PROVIDING ENTERPRISE PERFORMANCE, CAPACITY, AND DATA SERVICES FOR SPLUNK ENTERPRISE

November 2017

ABSTRACT

This white paper describes Dell EMC converged and hyper-converged infrastructures that would be appropriate for providing enterprise performance, capacity, and data services for Splunk Enterprise.

H16810

This document is not intended for audiences in China, Hong Kong, Taiwan, and Macao.

Copyright

The information in this publication is provided as is. Dell Inc. makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any software described in this publication requires an applicable software license.

Dell Inc. believes the information in this document is accurate as of its publication date. The information is subject to change without notice.

Contents

Introduction

IDC published their annual Digital Universe Survey in 2014 with the theme: *The expanding universe of business opportunities*. Like the physical universe, the digital universe is large – by 2020 it will contain nearly as many digital bits as there are stars in the universe. It is doubling in size every two years, and by 2020 the digital universe – the data we create and copy annually – will reach 44 zettabytes, or 44 trillion gigabytes, which is the equivalent of 1.7MB for every person every second of every day.

Much of this newly created information is machine-generated data, meaning it is automatically created from a computer process, application, or other machine without the intervention of a human. Sources of machine-generated data include compute servers, web and application servers, firewall and security devices, sensors, and mobile devices, just to name a few. In order for customers to extract the most business value from this data, they need an agile solution that can search, correlate, and perform advanced analytics and reporting at scale.

This white paper describes Dell EMC converged and hyper-converged infrastructures that would be appropriate for providing enterprise performance, capacity, and data services for Splunk Enterprise.

Splunk Enterprise Overview

Splunk Enterprise is the industry-leading platform for machine data. Splunk software provides the capability to mine machine-generated data to see valuable insights. Splunk Enterprise can take data from anywhere in your infrastructure and use its powerful Splunk search processing language (SPL) to help you extract meaningful insights about what's happening. This is called Operational Intelligence and has many use cases including:

- IT Operations: Utilization, capacity growth
- Security: Fraud detection, real-time detection of threats, forensics
- Internet of Things (IoT): Sensor data, machine-to-machine, human interactions

Splunk Enterprise enables the high-speed analysis of real-time and historic machine-generated data. This creates the need for a tiered storage infrastructure able to handle high-performance hot and warm data, as well as high capacity cold and frozen data.

The business-critical nature of Splunk Enterprise data necessitates high levels of security and data protection, as well as multiple copies of Splunk indexes in clustered deployments. In order to satisfy these requirements, many organizations are moving away from direct-attached storage (DAS) architecture in favor of all-flash-array (AFA) storage solutions combined with shared, networked storage to get scale-out high performance and capacity with rich functionality for data protection, security, compression and space-efficient copy services.

Splunk Enterprise Core Architecture

Splunk Enterprise collects, indexes, and harnesses machine data generated from almost any source, format, or location, including packaged and custom applications, application servers, web servers, databases, networks, virtual machines, hypervisors, and operating systems without requiring custom parsers, adapters, or a back-end database.

After data is indexed, you can correlate complex events that span diverse data sources and use the powerful search, analysis, and visualization capabilities of Splunk Enterprise. Splunk Enterprise provides you with a real-time understanding of what happened, why it happened, and what is happening across IT services, systems, and infrastructure. The following sections illustrate the various components of Splunk Enterprise.

Indexers

Indexers are Splunk instances that transform received raw data into events and store those events in an index. Indexers also search the indexed data in response to search requests.

Search heads

In a distributed environment, search heads handle search management functions, direct search requests across a set of indexers, and merge results back to the user. They are often both CPU- and memory-intensive.

Forwarders

Forwarders are instances that collect and forward data to indexers (or other forwarders) and are usually not resource-intensive.

Scale-out and tier data

The system resources needed to optimize search and index performance in Splunk Enterprise depend on the volume of data being indexed and the search load. In order to support multiple workloads and growth over time, Splunk Enterprise is designed to scale horizontally. If additional search or indexing performance is needed, a search head or an indexer can simply be added to the architecture without disrupting operations. This allows Splunk Enterprise to scale from hundreds of gigabytes of data per day to hundreds of terabytes and more.

Splunk indexes are stored in buckets according to the age of the data or overall size of the bucket. In general, data that has been recently indexed resides in "hot" buckets and can be placed on faster storage for faster index and search performance. As indexed data ages, Splunk automatically rolls data from one bucket type to the next—hot to warm, and warm to cold. After a set of conditions is met, such as a period of time, cold data rolls to frozen, at which point it is either archived or deleted. Data rolling, aging, and retention policies are governed by a set of configurable attributes.

DELL EMC Infrastructure for use with Splunk Enterprise

Dell EMC Converged Solutions are engineered, manufactured, managed, supported, and sustained as ONE product, giving IT leaders the foundation and flexibility to reduce costs, enhance service delivery, shift IT focus towards delivering business value instead of maintaining infrastructure, and meet the evolving expectations of a tech-savvy, mobile workforce.

With Dell EMC VxBlock® Systems[1], VxRack® Systems, and VxRail® Systems, organizations can simplify and optimize provisioning, deployment, and management of Splunk search and analytics workloads through multi-tenancy, virtualization, data protection, mission-critical availability, and performance. With technology extensions for compute and storage, existing VxBlock Systems can easily be expanded to:

- Incorporate large-scale user data lakes for Big Data and advanced Splunk architecture enterprise
- Provide a unified way to organize and extract real-time insights from massive amounts of machine data from virtually any source. This includes data from websites, business applications, social media platforms, app servers, hypervisors, sensors, traditional databases, and open source data stores.
- Standardize Splunk Enterprise infrastructure to scale across multiple business units while embracing existing applications and systems
- Streamline hardware sizing, deployment, and data management and protection tasks
- Achieve agility and lower operating expenses (OPEX) at lower risk with multi-tenancy, workload aggregation, and flexible use of virtual or bare metal implementations

All Dell EMC VxBlock, VxRack, and VxRail systems are factory-built and validated as unified systems, backed by the one-number to-call Dell EMC Support, and sustained with the Dell EMC Release Certification Matrix (RCM). This all-inclusive approach provides rapid deployment and a single point of support for all aspects of the platform (virtualization, network, compute, and storage).

DELL EMC Splunk Architecture Overview

As customers adopt Splunk Enterprise and take advantage of its compelling features, managing the underlying direct attached storage (DAS) infrastructure becomes challenging. Maintaining consistent performance and leveraging flash to ensure that end users get fast query and search capabilities from the Splunk dashboard begins to involve significant time in design and troubleshooting. In addition, enabling longer retention periods increases floor space usage in the data center, and adds to management overhead.

Dell EMC converged and hyper-converged infrastructures are optimized for the needs of a Splunk environment. This platform helps to consolidate, simplify, and protect all machine

[1] Dell EMC VxBlock Systems are also appropriate for Splunk workloads. For purposes of this paper, VxBlock Systems is used as the solution term. Dell EMC is now focused exclusively on positioning VxBlock Systems, as they can be configured similarly to a Vblock System, and can add additional levels of flexibility.

data while delivering the highest levels of performance and data retention.

The primary benefits Dell EMC provides to your Splunk Enterprise environments include:

- **Optimized Storage Data Tiering** – aligns storage to hot, warm, cold, and frozen data requirements with high retention and performance
- **Cost-Effective and Flexible Scale-Out** – provides scale-out capacity and compute independently or as a single, converged platform
- **Powerful Data Services** – include secure encryption, compression and deduplication, and fast, efficient snapshots for protection

The rapid extensibility of the physical infrastructure provides the ability to independently scale compute, hot/warm, and cold file systems as required, or allocate resources as needed to meet increase query demand.

Dell EMC's comprehensive infrastructure further enhances its data center flexibility, efficiency, and elasticity by leveraging VMware technology. Using a hypervisor, Splunk administrators are able to adjust (increase or decrease) each virtual machine core and memory based on specific search needs.

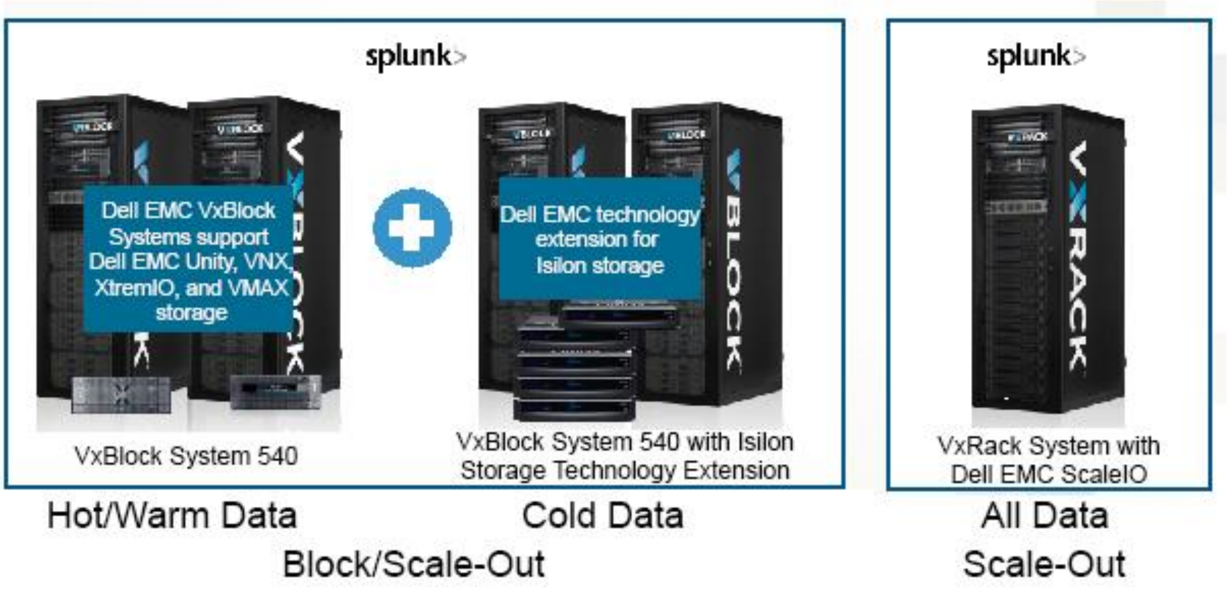
splunk>
Dell EMC VxBlock Systems support Dell EMC Unity, VNX, XtremIO, and VMAX storage
VxBlock System 540
Dell EMC technology extension for Isilon storage
VxBlock System 540 with Isilon Storage Technology Extension
splunk>
VxRack System with Dell EMC ScaleIO
Hot/Warm Data
Cold Data
Block/Scale-Out
All Data
Scale-Out

Figure 1. DELL EMC Converged and Hyper-converged architecture for Splunk

VxBlock System 540 and VxRack System FLEX for Hot and Warm Data

Typically, at higher daily volumes, a local disk does not provide cost-effective storage for a fast search. In these cases, a superior system design would deploy high performance, low-latency external networked storage, such as storage area networks (SAN) over fiber, to provide the required IOPS per indexer.

Specific cases where Dell EMC's technologies have large potential value are in I/O-bound workloads such as:

- **Super Sparse Searches**: These are searches that return a small number of results, but tax the I/O system due to the need to search the entire Index.
- **Rare Searches**: Similar to super sparse searches, these searches are also I/O bound, but leverage bloom filters to eliminate indexes for faster returns.

For accelerating super sparse and rare search performance of Splunk Enterprise at lowlatencies, Dell EMC recommends leveraging the Dell EMC VxBlock System 540 converged infrastructure with Dell EMC XtremIO® All-Flash Arrays, or the Dell EMC VxRack System FLEX hyper-converged infrastructure.

VxBlock System 540 Converged Infrastructure

The VxBlock System 540 is an all-flash based converged infrastructure, factory integrated and validated by Dell EMC. Powered by the next-generation Cisco Unified Computing System, an application centric infrastructure (ACI)-ready network, and XtremIO, the VxBlock System 540 was designed to fully leverage flash and delivers the breakthrough scale-out architecture, consistent performance, data reduction, thin provisioning, and manageability you would expect from an enterprise flash array.

Scale-out all-flash performance and simplicity

The VxBlock System 540's scalable, linear architecture easily accommodates expansion. Its performance-optimized data center footprint lets you consolidate multiple workloads without compromising performance or availability for maximum IOPS and the lowest TCO.

Specifically, the VxBlock System 540 with XtremIO can consistently scale out to >1M IOPS at <1ms latencies for all of your queries and workloads within your Splunk Enterprise environment.

The VxBlock System 540 all-flash architecture simplifies storage configuration and layout without the complex tuning or setup required by traditional DAS infrastructure. All workloads are configured the same way every time in three simple steps without setting RAID levels. XtremIO Data Protection (XDP) leverages in-memory data protection and does not rely on traditional disk-based RAID algorithms.

Inline compression and deduplication

The VxBlock System 540 with XtremIO compresses and deduplicates data inline without performance penalties. XtremIO compression operates all the time and inline, reducing the amount of capacity required for indexers. In an indexer cluster, data will be replicated to other indexers. Normally, this replication means greater data growth; however, with XtremIO inline deduplication, the data is written once, which saves space. All indexers have access to the data. XtremIO deduplication results in significant capacity savings for businesses.

VxRack System FLEX Hyper-converged Infrastructure

The VxRack System FLEX is a hyper-converged infrastructure that complements Dell EMC's converged infrastructure systems. These self-contained units of servers and networking are well-suited for Splunk implementations. The scalability, flexibility, and resilience demonstrated by VxRack System FLEX make it an ideal platform for Splunk.

The storage foundation of the VxRack System FLEX is based on the Dell EMC ScaleIO® software-only server-based SAN. ScaleIO converges storage and compute resources to form a single-layer, enterprise-grade storage product.

In VxRack System FLEX, VMware vSphere provides virtualization services. The core VMware vSphere components are the VMware vSphere ESXi and VMware vCenter Server for management.

VMware vCenter Server 6.0 simplifies planning and deployment by offering two deployment models. The embedded model deploys the new Platform Services Controller (PSC) and the vCenter Server instance on the same machine. The external model deploys the PSC and the vCenter Server instance on separate machines.

Converged scale-out architecture reduces costs

The VxRack System FLEX with ScaleIO software utilizes Dell EMC's integrated compute nodes' DAS and aggregates all disks into a global, shared, block storage pool. ScaleIO enables a single-layer compute and storage architecture without requiring additional hardware. Its scale-out server SAN architecture can expand to accommodate thousands of servers.

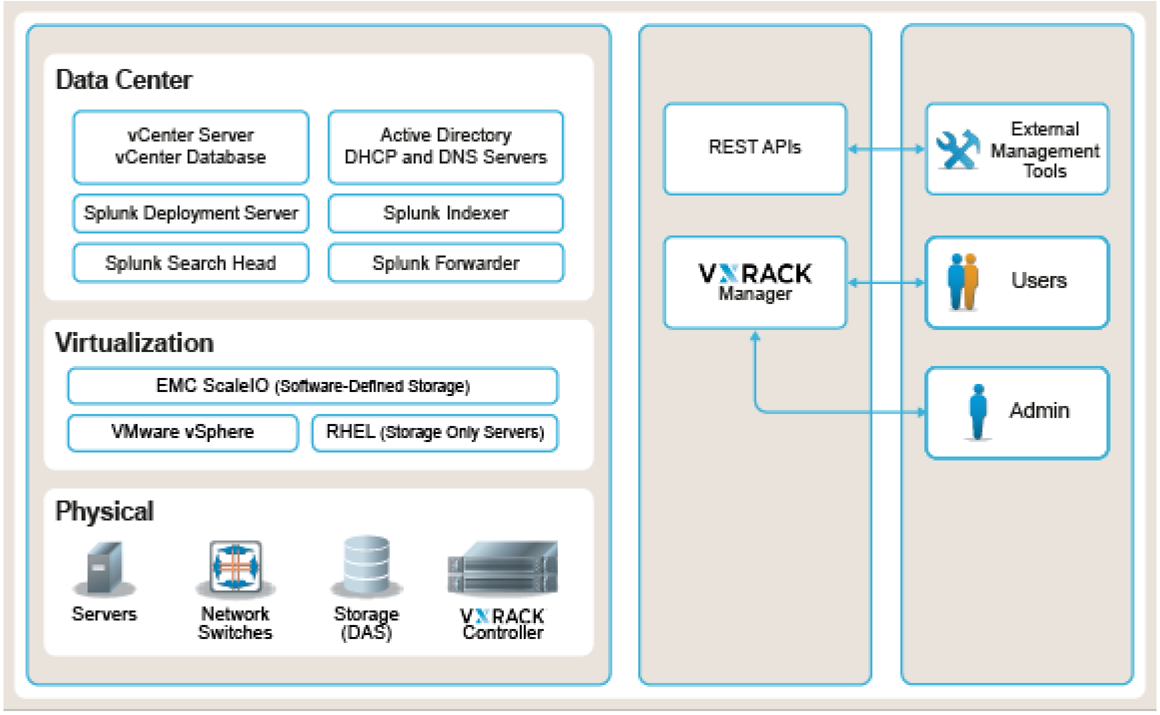

Figure 2. VxRack System FLEX Components

I/O parallelism

All servers running a Splunk Enterprise system in the ScaleIO cluster are also used for processing I/O operations, making all I/O and throughput accessible to any application within the cluster. Such massive I/O parallelism eliminates bottlenecks. Throughput and

IOPS scale in direct proportion to the number of servers and local storage devices added to the system.

Indexers in a distributed deployment configuration have equivalent disk I/O bandwidth requirements to indexers in a single-instance deployment. To compensate for the increase in search requests and higher data indexing volume in a distributed environment, distribute the work across many indexers.

Dell EMC Technology Extensions for Dell EMC Isilon for Cold and Frozen Data

Dell EMC Technology Extensions for Dell EMC Isilon® complements the Splunk Enterprise scale-out architecture by providing a powerful, cost-effective scale-out storage cluster for the retention of cold data in Splunk.

One file system

Dell EMC Technology Extensions for Isilon's single-volume, single-file system and simplified management typically require less than one full-time employee per petabyte (PB), reducing your overall storage administration costs.

"Bottomless" Cold Buckets - Powerful scale-out infrastructure

A Dell EMC Technology Extensions for Isilon cluster creates a unified pool of highly efficient storage, with a proven 80 percent storage utilization rate. There is no need to over-provision storage capacity or performance; simply purchase storage as needed and automatically scale it. You can scale-out your Splunk Enterprise environments to 50 PB in a single file system and tier your workloads in Splunk Enterprise across shared storage without the need for migration, essentially creating a "bottomless" cold bucket.

Furthermore, Splunk administrators can easily fine-tune their index containers by carving their "bottomless" bucket of storage into smaller shares per indexer for Splunk indexes. This is as simple as creating a new directory and does not require the typical heavy lifting of creating RAID groups, volumes, and aggregates as data scales. This flexibility and scalability allows organizations to use Isilon for a variety of workflows and bucket architectures with little management overhead.

Dedupe and snapshots

Dell EMC Isilon SmartDedupe™ technology reduces storage costs by using a post-process deduplication approach that stores new data before analyzing it for possible duplications.

Dell EMC Isilon SnapshotIQ™ provides a simple, scalable, and flexible way to enable enterprise-class, point-in-time data protection and recovery for scale-out storage. SnapshotIQ supports a highly scalable number of snapshots throughout an Isilon storage cluster and up to 1,024 snapshots per directory. With SnapshotIQ, you can retain the security of the file within the snapshot and automatically back up data as frequently as required to meet your recovery point objectives—all with little to no performance impact on users, regardless of the size of the file system or directory.

Lifecycle management and compliance

Dell EMC Isilon SmartLock® helps you protect your data against accidental, premature, or malicious alteration or deletion. Because SmartLock is a software-based approach to Write Once Read Many (WORM) data protection, you can store SmartLock-protected data alongside other data types in your Isilon scale-out storage environment with no effect on performance or availability.

SmartLock operates in either Enterprise mode or Compliance mode. Compliance mode provides an extra level of protection against malicious modification of data by disabling logins by the root user. This allows you to meet regulatory compliance requirements to provide absolute retention and protection of business critical data, including stringent SEC 17a-4 requirements.

Isilon HDFS for use with Hunk®: Splunk analytics for Hadoop

Hunk (Splunk Analytics for Hadoop) lets you configure remote Hadoop Distributed File System (HDFS) data stores as virtual indexes so that Splunk Enterprise software can natively report on data residing in Hadoop. Once your virtual index is properly configured, you can report and visualize data residing in remote Hadoop data stores.

Dell EMC Technology Extensions for Isilon includes native HDFS integration, helping you avoid the need to invest in a separate Hadoop infrastructure and eliminate the time and expense of moving large data sets. Isilon solutions support multiple instances of Apache Hadoop distributions from different vendors simultaneously.

As a result, organizations can improve time to results by augmenting mission-critical enterprise data warehousing (EDW). They can support environments with Hadoop by easily standardizing large Big Data analytic infrastructures on Dell EMC Technology Extensions for Isilon.

Conclusion

Machine data is one of the fastest growing and most complex areas of big data. It's also one of the most valuable, containing a definitive record of user transactions, customer behavior, sensor activity, machine behavior, security threats, fraudulent activity and more.

Making use of machine data, however, is challenging. It is difficult to process and analyze by traditional data management methods or in a timely manner. Splunk provides the leading software platform for real-time operational intelligence. Splunk Enterprise collects, analyzes and visualizes machine-generated data and is a robust platform for developing big data apps granting developers critical analytical insights.

Combined with VxBlock and VxRack Systems, and flexible options like Dell EMC technology extensions, organizations can easily, efficiently, and cost effectively incorporate enterprise level data analytics and search for real-time Operational Intelligence to an agile multi-tenant platform that aggregates workloads while modernizing and consolidating data center footprints.