operational intelligence platform that leverages ScaleIO software-defined storage and VMware for virtualizing the core Splunk components, while Isilon provides additional scale-out storage for colder data. Together, they provide the standardization and automation data centers require to leverage high-performance hot/warm data and manage low-cost cold data at the same time.

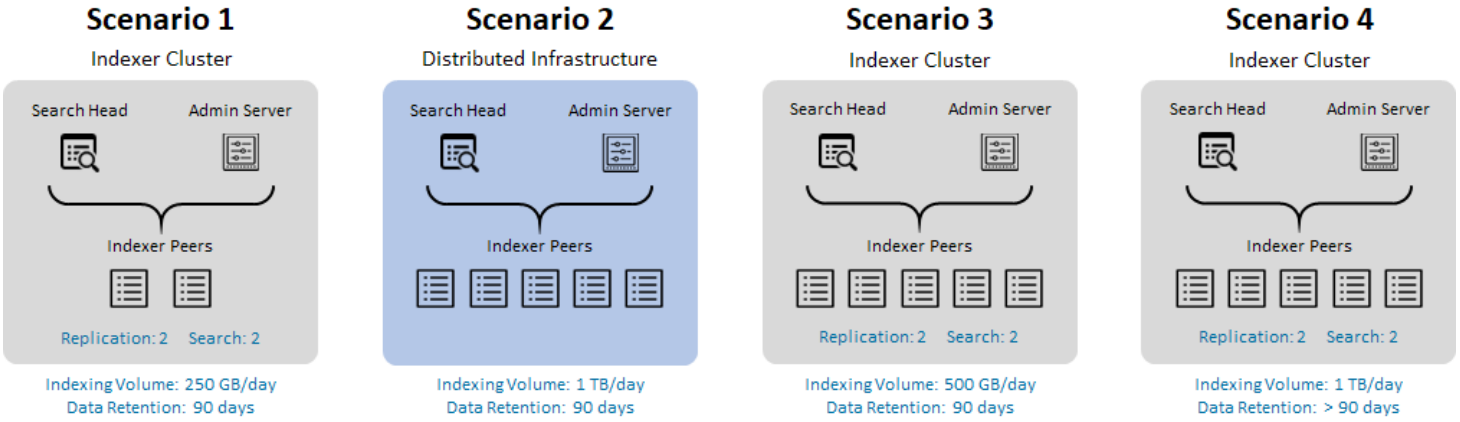
Dell EMC VxRack FLEX has created and validated Splunk reference architectures that meet or exceed the performance guidelines outlined by Splunk in its own best practice and implementation guides. The VxRack FLEX's hyperconverged architecture provides Splunk with integrated compute, networking, and ScaleIO software-defined storage. ESG Lab evaluated Splunk and VxRack FLEX together on two different application architectures that can be scaled appropriately to fit the size of dynamic and expanding Splunk deployments. The four-layer infrastructure consists of virtual storage, infrastructure, virtualization, and application layers. Thinking about it within the constructs of VxRack FLEX, the virtual storage layer is comprised of Dell EMC ScaleIO to handle the Splunk data buckets (hot, warm, and cold data). The infrastructure layer consists of a cluster of VxRack nodes with compute and all-flash SSD storage. The virtualization layer leverages VMware vSphere, enabling organizations to satisfy their requirements of evenly distributing VMs across a highly available infrastructure. Finally, the application layer provides dedicated VMs for Splunk, whether in clustered or distributed deployments.



The second architecture is like the first in that the application and virtualization layers remain the same. The added functionality comes from the addition of another important Dell EMC infrastructure component—a Dell EMC Isilon cluster. In this case, Splunk's hot and warm data buckets remain in the VxRack FLEX within ScaleIO's virtualized storage construct, while Splunk's cold data bucket shifts to the Isilon X410 cluster.

ESG Objectives

This technology review aimed to demonstrate the enterprise-level machine data analytics and real-time operational intelligence Splunk can deliver on VxRack System FLEX. Several points of focus included the ease of deployment, level of non-disruptive scalability, performance, and efficiency that Splunk Enterprise can achieve when virtualized on easily standardized hardware and software configurations. Four typical deployment scenarios illustrated below were validated.



The amount of data ingested by each scenario increased from 250 GB/day to 1 TB/day. The retention timeframe for the first three scenarios is 90 days, while the fourth is greater than 90 days. The replication factor specifies how many copies of the data the cluster maintains, and the search factor specifies how many copies of the data are searchable. Together, they give an indication of how resilient the cluster is against multiple node failures and how quickly the cluster can recover from a failed node. The gray shaded scenarios are indexer cluster deployments that demonstrate high data availability. The blue shaded scenario is a distributed cluster that enables faster search for larger instances.

Simplicity, Scale, and Flexibility

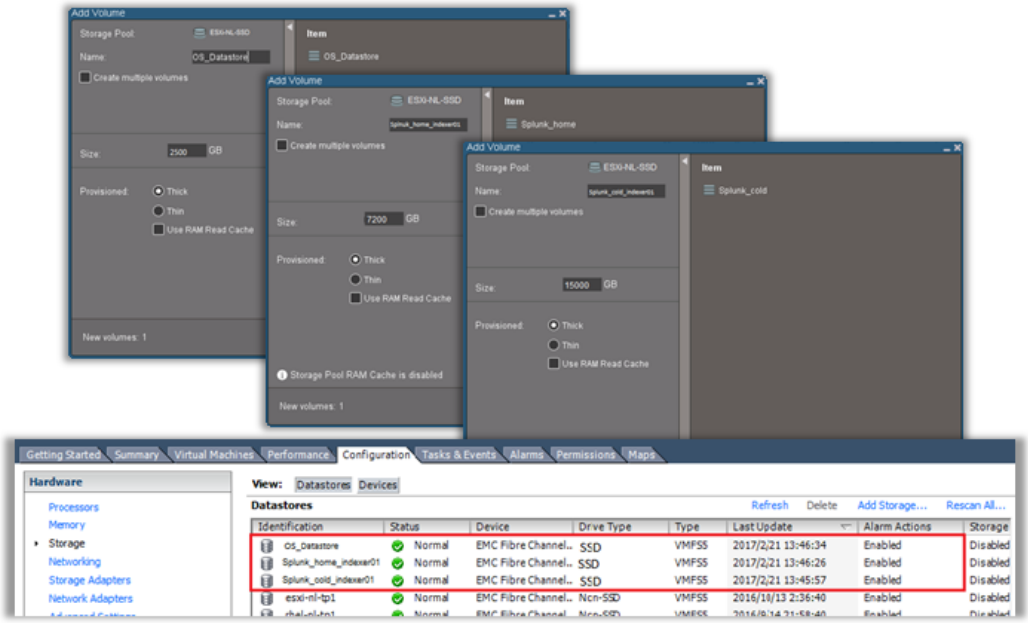
ESG Lab reviewed the deployment process of the aforementioned architectures and examined four validated infrastructure designs that showcased Splunk's ease of deployment and scalability, and the flexibility of its data retention capabilities on VxRack FLEX. The infrastructures varied by number of indexing peer nodes and daily data indexing volume. One infrastructure scenario also demonstrated the data retention capabilities of Dell EMC's configurable Isilon X410 storage clusters to house Splunk's cold data bucket.

Ease of Deployment

The first infrastructure scenario consisted of a Splunk deployment configured to handle 250 GB/day of indexed data with a 90-day retention period. This configuration featured four VxRack nodes—one search head, two indexers, and one admin server role to handle the storage of Splunk's hot/warm and cold data buckets. The process begins with Dell EMC implementing the four-node VxRack cluster. ScaleIO storage volumes are configured for VM operating systems and Splunk storage buckets. VMware vCenter is used to design VM templates to optimize performance and lay groundwork for how each Splunk component and node is ultimately deployed.

Each node is deployed using the Dell EMC deployment guide to ensure parameters are set appropriately for the component and overall configuration size. Splunk best practices are also followed, one of which is to configure and deploy the Master node as a universal forwarder connected to the clustered indexer peer nodes using indexer discovery. Setting the forwarder to use indexer discovery allows it to automatically update its list of peer nodes and use load balancing to distribute data by

peer depending on disk space, circumventing the need to reconfigure or manually adjust data loads via traditional methods. After enabling discovery, the entire implementation is validated by logging into the master node's web server to search for and validate each deployed VM/Splunk node. It should be noted that even though this scenario set out to achieve 250GB/day volume and 90-day retention, Splunk's high-performance and capacity pushes that retention capability up to 142 days (and up to 177 for the 1TB/day, five-indexer scenario). Additionally, ESG reviewed a second scenario that went through the same deployment process but handled 1 TB of indexed data per day with a 90-day retention. Aside from the storage requirements, the other difference was that the more demanding deployment leveraged five indexers rather than two.



Scalability

The next validated scenario focused on scaling an existing Splunk deployment. The initial Splunk deployment consisted of handling 250GB/day volume with two indexers. The goal was to increase the number of indexers from two to five to handle twice the amount of indexed volume per day (500 GB/day). This assumed three nodes were available and unused in the VxRack FLEX. Like the initial deployment, ScaleIO storage volumes must be provisioned for the name and size of the buckets within each of the new indexer nodes. In this scenario, three new indexer nodes were being added, accounting for three hot/warm buckets and three cold buckets. After specifying their names and sizes (7.2 TB for hot/warm and 15 TB for cold), the newly created storage buckets can be mapped to their ESXi hosts. Creating new datastores in the vCenter vSphere client allows you to fully deploy the three new peer nodes. Once they have been validated in the web server, the existing infrastructure has been successfully scaled.

Flexibility

The last phase of ESG Lab's infrastructure review focused on the flexibility of Splunk's data retention capabilities when running on VxRack FLEX. This deployment is configured for 1TB/day data indexing volume with greater than 90-day retention on seven VxRack nodes. While the hot/warm Splunk data buckets in this scenario remain on ScaleIO in the VxRack FLEX cluster, the cold buckets are stored on Dell EMC's Isilon X410 clusters.

Much of the deployment process is like the previous implementations. The main changes revolve around bucket sizing and configuring Isilon. Both the seven-node VxRack infrastructure (search head, five indexers, and server admin) and Isilon instance on the VxRack are configured independently. Then, ScaleIO storage volumes must be set for Splunk's hot/warm buckets to reflect the infrastructure's larger size, and can then be mapped to their ESXi hosts. Isilon NFS must then be configured for the VxRack FLEX cluster using the intuitive OneFS web service, where an access zone for Splunk in the Isilon instance is created and an NFS share export is created for each ESXi server. The process to do this takes just a few clicks through the Isilon dashboard. vCenter then allows you to add NFS storage to each server on the VxRack FLEX. The other difference comes during the storage deployment of the peer nodes. IT administrators must ensure that the disks for Splunk's cold storage bucket come from Isilon. This can all be done using the Add Hardware wizard from within vCenter. Once the directory of the cold bucket is established, the rest of the implementation follows the same steps from deploying the search head and forwarder before validating the entire implementation.

The Bigger Truth

As organizations look for solutions to help modernize their infrastructure, hyperconverged infrastructures are serving as go-to architectures due to their underlying ability to improve time to value. With VxRack FLEX, customers receive a pre-integrated, pretested, and pre-validated hyperconverged infrastructure that can be operational within hours of arriving on the loading dock. With software-defined storage from ScaleIO, organizations gain a flexible underlying storage solution that enables linear performance scalability, while the VxRack FLEX node configuration flexibility enables organizations to easily scale out just compute, just storage, or both.

Splunk Enterprise is a prime platform to leverage VxRack FLEX's integrated infrastructure to further operational intelligence capabilities. Splunk can gather and index machine data from virtually any source in real time and with Dell EMC, organizations gain a robust, scalable solution to meet the demands of any business looking to access its hot/warm and cold data at the same time with high performance and consistency.

ESG suggests exploring the VxRack FLEX offering as a way for organizations to future-proof their IT infrastructures while continuing to meet the performance, scalability, and protection requirements of their Splunk environments to improve time to value, time to insight, and overall operational intelligence.