Dell EMC Ready Bundle for Splunk

Ready Bundle on PowerEdge Servers with Isilon for Machine Data Analytics

December 2017

Abstract

This validation guide describes a Dell EMC Ready Bundle infrastructure solution that delivers flexible scaling options and tight integration with Splunk Enterprise and Isilon storage for analyzing large quantities of machine data.

H16193.1

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

VALIDATION GUIDE

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

Chapter 1 Executive Summary 5

Business case 6

Solution overview 6

Key results 8

Audience 9

We value your feedback 9

Chapter 2 Solution Architecture 10

Overview 11

Dell EMC PowerEdge Servers 13

Isilon storage 13

Splunk Enterprise 14

Chapter 3 Splunk Enterprise Design and Configuration 16

Overview 17

Compute design 17

Network design 18

Storage design 19

Splunk Enterprise design 21

Chapter 4 Splunk Enterprise Single Instance Infrastructure for 250 GB/day Indexing Volume 25

Overview 26

Implementation 26

Use case summary 35

Chapter 5 Splunk Enterprise Distributed Infrastructure for 250 GB/day Indexing Volume 36

Overview 37

Implementation 37

Use case summary 41

Chapter 6 Splunk Enterprise Clustered Infrastructure for 250 GB/day Indexing Volume 42

Overview 43

Implementation 43

Use case summary 47

Chapter 7 Splunk Enterprise High Performance Clustered Infrastructure for 250 GB/day Indexing Volume 48

Overview 49

Implementation 49

Use case summary 49

Chapter 8 Splunk Enterprise Clustered Infrastructure with Isilon for 250 GB/day Indexing Volume 50

Overview 51

Implementation 51

Use case summary 55

Chapter 9 Splunk Enterprise Single Instance Infrastructure for Development and Testing 56

Overview 57

Implementation 57

Use case summary 58

Chapter 10 Validated Configurations for Splunk Enterprise 59

Overview 60

Splunk-validated sizing configurations....60

Configuration 1: Single instance deployment for up to 250 GB/day indexing volume with 115-day retention 61

Configuration 2: Distributed deployment for up to 250 GB/day indexing volume with 115-day retention 61

Configuration 3: Clustered deployment for up to 250 GB/day indexing volume with 115-day retention 62

Configuration 4: High performance clustered deployment for up to 250 GB/day indexing volume with 115-day retention 62

Configuration 5: Isilon cold bucket expansion for up to 250 GB/day indexing volume with 210-day retention 63

Configuration 6: Single instance deployment for development and testing 63

Summary 63

Chapter 11 Conclusion 64

Summary 65

Findings 65

Conclusion 65

Chapter 12 References 66

Dell EMC documentation....67

Splunk Enterprise documentation....67

Other documentation....67

Chapter 1 Executive Summary

This chapter presents the following topics:

Business case 6

Solution overview 6

Key results 8

Audience 9

We value your feedback 9

Business case

All websites, communications, networking, and complex IT infrastructures generate massive streams of machine data every day. This machine-generated data (digital exhaust, log files, time-series data, or whatever name you give it) holds critical information about user behavior, security risks, capacity consumption, service levels, fraudulent activity, customer experience, and more. As the fastest growing segment of Big Data, machine data is extremely valuable, but is often one of the most underused and undervalued assets of an organization.

Machine data is a key asset that enables organizations to gain operational intelligence for IT, security, and the business. Delivering the real-time insights and business value from machine data is necessary to drive digital transformation. Making use of this data presents real challenges, however.

Traditional data analysis, management, and monitoring solutions are simply not engineered for this high-volume, high-velocity, and highly diverse data. Machine data comes in an array of unpredictable formats that are difficult to process and analyze through methods designed for structured data with rigid schemas, such as business intelligence and data warehouse tools.

Business leaders seek to accelerate innovation while reducing costs and risk with optimized and validated solutions that power new or expanded operational intelligence capabilities. From fully built ready systems validated for specific use cases and delivered with full lifecycle support, to ready bundles and reference architectures that serve as starting points for your own custom-built solutions, you can count on Dell EMC™ and Splunk to help you deliver better outcomes.

Dell EMC and Splunk have partnered to provide a menu of standardized reference architectures for non-disruptive scalability and performance to aid an organization’s digital transformation. Together, Dell EMC and Splunk combine the analytics provided by the Splunk platform with the cost-effective, scalable, and flexible infrastructure of Dell EMC systems to deliver the right operational intelligence for your organization.

Solution overview

This validation guide shows how the Dell EMC Ready Bundle for Splunk built with Splunk Enterprise software, Dell EMC PowerEdge™ servers, and our Dell EMC Isilon™ scale-out network-attached storage (NAS) platform can easily, efficiently, and cost-effectively scale to support enterprise-level machine data analytics and real-time operational intelligence.

Splunk Enterprise is the industry-leading platform for generating business value from machine data. It gives you real-time visibility, insight, and understanding across your IT infrastructure and the application and services that run on top of it. The Splunk platform also:

- Seamlessly blends metrics and events from both structured and unstructured data sources
- Collects and correlates multiple data sources to rapidly pinpoint service degradations and reduce mean-time-to-resolution (MTTR)

- Monitors end-to-end infrastructure to detect anomalies and prevent problems in real time
- Delivers powerful visualizations to understand relationships, track trends, and accelerate investigations

As the foundation for a complete, adaptive Ready Bundle for Splunk solution, PowerEdge servers deliver superior agility and reliability, outstanding operational efficiencies, and top performance at any scale. The flexible, adaptable portfolio of PowerEdge servers can be used as modular building blocks to create an agile, future ready infrastructure. Versatile, powerful in-server storage allows you to accelerate performance of targeted applications with flexible configurations designed to maximize your data center efficiency. Dell EMC makes server innovations more affordable and accessible, putting more power into the hands of more people than ever before.

The Dell EMC Isilon OneFS™ operating system (OS) provides the intelligence behind all Isilon scale-out storage systems. It combines the three layers of traditional storage architectures—file system, volume manager, and data protection (DP) — into one unified software layer, creating a single intelligent file system that spans all nodes within a cluster. It meets the rapidly converging requirements of both Big Data and enterprise IT that support a fundamentally different way of meeting data storage needs in the future. Scale your capacity and performance as needed with unmatched efficiency, high utilization rates, and automated storage tiering.

The Dell EMC Ready Bundle for Splunk is rigorously tested with Splunk Enterprise software, and is designed based on extensive customer experience with real-world Splunk production installations. These solutions include all the hardware, software, resources, and services needed to deploy and manage Splunk Enterprise in a production environment. This validation guide describes the design, deployment, and configuration of the Ready Bundle for Splunk Enterprise for six representative use cases covering a range of customer needs, as described in Table 1.

Table 1. Six use cases of Splunk Enterprise on PowerEdge servers with Isilon storage

Use case	Description	Daily ingest (GB/day)	Retention (days)	Equipment
1	One PowerEdge R730xd Server "starter kit" meets the indexing and search needs of a single departmental environment.	250	115	1 PowerEdge R730xd server
2	Three PowerEdge servers distribute Splunk roles/instances to meet the indexing and search needs of a small enterprise.	250	115	3 PowerEdge servers • 1 search head • 1 indexer • 1 admin server
3	Four PowerEdge servers with Splunk Enterprise clustered deployment meet the indexing and search needs of a small enterprise with high availability requirements.	250	115	4 PowerEdge servers • 1 search head • 2 indexers • 1 admin server

Use case	Description	Daily ingest (GB/day)	Retention (days)	Equipment
4	Four PowerEdge servers with Splunk Enterprise clustered deployment meet the indexing and search needs of a small enterprise with high performance requirements for complex, low-latency query support.	250	115	4 PowerEdge servers • 1 search head • 2 indexers with more CPU and memory • 1 admin server
5	Deploying Isilon with a Splunk Enterprise clustered infrastructure on four PowerEdge servers provides configurable retention of Splunk cold bucket data.	250	210	4 PowerEdge servers • 1 search head • 2 indexers • 1 admin server • 3-node Isilon cluster
6	One PowerEdge R430 server meets the indexing and search needs of a single instance for development and testing	100	70	1 PowerEdge R430 server

Key results

Dell EMC and Splunk jointly validated a wide range of data ingestion rates and customer use cases that meet or exceed the performance of Splunk Enterprise running on Splunk's reference hardware. The configuration flexibility of Splunk Enterprise software together with the Ready Bundle options provides an integrated technology platform for analyzing machine-generated data.

This validation guide helps you understand the work that the Dell EMC and Splunk engineering teams performed to test and document the following best practices:

- PowerEdge servers configured with SSD disks and traditional HDD disks provide rapid read and write disk I/O and low latency for Splunk hot/warm bucket data and high capacity for Splunk cold bucket data.
- The Splunk distributed deployment is highly scalable.
- Splunk indexer offers high availability and scalability using the Splunk Cluster Master to automatically add additional PowerEdge servers.
- PowerEdge SSD direct-attached storage for Splunk hot/warm buckets with Isilon storage is used for long-term data retention of Splunk cold bucket data.
- Isilon Smart Pools, Smart Connect, and Smart Cache provide Splunk cold data storage and access.
- Linux configuration parameter settings provide optimal Splunk Enterprise performance.
- Splunk Enterprise can be expanded from a single instance to a clustered deployment.

Chapter 12 provides more background information and resources for additional research.

Audience

This guide is intended for IT administrators, storage administrators, virtualization administrators, system administrators, IT managers, and personnel 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: Eric Wang, Tao Guo, Phil Hummel, Reed Tucker

Chapter 2 Solution Architecture

This chapter presents the following topics:

Overview 11

Dell EMC PowerEdge Servers 13

Isilon storage 13

Splunk Enterprise 14

Overview

Reference architecture

This section describes the reference architecture of the Ready Bundle for Splunk with Isilon storage. Dell EMC and Splunk jointly validated this reference architecture to meet or exceed the performance of Splunk Enterprise running on Splunk's reference hardware.

Figure 1 shows the reference architecture of the Ready Bundle for Splunk with Isilon storage. This reference architecture provides rapid read and write disk I/O and low latency through the use of SSD disks for Splunk hot/warm bucket data and high capacity through the use of HDD disks or Isilon storage for Splunk cold bucket data. Splunk can be deployed in single instance mode, distributed mode, or indexer cluster mode in this reference architecture.

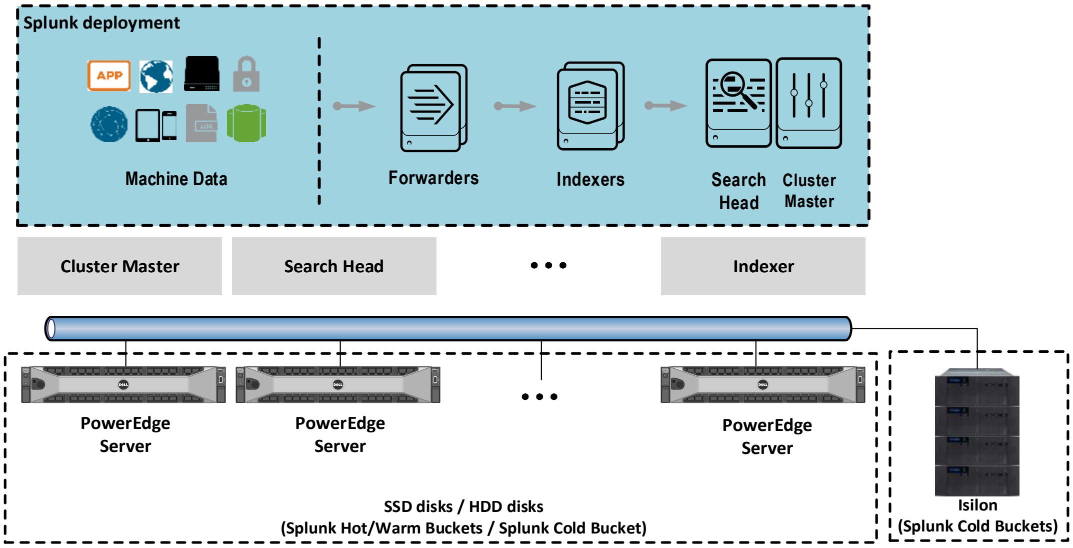


Figure 1. **Reference architecture of Dell EMC Ready Bundle for Splunk with Isilon storage**

Note: For an explanation of the hot/warm and cold bucket concepts, see Splunk core architecture.

Hardware components

Table 2 lists the hardware components referenced in this validation guide:

Table 2. Hardware configuration

Component	Hardware
PowerEdge R430 server	• 2 x Intel Xeon E5-2603 v4 12c • 32 GB RAM • 2 x 300 GB 10k RPM SAS HDD 2.5" Drives (OS) RAID 1 • 2 x 960 GB SATA SSD Mixed use 2.5" Drives RAID 1 • 4 x 1.8 TB 10k RPM SAS HDD 2.5" Drives RAID 10 • 1 x 1 GbE NIC • R430 8 2.5" Drive Bay Chassis

Component	Hardware
PowerEdge R630 server	• 2 Intel Xeon E5-2620 v4 16c • 64 GB RAM • 2 x 300 GB 10k RPM SAS HDD 2.5” Drives (OS) RAID 1 • 2 x 1GbE, 2x 10GbE NIC • R630 8 2.5" Drive Bay Chassis
PowerEdge R730xd server	• 2 Intel Xeon E5-2650 v4 24c • 64 GB RAM • 2 x 300 GB 10k RPM SAS HDD 2.5" Drives (OS) RAID 1 • 8 x 800 GB SATA SSD Mix use 2.5" Drives RAID 6 • 14 x 2 TB 7.2k RPM SAS HDD 2.5" Drives RAID 10 • 2 x 1GbE, 2 x 10GbE NIC • R730XD 24 2.5" Drive Bay Chassis
PowerEdge R930 server	• 2 Intel Xeon E7-8890 v4 48c • 128 GB RAM • 2 x 300 GB 10k RPM SAS HDD 2.5” Drives (OS) RAID 1 • 8 x 800 GB SATA SSD Mix use 2.5" Drives RAID 6 • 14 x 2 TB 7.2k RPM SAS HDD 2.5" Drives RAID 10 • 2 x 1GbE, 2 x 10GbE NIC • R930 24 2.5" Drive Bay Chassis
Network switch	• Dell Networking 10GbE • Dell Networking 1GbE
Isilon X210 storage server	• 1 Intel Xeon E5-2407 • 48 GB RAM per node • 4 x 800 GB SSD storage • 8 x 4 TB HDD storage • 2 x 10 GbE, 2 x 1 GbE per node

Software components

Table 3 lists the software versions referenced in this validation guide.

Table 3. Software configuration

Software	Version
Splunk Enterprise	6.5.0
Splunk Universal Forwarder	6.5.0
Red Hat Linux 64-bit	7.2
OneFS	8.0.0.3

Dell EMC PowerEdge Servers

PowerEdge servers are built to support the work of IT organizations. They are engineered to handle the most demanding business applications and are designed with specific features to better run workloads like High Performance Computing (HPC), collaboration, database, Enterprise Resource Planning (ERP), business intelligence, and data warehousing and data analytics.

As the foundation for a complete and adaptive Ready Bundle solution, PowerEdge servers deliver exceptional performance and management advantages that power the business applications that our customers run most.

Combined with the innovative Dell EMC OpenManage™ systems management portfolio and industry-leading workload solutions, PowerEdge servers provide technology that is intelligent, yet simple, giving you the power to do more in even the most complex environments.

The latest generation of servers responds to customer needs in the following areas:

- **Memory capacity and scalability**—Much larger memory footprints
- **Virtualization performance**—More processor cores and denser memory
- **System management**—Complete lifecycle management by using the integrated Dell Remote Access Controller (iDRAC) with Lifecycle Controller and monitoring and updating capabilities by using OpenManage Essentials
- **Energy efficiency**—Comprehensive optimizations, including Dell EMC OpenManage Power Center
- **Infrastructure flexibility**—Innovations like select network adapters, offering more and better I/O options
- **Reliability**—Additional Remote Access Service (RAS) features, including a failsafe hypervisor option on most servers

Isilon storage

The Isilon X-Series is a flexible and comprehensive storage product that provides large capacity and high performance. The Ready Bundle for Splunk supports 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 OS, offers a single-clustered file system.

OneFS provides value by incorporating parallelism at a deep level in 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 multipetabyte scale while providing greater reliability than traditional systems.

OneFS runs on Isilon scale-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.

Splunk Enterprise

Splunk Enterprise is a software platform that enables you to collect, index, and visualize machine-generated data gathered from different sources in your IT infrastructure. These sources include applications, networking devices, host and server logs, mobile devices, and more.

Splunk turns silos of data into operational insights and provides end-to-end visibility across your IT infrastructure to enable faster problem solving and informed, data-driven decisions.

Splunk core architecture

Figure 2 provides a graphic overview of Splunk system architecture. A Splunk Enterprise instance can perform the role of a search head, an indexer, or both for small deployments. When 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, see the [Splunk Capacity Planning Manual](#).

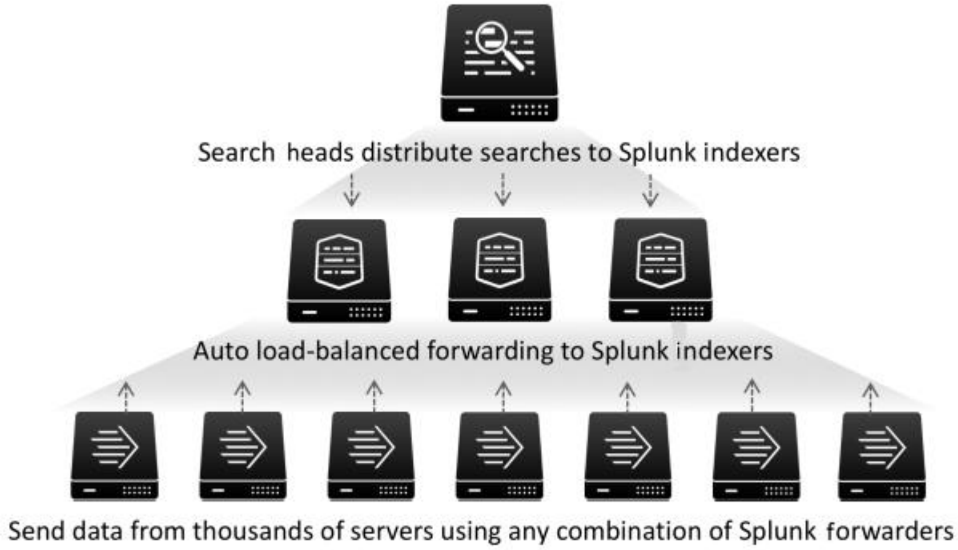

Search heads distribute searches to Splunk indexers
Auto load-balanced forwarding to Splunk indexers
Send data from thousands of servers using any combination of Splunk forwarders

Figure 2. **Splunk architecture overview**

When a Splunk Enterprise indexer receives data, 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 tier involving hot/warm and cold buckets of data to optimize performance for newly indexed data and to 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 when 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. It is important that the storage that is 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 (Network File System) mount if the latency is less than 5 ms (ideally) and not more than 200 ms. 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 by Splunk search heads. Manual user action is required to bring the data back into Splunk Enterprise buckets to be searchable. While you might choose to use frozen buckets to meet compliance retention requirements, this paper shows how Isilon's massive scalability and competitive cost of ownership can empower you to retain more data in the cold bucket, where it remains searchable. Figure 3 provides more information about Splunk bucket concepts.

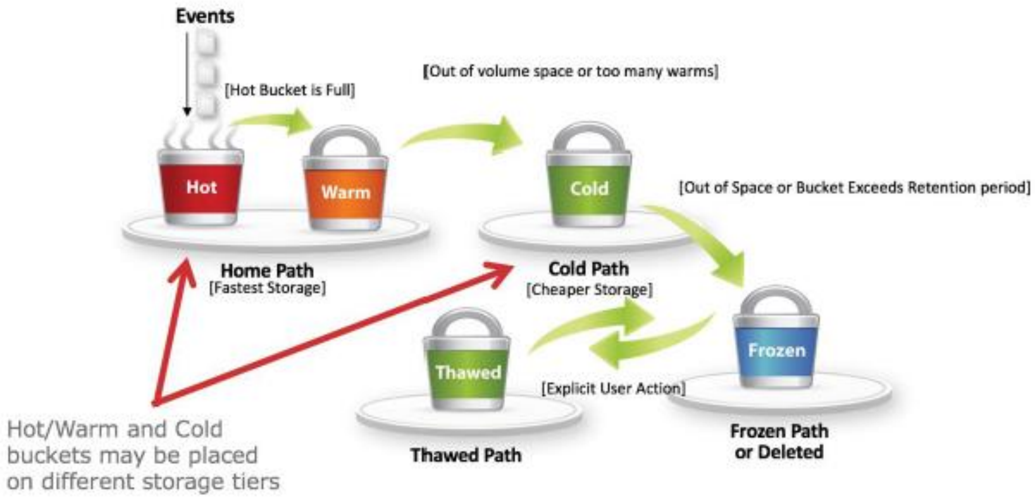


Figure 3. **Splunk index buckets**

Chapter 3 Splunk Enterprise Design and Configuration

This chapter presents the following topics:

Overview 17

Compute design 17

Network design 18

Storage design 19

Splunk Enterprise design 21

Overview

This chapter provides details about the design, and configuration of the Ready Bundle for Splunk with Isilon storage. This validation guide covers six use cases:

- Use Case 1: Single instance deployment for 250 GB/day indexing volume, for departmental use
- Use Case 2: Distributed deployment for 250 GB/day indexing volume, for small enterprises and customers interested in a distributed configuration
- Use Case 3: Clustered deployment for 250 GB/day indexing volume, for small enterprises interested in data loss protection
- Use Case 4: High performance clustered deployment for 250 GB/day indexing volume, for small enterprises requiring high-data availability, and complex, low-latency query support
- Use Case 5: Isilon cold bucket expansion for 250 GB/day indexing volume, for cold-bucket retention
- Use Case 6: Single instance deployment for development and testing for 100 GB/day indexing volume

Compute design

Table 4, Table 5, Table 6, Table 7, Table 8, and Table 9 show the details of the compute design of the six use cases.

Table 4. Compute design of Use Case 1

Instance role	Quantity	PowerEdge model	Physical cores	Memory
Single Instance combined search head and indexer	1	PowerEdge R730xd	24	64 GB

Table 5. Compute design of Use Case 2

Instance role	Quantity	PowerEdge model	Physical cores	Memory
Indexer	1	PowerEdge R730xd	24	64 GB
Search head	1	PowerEdge R630	16	64 GB
Admin server	1	PowerEdge R630	16	64 GB

Table 6. Compute design of Use Case 3

Instance role	Quantity	PowerEdge model	Physical cores	Memory
Indexer	2	PowerEdge R730xd	24	64 GB
Search head	1	PowerEdge R630	16	64 GB
Admin server	1	PowerEdge R630	16	64 GB

Table 7. Compute design of Use Case 4

Instance role	Quantity	PowerEdge model	Physical cores	Memory
Indexer	2	PowerEdge R930	48	128 GB
Search head	1	PowerEdge R630	16	64 GB
Admin server	1	PowerEdge R630	16	64 GB

Table 8. Compute design of Use Case 5

Instance role	Quantity	PowerEdge model	Physical cores	Memory
Indexer	2	PowerEdge R730xd	24	64 GB
Search head	1	PowerEdge R630	16	64 GB
Admin server	1	PowerEdge R630	16	64 GB

Table 9. Compute design of Use Case 6

Instance role	Quantity	PowerEdge model	Physical cores	Memory
Single Instance combined search head and indexer	1	PowerEdge R430	12	32 GB

Network design

For the reference architecture documented in this validation guide, we designed the network as follows:

- 1 GbE PowerEdge System Management (iDRAC) network
- 10 GbE Splunk Enterprise network

As a best practice, Dell EMC recommends using dual Top-of-the-Rack (ToR) switches to eliminate the switch as a single point of failure. Figure 4 shows the PowerEdge server network design.

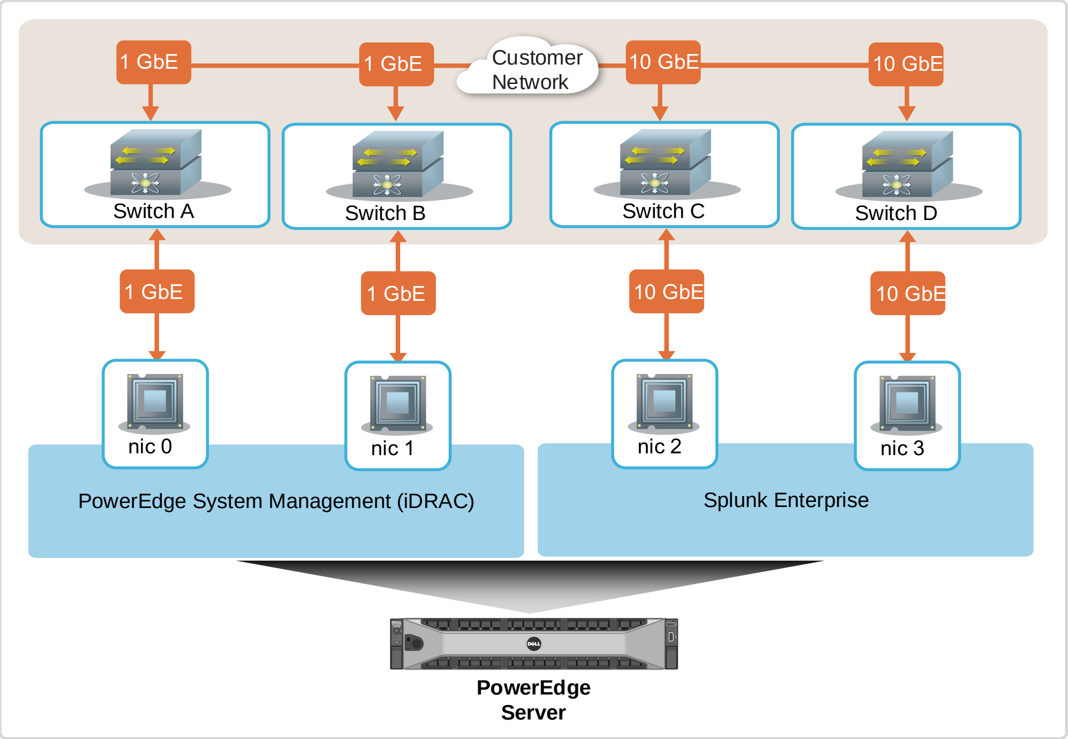


Figure 4. **Dell EMC PowerEdge server network design**

Storage design

PowerEdge storage design

Table 10 shows the PowerEdge storage design for the six use cases.

Table 10. PowerEdge storage design for Splunk

Use Case	Instance role	Quantity	OS storage	Hot/warm bucket storage	Cold bucket storage
Use Case 1	Single instance combined search head and indexer	1	300 GB	4,800 GB	14 TB
Use Case 2	Indexer	1	300 GB	4,800 GB	14 TB
	Search head	1	300 GB	0	0
	Admin server	1	300 GB	0	0
Use Case 3	Indexer	2	300 GB	4,800 GB	14 TB
	Search head	1	300 GB	0	0
	Admin server	1	300 GB	0	0

Use Case	Instance role	Quantity	OS storage	Hot/warm bucket storage	Cold bucket storage
Use Case 4	Indexer	2	300 GB	4,800 GB	14 TB
	Search head	1	300 GB	0	0
	Admin server	1	300 GB	0	0
Use Case 5	Indexer	2	300 GB	4,800 GB	0
	Search head	1	300 GB	0	0
	Admin server	1	300 GB	0	0
Use Case 6	Single instance combined search head and indexer	1	300 GB	960 GB	3.6 TB

Isilon storage design

In this validation guide, a three-node Isilon X210 cluster is used for Use Case 5 to provide configurable retention for cold buckets. Table 11 and Table 12 show the detailed configurations of Isilon nodes and Isilon storage design for Splunk Enterprise.

Table 11. Isilon node configuration

CPU	CPU cores	RAM	SSD capacity	HDD capacity	Network
1 Intel Xeon E5-2407 2.4 GHz	4 cores	48 GB	3.2 TB	32 TB	2 x 10 GbE 2 x 1 GbE

Table 12. Isilon storage design for Splunk Enterprise for Use Case 5

Use Case	Instance role	Quantity	Indexer cold bucket storage
Use Case 5	Indexer	2	63 TB

For the overall Isilon configuration, we followed these best practices:

- Enable SmartPools settings across all Isilon nodes and use an SSD as L3 cache for random read acceleration
- Enable SmartConnect to provide automatic client connection load balancing and failover capabilities
- Enable SmartCache for write performance
- Use optimization for concurrent data access pattern
- Use a 10 Gb/s external network for data connection
- Increase network MTU to 9000 (Jumbo Frames)

Splunk and Dell EMC recommend that NFS storage, including Isilon storage, is only used for cold and frozen data, never for hot/warm data. For information about system requirements, see the *Splunk Enterprise Installation Manual*.

Splunk Enterprise design

Splunk Enterprise deployment design

Splunk Enterprise supports three types of deployment: single instance deployment, distributed deployment, and clustered deployment.

Figure 5 shows the Splunk Enterprise single instance deployment that combines the indexer and search head, as in Use Case 1.

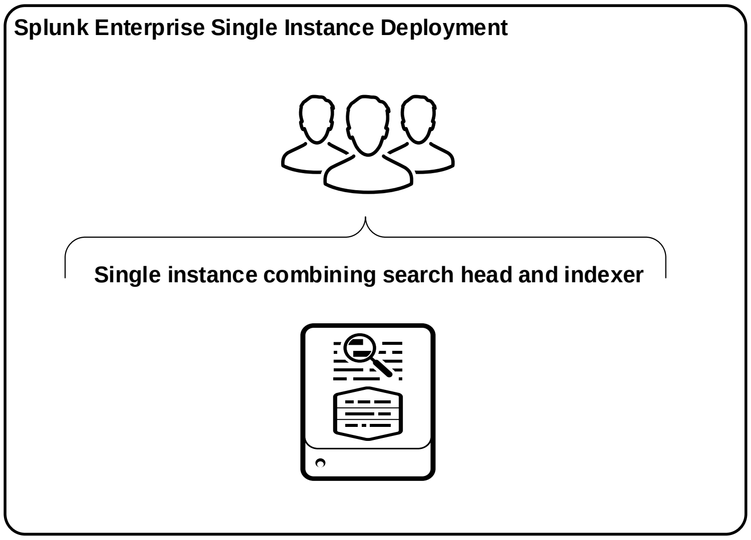

Splunk Enterprise Single Instance Deployment
Single instance combining search head and indexer

Figure 5. **Splunk Enterprise single instance deployment**

Figure 6 shows the Splunk Enterprise distributed deployment with one search head, one indexer, and one master node (admin server), as in Use Case 2.

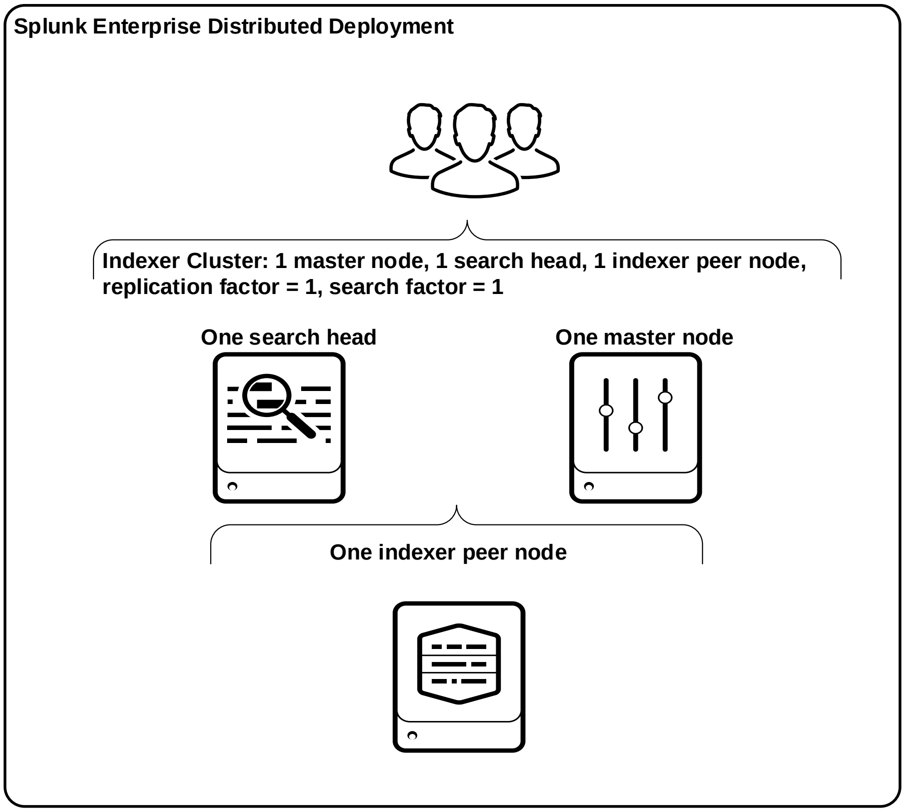

Splunk Enterprise Distributed Deployment
Indexer Cluster: 1 master node, 1 search head, 1 indexer peer node, replication factor = 1, search factor = 1
One search head
One master node
One indexer peer node

Figure 6. **Splunk Enterprise distributed deployment**

Figure 7 shows the Splunk Enterprise clustered deployment with one search head, two indexers, and one master node (admin server). This type of deployment targets clients requiring high data availability, as in Use Cases 3, 4, and 5.

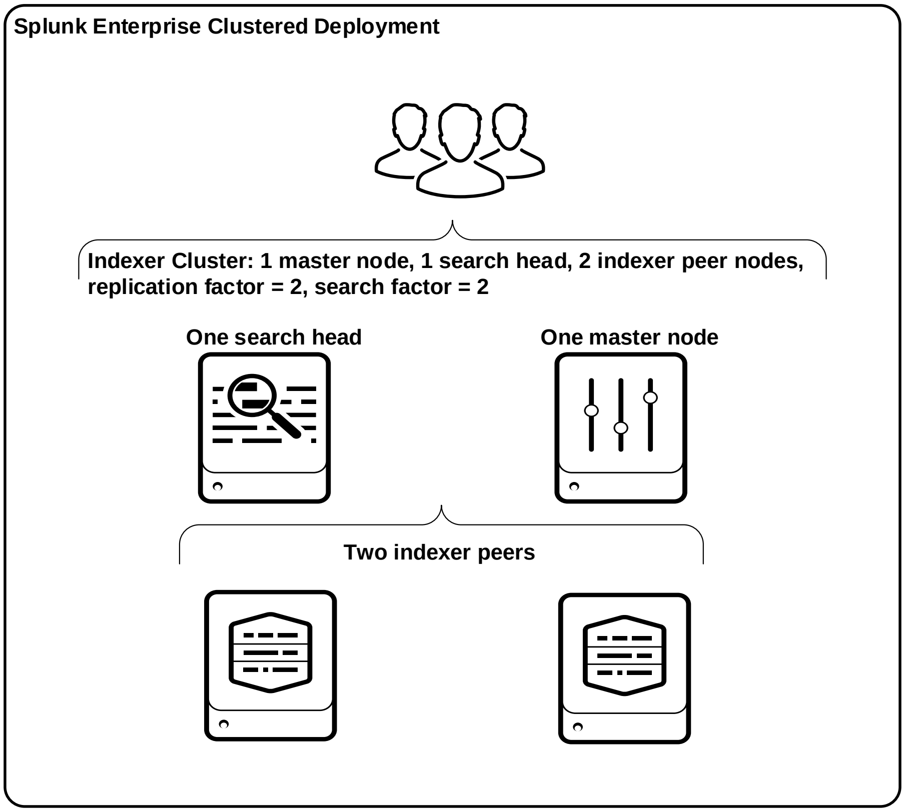

Splunk Enterprise Clustered Deployment
Indexer Cluster: 1 master node, 1 search head, 2 indexer peer nodes, replication factor = 2, search factor = 2
One search head
One master node
Two indexer peers

Figure 7. **Splunk Enterprise clustered deployment**

Splunk Enterprise Linux configuration

In this validation guide, we implement the following Linux configuration parameter settings to provide optimal Splunk Enterprise performance for all use cases:

- Disable Transparent Huge Pages (THP) to avoid the degradation of Splunk Enterprise performance on Red Hat Enterprise Linux. For more information, ~~refer to~~see *Transparent huge memory pages and Splunk performance.*
- Disable SELinux, so that enhanced system security does not add overhead to the performance.

Increase the maximum number of open file descriptors and processes by configuring ulimit values to avoid the “Too Many Open Files” exception. Table 13 shows the recommended minimum values. For more information, see *System requirements for use of Splunk Enterprise on-premises*.

Table 13. Recommended ulimit values

System-wide resources	Ulimit invocation	Recommended minimum value
Open files	ulimit -n	8,192
User processes	ulimit -u	1,024
Data segment size	ulimit -d	1,073,741,824

- Tune the kernel to optimize the network for high throughput over a 10 Gb Ethernet by adding the following command string to `/etc/sysctl.conf`:

```
net.core.rmem_max = 67108864
net.core.wmem_max = 67108864
net.ipv4.tcp_rmem = 4096 87380 33554432
net.ipv4.tcp_wmem = 4096 65536 33554432
net.ipv4.tcp_congestion_control=htcp
net.core.default_qdisc = fq
```

Chapter 4 Splunk Enterprise Single Instance Infrastructure for 250 GB/day Indexing Volume

This chapter presents the following topics:

Overview..26

Implementation 26

Use case summary..35

Overview

This chapter describes the implementation procedure for Use Case 1, a Splunk Enterprise single instance serving as both an indexer and a search head on one PowerEdge R730xd server. This use case, which can index up to 250 GB/day indexing volume with 115-day retention, usually meets the indexing and search needs of a single department.

Implementation

Table 14 lists the implementation process flow of Use Case 1.

Table 14. Implementation process flow of Use Case 1

Step	Action	Description
1	Preparing the PowerEdge server	Prepare PowerEdge servers for use.
2	Configuring storage	Configure disk storage on PowerEdge servers for the OS and Splunk index data.
3	Installing the OS	Install the OS on PowerEdge servers and configure it for Splunk Enterprise installation and operations.
4	Installing Splunk Enterprise	Install the Splunk Enterprise instance and perform initial configuration tasks.
5	Configuring Splunk Enterprise indexer	Configure the location and data size of the Splunk index bucket and enable the receiver to receive data.
6	Validating the implementation	Validate the implementation of Splunk Enterprise.

Preparing the PowerEdge server

One Dell EMC PowerEdge R730xd server is deployed in this use case. Before starting the implementation, ensure that the PowerEdge servers are racked, cabled, powered, and ready for use. For information about the PowerEdge R730xd server, see *Dell EMC PowerEdge R730xd Manuals & Documents*.

Configuring storage

When the PowerEdge servers are ready for use, configure the storage on the servers for OS and Splunk index data. Table 15 lists the virtual disks configuration on PowerEdge servers.

Table 15. Virtual disks configuration of Use Case 1

Server role	Storage	Virtual disk name	RAID level	Physical disks
Single instance combined search head and indexer	OS	RAID1_HDD	RAID-1	2 X 300 GB SAS HDD
	Splunk hot/warm bucket	RAID6_SSD	RAID-6	8 X 800 GB SATA SSD
	Splunk cold bucket	RAID10_HDD	RAID-10	14 X 2 TB SAS HDD

There are multiple methods to create the virtual disks on a PowerEdge server. The procedure described in this validation guide uses the iDRAC console.

Creating virtual disk for OS

1. Log in to the iDRAC web interface and go to **Overview > Storage > Virtual Disks > Create.** The **Create Virtual Disk** page is displayed.
2. In the **Settings** section, fill in the required information as follows:

   ```
   Name: RAID1_HDD
   Controller: PERC H730P Mini (Embedded)
   Layout: RAID-1
   Media Type: HDD
   Stripe Element Size: 64KB
   Capacity: Leave the default value
   Read Policy: Adaptive Read Ahead
   Write Policy: Write Back
   Disk Cache Policy: Default
   T10 PI Capacity: Disabled
   Span Count: Leave the default value
   ```
3. In the **Select Physical Disks** section, select the required disks as shown:

   ```
   Physical Disks 0:1:0
   Physical Disks 0:1:1
   ```
4. From the **Apply Operation Mode** drop-down list, select **Apply Now**, and click **Create Virtual Disk** to apply the settings.
5. Go to **Overview > Server > Job Queue** to ensure the “create virtual disk” job is completed successfully.

Creating virtual disk for Splunk hot/warm bucket

1. Log in to the iDRAC web interface and go to **Overview > Storage > Virtual Disks > Create.** The **Create Virtual Disk** page is displayed.
2. In the **Settings** section, fill in the required information as follows:

   ```
   Name: RAID6_SSD
   Controller: PERC H730P Mini (Embedded)
   Layout: RAID-6
   Media Type: SSD
   Stripe Element Size: 64KB
   Capacity: Leave the default value
   Read Policy: Adaptive Read Ahead
   Write Policy: Write Back
   Disk Cache Policy: Default
   T10 PI Capacity: Disabled
   Span Count: Leave the default value
   ```
3. In the **Select Physical Disks** section, select the required disks as shown:

   ```
   Solid State Disks 0:1:2
   ```

```
Solid State Disks 0:1:3
Solid State Disks 0:1:4
Solid State Disks 0:1:5
Solid State Disks 0:1:6
Solid State Disks 0:1:7
Solid State Disks 0:1:8
Solid State Disks 0:1:9
```

4. From the **Apply Operation Mode** drop-down list, select **Apply Now**, and click **Create Virtual Disk** to apply the settings.

5. Go to **Overview > Server > Job Queue** to ensure virtual disk creation is completed successfully.

Creating virtual disk for Splunk cold bucket

1. Log in to the iDRAC web interface and go to **Overview > Storage > Virtual Disks > Create**. The **Create Virtual Disk** page is displayed.

2. In the **Settings** section, fill in the required information as shown:

```
Name: RAID10_HDD
Controller: PERC H730P Mini (Embedded)
Layout: RAID-10
Media Type: HDD
Stripe Element Size: 64KB
Capacity: Leave the default value
Read Policy: Adaptive Read Ahead
Write Policy: Write Back
Disk Cache Policy: Default
T10 PI Capacity: Disabled
Span Count: Leave the default value
```

3. In the **Select Physical Disks** section, select the required disks as shown:

```
Physical Disks 0:1:10
Physical Disks 0:1:11
Physical Disks 0:1:12
Physical Disks 0:1:13
Physical Disks 0:1:14
Physical Disks 0:1:15
Physical Disks 0:1:16
Physical Disks 0:1:17
Physical Disks 0:1:18
Physical Disks 0:1:19
Physical Disks 0:1:20
Physical Disks 0:1:21
Physical Disks 0:1:22
Physical Disks 0:1:23
```

4. From the **Apply Operation Mode** drop-down list, select **Apply Now**, and click **Create Virtual Disk** to apply the settings.
5. Go to **Overview > Server > Job Queue** to ensure the virtual disk is created successfully.

Installing the OS

Install the OS on the PowerEdge servers. This use case uses Red Hat Enterprise Linux 7.2 as the OS for Splunk Enterprise. The procedure described in this validation guide uses the virtual console in iDRAC web interface method.

Installing Red Hat Enterprise Linux 7.2

1. Log in to the iDRAC web interface and go to **Overview > Server > Virtual Console > Launch Virtual Console.** The **Virtual Console** is displayed.
2. Click **Launch Virtual Console** to connect the console of the server.
3. Click **Virtual Media > Connect Virtual Media** to activate the virtual media function.
4. Click **Virtual Media > Map CD/DVD** to open the image ISO file selection prompt.
5. Click **Browse** and select the **Red Hat Enterprise Linux 7.2 image ISO file.**
6. Click **Map Device** to mount the ISO image file.
7. Click **Next Boot > Virtual CD/DVD/ISO** to allow the server to boot from the virtual image.
8. Click **Power** and Choose **Reset System (warm boot).**

   After a restart, the server starts the Red Hat Enterprise Linux installation automatically.
9. Follow the installation wizard to install Red Hat Enterprise Linux 7.2 on the PowerEdge server. For more details about the installation, see the *Red Hat Enterprise Linux 7 Installation Guide*.

Performing post OS installation configuration

After installing Red Hat Enterprise Linux 7.2, perform the following steps to make the system ready for Splunk Enterprise installation and to optimize performance:

1. Log in to the iDRAC web interface and go to **Overview > Server > Virtual Console > Launch Virtual Console.** The **Virtual Console** is displayed.
2. At the Red Hat Enterprise Linux login prompt, type the root account username and password to log in to the system.
3. Edit `/etc/sysconfig/network-scripts/ifcfg-<`*`Network Interface Card Name`*`>` to configure the IP address.
4. Edit `/etc/hostname` to configure the hostname.
5. Edit `/etc/resolv.conf` to configure the name server and search list.
6. Edit `/etc/sysctl.conf` and add these lines to tune the network for the 10 GbE network interface card (NIC) for Red Hat Enterprise Linux 7.X:

   ```
   net.core.rmem_max = 67108864
   ```

```
net.core.wmem_max = 67108864
net.ipv4.tcp_rmem = 4096 87380 33554432
net.ipv4.tcp_wmem = 4096 65536 33554432
net.ipv4.tcp_congestion_control=htcp
net.core.default_qdisc = fq
```

7. Stop and disable the firewall daemon:

```
systemctl stop firewalld
systemctl disable firewalld
```

8. Disable SELinux in the configuration file `/etc/selinux/config`:

```
SELINUX=disabled
```

9. Disable the tuned service:

```
systemctl stop tuned
systemctl disable tuned
```

10. Disable THP by appending the "**transparent_hugepage=never**" kernel parameter on the **GRUB_CMDLINE_LINUX** option in the `/etc/sysconfig/grub` file and run the `grub2-mkconfig` command to regenerate the `grub.cfg` file:

```
grub2-mkconfig -o /boot/grub2/grub.cfg
```

11. Create a configuration file `/etc/security/limits.d/99-splunk.conf` with these lines to set the limitation of the maximum file descriptors for the Splunk user:

```
root        -       nofile    65535
root        -       nproc     65535
splunk      -       nofile    65535
splunk      -       nproc     65535
```

12. Restart the system.

Installing Splunk Enterprise

To install Splunk Enterprise on the PowerEdge servers, follow these steps:

1. Log in to the PowerEdge server system as the root account with SSH.
2. Upload the Splunk Enterprise rpm package to directory `/root/`.
3. (Optional) Create the Splunk user and group. By default, the Splunk user and group is created automatically during Splunk rpm package installation in the next step. But, if an Isilon cluster is used for the Splunk index cold bucket data, the user ID and group ID of Splunk user must be consistent across all indexers. To ensure that the user ID and group ID are consistent, create the Splunk user and group manually with these commands:

```
groupadd splunk -g <splunk group ID>
useradd splunk -u <splunk user ID> -g splunk -d /opt/splunk
```

Note: Linux user ID and group ID below 1000 are reserved for system accounts.

4. Install the Splunk Enterprise rpm package.

```
rpm -ivh splunk-6.5.0-*.rpm
```

5. Create a file `/etc/profile.d/splunk.sh` with these lines to set Splunk Enterprise environment variables at shell startup.

```
export SPLUNK_HOME=/opt/splunk
export PATH=${PATH}:${SPLUNK_HOME}/bin
```

6. Initialize Splunk Enterprise environment variables in the current session.

```
source /etc/profile.d/splunk.sh
```

7. Enable Splunk Enterprise to start using the Splunk account when the system starts.

```
splunk enable boot-start --accept-license -user splunk
```

8. Run this command to set service level limits for the systemd service manager in Red Hat Enterprise Linux 7.X:

```
mkdir -p /etc/systemd/system/splunk.service.d/
cat >/etc/systemd/system/splunk.service.d/filelimit.conf
<<EOF
[Service]
LimitNOFILE=65535
LimitNPROC=65535
EOF
systemctl daemon-reload
```

Note: The systemd service manager in Red Hat Enterprise Linux 7.X ignores the limits set in `/etc/security/limits.conf` or `/etc/security/limits.d/*.conf`.

Performing initial configuration tasks

1. Log in to the PowerEdge server system as the root account with SSH.

2. Start the Splunk Enterprise instance.

```
service splunk start
```

3. Upload a license file to `/opt/splunk/etc/licenses/enterprise/enterprise.lic` and install it.

```
sudo -H -u splunk $SPLUNK_HOME/bin/splunk add licenses
/opt/splunk/etc/licenses/enterprise/enterprise.lic -auth
admin:changeme
```

4. Change the default password of the Splunk admin account.

```
sudo -H -u splunk $SPLUNK_HOME/bin/splunk edit user admin -
password 'password' -role admin -auth admin:changeme
```

5. Delete the change password command from the Linux shell history to avoid password disclosure.

```
history
history -d <sequence of splunk change password command>
history -w
```

Configuring Splunk Enterprise indexer

In a Splunk Enterprise single instance deployment, the indexer and search head are combined on a single instance. The following sections provide detailed procedures for configuring a Splunk Enterprise indexer.

Changing Splunk index location

1. Log in to the Splunk Enterprise indexer as the root account with SSH.
2. Partition and format the two virtual disks used for Splunk index hot/warm and cold buckets.

   ```
   parted -a optimal /dev/sdb --script mklabel gpt mkpart
   primary 2048s 100%
   mkfs.xfs -f /dev/sdb1
   parted -a optimal /dev/sdc --script mklabel gpt mkpart
   primary 2048s 100%
   mkfs.xfs -f /dev/sdc1
   ```

3. Mount the two disk partitions into the file system and change the mount point permission.

   ```
   mkdir -p /data/splunk/homedb
   mkdir -p /data/splunk/colddb
   mount /dev/sdb1 /data/splunk/homedb
   mount /dev/sdc1 /data/splunk/colddb
   chmod 777 /data
   chown -R splunk:splunk /data/splunk
   chmod -R 750 /data/splunk
   ```

4. Add the two disk partitions in the file systems table **/etc/fstab** to ensure that they can be mounted automatically on system startup.

   ```
   /dev/sdb1           /data/splunk/homedb      xfs
   defaults   0 0
   /dev/sdc1           /data/splunk/colddb      xfs
   defaults   0 0
   ```

5. Create a new configuration file `$SPLUNK_HOME/etc/system/local/indexes.conf` with these lines to point the index bucket to the new location and set the maximum size of each index storage:

   ```
   [main]
   homePath = /data/splunk/homedb/$_index_name
   coldPath = /data/splunk/colddb/$_index_name
   homePath.maxDataSizeMB = 3840000
   coldPath.maxDataSizeMB = 11200000
   maxWarmDBCount = 4294967295
   ```

6. Change the owner and permission of this configuration file and restart Splunk Enterprise Instance.

   ```
   chown splunk:splunk
   $SPLUNK_HOME/etc/system/local/indexes.conf
   chmod 600 $SPLUNK_HOME/etc/system/local/indexes.conf
   ```

```
service splunk restart
```

Enabling a receiver

1. Log in to the Splunk Enterprise indexer as the root account with SSH.
2. Enable a receiver.

```
sudo -H -u splunk $SPLUNK_HOME/bin/splunk enable listen
9997 -auth admin:password
```

Validating the implementation

Validate the Splunk Enterprise instance to ensure it is ready to run. This process requires installing Splunk Universal Forwarder on one machine and indexing some testing data for the validation. The following sections provide detailed procedures for validating the Splunk Enterprise implementation.

Installing Splunk Universal Forwarder

1. Install Red Hat Enterprise Linux 7.2 on one physical or virtual machine.
2. Log in to that machine's system as the root account with SSH.
3. Upload the Splunk Universal Forwarder rpm package to directory `/root/`.
4. Install the Splunk Universal Forwarder rpm package.

```
rpm -ivh splunkforwarder-6.5.0-*.rpm
```

5. Create a file `/etc/profile.d/splunk.sh` with these lines to set Splunk Enterprise environment variables at shell startup~~.~~:

```
export SPLUNK_HOME=/opt/splunkforwarder
export PATH=${PATH}:${SPLUNK_HOME}/bin
```

6. Initialize Splunk Enterprise environment variables in the current session.

```
source /etc/profile.d/splunk.sh
```

7. Enable Splunk Universal Forwarder to start using the Splunk account when the system starts.

```
splunk enable boot-start --accept-license -user splunk
```

8. Run these commands to set service level limits for the systemd service manager in Red Hat Enterprise Linux 7.X:

```
mkdir -p /etc/systemd/system/splunk.service.d/
cat >/etc/systemd/system/splunk.service.d/filelimit.conf
<<EOF
[Service]
LimitNOFILE=65535
LimitNPROC=65535
EOF
systemctl daemon-reload
```

Note: The systemd service manager in Red Hat Enterprise Linux 7.X ignores limits set in `/etc/security/limits.conf` or `/etc/security/limits.d/*.conf.`

9. Start Splunk Universal Forwarder.

```
service splunk start
```

10. Change the default password of the Splunk admin account.

    ```
    sudo -H -u splunk $SPLUNK_HOME/bin/splunk edit user admin -
    password 'password' -role admin -auth admin:changeme
    ```

11. Delete the `change password` command from the Linux shell history to avoid password disclosure.

    ```
    history
    history -d <sequence of splunk change password command>
    history -w
    ```

Forwarding data to the indexer

1. Log in to the Splunk Universal Forwarder (installed in previous step) as the root account with SSH.

2. Create a monitor directory `/data/forwarder` and upload Splunk tutorial data `tutorialdata.zip` to this directory.

    Note: You can download the `tutorialdata.zip` file from the Splunk website.

3. Change the file owner and permission of the `/data/forwarder/tutorialdata.zip` file.

    ```
    chmod 777 /data
    chown -R splunk:splunk /data/forwarder
    chmod -R 750 /data/forwarder
    ```

4. Configure the Splunk universal forwarder to connect to a Splunk Enterprise indexer.

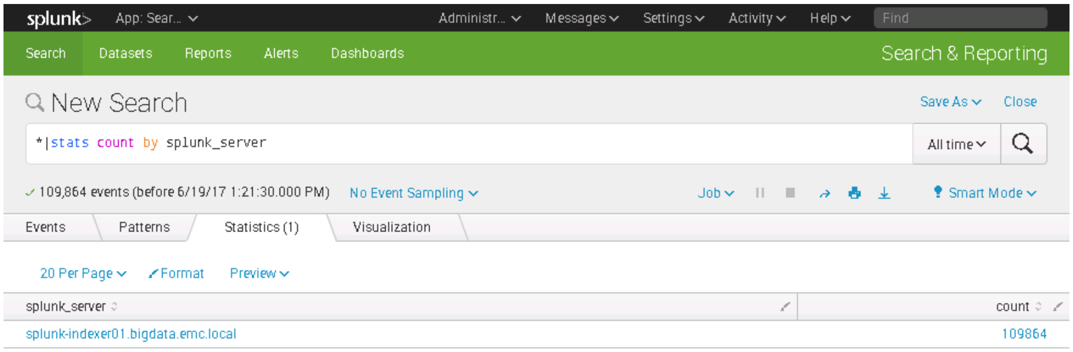
    ```
    sudo -H -u splunk $SPLUNK_HOME/bin/splunk add forward-
    server <server name of the splunk single instance>:9997
    ```

5. Configure data input on the Splunk Universal Forwarder.

    ```
    sudo -H -u splunk $SPLUNK_HOME/bin/splunk add monitor
    /data/forwarder -auth admin:password
    ```

Validating Splunk Enterprise implementation

1. Log in to the Splunk Enterprise search head web interface **http://*<server name of splunk single instance>*:8000/** to view the indexed data Splunk indexes all the tutorial data, as shown in Figure 8.



```
*|stats count by splunk_server
```

✓ 109,864 events (before 6/19/17 1:21:30.000 PM)

splunk_server	count
splunk-indexer01.bigdata.emc.local	109864

Figure 8. **Validation result of Splunk Enterprise in Use Case 1**

Removing tutorial data

1. Log in to Splunk Enterprise indexer as the root account with SSH.
2. Stop Splunk service, remove the tutorial data, and restart Splunk service.

```
service splunk stop
sudo -H -u splunk $SPLUNK_HOME/bin/splunk clean eventdata -index main
service splunk start
```

Important: This step removes all indexed data from the indexer. Do not perform this step if you have production data in the indexer.

Use case summary

In this use case, we implemented a Splunk Enterprise single instance on one Dell EMC PowerEdge R730xd server that combined a Splunk indexer and search head into a single instance. This use case, which can index up to 250 GB/day indexing volume with 115-day retention, usually meets the indexing and search needs of a single department. This use case shows that a single PowerEdge R730xd server can be a starter kit for a Splunk Enterprise departmental environment.

Chapter 5 Splunk Enterprise Distributed Infrastructure for 250 GB/day Indexing Volume

This chapter presents the following topics:

Overview..37

Implementation 37

Use case summary..41

Overview

This chapter describes the implementation procedure for Use Case 2 which can index up to 250 GB/day indexing volume with 115-day retention and usually meets the indexing and search needs of a small enterprise. It provides detailed instructions for expanding Splunk Enterprise from a single instance to a distributed deployment. The procedure includes separating the search management function from a Splunk Enterprise single instance by adding one dedicated search head, and adding one admin server as the master node to manage the indexers or search peers.

Implementation

Table 16 lists the implementation process flow of Use Case 2.

Table 16. Implementation process flow of Use Case 2

Step	Action	Description
1	Preparing PowerEdge servers	Prepare PowerEdge servers for use.
2	Configuring storage	Configure disk storage on PowerEdge servers for the OS and Splunk index data.
3	Installing the OS	Install the OS on PowerEdge servers and configure it for Splunk Enterprise installation and operation.
4	Installing Splunk Enterprise	Install Splunk Enterprise instance and perform initial configuration tasks.
5	Performing migration	Migrate the existing set of knowledge objects and applications from the former instance to the new search head.
6	Configuring Splunk Enterprise Indexer	Configure the location and data size of the Splunk index bucket and enable the receiver to receive data.
7	Configuring Splunk Enterprise admin server	Enable the master node and index discovery on Splunk Enterprise admin server.
8	Configuring Splunk Enterprise search peers	Enable the peer node on Splunk Enterprise indexers.
9	Configuring Splunk Enterprise search head	Enable the search head on Splunk Enterprise search head.
10	Validating implementation	Validate the implementation of Splunk Enterprise.

Preparing PowerEdge servers

Two new PowerEdge servers are added to expand Splunk Enterprise from a single instance to a distributed deployment. The former PowerEdge R730xd server takes the indexer role in the distributed deployment. One newly added PowerEdge R630 server takes the search head role and another newly added PowerEdge R630 server takes the admin server role.

Before starting the implementation, ensure that these newly added PowerEdge servers are racked, cabled, powered, and ready for use. For detailed instructions about the installation of a PowerEdge R730xd server, see *Dell EMC PowerEdge R730xd Manuals & Documents*. For detailed instructions about the installation of a PowerEdge R630 server, see *Dell EMC PowerEdge R630 Manuals & Documents*.

Configuring storage

After the newly added PowerEdge servers are set up correctly and ready for use, configure storage on the servers for the OS and Splunk index data. Table 17 lists the virtual disks configuration on PowerEdge servers.

Table 17. Virtual disks configuration of Use Case 2

Server role	Storage	Virtual disk name	RAID level	Physical disks
Indexer, Search Head, Admin Server	OS	RAID1_HDD	RAID-1	2 x 300 GB SAS HDD
Indexer	Splunk hot/warm bucket	RAID6_SSD	RAID-6	8 x 800 GB SATA SSD
Indexer	Splunk cold bucket	RAID10_HDD	RAID-10	14 x 2 TB SAS HDD

For details about how to configure storage on these newly added PowerEdge servers, see Configuring Storage in Chapter 4.

Installing the OS

Next, install the OS on these newly added PowerEdge servers. For detailed procedures, see Installing the OS in Chapter 4.

Installing Splunk Enterprise

Next, install Splunk Enterprise on these newly added PowerEdge servers. For detailed procedures, see Installing Splunk Enterprise in Chapter 4.

Performing migration

Migrate from a Splunk Enterprise single instance to a distributed deployment. As the distributed deployment is expanded from a standalone instance, the former standalone instance continues to take the indexer role. The data already indexed on the former standalone instance continues to be available in the new environment. Therefore, there is no need to migrate the data. The only remaining task is to migrate the existing set of knowledge objects and applications from the former standalone instance to the newly added search head instance. Follow these steps:

1. Log in to the former Splunk standalone instance as the root account with SSH.
2. Copy the set of knowledge objects to the new search head instance.

```
tar cpf - -C $SPLUNK_HOME/etc/apps . | (ssh <server name of
new search head> "tar xpf - -C $SPLUNK_HOME/etc/apps")
```

```
tar cpf - -C $SPLUNK_HOME/etc/users . | (ssh <server name
of new search head> "tar xpf - -C $SPLUNK_HOME/etc/users")
```

Configuring the Splunk Enterprise indexer

In the distributed deployment, we separate the search management function from the Splunk Enterprise single instance. The former standalone instance continues to take the indexer role and the data already indexed on the instance continues to be available. No configuration is required on the Splunk Enterprise indexer.

Configuring the Splunk Enterprise admin server

Configure the Splunk Enterprise admin server to get it ready to run. This section provides detailed procedures for enabling master node and indexer discovery on the Splunk Enterprise admin server. These procedures enable the admin server to act as a master node to coordinate the activities of the search peer nodes (indexers), manage common configurations across all search peers, and enable forwarders to connect dynamically to the full set of available search peers.

1. Log in to the Splunk Enterprise admin server as the root account with SSH.
2. Enable the admin server as the master node with a replication factor of 1 and a search factor of 1.

   ```
   sudo -H -u splunk $SPLUNK_HOME/bin/splunk edit cluster-
   config -mode master -replication_factor 1 -search_factor 1
   -secret your_secret_key -cluster_label indexer_cluster_01 -
   auth admin:password
   ```

3. Add these lines in configuration file `$SPLUNK_HOME/etc/system/local/server.conf` to enable mast node indexer discovery:

   ```
   [indexer_discovery]
   pass4SymmKey = your_secret_key
   polling_rate = 10
   indexerWeightByDiskCapacity = true
   ```

4. Restart the Splunk Enterprise instance.

   ```
   service splunk restart
   ```

Configuring Splunk Enterprise search peers

Configure Splunk Enterprise search peers to get them ready to run. A search peer has the role of an indexer in a Splunk distributed deployment. It performs indexing and responds to search requests from the search head. In this distributed deployment, we separate the search management function from the Splunk Enterprise single instance, and the former standalone instance continues to take the indexer role.

To enable the peer node on the indexer:

1. Log in to the Splunk Enterprise indexer as the root account with SSH.
2. Enable the indexer as the peer node.

   ```
   sudo -H -u splunk $SPLUNK_HOME/bin/splunk edit cluster-
   config -mode slave -master_uri https://<server name of the
   admin server>:8089 -replication_port 9887 -secret
   your_secret_key -auth admin:password
   service splunk restart
   ```

3. Repeat this procedure on all Splunk Enterprise indexer instances.

Configuring Splunk Enterprise search head

Configure the Splunk Enterprise search head to get it ready to run. A search head handles search management functions in a distributed deployment. It directs search requests to a set of search peers and then merges the results and send them back to the user. In this use case, a new search head instance acting as a dedicated search head is added. To configure the Splunk Enterprise search head:

1. Log in to the Splunk Enterprise search head as the root account with SSH.

2. Enable a search head and restart the Splunk service.

```
sudo -H -u splunk $SPLUNK_HOME/bin/splunk edit cluster-
config -mode searchhead -master_uri https://<server name of
the admin server>:8089 -secret your_secret_key -auth
admin:password
service splunk restart
```

Validating implementation

Validate Splunk Enterprise instances to ensure they are ready to run. This process requires installing Splunk Universal Forwarder on one machine and indexing some testing data for validation. The following sections describe how to validate the Splunk Enterprise implementation.

Installing Splunk Universal Forwarder

On one physical machine or virtual machine, install Red Hat Enterprise Linux 7.2, and then install Splunk Universal Forwarder. See Installing Splunk Universal Forwarder in Chapter 4 for details.

Forwarding data to the indexer

1. Log in to the Splunk universal forwarder (installed in previous step) as the root account with SSH.
2. Create a monitor directory `/data/forwarder` and upload the Splunk tutorial data file `tutorialdata.zip` to this directory.

Note: You can download the `tutorialdata.zip` file from the Splunk website.

3. Change the file owner and permission of `/data/forwarder/tutorialdata.zip` file:

```
chmod 777 /data
chown -R splunk:splunk /data/forwarder
chmod -R 750 /data/forwarder
```

4. Add these lines in configuration file `$SPLUNK_HOME/etc/system/local/outputs.conf` to configure the universal forwarder to use master node indexer discovery.

```
[indexer_discovery:master1]
pass4SymmKey = your_secret_key
master_uri = https://<server name of admin server>:8089

[tcpout:group1]
autoLBFrequency = 30
indexerDiscovery = master1
useACK=true

[tcpout]
defaultGroup = group1
```

5. Restart the Splunk Universal Forwarder instance.

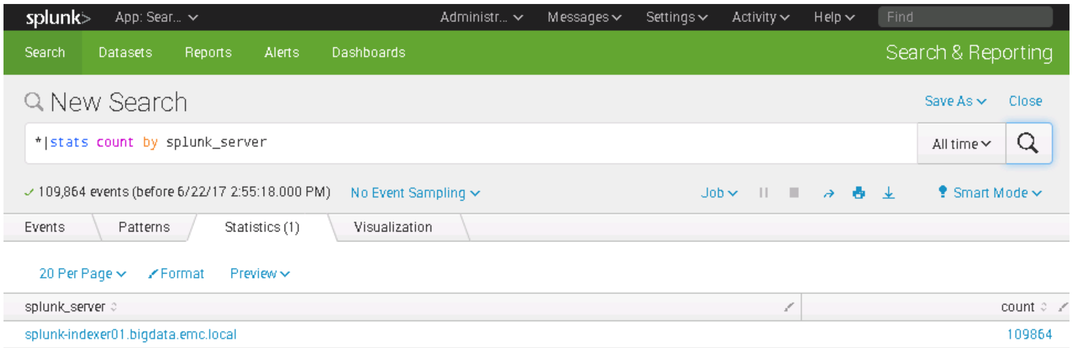
```
service splunk restart
```

6. Configure data input on the Splunk universal forwarder.

```
sudo -H -u splunk $SPLUNK_HOME/bin/splunk add monitor
/data/forwarder -auth admin:password
```

Validating Splunk Enterprise implementation

1. Log in to the Splunk Enterprise search head web interface **http://*<server name of splunk search head>*:8000/** to view the indexed data. Splunk indexes all the tutorial data and distributes it across all indexers, as shown in Figure 9.



```
*|stats count by splunk_server
```

✓ 109,864 events (before 6/22/17 2:55:18.000 PM)

splunk_server	count
splunk-indexer01.bigdata.emc.local	109864

Figure 9. **Validation result of Splunk Enterprise in Use Case 2**

Removing tutorial data

1. Log in to the Splunk Enterprise indexer as the root account with SSH.
2. Remove the tutorial data.

```
service splunk stop
sudo -H -u splunk $SPLUNK_HOME/bin/splunk clean eventdata -
index main
service splunk start
```

Important: This step removes all indexed data from the indexer. Do not perform this step if you have production data in the indexer.

3. Repeat this procedure on all Splunk Enterprise indexer instances.

Use case summary

In this use case, we implemented a Splunk Enterprise distributed deployment on three PowerEdge servers with one search head, one indexer, and one admin server. This use case, which can index up to 250 GB/day indexing volume with 115-day retention, usually meets the indexing and search needs of a small enterprise. This use case shows that PowerEdge servers can scale a Splunk deployment by distributing Splunk Enterprise instances across multiple servers.

Chapter 6 Splunk Enterprise Clustered Infrastructure for 250 GB/day Indexing Volume

This chapter presents the following topics:

Overview 43

Implementation 43

Use case summary...47

Overview

This chapter describes the implementation procedure for Use Case 3, which can index up to 250 GB/day indexing volume with 115-day retention and usually meets the indexing and search needs of a small enterprise requiring high data availability. It provides detailed step-by-step instructions for expanding Splunk Enterprise from a single instance to a clustered deployment. We separated the search management function from a Splunk Enterprise single instance by adding one dedicated search head. We increased the indexing capacity by adding one indexer to build an indexer cluster to replicate data and ensure data high availability. Finally, we added one admin server as the master node to manage these two indexers or search peers.

Implementation

Table 18 lists the implementation process flow of Use Case 3.

Table 18. Implementation process flow of Use Case 3

Step	Action	Description
1	Preparing the PowerEdge server	Prepare PowerEdge servers for use.
2	Configuring storage	Configure disk storage on PowerEdge servers for the OS and Splunk index data.
3	Installing the OS	Install the OS on PowerEdge servers and configure it for Splunk Enterprise installation and operation
4	Installing Splunk Enterprise	Install the Splunk Enterprise instance and perform initial configuration tasks.
5	Performing migration	Migrate the existing set of knowledge objects and applications from the former instance to the new search head.
6	Configuring Splunk Enterprise indexer	Configure the location and data size of the Splunk index bucket and enable the receiver to receive data.
7	Configuring Splunk Enterprise admin server	Enable the master node and index discovery on Splunk Enterprise admin server.
8	Configuring Splunk Enterprise search peers	Enable the peer node on Splunk Enterprise indexers.
9	Configuring Splunk Enterprise search head	Enable the search head on Splunk Enterprise search head.
10	Validating the implementation	Validate the implementation of Splunk Enterprise.

Preparing the PowerEdge server

In this use case, three PowerEdge servers are added to expand Splunk Enterprise from a single instance to a clustered deployment. The former PowerEdge R730xd server and one newly added PowerEdge R730xd server take the indexer role in the clustered deployment. One newly added PowerEdge R630 server takes the search head role. Another newly added PowerEdge R630 server takes the admin server role.

Before starting the implementation, ensure that these newly added PowerEdge servers are racked, cabled, powered, and made ready for use. See *Dell EMC PowerEdge R730xd Manuals & Documents* or the *Dell EMC PowerEdge R630 Manuals & Documents* for instructions.

Configuring storage

Configure storage on the servers for OS and Splunk index data. Table 19 lists the virtual disks configuration on PowerEdge servers.

Table 19. Virtual disks configuration of Use Case 3

Server role	Storage	Virtual disk name	RAID level	Physical disks
Indexer, Search Head, Admin Server	OS	RAID1_HDD	RAID-1	2 x 300 GB SAS HDD
Indexer	Splunk hot/warm bucket	RAID6_SSD	RAID-6	8 x 800 GB SATA SSD
Indexer	Splunk cold bucket	RAID10_HDD	RAID-10	14 x 2 TB SAS HDD

For details about configuring storage on these newly added PowerEdge servers, see Configuring storage in Chapter 4.

Installing the OS

Next, install the OS on the newly added PowerEdge servers. For details, see Installing the OS in Chapter 4.

Installing Splunk Enterprise

Next, install Splunk Enterprise on the newly added PowerEdge servers. For details, see Installing Splunk Enterprise in Chapter 4.

Performing migration

Migrate from a Splunk Enterprise single instance to a distributed deployment. As the distributed deployment is expanded from a standalone instance, the former standalone instance continues to take the indexer role. The data already indexed on the former standalone instance continues to be available in the new environment, so there is no need to migrate the data. The only remaining task is to migrate the existing set of knowledge objects and applications from the former standalone instance to the newly added search head instance. For details, see Performing Migration in Chapter 5.

Configuring Splunk Enterprise indexer

Configure the newly added Splunk Enterprise indexer instances to get them ready to run. For details, see Configuring Splunk Enterprise Indexer in Chapter 4.

Configuring Splunk Enterprise admin server

Configure the Splunk Enterprise admin server to get it ready to run. This procedure involves enabling the master node and indexer discovery on the Splunk Enterprise admin server. It enables the admin server to act as a master node to coordinate the activities of the search peer nodes (indexer), manage common configurations across all search peers, and enable forwarders to connect dynamically to the full set of available search peers.

1. Log in to the Splunk Enterprise admin server as root account with SSH.
2. Enable the admin server as the master node with a replication factor of 2 and a search factor of 2.

   ```
   sudo -H -u splunk $SPLUNK_HOME/bin/splunk edit cluster-
   config -mode master -replication_factor 2 -search_factor 2
   -secret your_secret_key -cluster_label indexer_cluster_01 -
   auth admin:password
   ```

3. Add these lines in the configuration file `$SPLUNK_HOME/etc/system/local/server.conf` to enable master node indexer discovery.

   ```
   [indexer_discovery]
   pass4SymmKey = your_secret_key
   polling_rate = 10
   indexerWeightByDiskCapacity = true
   ```

4. Restart the Splunk Enterprise instance.

   ```
   service splunk restart
   ```

Configuring Splunk Enterprise search peers

Configure Splunk Enterprise search peers to get them ready to run. A search peer has the role of an indexer in a Splunk clustered deployment. It indexes and responds to search requests from the search head. In this use case, two indexers must be configured to enable the peer node role: One is the former standalone instance, and the other one is the newly added indexer instance. For detailed procedures, see Configuring Splunk Enterprise Search Peers in Chapter 5.

Configuring Splunk Enterprise search head

Configure a Splunk Enterprise search head to get it ready to run. A search head handles search management functions in a clustered deployment. It directs search requests to a set of search peers and then merges the results back to the user. In this use case, a new search head instance acting as a dedicated search head is added. For detailed procedures, see Configuring Splunk Enterprise Search Head in Chapter 5.

Validating implementation

Validate the Splunk Enterprise instances to ensure they are ready to run. This process requires installing a Splunk Universal Forwarder on one machine and indexing some test data to do the validation. The following sections provide detailed procedures for validating Splunk Enterprise implementation.

Installing the Splunk Universal Forwarder

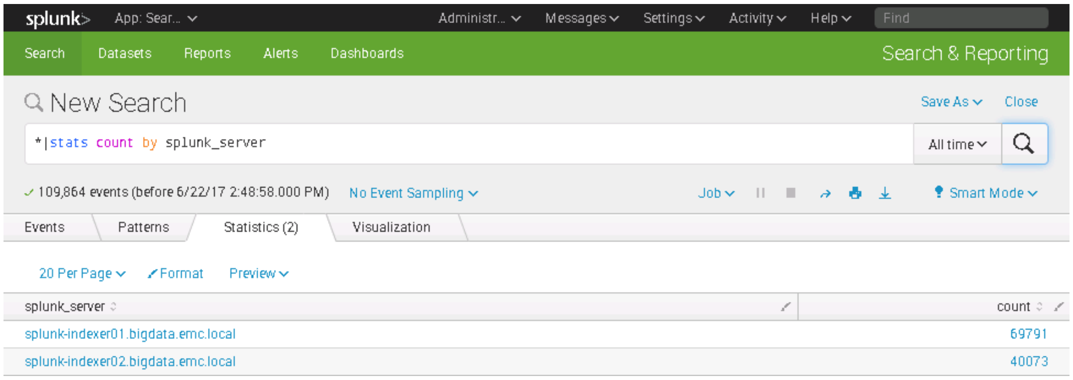
Install Red Hat Enterprise Linux 7.2 on aphysical machine or virtual machine and then install Splunk Universal Forwarder. For detailed procedures, see Installing Splunk Universal Forwarder in Chapter 4.

Forwarding data to indexer

Forward tutorial data to the Splunk indexer cluster. For detailed procedures, see Forwarding Data to Indexer in Chapter 5.

Validating the Splunk Enterprise implementation

1. Log in to the Splunk Enterprise search head web interface **http://<*server name of splunk search head*>:8000/** to view the indexed data. Splunk indexes all the tutorial data and distributes it across all indexers, as shown in Figure 10.



```
*|stats count by splunk_server
```

✓ 109,864 events (before 6/22/17 2:48:58.000 PM)

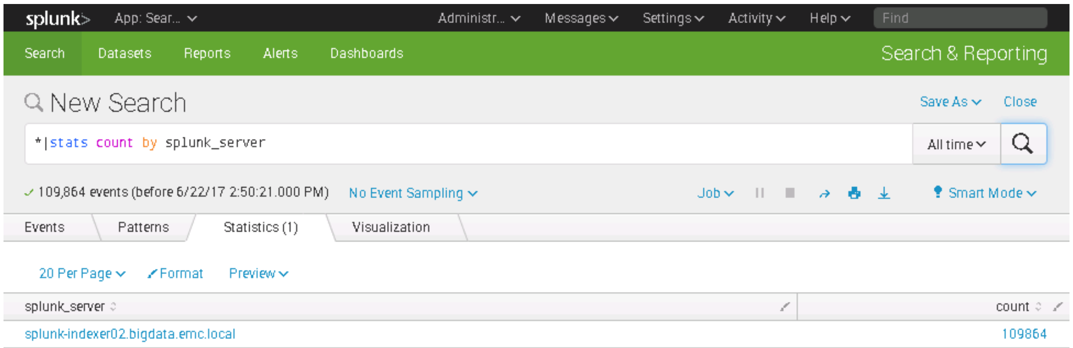
splunk_server	count
splunk-indexer01.bigdata.emc.local	69791
splunk-indexer02.bigdata.emc.local	40073

Figure 10. **Validation result step 1 of Splunk Enterprise in Use Case 3**

2. Log in to one of the Splunk Enterprise indexers and stop the Splunk Enterprise instance.

```
service splunk stop
```

3. Log in to the Splunk Enterprise search head web interface **http://<*server name of splunk search head*>:8000/** to view the indexed data again. The indexed events are still accessible on the rest of the indexers, as shown in Figure 11.



```
*|stats count by splunk_server
```

✓ 109,864 events (before 6/22/17 2:50:21.000 PM)

splunk_server	count
splunk-indexer02.bigdata.emc.local	109864

Figure 11. **Validation result step 2 of Splunk Enterprise in Use Case 3**

Removing tutorial data

Remove the tutorial data from Splunk Enterprise indexers to get them ready for production. For the detailed procedures, see Removing Tutorial Data in Chapter 5.

Use case summary

In this use case, we implemented Splunk Enterprise clustered deployment on four PowerEdge servers with one search head, two indexers, and one admin server. This use case, which can index up to 250 GB/day indexing volume with 115-day retention, usually meets the indexing and search needs of a small enterprise requiring high data availability. This use case shows that PowerEdge servers can scale Splunk by distributing Splunk Enterprise instances across multiple servers to build an indexer cluster that can prevent data loss while promoting data availability for searching.

Chapter 7 Splunk Enterprise High Performance Clustered Infrastructure for 250 GB/day Indexing Volume

This chapter presents the following topics:

Overview 49

Implementation 49

Use case summary 49

Overview

This validation guide provides two use cases for Splunk Enterprise clustered deployment, including the use case described in Compute Design in Chapter 3. One of the cases is a high-performance clustered deployment, which has a more powerful CPU and more memory for the indexers to provide complex and low-latency query support. The difference between these two cases is the hardware configuration of the Splunk Enterprise indexer; there is no difference in the implementation procedure.

This chapter describes the implementation procedure for Use Case 4, a high-performance clustered deployment. In this use case, the cluster is also expanded from a single instance. We separated the search management function from Splunk Enterprise single instance by adding one dedicated search head. We increased the indexing capacity by adding one indexer to build an indexer cluster to replicate data and ensure data high availability. Finally, we added one admin server as the master node to manage these two indexers or search peers.

Implementation

The procedure for implementing this high-performance clustered deployment is the same as for the balanced clustered deployment in Chapter 6. See Implementation in Chapter 6 for details.

Use case summary

In this use case, we implemented a Splunk Enterprise high-performance clustered deployment on four PowerEdge servers with one search head, two indexers, and one admin server. This use case, which can index up to 250 GB/day indexing volume with 115-day retention, usually meets the indexing and search needs of a small enterprise requiring high data availability, and complex, low-latency query support. This use case shows that PowerEdge servers can support Splunk Enterprise high-performance workload requirements and can scale Splunk by distributing Splunk Enterprise instances across multiple servers to build an indexer cluster that can prevent data loss while promoting data availability for searching.

Chapter 8 Splunk Enterprise Clustered Infrastructure with Isilon for 250 GB/day Indexing Volume

This chapter presents the following topics:

Overview..51

Implementation 51

Use case summary..55

Overview

This chapter describes the implementation procedure for Use Case 5. It provides detailed instructions for integrating Isilon storage into the Splunk Enterprise clustered infrastructure on PowerEdge servers to provide a cost-effective, scale-out NAS for the configurable retention of Splunk cold bucket data.

Isilon storage is not limited to integration in a Splunk Enterprise clustered deployment. It can also be attached to a distributed deployment to provide the configurable retention of cold bucket data. A general recommendation is to add Isilon storage when the cold bucket data is larger than 100 TB.

Implementation

Table 20 lists the implementation process flow of Use Case 5.

Table 20. Implementation process flow of Use Case 5

Step	Action	Description
1	Preparing the Isilon cluster	Prepare the Isilon cluster for use.
2	Configuring Isilon OneFS	Configure Isilon OneFS for storing Splunk cold bucket data, including access zone creation, IP address pool creation, SmartConnect configuration, and so on.
3	Mounting Isilon NFS export	Mount the Isilon Network File System share to the Splunk Enterprise Indexer Linux file system to store Splunk cold bucket data.
4	Performing migration	Migrate Splunk cold bucket data from local Direct Attached Storage (DAS) to Isilon NAS on Splunk Enterprise Indexers.
5	Validating implementation	Validate the implementation of Splunk.

Preparing the Isilon cluster

In this use case, one three-node Isilon X210 cluster is integrated into the existing Splunk Enterprise clustered infrastructure to store Splunk cold bucket data.

Before starting the implementation, ensure that the Isilon cluster is racked, cabled, powered, and initially configured. For detailed instructions about installing the new Isilon cluster, see the *Isilon Site Preparation and Planning Guide*.

Configuring Isilon OneFS

When the new Isilon cluster is correctly set up and ready for use, configure the Isilon OneFS file system to store Splunk cold bucket data.

Creating an access zone, user, group, and directory

1. Log in to one of the Isilon nodes as the root account with SSH.
2. Create a new access zone for Splunk.

```
isi zone zones create --name=splunk --path=/ifs/data/splunk
--create-path
```

3. Create a Splunk user and group in the access zone. The NFS client uses this user to access Isilon.

   ```
   isi auth groups create splunk --provider local --gid
   <Splunk group ID> --zone splunk
   ```

   ```
   isi auth users create splunk --password <Splunk user
   password> --primary-group splunk --provider local --uid
   <Splunk user ID> --enabled yes --zone splunk
   ```

   Note: The user ID and group ID of the Splunk user in OneFS must be the same as the one used for the Splunk Enterprise indexers.

4. Create a base directory `/ifs/data/splunk/colddb` and one dedicated subdirectory for each Splunk Enterprise indexer to store Splunk cold bucket data in the access zone. This base directory is configured as an Isilon NFS export later.

   ```
   mkdir -p /ifs/data/splunk/colddb
   mkdir -p /ifs/data/splunk/colddb/indexer01
   mkdir -p /ifs/data/splunk/colddb/indexer02
   ```

5. Grant read/write permission to the Splunk user on these directories.

   ```
   chown -R 1000:1000 /ifs/data/splunk/colddb
   chmod -R 755 /ifs/data/splunk/colddb
   ```

Creating NFS export

1. Log in to one of the Isilon nodes as the root account with SSH.

2. Create an NFS export to allow the Splunk indexer to access the Splunk access zone on Isilon.

   ```
   isi nfs exports create /ifs/data/splunk/colddb --all-
   dirs=yes --zone splunk
   ```

Creating IP address pool

1. Log in to one of the Isilon nodes as the root account with SSH.

2. Create an IP address pool and associate it with the Splunk access zone.

   ```
   isi network pools create groupnet0:subnet0:pool1 --
   ranges=<ip range for splunk access zone> --access-
   zone=splunk --alloc-method=static --ifaces=1-3:ext-1 --sc-
   subnet=subnet0 --sc-dns-zone=<splunk SmartConnect zone
   name> --description="splunk access zone"
   ```

   Note: The default Isilon SmartConnect license does not allow creating multiple IP address pools on a single subnet. Install the SmartConnect Advance license to allow this feature.

Configuring SmartConnect

1. Log in to one of the Isilon nodes as the root account with SSH.

2. Check the SmartConnect service IP address of the Isilon external network subnet0.

   ```
   isi network subnets view groupnet0.subnet0
   ```

3. If necessary, configure the SmartConnect service IP address:

   ```
   isi network subnets modify groupnet0.subnet0 --sc-service-
   addr=<SmartConnect service ip address>
   ```

Validating SmartConnect

1. Log in to one of the Splunk Enterprise indexers as the root account with SSH.
2. Run the `nslookup` command on the Splunk SmartConnect zone name several times. When you view the output of this command, note that different IP addresses are returned for each run.

   ```
   nslookup <splunk SmartConnect zone name>
   ```

   Note: The Splunk SmartConnect zone name is defined in the step Creating IP address Pool.

3. If `nslookup` does not work, check the NS record configuration in the Domain Name System (DNS). The Isilon SmartConnect service IP address must be configured as the delegation DNS server for Splunk SmartConnect zone name.

Mounting Isilon NFS export

Mount the Isilon NFS export to the Linux File System on the Splunk Enterprise indexers to allow Splunk to store cold bucket data on the Isilon storage. To mount the Isilon NFS export:

1. Log in to the Splunk Enterprise indexer as the root account with SSH.
2. Install the nfs-utils package if it is not already installed in the system, and start the related services. The `nfs-utils` command provides NFS-related utilities to mount to the NFS share.

   ```
   yum install nfs-utils
   systemctl enable  rpcbind.service
   systemctl start  rpcbind.service
   systemctl restart  rpcbind.service
   systemctl restart  nfs.service
   ```

3. Create a mount point for the Isilon export and mount it.

   ```
   mkdir -p /data/splunk/isilon_colddb
   mount <Splunk SmartConnect zone
   name>:/ifs/data/splunk/colddb/<[indexer01, indexer02]>
   /data/splunk/isilon_colddb
   ```

4. Add the Isilon NFS export in the file systems table **/etc/fstab** to ensure that it can be mounted automatically on system startup.

   ```
   <Splunk SmartConnect zone
   name>:/ifs/data/splunk/colddb/<[indexer01, indexer02]>
   /data/splunk/isilon_colddb    nfs    defaults 0    0
   ```

5. Repeat this process on all Splunk Enterprise indexer instances.

Migrate data

Migrate Splunk cold bucket data from the DAS storage system to Isilon storage. If there is no cold bucket data to migrate, skip these steps:

1. Log in to the Splunk Enterprise indexer as the root account with SSH.

2. Stop the Splunk service.

   ```
   service splunk stop
   ```

3. Copy all data from the existing cold bucket data directory on the DAS storage to the new cold bucket data directory on Isilon storage.

   ```
   sudo -H -u splunk cp -R -p /data/splunk/colddb/*
   /data/splunk/isilon_colddb
   ```

4. Update the configuration file `$SPLUNK_HOME/etc/system/local/indexes.conf` to point the index cold bucket to the Isilon mount point and change the maximum data size of the cold bucket.

   ```
   [main]
   homePath = /data/splunk/homedb/$_index_name
   coldPath = /data/splunk/isilon_colddb/$_index_name
   homePath.maxDataSizeMB = 3840000
   coldPath.maxDataSizeMB = 25000000
   maxWarmDBCount = 4294967295
   ```

5. Start Splunk service.

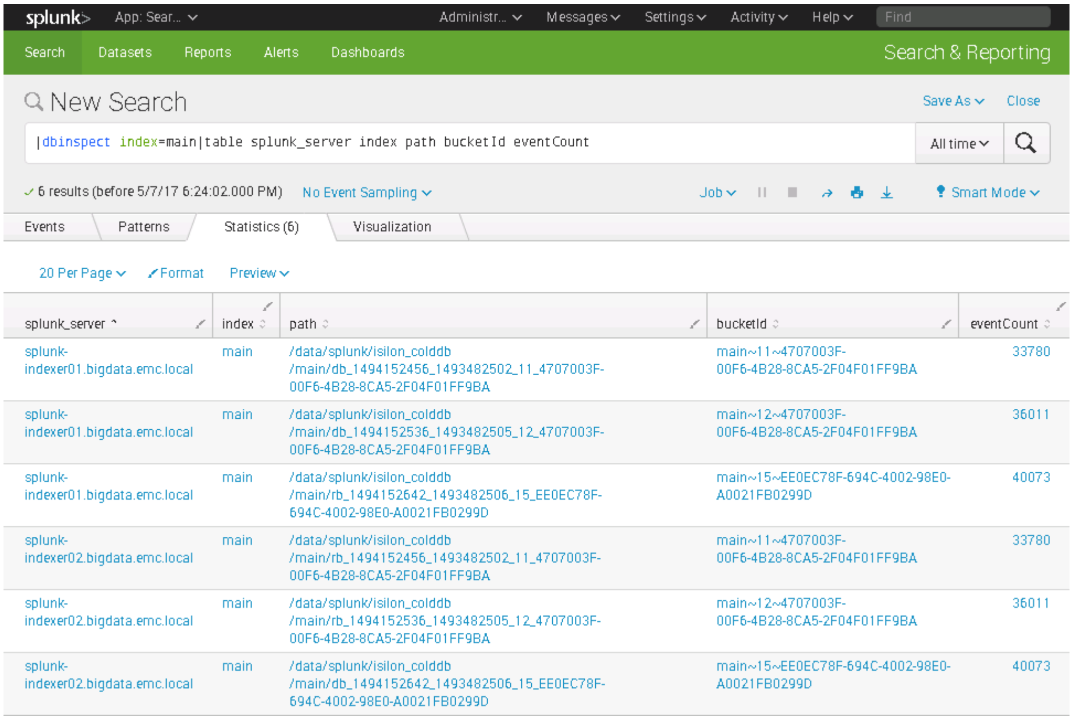
   ```
   service splunk start
   ```

6. Repeat this process on all Splunk Enterprise indexer instances.

Validating implementation

Validate the Splunk Enterprise instance to ensure it works properly.

1. Log in to the Splunk Enterprise search head web interface **http://<*server name of splunk search head*>:8000/** to view the indexed data. The output of the Splunk `dbinspect` command shows that the index path is changed from local DAS storage to Isilon storage, as shown in Figure 12.



```
|dbinspect index=main|table splunk_server index path bucketId eventCount
```

✓ 6 results (before 5/7/17 6:24:02.000 PM)

splunk_server	index	path	bucketId	eventCount
splunk-indexer01.bigdata.emc.local	main	/data/splunk/isilon_colddb/main/db_1494152456_1493482502_11_4707003F-00F6-4B28-8CA5-2F04F01FF9BA	main~11~4707003F-00F6-4B28-8CA5-2F04F01FF9BA	33780
splunk-indexer01.bigdata.emc.local	main	/data/splunk/isilon_colddb/main/db_1494152536_1493482505_12_4707003F-00F6-4B28-8CA5-2F04F01FF9BA	main~12~4707003F-00F6-4B28-8CA5-2F04F01FF9BA	36011
splunk-indexer01.bigdata.emc.local	main	/data/splunk/isilon_colddb/main/rb_1494152642_1493482506_15_EE0EC78F-694C-4002-98E0-A0021FB0299D	main~15~EE0EC78F-694C-4002-98E0-A0021FB0299D	40073
splunk-indexer02.bigdata.emc.local	main	/data/splunk/isilon_colddb/main/rb_1494152456_1493482502_11_4707003F-00F6-4B28-8CA5-2F04F01FF9BA	main~11~4707003F-00F6-4B28-8CA5-2F04F01FF9BA	33780
splunk-indexer02.bigdata.emc.local	main	/data/splunk/isilon_colddb/main/rb_1494152536_1493482505_12_4707003F-00F6-4B28-8CA5-2F04F01FF9BA	main~12~4707003F-00F6-4B28-8CA5-2F04F01FF9BA	36011
splunk-indexer02.bigdata.emc.local	main	/data/splunk/isilon_colddb/main/db_1494152642_1493482506_15_EE0EC78F-694C-4002-98E0-A0021FB0299D	main~15~EE0EC78F-694C-4002-98E0-A0021FB0299D	40073

Figure 12. **Validation result of Splunk Enterprise in Use Case 5**

Use case summary

In this use case, we integrated Isilon storage into a Splunk Enterprise clustered infrastructure on PowerEdge servers to provide configurable retention of Splunk cold bucket data. This use case, which can index up to 250 GB/day indexing volume with 210-day retention, usually meets the indexing and search needs of a medium-sized enterprise requiring high data availability and retention.

Chapter 9 Splunk Enterprise Single Instance Infrastructure for Development and Testing

This chapter presents the following topics:

Overview 57

Implementation 57

Use case summary...58

Overview

This chapter describes the implementation procedure for Use Case 6, an entry level clustered deployment for development and testing that is not intended to expand or scale. In this use case, a Splunk Enterprise single instance serving as both an indexer and a search head on one PowerEdge R430 server.

Implementation

Table 21 lists the implementation process flow of Use Case 6.

Table 21. Implementation process flow of Use Case 6

Step	Action	Description
1	Preparing the PowerEdge server	Prepare PowerEdge servers for use.
2	Configuring storage	Configure disk storage on PowerEdge servers for the OS and Splunk index data.
3	Installing the OS	Install the OS on PowerEdge servers and configure it for Splunk Enterprise installation and operations.
4	Installing Splunk Enterprise	Install the Splunk Enterprise instance and perform initial configuration tasks.
5	Configuring Splunk Enterprise indexer	Configure the location and data size of the Splunk index bucket and enable the receiver to receive data.
6	Validating implementation	Validate the implementation of Splunk Enterprise.

Preparing PowerEdge servers

One single Dell EMC PowerEdge R430 server is used in this use case. Before starting the implementation, ensure that the PowerEdge servers are racked, cabled, powered, and ready for use. For information about the PowerEdge R430 server, refer to *Dell EMC PowerEdge R430 Manuals & Documents*.

Configuring storage

Once the PowerEdge server is properly set up and ready for use, the next step is to configure the storage on the server for OS and Splunk index data. Table 22 lists the virtual disks configuration on PowerEdge server.

Table 22. Virtual disks configuration of Use Case 6

Server role	Storage	Virtual disk name	RAID level	Physical disks
Single Instance combined search head and indexer	OS	RAID1_HDD	RAID-1	2 x 300 GB SAS HDD
	Splunk hot/warm bucket	RAID10_SSD	RAID-10	2 x 960 GB SATA SSD
	Splunk cold bucket	RAID10_HDD	RAID-10	4 x 1.8 TB SAS HDD

For details about how to configure storage on these newly added PowerEdge servers, see Configuring Storage in Chapter 4.

Installing the OS

Next, install the OS on the PowerEdge server. For detailed procedures, see Installing the OS in Chapter 4.

Installing Splunk Enterprise

Next, install Splunk Enterprise on the PowerEdge server. For detailed procedures, see Installing Splunk Enterprise in Chapter 4.

Configuring Splunk Enterprise indexer

Next, configure Splunk Enterprise Indexer on the PowerEdge server. For detailed procedures, see Configuring Splunk Enterprise Indexer in Chapter 4.

Validating implementation

Next, validate Splunk Enterprise implementation. For detailed procedures, see Validating the implementation in Chapter 4.

Use case summary

In this use case, we implemented a Splunk Enterprise single instance on one Dell EMC PowerEdge R430 server that combined a Splunk indexer and search head into a single instance. This case usually meets the development and testing needs within an organization.

Chapter 10 Validated Configurations for Splunk Enterprise

This chapter presents the following topics:

Overview 60

Splunk-validated sizing configurations 60

Configuration 1: Single instance deployment for up to 250 GB/day indexing volume with 115-day retention 61

Configuration 2: Distributed deployment for up to 250 GB/day indexing volume with 115-day retention 61

Configuration 3: Clustered deployment for up to 250 GB/day indexing volume with 115-day retention 62

Configuration 4: High performance clustered deployment for up to 250 GB/day indexing volume with 115-day retention 62

Configuration 5: Isilon cold bucket expansion for up to 250 GB/day indexing volume with 210-day retention 63

Configuration 6: Single instance deployment for development and testing.63

Summary 63

Overview

In this chapter, we provide the Splunk-validated configurations of the Dell EMC Ready Bundle hardware that meet or exceed the performance of Splunk's documented reference hardware.

Splunk-validated sizing configurations

Splunk validated the following configurations for Dell EMC to meet or exceed the performance of Splunk's documented reference hardware:

- Configuration 1: Single instance deployment for up to 250 GB/day indexing volume with 115-day retention
- Configuration 2: Distributed deployment for up to 250 GB/day indexing volume with 115-day retention
- Configuration 3: Clustered deployment for up to 250 GB/day indexing volume with 115-day retention
- Configuration 4: High performance clustered deployment for up to 250 GB/day indexing volume with 115-day retention
- Configuration 5: Isilon cold bucket expansion for up to 250 GB/day indexing volume with 210-day retention
- Configuration 6: Single instance deployment for development and testing

These configurations represent typical uses in the current marketplace.

PowerEdge servers description

We use PowerEdge R430, R630, R730xd and R930 servers to provide a cost-optimized, highly available infrastructure for Splunk Enterprise in these configurations. Chapter 2 lists the attributes of the PowerEdge server and Table 23 describes the physical characteristics of the PowerEdge server based on Splunk's reference hardware recommendations in these five configurations.

Table 23. PowerEdge servers description

Components	R430: single instance	R730xd: indexer	R930: high performance indexer	R630: search head and admin server
Processors	2 x E5-2603 v4	2 x E5-2650 v4	2 x E7-8890 v4	2 x E5-2620 v4
Processor cores	12	24	48	16
Memory/RAM	32 GB	64 GB	128 GB	64 GB
SAS (OS)	2 x 300 GB RAID 1			
SSD	2 x 960 GB RAID 10	8 x 800 GB RAID 6		NA
HDD	4 x 1.8 TB RAID 10	14 x 2 TB RAID 10		NA

Configuration 1: Single instance deployment for up to 250 GB/day indexing volume with 115-day retention

This configuration is a Splunk Enterprise single instance deployment on one PowerEdge server that can index up to 250 GB/day data with 115-day retention.

Table 24. Configuration of single instance deployment for up to 250 GB/day indexing volume with 115-day retention (can scale out)

Instance role	Qty	PowerEdge model	CPU cores	Memory	OS storage (RAID1)	Hot/warm storage (RAID6)	Cold storage (RAID10)
Single Instance combined search head and indexer	1	PowerEdge R730xd	24	64 GB	300 GB	4,800 GB	14 TB

Configuration 2: Distributed deployment for up to 250 GB/day indexing volume with 115-day retention

This configuration is a Splunk Enterprise distributed deployment on three PowerEdge servers that can index up to 250 GB/day data with 115-day retention.

Table 25. Configuration of distributed deployment for up to 250 GB/day indexing volume with 115-day Retention

Instance role	Qty	PowerEdge model	CPU cores	Memory	OS storage (RAID1)	Hot/warm storage (RAID6)	Cold storage (RAID10)
Indexer	1	PowerEdge R730xd	24	64 GB	300 GB	4,800 GB	14 TB
Search head	1	PowerEdge R630	16	64 GB	300 GB	0	0
Admin server	1	PowerEdge R630	16	64 GB	300 GB	0	0

Configuration 3: Clustered deployment for up to 250 GB/day indexing volume with 115-day retention

This configuration is a Splunk Enterprise clustered indexer deployment on four PowerEdge servers that can index up to 250 GB/day data with 115-day retention.

Table 26. Configuration of clustered deployment for up to 250 GB/day indexing volume with 115-day retention

Instance role	Qty	PowerEdge model	CPU cores	Memory	OS storage (RAID1)	Hot/warm storage (RAID6)	Cold storage (RAID10)
Indexer	2	PowerEdge R730xd	24	64 GB	300 GB	4,800 GB	14 TB
Search head	1	PowerEdge R630	16	64 GB	300 GB	0	0
Admin server	1	PowerEdge R630	16	64 GB	300 GB	0	0

Configuration 4: High performance clustered deployment for up to 250 GB/day indexing volume with 115-day retention

This configuration is a Splunk Enterprise high-performance clustered indexer deployment on four PowerEdge servers that can index up to 250 GB/day data with 115-day retention.

Table 27. Configuration of high performance clustered deployment for up to 250 GB/day indexing volume with 115-day retention

Instance role	Qty	PowerEdge model	CPU cores	Memory	OS storage (RAID1)	Hot/warm storage (RAID6)	Cold storage (RAID10)
Indexer	2	PowerEdge R930	48	128 GB	300 GB	4,800 GB	14 TB
Search head	1	PowerEdge R630	16	64 GB	300 GB	0	0
Admin server	1	PowerEdge R630	16	64 GB	300 GB	0	0

Configuration 5: Isilon cold bucket expansion for up to 250 GB/day indexing volume with 210-day retention

This configuration is a Splunk Enterprise clustered indexer deployment on four PowerEdge servers with Isilon storage. This configuration is capable of indexing up to 250 GB/ day of data with 210-day retention.

For configuration guidance about Isilon scale-out storage, ~~refer to~~ see the *EMC Isilon Scale-Out Storage and VMware vSphere Sizing Guide.*

Table 28. Configuration of Isilon cold bucket expansion for up to 250 GB/day indexing volume with 210-day retention

Instance role	Qty	PowerEdge model	CPU cores	Memory	OS storage (RAID1)	Hot/warm storage (RAID6)	Cold Storage (Isilon)
Indexer	2	PowerEdge R730xd	24	64 GB	300 GB	4,800 GB	63 TB
Search head	1	PowerEdge R630	16	64 GB	300 GB	0	0
Admin server	1	PowerEdge R630	16	64 GB	300 GB	0	0

Configuration 6: Single instance deployment for development and testing

This configuration is a Splunk Enterprise single instance deployment on one PowerEdge server for development and testing.

Table 29. Configuration of single instance deployment for development and testing

Instance role	Qty	PowerEdge model	CPU cores	Memory	OS storage (RAID1)	Hot/warm storage (RAID10)	Cold storage (RAID10)
Single Instance combined search head and indexer	1	PowerEdge R430	12	32 GB	300 GB	960 GB	3.6 TB

Summary

Splunk Enterprise software together with Dell EMC PowerEdge servers provide an integrated technology for analyzing machine-generated Big Data across a wide range of data ingestion rates and customer use cases. Customers can be confident that the jointly validated systems described in this document will meet current customer needs and flexibly scale when the need arises.

Chapter 11 Conclusion

This chapter presents the following topics:

Summary 65

Findings...65

Conclusion 65

Summary

Business leaders are ready to increase their analytics capability, realize lower operational expenses, and improve customer experiences. Most enterprises cannot afford to risk success by implementing homegrown solutions. Splunk, in partnership with Dell EMC, offers a documented set of proven solutions that accommodate the needs of a variety of customers, including small businesses, enterprise departmental solutions, and medium enterprise full-scale deployments, all with an integrated set of technologies featuring detailed deployment and implementation guidance. Our approach provides a low-risk, fast time-to-value, fully supported option for machine-generated data analytics.

Findings

The ongoing partnership between Splunk and Dell EMC makes investing in new or expanded machine data analytics less risky and more cost-effective for businesses of all sizes. The jointly validated system configurations described in this document together with the detailed guidance on configuration and implementation provide prospective customers with the information they need to match their investment in equipment and necessary people skills so that they can confidently commit to meeting a wide range of use case goals.

Conclusion

Big Data analytics, specifically the analysis of machine data, can help business of all sizes drive critical decisions, reduce costs, and maximize operational efficiencies to overcome these challenges. This validation guide provides detailed information for evaluating the applicability of the Ready Bundle for Splunk Enterprise with Isilon storage offerings for a Splunk implementation. Splunk and Dell EMC have validated multiple use case configurations that meet or exceed the performance of Splunk’s documented reference hardware. Potential customers can match almost any current needs with an approved configuration. Customers can also be confident that all Dell EMC validated offerings together with the flexibility of Splunk Enterprise configuration options can be scaled out to handle future needs without the need for extensive upgrades or expensive re-platforming.

Chapter 12 References

This chapter presents the following topics:

Dell EMC documentation 67

Splunk Enterprise documentation 67

Other documentation 67

Dell EMC documentation

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

- *Dell EMC Ready Bundle for Splunk: Ready Bundle on PowerEdge Servers with Isilon for Machine Data Analytics Validation Guide*
- Dell EMC PowerEdge R730xd Spec Sheet
- *Dell EMC PowerEdge R430 Manuals & Documents*
- *Dell EMC PowerEdge R630 Manuals & Documents*
- *Dell EMC PowerEdge R730xd Manuals & Documents*
- *Dell EMC PowerEdge R930 Manuals & Documents*
- *Isilon Site Preparation and Planning Guide*
- *EMC Isilon Scale-Out storage and VMware vSphere Sizing Guide*

Splunk Enterprise documentation

The following documentation on the Splunk website provides additional and relevant information:

- *Splunk Capacity Planning Manual*
- *Splunk Enterprise Installation Manual*
- *Transparent huge memory pages and Splunk performance*
- *System requirements for use of Splunk Enterprise on-premises*

Other documentation

The following documentation on the internet provides additional and relevant information:

- *Red Hat Enterprise Linux 7 Installation Guide*