DELL EMC VBLOCK SYSTEM 540 INFRASTRUCTURE FOR SPLUNK ENTERPRISE

Scale-out Capacity and Performance for Varied Splunk Enterprise Workloads with XtremIO and Dell EMC Technical Extension for Isilon

April 2017

Abstract

This solution guide describes a Dell EMC Vblock® System 540 converged infrastructure solution that highlights flexible scaling options and tight integration with Splunk Enterprise for analyzing large quantities of machine data.

H15264.2

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

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Executive Summary</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Business case</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Solution overview</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Key results</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Audience</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>We value your feedback</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>Solution Architecture</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Overview</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Vblock 540 architecture</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Dell EMC Technology Extension for Isilon</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>VMware vSphere</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Splunk Enterprise</td>
<td>13</td>
</tr>
<tr>
<td>3</td>
<td>Performance Testing Design for Splunk Enterprise</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Overview</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Compute design</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Network design</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Storage design</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Virtualization design</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Splunk Enterprise deployment and configuration</td>
<td>22</td>
</tr>
<tr>
<td>4</td>
<td>Performance Results</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>Test scenario</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Test methodology</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Test results</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>Use case summary</td>
<td>31</td>
</tr>
<tr>
<td>5</td>
<td>Data Reduction Savings for Splunk Enterprise</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>Test scenario</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>Test methodology</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>Test results</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>Use case summary</td>
<td>34</td>
</tr>
<tr>
<td>6</td>
<td>XtremIO Virtual Copy Backup and Restore</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Test scenario</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Test methodology</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Test results</td>
<td>37</td>
</tr>
<tr>
<td>Contents</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use case summary</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chapter 7  Validated Configurations for Splunk Enterprise</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overview</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Splunk-validated sizing configurations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chapter 8  Conclusion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Findings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conclusion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chapter 9  References</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dell EMC documentation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VMware documentation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Splunk Enterprise documentation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter 1  Executive Summary

This chapter presents the following topics:

- Business case .................................................................6
- Solution overview ..........................................................7
- Key results .....................................................................7
- Audience ......................................................................8
- We value your feedback .............................................8
Business case

Access to and analysis of machine data is one of the fastest growing and complex areas of big data. It is also one of the most valuable sources of data, containing a definitive record of events that can reveal information about user transactions, customer behavior, machine behavior, security threats, fraudulent activity, and more.

Making use of this data, however, presents real challenges. Traditional data analysis, management, and monitoring solutions are not engineered to handle such high-volume, high-velocity, and highly diverse data.

Splunk Enterprise is the industry-leading platform for machine data. It gives you realtime insight and understanding into what is happening, and provides end-to-end visibility across your IT infrastructure to enable informed, data-driven decisions. Splunk Enterprise has many critical uses across IT and the business, including:

- IT operations management
- Security and fraud
- Business analytics
- Application delivery

Dell EMC converged infrastructure offers scalable solutions that meet and exceed the performance and capacity requirements for a high-performance Splunk deployment\(^1\). These converged solutions enable customers to simplify purchasing and deployment, reduce their hardware management overhead, and accelerate their time to value. Dell EMC offers a wide portfolio of options to match the performance and pricing needs of large and small Splunk Enterprise deployments.

The primary benefits of running Splunk Enterprise on Dell EMC converged infrastructure are:

- **Splunk-validated configurations**—Jointly validated with Splunk to meet or exceed the performance of Splunk’s documented reference hardware

- **Optimal storage tier alignment for Splunk Enterprise**—Flexible sizing options to achieve the desired retention and performance profile for indexing and searching data in Splunk Enterprise

- **Cost-effective and flexible scale-out**—Scale-out compute and storage management in a single converged platform package

- **Powerful data services**—Out-of-the-box secure data encryption and data reduction services, along with integrated copy data management for efficient backup and restore capabilities

\(^1\) When configured according to the sizing guidance provided by Splunk and following the documented best practices for Splunk Enterprise, XtremIO, Isilon, and VMware.
Solution overview

Dell EMC Vblock systems easily, efficiently, and cost-effectively scale to support an enterprise-level, high-performance Splunk Enterprise environment. The Splunk-validated sizing guidance and performance results provided in this document are based on a Vblock™ System 540 with the Dell EMC Technology Extension for the Dell EMC Isilon™ storage system.

This solution guide discusses the following:

- Hardware architecture overview of a Vblock System 540 with the Technology Extension for Isilon
- Splunk architecture overview and an explanation of how a Vblock System 540 with the Technology Extension for Isilon aligns to Splunk Enterprise’s indexing and data management implementation
- Splunk-validated sizing guidance and best practices for a Vblock System 540 with the Technology Extension for Isilon to achieve the best performance for distributed or clustered indexer Splunk Enterprise deployments
- Overview of the test methodology used to measure storage performance and latency
- Additional benefits of using a Vblock System 540 for Splunk Enterprise, including:
  - Dell EMC XtremIO™ data reduction storage savings
  - XtremIO Virtual Copy (XVC) for space-efficient and fast backup and restore of Splunk Enterprise hot/warm indexed data

Key results

Our testing of this solution yielded the following:

- Dell EMC and Splunk jointly validated a Vblock System 540 with the Technology Extension for Isilon\(^\text{2}\) to comfortably scale to support a Splunk Enterprise deployment of up to 3 TB per day.
- Splunk provided and validated sizing guidance for a Vblock System 540 with Isilon providing a clear path for scaling from 100 GB per day to 1 TB per day.
- Data reduction savings for Splunk Enterprise data on XtremIO range from 1.5:1 to 2:1 for a clustered indexer Splunk environment.
- A Vblock System 540 converged infrastructure with Technology Extension for Isilon provides backup and restore capabilities for Splunk Enterprise hot/warm and cold indexed data using XVC and Isilon snapshot functionality.
- Using Isilon as scale-out storage for Splunk Enterprise cold data enables Splunk customers to increase their data retention period. Customers can search and continue to get value from their indexed data for much longer periods of time than direct-attached storage can typically accommodate.

\(^\text{2}\) Based on hardware and indexer configuration requirements documented in Table 1.
Chapter 1: Executive Summary

Audience

This guide is intended for IT administrators, storage administrators, virtualization administrators, system administrators, IT managers, and those who evaluate, acquire, manage, maintain, or operate Splunk Enterprise environments.

We value your feedback

Dell EMC and the authors of this document welcome your feedback on the solution and the solution documentation. Contact EMC.Solution.Feedback@emc.com with your comments.

Authors:
Dell EMC: Kirankumar Bhusanurmath, Lance Fang, Vincent Shen, Reed Tucker
Splunk: Jenny Hollfelder
This chapter presents the following topics:

- **Overview** ................................................................. 10
- **Vblock 540 architecture** .................................................. 11
- **Dell EMC Technology Extension for Isilon** ........................ 12
- **VMware vSphere** ............................................................ 13
- **Splunk Enterprise** .......................................................... 13
Overview

The following reference architecture describes a Vblock 540 with the Technology Extension for Isilon for a virtualized Splunk Enterprise environment. This reference architecture was jointly tested and validated by Splunk and Dell EMC to meet or exceed the performance of Splunk Enterprise running on Splunk’s reference hardware.

The base Vblock 540 provides compute resources, networking, and all-flash storage with XtremIO, and a virtualization and management layer in a converged configuration. This architecture enables simplified management and provisioning of resources. The Technology Extension for Isilon provides optimized network connectivity between the compute resources located within the Vblock 540 to a scale-out Isilon storage cluster over NFS.

Figure 1 shows how we deployed the Splunk instances as virtual machines on a VMware vSphere 5.5 cluster following Splunk’s documented virtualization best practices. The XtremIO storage that is part of the Vblock 540 System hosts the Splunk indexer virtual machines and hot/warm buckets. We configure the Isilon storage as NFS exports that are mounted on the Splunk indexers for long-term retention of cold buckets.

**Note:** For an explanation of the hot/warm and cold bucket concept, refer to Splunk core architecture.

---

Figure 1. Vblock 540 reference architecture for Splunk Enterprise
Table 1 lists the hardware components in this solution.

**Table 1. Hardware components**

<table>
<thead>
<tr>
<th>Layer</th>
<th>Hardware</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compute</td>
<td>Cisco UCS B200 M3 blade server</td>
</tr>
<tr>
<td></td>
<td>20 CPU cores at 2.8 GHz</td>
</tr>
<tr>
<td></td>
<td>256 GB memory</td>
</tr>
<tr>
<td>Network</td>
<td>Cisco UCS 6248UP fabric interconnect</td>
</tr>
<tr>
<td></td>
<td>Cisco MDS 9148 Fibre Channel switch</td>
</tr>
<tr>
<td>Storage</td>
<td>XtremIO storage system (2 X-Bricks)</td>
</tr>
<tr>
<td>Isilon</td>
<td>X410</td>
</tr>
</tbody>
</table>

Table 2 lists the versions of software used in testing this solution.

**Table 2. Software components**

<table>
<thead>
<tr>
<th>Software</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Splunk Enterprise</td>
<td>6.4</td>
</tr>
<tr>
<td>Splunk Universal Forwarder</td>
<td>6.4</td>
</tr>
<tr>
<td>CentOS 6.7 64bit</td>
<td>6.7</td>
</tr>
<tr>
<td>VMware vSphere</td>
<td>5.5</td>
</tr>
<tr>
<td>XIOS</td>
<td>4.0.0-58</td>
</tr>
<tr>
<td>OneFS®</td>
<td>8.0.0.1</td>
</tr>
</tbody>
</table>
Chapter 2: Solution Architecture

Vblock System 540 uses Cisco Unified Computing System (UCS) blade enclosures, interconnects, and blade servers.

The UCS data center platform combines x86-architecture blade and rack servers with networking and storage access in a single system. Innovations in the platform include a standards-based, unified network fabric; a Cisco Virtual Interface Card (VIC); and Cisco UCS Extended Memory Technology. A wire-once architecture with a self-aware, self-integrating, intelligent infrastructure eliminates the need to manually assemble components into systems.

Cisco UCS B-Series 2-socket blade servers deliver optimized performance to a wide range of workloads. Based on Intel Xeon processor E7 and E5 product families and designed for virtualized applications, these servers deliver fast performance and reduce expense by integrating systems management and converging network fabrics.

Networking components

The networking components in the Vblock System 540 include Cisco Nexus 5548UP switches, fabric interconnects, and Cisco Nexus 3064-T Ethernet switches.

A pair of Cisco Nexus 5548UP switches provides 10 GbE connectivity to the Vblock System 540 components as well as connectivity to the external network through the customer's core network. A pair of Cisco Nexus 3064-T switches connects the Advanced Management Pod (AMP) to the external network, supporting the Vblock System management infrastructure with redundancy.

The Cisco Nexus 5548UP switches provide 10 GbE connectivity as follows:

- Between the Vblock System internal components
- From those internal components to the AMP
- From the internal components to the external network

Dell EMC Technology Extension for Isilon

The Isilon X-Series is a flexible and comprehensive storage product that provides large capacity and high performance. Vblock 540 supports the Dell EMC Technology Extension for Isilon storage.

Isilon storage uses intelligent software to scale data across a large number of commodity hardware units, enabling explosive growth in performance and capacity. The product's revolutionary storage architecture—the OneFS® operating system (OS)—offers a single clustered file system.

OneFS provides its value by incorporating parallelism at a deep level of the OS. Virtually, the system is distributed across multiple hardware units. This parallelism allows OneFS to scale in every dimension as the infrastructure is expanded. By providing multiple redundancy levels, the system has no single point of failure. As a result, OneFS can grow to a multi-petabyte scale while providing greater reliability than traditional systems.

OneFS runs on Isilon scale-out NAS hardware, ensuring that Isilon benefits from the ever-improving cost and efficiency curves of commodity hardware. OneFS allows you to add hardware to or remove hardware from the cluster at any time. The data is protected from...
hardware changes. This feature alleviates the cost and burden of data migrations and hardware refreshes.

**VMware vSphere**

VMware vSphere is a widely adopted virtualization platform. The technology increases server utilization so that a firm can consolidate its servers and spend less on hardware, administration, energy, and floor space. The vSphere platform enables its installations to respond to user requests reliably while giving administrators the tools to respond to their changing needs.

The components of particular importance in this solution are vSphere ESXi and vCenter.

**VMware vSphere ESXi**

VMware vSphere ESXi is a bare-metal hypervisor. It installs directly on a physical server, and partitions that server into multiple virtual machines. The phrase *ESXi host* refers to the physical server.

vSphere ESXi hosts and their resources are pooled together into clusters that contain the CPU, memory, network, and storage resources that are available for allocation to the virtual machines. Clusters scale up to a maximum of 32 hosts and can support thousands of virtual machines.

**VMware vSphere vCenter**

VMware vCenter Server is the management software that runs on a virtual or physical server to oversee multiple ESXi hypervisors as a single cluster. An administrator can interact directly with vCenter Server or use vSphere Client to manage virtual machines from a browser window anywhere in the world. For example, the administrator can capture the detailed blueprint of a known, validated configuration—a configuration that includes networking, storage, and security settings—and then deploy that blueprint to multiple ESXi hosts.

**Splunk Enterprise**

Splunk Enterprise is a software product that enables you to search, analyze, and visualize the machine-generated data gathered from different sources in your IT infrastructure, including applications, networking devices, host and server logs, mobile devices, and more.

It gives you realtime insight and understanding into what is happening and provides end-to-end visibility across your IT infrastructure to enable informed, data-driven decisions.

**Splunk core architecture**

Figure 2 provides a graphic overview of the Splunk system architecture. A Splunk Enterprise instance can perform the role of a search head, an indexer, or both in the case of small deployments. Once the daily ingest rate or search load exceeds the sizing recommendations for a combined instance environment, Splunk Enterprise scales horizontally by adding additional indexers and search heads. For more information, refer to the *Splunk Capacity Planning Manual*. 
This solution guide focuses on the indexing tier, where the bulk of the work is done, including parsing, transforming, indexing, and retrieving data when a search is issued from the search head.

![Splunk architecture overview](image)

**Figure 2. Splunk architecture overview**

When data is received by a Splunk Enterprise indexer, the indexer parses the raw data into distinct events based on the timestamp of the event and writes them to the appropriate index. Splunk implements a form of storage tiering called hot/warm and cold buckets of data to optimize performance for newly indexed data and provide an option to keep older data for longer periods on higher capacity storage.

Newly indexed data lands in a hot bucket, where it is actively read and written by Splunk. When the number of hot buckets is reached, or the size of the data in the hot buckets exceeds the specified threshold, the hot bucket is rolled to a warm bucket. Warm buckets reside on the same tier of storage as hot buckets. The only difference is that warm buckets are read-only. This means that they can be backed up with traditional, file-based backup tools unlike hot buckets, which can only be backed up with something like XVC. It is important that the storage identified for hot/warm data is your fastest storage tier because it has the biggest impact on the performance of your Splunk Enterprise deployment.

When the number of warm buckets or volume size is exceeded, data is rolled into a cold bucket, which can optionally reside on another tier of storage. Cold data may reside on an NFS mount if the latency is less than 5ms (ideally) and no more than 200ms at most. Network-attached storage (NAS) technologies offer an acceptable blend of performance and lower cost per TB making them a good choice for longer-term retention of cold data.

Data can also be archived or frozen, but such data is no longer searchable. Manual user action is required to bring the data back into Splunk Enterprise buckets to be searchable. While customers sometimes choose to use frozen buckets to meet compliance retention requirements, this paper seeks to show how Isilon’s massive scalability and competitive cost of ownership can empower you to retain more data in their cold bucket, where it remains searchable. Figure 3 provides more detail about Splunk bucket concepts.
Splunk Indexer Buckets

Figure 3. Splunk index buckets
This chapter presents the following topics:

Overview .............................................................................................................. 17
Compute design ................................................................................................... 17
Network design ................................................................................................. 18
Storage design ................................................................................................... 19
Virtualization design ......................................................................................... 20
Splunk Enterprise deployment and configuration ........................................... 22
## Overview

Dell EMC executed the following tests to measure the performance of a Vblock System 540 with the Technology Extension for Isilon running a synthetic workload for Splunk Enterprise.

## Compute design

Table 3 and Table 4 provide details on the configuration of the ESXi and virtual machines for the Splunk infrastructure:

**Table 3. ESXi and virtual machines configuration for a Splunk distributed environment**

<table>
<thead>
<tr>
<th>ESXi cluster</th>
<th>ESXi quantity</th>
<th>Virtual machine role</th>
<th>Virtual machine quantity</th>
<th>vCPUs</th>
<th>RAM (GB)</th>
<th>Operating system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Splunk distributed environment</td>
<td>7</td>
<td>Splunk search head</td>
<td>1</td>
<td>24</td>
<td>64</td>
<td>Splunk Enterprise 6.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Splunk indexer</td>
<td>3</td>
<td>32</td>
<td>64</td>
<td>CentOS 6.7 64 bit</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Splunk forwarder</td>
<td>3</td>
<td>8</td>
<td>32</td>
<td>Splunk Universal Forwarder 6.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CentOS 6.7 64 bit</td>
</tr>
</tbody>
</table>

**Table 4. ESXi and virtual machines configuration for a Splunk clustered environment**

<table>
<thead>
<tr>
<th>ESXi cluster</th>
<th>ESXi quantity</th>
<th>Virtual machine role</th>
<th>Virtual machine quantity</th>
<th>vCPUs</th>
<th>RAM (GB)</th>
<th>Operating system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Splunk clustered indexer environment</td>
<td>3</td>
<td>Master node</td>
<td>1</td>
<td>8</td>
<td>32</td>
<td>Splunk Enterprise 6.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Splunk search head</td>
<td>1</td>
<td>24</td>
<td>64</td>
<td>CentOS 6.7 64 bit</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Splunk indexer</td>
<td>4</td>
<td>32</td>
<td>64</td>
<td>Splunk Universal Forwarder 6.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Splunk forwarder</td>
<td>2</td>
<td>8</td>
<td>32</td>
<td>CentOS 6.7 64 bit</td>
</tr>
</tbody>
</table>
Network design

**IP network**

We designed the IP network for this solution as follows:

- Configured the two Cisco Nexus 6248UP switches to provide 10 Gb Ethernet connectivity to the Splunk infrastructure
- Used virtual local area networks (VLANs) to logically group devices that were on different network segments or subnetworks
- Deployed separate network adapters/networks for vMotion and VMkernel management

In this solution, we deployed the Splunk infrastructure on seven Cisco B200 M3 blades connected through the vSphere 5.5 distributed switches. We deployed vMotion and VMkernel management on a vSphere 5.5 standard switch.

Figure 4 shows the Splunk infrastructure network design.

![Splunk infrastructure network design](image-url)
Dell EMC recommends the following as best practices for configuring the storage area network (SAN):

- Do not use more than 16 paths per device.
- Keep a consistent link speed and duplex across all paths between the host and the XtremIO cluster.

In the Vblock system, data moves from the compute layer through a pair of fabric interconnects to the SAN switches. Each fabric has two port channels at 64 Gbps (using all eight FC ports at 8 Gbps). The total aggregate bandwidth available to the XtremIO array from the Cisco UCS is 128 Gbps, which theoretically provides approximately 11 GBps in bandwidth.

**Storage design**

**XtremIO storage design**

XtremIO uses its multi-controller scale-out design and RDMA fabric to maintain all metadata in memory. The performance is always consistent and predictable. With built-in thin provisioning, XtremIO storage is allocated only when it is needed. This enables you to create larger LUNs to accommodate future or unexpected growth for indexes, without wasting any physical space on storage.

On the Vblock 540, we configured a two-XtremIO X-Brick array, which provides 30 TB of usable physical space. As Figure 5 shows, we created and allocated seven 2 TB volumes on XtremIO to support:

- Splunk forwarders that store raw data
- Splunk indexers that store hot and warm buckets
- Linux OS, Splunk Enterprise, and Splunk forwarder installation binaries

![Figure 5. Cluster and storage design for Splunk Enterprise](image-url)
Table 5 lists the volume sizes that we set up for this solution. XtremIO uses thin provisioning to avoid wasting physical storage and provide growth capacity for future business requirements.

Table 5. XtremIO storage design for Splunk infrastructure

<table>
<thead>
<tr>
<th>Environment</th>
<th>Volume name</th>
<th>Description</th>
<th>Volume number</th>
<th>Volume capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Splunk distributed deployment</td>
<td>Splunk indexer</td>
<td>Splunk indexer volume</td>
<td>3</td>
<td>2 TB</td>
</tr>
<tr>
<td></td>
<td>Splunk forwarder</td>
<td>Splunk forwarder volume</td>
<td>3</td>
<td>2 TB</td>
</tr>
<tr>
<td>OS</td>
<td></td>
<td>Linux OS, Splunk Enterprise volume</td>
<td>1</td>
<td>2 TB</td>
</tr>
<tr>
<td>Splunk clustered indexer deployment</td>
<td>Splunk indexer</td>
<td>Splunk indexer volume</td>
<td>2</td>
<td>2 TB</td>
</tr>
<tr>
<td></td>
<td>Splunk forwarder</td>
<td>Splunk forwarder volume</td>
<td>4</td>
<td>2 TB</td>
</tr>
<tr>
<td>OS</td>
<td></td>
<td>Linux OS, Splunk Enterprise volume</td>
<td>1</td>
<td>2 TB</td>
</tr>
</tbody>
</table>

Isilon provides a highly scalable, flexible storage system that protects data and optimizes the flow of information within an organization. The Isilon OneFS operating environment provides the specialized data protection, data security, compliant retention, and simple, massive scalability required for long-term retention.

On the Vblock 540, the Technology Extension for 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. In this solution, we allocated 1.5 TB of SAS storage from Isilon to each indexer for cold buckets, which store data rolled out from the hot or warm buckets from XtremIO, as Figure 5 shows.

Table 6 shows the volume size that we set up for this solution.

Table 6. Isilon storage design for Splunk infrastructure

<table>
<thead>
<tr>
<th>Volume name</th>
<th>Description</th>
<th>Volume number</th>
<th>Volume capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Splunk indexer</td>
<td>Splunk indexer volume</td>
<td>3</td>
<td>1.5 TB</td>
</tr>
</tbody>
</table>

Splunk and Dell EMC recommend that you use only Isilon NFS for Splunk cold and frozen data. Refer to the System Requirements section in the Splunk Installation Manual.

Virtualization design

Virtual machine configuration Virtualization improves resource sharing and utilization, includes high availability, and simplifies provisioning and management. VMware vSphere is the virtualization platform that provides the foundation for this solution. The solution uses VMware vSphere ESXi hypervisor.
In this solution, we installed ESXi hypervisors on seven hosts. These vSphere ESXi hosts and their resources are pooled together into one cluster. The cluster contains the CPU, memory, network, and storage resources available for allocation to virtual machines. We deployed eight virtual machines for our Splunk distributed deployment, and eight more for the Splunk clustered indexer deployment, for the roles listed in Table 7.

<table>
<thead>
<tr>
<th>Physical ESXi cluster</th>
<th>Splunk environment</th>
<th># of VMs</th>
<th>Roles of virtual machines</th>
</tr>
</thead>
<tbody>
<tr>
<td>vSphere cluster of seven ESXi servers</td>
<td>Splunk distributed deployment</td>
<td>1</td>
<td>Splunk search head</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>Splunk indexer</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>Splunk forwarder</td>
</tr>
<tr>
<td></td>
<td>Splunk clustered indexer deployment</td>
<td>1</td>
<td>Splunk master node</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>Splunk search head</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>Splunk indexer</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>Splunk forwarder</td>
</tr>
</tbody>
</table>

This solution implements the following Dell EMC and VMware best practices to provide the best performance for all the Splunk virtual machines running on ESXi hosts:

- Install PowerPath®/VE to provide efficient path management and load balancing.
- Use a single virtual socket for each virtual machine. With virtual Non-Uniform Memory Access (NUMA) topology, we recommend a single virtual socket that has fewer virtual CPU cores than the physical CPU cores of a socket in the physical ESXi host.
- Use a VMware Paravirtual SCSI controller to increase throughput with significant CPU utilization reduction in the SAN environment.
- Use a VMware VMXNET 3 network adapter to optimize network performance.
- Use Thick Provision Eager Zeroed disk provisioning to optimize virtual disk performance on the XtremIO array.
- Set the ESXi host parameters as follows to maximize Splunk performance on the XtremIO array:
  - lpfc_lun_queue_depth=128
  - Disk.SchedQuantum=64
  - Disk.DiskMaxIOMax=4096

For more information, refer to *Performance Best Practices for VMware vSphere 5.5* and the *Dell EMC XtremIO Storage Array Host Configuration Guide*. 

**Virtualization configuration**
Splunk Enterprise deployment and configuration

**Splunk distributed environment**

In this solution, we installed and configured the following components to implement a multi-tiered distributed environment:

- One Splunk search header
- Three Splunk forwarders
- Three Splunk indexers

Figure 6 outlines the Splunk distributed environment.

---

**Splunk clustered indexer environment**

In this solution, we installed and configured the following components to implement the Splunk clustered indexer environment:

- One Splunk search header
- Two Splunk forwarders
- Four Splunk indexers
- One Splunk master node

Figure 7 outlines the Splunk clustered indexer environment.

---

We provisioned the hot/warm bucket storage in XtremIO on the Vblock 540. The Technology Extension for Isilon provides storage for the cold buckets.
Dell EMC recommends that you tune the Linux OS for shared storage on all Splunk indexers to support better performance of the Splunk infrastructure, as follows:

- Configure a tuned daemon to use the virtual-host profile. This statically and dynamically tunes the system settings for high throughput and low latency. Use the following command:
  ```
  tuned-adm profile virtual-host
  ```

- Disable SELinux, so that enhanced system security will not add overhead to the performance.

- Tune the kernel to optimize the network for high throughput over a 10-Gigabit Ethernet by adding the following command string to `/etc/sysctl.conf`:
  ```
  net.ipv4.tcp_timestamps=0
  net.ipv4.tcp_sack=1
  net.core.netdev_max_backlog=250000
  net.core.rmem_max=4194304
  net.core.wmem_max=4194304
  net.core.rmem_default=4194304
  net.core.wmem_default=4194304
  net.core.optmemp_max=4194304
  net.ipv4.tcp_rmem=4096 87380 4194304
  net.ipv4.tcp_wmem=4096 65536 4194304
  net.ipv4.tcp_low_latency=1
  ```
This chapter presents the following topics:

- **Test scenario** ................................................................. 25
- **Test methodology** ............................................................ 25
- **Test results** ........................................................................ 26
- **Use case summary** ............................................................ 31
Test scenario

In this test scenario, Splunk and Dell EMC performed a baseline validation of a 3 TB virtualized distributed Splunk Enterprise infrastructure. The test demonstrates that a virtualized distributed Splunk Enterprise environment provides both a high performance and data retention capability using XtremIO for hot/warm buckets and Isilon for cold buckets.

We compared the baseline results in the third scenario to show that the intensive XVC imposes a minimum performance impact to the whole system.

Test methodology

In this test scenario, we deployed a distributed Splunk Enterprise system on a VMware ESXi cluster in the following configuration:

- One Splunk search header
- Three Splunk forwarders
- Three Splunk indexers

The indexers and the forwarders have 1 to 1 mapping. Table 8 shows the Splunk index allocation.

Table 8.  Splunk index allocation on Splunk indexer virtual machine

<table>
<thead>
<tr>
<th>Splunk bucket type</th>
<th>Backend array</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hot/warm</td>
<td>XtremIO in a Vblock 540 system</td>
</tr>
<tr>
<td>Cold</td>
<td>Vblock technology extension for Isilon</td>
</tr>
</tbody>
</table>

For more information, refer to Splunk Enterprise deployment and configuration.

For instructions on how to set different sizes for hot/warm and cold buckets, refer to the Splunk Admin Manual.

Table 9 describes the test stages.

Table 9.  Test stages

<table>
<thead>
<tr>
<th>Stage</th>
<th>Test stage name</th>
<th>Details</th>
<th>Test duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Index stage</td>
<td>3 TB data from 3 forwarders transformed into events and stored as indexes.</td>
<td>8 hours</td>
</tr>
<tr>
<td>2</td>
<td>Static search stage</td>
<td>Search test to run over data in the static index without indexing workload</td>
<td>8 hours</td>
</tr>
<tr>
<td>3</td>
<td>Streaming search stage</td>
<td>Streaming search stage to introduce the indexing load to the system and at the same time run the search test</td>
<td>10 hours</td>
</tr>
</tbody>
</table>
Chapter 4: Performance Results

For both static search and streaming search, the test methodology introduces various search types with a different concurrency, as Table 10 explains.

Table 10. Details of static search and streaming search

<table>
<thead>
<tr>
<th>Test stage name</th>
<th>Search type</th>
<th>Details of search type</th>
<th>Concurrency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static search stage and streaming search stage</td>
<td>Dense search</td>
<td>This search returns a large set of matching results for a given set of data in a given period of time.</td>
<td>4,8,12,…,64</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Note: Dense searches are primarily CPU bound</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rare search</td>
<td>Returns a very small number of result set from each index bucket that matches the search.</td>
<td>4,8,12,…,64</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Note: Rare searches are primarily I/O bound</td>
<td></td>
</tr>
</tbody>
</table>

Note: For more information about dense and rare searches and why they are CPU-bound and I/O-bound, refer to the Splunk Capacity Planning Manual.

Table 11 lists the metrics that we recorded for this use case.

Table 11. System metrics

<table>
<thead>
<tr>
<th>Data source</th>
<th>Metrics</th>
<th>Collection method</th>
</tr>
</thead>
<tbody>
<tr>
<td>XtremIO</td>
<td>XtremIO Volume IOPS</td>
<td>XMS Report</td>
</tr>
<tr>
<td></td>
<td>XtremIO Volume latency (us)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>XtremIO Volume bandwidth(Gb/s)</td>
<td></td>
</tr>
<tr>
<td>vCenter</td>
<td>Vblock ESXi host CPU percentage</td>
<td>vSphere PowerCLI and vCenter performance panel</td>
</tr>
<tr>
<td></td>
<td>Vblock ESXi host memory usage</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Splunk virtual machine CPU percentage</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Splunk virtual machine memory usage</td>
<td></td>
</tr>
<tr>
<td>Insight IQ</td>
<td>NFS throughput (Gb/s)</td>
<td>Performance panel in Insight IQ</td>
</tr>
<tr>
<td></td>
<td>Avg. disk latency (us)</td>
<td></td>
</tr>
<tr>
<td>Splunk</td>
<td>Splunk reference hardware. Refer to the Splunk Capacity Planning Manual.</td>
<td>Splunk Enterprise Distributed Management Console (DMC)</td>
</tr>
</tbody>
</table>

Test results

This section describes and analyzes the results from different perspectives, including overall Splunk test results, Vblock test results, and Technology Extension with Isilon test results.
Chapter 4: Performance Results

Splunk test results

All the test results indicate that the configuration in this solution exceeds the performance of Splunk reference hardware. For more information, refer to Reference Hardware in the Splunk Capacity Planning Manual.

Vblock 540 test results

For Vblock 540, we analyzed the results from the following two perspectives:

- Backend storage – XtremIO
- Compute resource

Backend storage: XtremIO

Figure 8 shows the XtremIO performance during the Splunk index stage. In this stage, the read/write ratio is 1:1, with an average size of 320 KB. The average throughput is 2.8 Gb/s, with 3 ms as the average XtremIO latency.

![XtremIOAvg. throughput and latency during Splunk index stage](image)

**Figure 8.** XtremIO performance in the Splunk index stage

Figure 9 illustrates the XtremIO performance during the Splunk static search stage. This stage involves all read I/O with an average size of 16 KB. The average latency is 0.5 ms, which is consistently low.

The rare search is I/O-bound, while the dense search is Splunk indexer CPU-bound. During a rare search, IOPS grows together with concurrency configurations. The extremely low latency and the predictable performance of XtremIO give the system administrator the capacity to scale out the system while maintaining the expected performance.
We observed similar performance behavior during the Splunk streaming search stage, as shown in Figure 10. The average latency on XtremIO during a rare search is 1ms and 2.3ms during a dense search. This is due to the difference in the read/write ratio between a dense search and a rare search, as shown in Table 12.
Figure 10. XtremIO performance in Splunk streaming search stage

Table 12. Read/write ratio for dense search and rare search

<table>
<thead>
<tr>
<th></th>
<th>Dense search</th>
<th>Rare search</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read/write ratio</td>
<td>1:2.25</td>
<td>11.4:1</td>
</tr>
</tbody>
</table>

Compute resource

Figure 11 shows the overall CPU and memory utilization of the Vblock 540 using seven ESXi hosts. The overall CPU utilization is 39.6 percent and memory utilization of the Vblock system is 36.2 percent, indicating that there is room for future growth for Splunk configuration in this test scenario. The Vblock system also enables you to dynamically add servers so that you can allocate resources.
Chapter 4: Performance Results

VCE technology extension with Isilon test results

The VCE technology extension with Isilon provides the NFS mount for the cold bucket data in each Splunk indexer virtual machine.

Figure 12 shows the performance results for the cold bucket data. The average disk latency is extremely low: less than 1ms through all three stages. This unique consistency of response time is enabled by eight Isilon nodes and SSD because the L3 cache tier enables system administrators to confidently store their cold bucket data in Isilon with stable and predictable performance.

Figure 11. CPU and memory usage for Vblock

Figure 12. Test results for the cold bucket data on Isilon
Use case summary

Combined with the Technology Extension for Dell EMC Isilon storage, the Vblock System 540 is an excellent scale-out solution for a high performance Splunk Enterprise deployment. In this test scenario, the Vblock 540 provides predictable performance, low latency, flexibility, and room for future growth for mission-critical big data applications like Splunk Enterprise.
This chapter presents the following topics:

- **Test scenario** ................................................................. 33
- **Test methodology** ......................................................... 33
- **Test results** ................................................................. 34
- **Use case summary** ....................................................... 34
Test scenario

In this test scenario, Splunk and Dell EMC performed a baseline validation of data reduction savings for a virtualized Splunk cluster. The test shows how XtremIO drives exceptional data efficiencies with a reduced total cost of ownership (TCO) on a virtualized Splunk cluster.

Test methodology

In this test scenario, we deployed Splunk Enterprise in a clustered configuration on a VMware ESXi cluster. The clustered Splunk Enterprise test environment included:

- One Splunk search head
- Four Splunk indexers
- One Splunk cluster master
- Two Splunk forwarders

We configured Splunk Enterprise to use a replication and search factor of two (RF=2/SF=2) to achieve index redundancy. Although the Splunk default replication settings are RF=3/SF=2, Dell EMC recommends a replication factor of RF=2/SF=2 to make use of the reliability of the underlying storage without incurring the storage penalty of having Splunk Enterprise store additional copies of the data. This is an added storage savings benefit that Vblock 540 customers enjoy for clustered Splunk Enterprise environments. For more information, refer to Managing Indexers and Clusters of Indexers: Search factor and Managing Indexers and Clusters of Indexers: Replication factor in the Splunk documentation.

The high-level test procedure is as follows:

- Set up the Splunk cluster with a replication factor of RF=2/SF=2.
- Index a total of 2 TB data from the forwarders, and monitor performance.
- Measure the data reduction ratio and storage efficiency.

Table 13 lists the metrics that we obtained for this use case scenario.

Table 13. System metrics

<table>
<thead>
<tr>
<th>Data source</th>
<th>Metrics</th>
<th>Collection method</th>
</tr>
</thead>
<tbody>
<tr>
<td>XtremIO</td>
<td>XtremIO overall efficiency and data reduction ratio</td>
<td>XMS report</td>
</tr>
</tbody>
</table>
Test results

XtremIO overall efficiency

When Splunk indexes data, the resulting files written to disk are about 50 percent of the original ingested data size. Of the data written to disk, approximately 70 percent is an index file and 30 percent is the raw data written out in a compressed format.

Figure 13 shows the capacity savings for the hot/warm data when Splunk Enterprise is configured for clustered indexing with a replication factor set to RF=2/SF=2. Based on our testing, the data reduction ratio, which is used to size XtremIO storage for the hot/warm buckets, was 1.9:1. When you add the storage savings from the XtremIO thin provisioning feature, the overall efficiency for the XtremIO storage is 4.2:1.

The XtremIO data reduction technology, including compression, inline data deduplication, and thin provisioning, can effectively lower the TCO of the whole system.

<table>
<thead>
<tr>
<th>Storage</th>
<th>Overall Efficiency</th>
<th>4.2:1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Efficiency</td>
<td>Data Reduction Ratio</td>
<td>1.9:1</td>
</tr>
<tr>
<td>Deduplication</td>
<td>1:1</td>
<td></td>
</tr>
<tr>
<td>Compression</td>
<td>1.7:1</td>
<td></td>
</tr>
<tr>
<td>Thin Provisioning Saving</td>
<td>55%</td>
<td></td>
</tr>
<tr>
<td>Volume Capacity</td>
<td>14TB</td>
<td></td>
</tr>
<tr>
<td>Physical Capacity</td>
<td>7.587TB</td>
<td></td>
</tr>
</tbody>
</table>

A Vblock 540 system with XtremIO provides a significant reduction in the TCO of a Splunk clustered deployment by taking advantage of thin provisioning, compression, and inline data deduplication.
Chapter 6  XtremIO Virtual Copy Backup and Restore

This chapter presents the following topics:

Test scenario ........................................................................................................................................36
Test methodology .............................................................................................................................36
Test results .........................................................................................................................................37
Use case summary ............................................................................................................................40
Chapter 6: XtremIO Virtual Copy Backup and Restore

Test scenario

In this test scenario, Dell EMC validated a Vblock 540 XVC for Splunk backup and restore. The test shows that the intensive XtremIO XVC imposes a minimum performance impact to the whole system, with a minimum storage footprint. It demonstrates how using XtremIO XVC in a Vblock 540 system dramatically helps to recover from a Splunk Enterprise component corruption.

Test methodology

In this test scenario, we configured Splunk Enterprise in a distributed configuration using:

- One Splunk search head
- Three Splunk indexers
- Three Splunk universal forwarders

We used the same configuration, distributed Splunk Enterprise, as in the first test scenario.

We conducted the overall test procedure in two stages: performance validation and backup and restore.

The performance validation stage demonstrates the minimum performance impact and minimum additional storage footprint with intensive XVC in a Splunk distributed environment, compared with the baseline performance test in the first test scenario. We used the following process:

- Start the workload.
- During the index stage, configure the XVC interval to be 30 minutes and retain 20 copies. Measure the performance.
- During the static and streaming search stages, configure the XVC interval to be 1 minute and retain 100 copies. Measure the performance and storage efficiency.
- Measure the performance and storage footprint compared with the first test scenario.

The backup and restore stage simulated an all-Splunk indexer failure. We used the following process:

- Start the workload.
- Align the XVC and Isilon snapshot scheduler so that they both have a 20-minute interval for the Splunk hot/warm bucket and the cold bucket while the benchmark tool is running
- Simulate a failure by directly powering off all three Splunk indexers.
- Create three new Splunk indexers and configure them the same way as the previous ones. Then mount the most recent XVC for the hot bucket and the Isilon snapshot for the cold bucket data to recover the newly created Splunk indexers and check the data consistency.

Table 14 shows the metrics that we recorded for this use case scenario.
Table 14. System metrics

<table>
<thead>
<tr>
<th>Data source</th>
<th>Metrics</th>
<th>Collection method</th>
</tr>
</thead>
<tbody>
<tr>
<td>XtremIO</td>
<td>XtremIO volume IOPS</td>
<td>XMS report</td>
</tr>
<tr>
<td></td>
<td>XtremIO volume latency (us)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>XtremIO volume bandwidth (Gb/s)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>XtremIO overall efficiency and data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>reduction ratio</td>
<td></td>
</tr>
<tr>
<td>Splunk</td>
<td>Refer to the Reference Hardware section in</td>
<td>Splunk distributed console</td>
</tr>
<tr>
<td></td>
<td>the Splunk Capacity Planning Manual.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Event count for the last 7 days</td>
<td></td>
</tr>
</tbody>
</table>

**Test results**

**Splunk test results**

Test results from the first stage indicate that the configuration in this solution exceeds that of the Splunk reference hardware with a minimum compute resource overhead and shows no performance impact from Splunk’s perspective. For more information on Splunk reference hardware, refer to the Splunk Capacity Planning Manual.

For the second stage, we used the event count for a given period of time to determine if the data was consistent before and after the outage. We examined the values for both the static search stage and the streaming search stage. Table 15 shows that the data is consistent after we mounted the XVCs in the new Splunk indexers.

Table 15. Event count comparison

<table>
<thead>
<tr>
<th></th>
<th>Static search stage: Event count for given time period</th>
<th>Streaming search stage: Event count for given time period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before the outage</td>
<td>80,155,710</td>
<td>78,001,110</td>
</tr>
<tr>
<td>After the XVCs are mounted in</td>
<td>80,155,710</td>
<td>78,001,110</td>
</tr>
<tr>
<td>three new Splunk indexers and</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the indexers are synchronized</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter 6: XtremIO Virtual Copy Backup and Restore

XtremIO Virtual Copy efficiency

Minimal storage footprint

Figure 15 shows the XtremIO overall efficiency before and after 120 XVCs are taken. At the beginning, with no XVC, the overall efficiency is 13.4:1. After 120 XVCs are taken, the number increases to 112.4:1 with the same occupied physical capacity.

![XtremIO event count comparison](image1.png)

**Figure 14.** Event count comparison during static search stage

![XtremIO overall efficiency](image2.png)

**Figure 15.** Overall efficiency before and after the XVC
Minimal performance impact

Figure 16 shows that there is no performance impact during the index stage when the XtremIO Virtual Copy is enabled. The average latency is 3ms and the average throughput remains at 2.8 Gb/s for both scenarios.

![XtremIO XVC - Index stage](image)

**Figure 16. Performance comparison during the index stage**

Figure 17 and Figure 18 show that there is no performance impact during a static search and a streaming search when XVC is enabled. For the static search, the average latency is 0.5ms and the latency for a streaming search is 1ms, regardless of whether XVC is enabled or disabled.

These two graphs depict a rare search performance in both stages. Only rare search results are shown because a rare search is I/O-bound and a dense search is CPU-bound.

![XtremIO XVC - Static search stage](image)

**Figure 17. Performance comparison during static search stage**
Use case summary

With no additional storage footprint or performance impact, the Vblock 540 with the Technology Extension for Isilon provides full backup and restore capabilities for hot/warm and cold indexed data in Splunk Enterprise.
This chapter presents the following topics:

**Overview** ............................................................................................................................................. 42

**Splunk-validated sizing configurations** .................................................................................................. 42

**Summary** ................................................................................................................................................ 44
Overview

Splunk Enterprise can scale to meet almost any capacity requirement. Determining the best hardware and software configuration to achieve optimal performance at scale requires careful planning. Dell EMC offers the following Splunk-validated configurations for this Vblock 540 with the Technology Extension for Isilon solution to meet the challenge.

The hardware configurations and Splunk deployment recommendations provided are based on extensive testing, research, and collaboration with engineers at Splunk. The results exceed the performance of Splunk’s reference hardware at scale.

Splunk-validated sizing configurations

The sizing configurations described in this section provide guidance for daily ingest rates ranging from 100 GB per day to 1 TB per day for both distributed and clustered indexer Splunk Enterprise environments. Ensure that you have read all the information in this section before proceeding to the sizing tables.

We calculated the storage capacities listed in Table 16 and Table 17 using Splunk’s standard formula, which assumes that data written to disk is approximately half of the ingest rate\(^3\), and are rounded up to the nearest terabyte.

\[
\text{Daily Ingest Rate} \times \frac{x \text{ Retention}}{2}
\]

The retention values 90 and 180 listed in Table 16 and Table 17 represent the total retention time before data is deleted or optionally archived. Therefore, to calculate the amount of storage required for cold data given a particular hot/warm retention period, use the formula.

\[
\text{Daily Ingest Rate} \times \frac{x (\text{Total Retention} - \text{Hot/Warm Retention})}{2}
\]

With data reduction savings on XtremIO ranging from 1.5:1 and 2:1, you will be able to keep indexed data on your hot/warm storage tier longer. This means better performance for searches with a window exceeding the last 20 days. Using XtremIO provides a significant benefit for your hot/warm tier.

Important: When looking at the storage capacity requirements in Table 16 and Table 17 for a specific daily ingestion rate and retention period, we recommend adding at least 15-20 percent to the minimum amount of storage required to ensure that the filesystem has adequate free space to operate efficiently.

Distributed deployment

In a distributed deployment, each indexer stores a unique portion of the indexed data and no data is replicated. If an indexer is inaccessible for any reason, including OS reboots and Splunk restarts, the portion of data residing on that indexer will be unsearchable until

\[^3\text{The amount of data written to disk by Splunk Enterprise is approximately half the size of the original data, based on syslog-like data.}\]
the indexer comes back online. For this reason, even though the underlying infrastructure is highly reliable, we recommend that you use clustered indexer deployments to achieve high availability at the application level and to ensure that your data is always searchable. Table 16 shows the sizing guide for a distributed Splunk Enterprise deployment on a Vblock 540.

Table 16. Sizing guidelines for a distributed Splunk Enterprise deployment

<table>
<thead>
<tr>
<th>Ingest rate (GB/Day)</th>
<th>Hot/warm capacity required for 20+ days (TB)</th>
<th>Cold capacity required for 90 days (TB)</th>
<th>Cold capacity required for 180 days (TB)</th>
<th>Physical servers</th>
<th>Splunk indexer instances</th>
<th>Splunk search head instances</th>
<th>Splunk admin instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>1</td>
<td>4</td>
<td>8</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>200</td>
<td>2</td>
<td>7</td>
<td>16</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>500</td>
<td>5</td>
<td>18</td>
<td>40</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1,000</td>
<td>10</td>
<td>35</td>
<td>79</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Clustering deployment

Unlike distributed deployments, clustered indexer deployments replicate data to one or more indexers to achieve high availability at the application level. If an indexer goes down, Splunk is still able to search all the data. If an indexer fails and has to be replaced, a clustered indexer deployment can rebuild the missing data from a replica.

Splunk provides the ability to set different replication factors for replicating compressed raw data (RF) and indexed data, also known as searchable data (SF). We chose a replication factor of RF=2/SF=2 instead of the Splunk default of RF=3/SF=2 for these configurations because of the reliability of the underlying storage. This leads to a net storage savings for Splunk deployments running on a Vblock 540 with the Technology Extension for Isilon.

Table 17 shows the sizing guidelines for a clustered indexer Splunk Enterprise deployment on a Vblock 540.

Table 17. Sizing guide for a clustered indexer Splunk Enterprise deployment

<table>
<thead>
<tr>
<th>Ingest rate (GB/Day)</th>
<th>Hot/warm capacity required for 20+ days (TB)</th>
<th>Cold capacity required for 90 days (TB)</th>
<th>Cold capacity required for 180 days (TB)</th>
<th>Physical servers</th>
<th>Splunk indexer instances</th>
<th>Splunk search head instances</th>
<th>Splunk admin instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>2</td>
<td>8</td>
<td>16</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>200</td>
<td>4</td>
<td>14</td>
<td>32</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>500</td>
<td>10</td>
<td>36</td>
<td>80</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1,000</td>
<td>20</td>
<td>70</td>
<td>160</td>
<td>8</td>
<td>6</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Summary

The Splunk-validated configurations on the Vblock 540 with the Technology Extension for Isilon solution provide concrete examples that work well with, and exceed the performance of, Splunk’s reference hardware.

Scaling beyond 1 TB per day was achieved without any problems. Several Splunk customers ingest more than 1 TB per day on this infrastructure. For information on scaling beyond 1 TB per day or hosting Splunk Enterprise on a consolidated infrastructure, contact the Dell EMC Ninja team at splunk.ninja.team@dell.com.
This chapter presents the following topics:

**Summary** ................................................................. 46

**Findings** ........................................................................ 46

**Conclusion** ........................................................................ 47
Summary

This solution guide demonstrates that the Vblock 540 and the Technology Extension for Isilon can easily, efficiently, and cost-effectively scale to support enterprise-level data analytics and search for Splunk Enterprise.

Dell EMC XtremIO with Isilon architectures is well-suited to managing the resource-intensive demands of a Splunk Enterprise environment, including I/O-intensive workloads for hot/warm data and the ever-expanding capacity requirements for cold data. Figure 19 provides an illustration.

![Figure 19. Servers and buckets](image)

By addressing these key Splunk priorities, this guide enables you to implement the solution that best fits your needs without the contention across data tiers that would arise in a DAS or a single-storage-tier solution. This solution provides the flexibility to expand hot/warm and cold storage tiers independently and protects you against the limitations and bottlenecks found in traditional architectures at scale.

Findings

The key findings of this solution are:

- As a converged infrastructure, the Vblock 540 with the Technology Extension for Isilon is easy to order and fast to deploy. It requires little storage tuning and service in a virtualized environment for Splunk distributed and clustered indexer deployments.
- Performance testing with a synthetic Splunk Enterprise workload showed an XtremIO latency of 0.5ms during a Splunk static search, which significantly exceeds the recommendation of 5ms or less.
- Data reduction savings for Splunk Enterprise data on XtremIO were observed to be 1.9:1 during testing. Based on the type of data being ingested, the actual data reduction savings may vary from 1.5:1 to 2:1 for a clustered Splunk environment.
• The Vblock 540 with XtremIO Virtual Copy and Isilon snapshots provides instant recovery and data consistency in Splunk implementations with a minimum recovery time objective (RTO).

• The Technology Extension for Isilon is quick and easy to deploy and service in a virtualized environment for Splunk.

• Isilon provides a highly efficient, massively scalable storage of Splunk cold buckets with consistent throughput and excellent response times (less than 1ms).

### Conclusion

Deep insight into new or previously ignored data sources has resulted in increased competitive advantages for corporations because they can improve productivity, profitability, customer experience, and customer retention. Splunk is the leading platform in this space, enabling collection and analysis of data and realtime insights into data sources. As customers take advantage of these capabilities and increase the volume of their analyzed data, supporting the performance, reliability, and security of the underlying infrastructure becomes critical. The Splunk Enterprise on Vblock 540 architecture including XtremIO and Isilon meets these requirements.

XtremIO provides unmatched consistent performance and efficiency for hot and warm buckets, and Isilon scale-out NAS creates a long-term and powerful storage solution for cold buckets. Both platforms leverage the underlying scale-out architecture to easily support scale without added overhead. This approach enables organizations to avoid resource-intensive complexity and illustrates a simple environment for Splunk that can leverage existing investments and net-new investments in VMware, Vblock 540, and Isilon.

If you are interested in workload consolidation, you might also consider using Dell EMC Vblock systems for multiple application workloads, including Splunk.
This chapter presents the following topics:

**Dell EMC documentation** .......................................................... 49
**VMware documentation** ............................................................. 49
**Splunk Enterprise documentation** ............................................. 49
Dell EMC documentation

The following documents on EMC.com or EMC Online Support provide additional relevant information. Access to these documents depends on your login credentials. If you do not have access to a document, contact your EMC representative.

- Introduction to XtremIO Virtual Copies
- XtremIO Data Protection (XDP)
- Introduction to the EMC XtremIO Storage Array (Version 4.0)
- Vblock-540-overview
- VCE Solutions for Enterprise Mixed Workload on Vblock System 540 Solutions Guide
- Providing Enterprise Performance, Capacity, and Data Services for Splunk Enterprise
- EMC Isilon OneFS: A Technical Overview
- EMC Isilon OneFS Operating System
- EMC XtremIO Storage Array Host Configuration Guide.

VMware documentation

The following documentation provides additional relevant information:

- Performance Best Practices for VMware vSphere 5.5

Splunk Enterprise documentation

The following documentation provides additional relevant information:

- Welcome to Splunk Enterprise
- Download Splunk Enterprise
- Splunk Enterprise Distributed Deployment Manual
- Managing Indexers and Clusters of Indexers
- Splunk’s Capacity Planning Manual
- Splunk Admin Manual