SUSE® Enterprise Storage

Implementation Guide for VMware ESXi

David Byte

Publication Date: 2020-04-16

Contents

1 Solution Components

SUSE

SUSE Enterprise Storage 5.5 or 6

VMware

VMware ESXi 6.7

2 Introduction

The objective of this guide is to present a step-by-step guide on how to implement SUSE Enterprise Storage 6 as an iSCSI storage target for VMware ESXi. The settings in this guide reflect the parameters necessary to achieve listing in the VMware Hardware Compatibility Guide (https://www.vmware.com/resources/compatibility/search.php?deviceCategory=san&details=1&partner=465&isSVA=0&page=1&display_interval=10&sortColumn=Partner&sortOrder=Asc) for SUSE Enterprise Storage 5.5. While the certification is on the prior version, the underlying iSCSI technology is the same with version 6, and thus the information contained herein applies equally to either version. It is suggested that the document be read in its entirety, along with the supplemental appendix information before attempting the implementation process.

The deployment presented in this guide aligns with the architectural best practices of both SUSE and VMware.

Upon completion of the steps in this document, a working SUSE Enterprise Storage 6 cluster will be operational, as described in the SUSE Enterprise Storage Deployment Guide (https://documentation.suse.com/ses/6/html/ses-all/book-storage-deployment.html) acting as a storage target VMware ESXi.

3 Solution Description

The solution outlined in this guide enables a customer to deploy SUSE Enterprise Storage as an iSCSI storage device for VMware ESXi.

4 Business Value

SUSE Enterprise Storage

SUSE Enterprise Storage delivers value in the following ways:

- Commodity hardware for minimal hardware cost
- Open source software, for minimal software cost and maximum flexibility
- A self-managing, self-healing architecture for minimal management cost
- A flexible, cluster-based design for graceful and inexpensive upgrade and innovative licensing model that avoids per-gigabyte storage charges, so you won't owe more for saving more data.
- With minimal acquisition cost, management cost, and upgrade cost, SUSE Enterprise Storage is the lowest-price solution for enterprise archive and backup implementations

VMware ESXi

VMware ESXi provides an Enterprise-Ready Virtualization infrastructure suite that is widely deployed.

Solution Value Propositions

Issue: Customer needs scalable, yet affordable storage for golden images, archive images, and software repositories. Solution: SUSE Enterprise Storage provides erasure-coded pools with exported iSCSI targets yielding low $/GB with adequate performance.

Issue: Customer needs storage with reasonable performance for virtual machines in addition to object storage. Solution: SUSE Enterprise Storage provides a single storage cluster that can service both the higher performance requirements of virtual machines and the lower performance requirements of Object Storage

5 Requirements

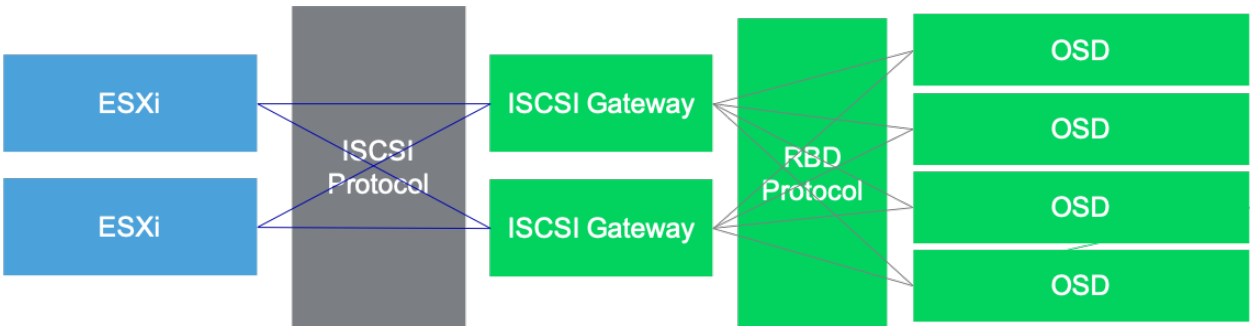
The solution has the following criteria:

- Simple to set up and deploy and meets the documented guidelines for system hardware, networking, and environmental prerequisites.
- Adaptable to the physical and logical constraints needed by the business, both initially and as required over time for performance, security, and scalability concerns.
- Resilient to changes in physical infrastructure components, caused by failure or required maintenance.

- Capable of providing block services via iSCSI to VMware ESXi
- Is certified and present in the VMware Hardware Compatibility Guide
- Data protection is configurable to meet the customer's individual needs at a granular level.

6 Architectural Overview

SUSE Enterprise Storage provides everything that VMware needs for storage in a scalable and certified manner. The solution utilizes multiple gateways and network paths in an active-active manner with the VMware multipath I/O stack in a round-robin configuration to ensure that performance scales while providing fault-tolerance.



6.1 Architectural Notes & Discussion

7 Pool Configuration

When configuring the SUSE Enterprise Storage cluster for use as a virtual machine datastore target, the data protection scheme is an important consideration. There are two main options for data protection, each with advantages and disadvantages.

7.1 Replication

The first is replication. Replication works by replicating each data chunk on each of the specified numbers of unique devices. The default is three. If the failure domain is assumed to be at the storage host level, this means the cluster could survive the loss of two storage servers without data loss. The downside of replication is the space overhead, which is 200% or two-thirds of the total cluster capacity.

The performance characteristics of replication are that it has lower latency than erasure coding and is generally the most performant data protection scheme. This is especially true where the I/O pattern is that of small random I/O.

7.2 Erasure Coding

The second scheme is erasure coding (EC). While there may be use cases such as archive data-stores and software depots where the space efficiency of erasure coding makes sense, *it is not recommended for use with pools hosting active VM images.*

An exception to this policy might be made when the virtual machine image is hosted on a replicated pool, and the application itself performs large writes. In this case, it is possible to host the data volume on an erasure coded pool.

One of the most important factors in determining whether a write is large in this scenario is the number of data chunks (K) in the erasure coding scheme. If K * 4KB is smaller than the I/O size used by the application, then a costly read-modify-write operation is avoided, resulting in reasonable performance. However, every I/O below that threshold incurs a penalty.

7.3 iSCSI

The iSCSI protocol provides SCSI commands and data transfers over a network interface. This protocol provides an option that is easier to consume for many end users than the competing Fibre Channel protocol due to iSCSI's use of standardized ethernet hardware and topologies.

In relation to SUSE Enterprise Storage, iSCSI is made available by means of one or more protocol gateways. These gateways translate between RBD and iSCSI, thus enabling clients that lack support for the more robust RBD, to access data stored on a SUSE Enterprise Storage cluster. The particular implementation used for certification utilizes the LIO kernel features and enables SUSE Enterprise Storage to act as an ALUA target. ALUA stands for Asymmetric Logical Unit Access and is the standardized set of functionality that enables a target to present multiple

paths and to communicate back to the initiator. The net outcome being that the SUSE Enterprise Storage iSCSI stack enables active-active multi-pathing functionality and supports the round-robin policy, thus providing bandwidth aggregation and resilience against port failures.

8 Deployment Recommendations

This deployment section should be seen as a supplement to available online documentation. (https://documentation.suse.com) Specifically, the SUSE Enterprise Storage 6 Deployment, Administration, and Tuning Guides and the SUSE Linux Enterprise Server Administration Guide.

The following recommendations should be strongly considered:

- Enterprise-class SSD or NVMe media is preferable for performance-sensitive environments
- Spinning media should be 7200RPM or faster, enterprise-class devices
    - When deploying spinning media, RocksDB and WAL should be located on a flash device
- It is recommended that the network leverage jumbo-frame ethernet, including the network between the iSCSI gateways and the ESXi nodes. This allows for more efficient data transport.

9 iSCSI Deployment

This section outlines the steps required to deploy an environment similar in architecture to the tested environment.

9.1 Deploy and prepare SUSE Enterprise Storage Environment

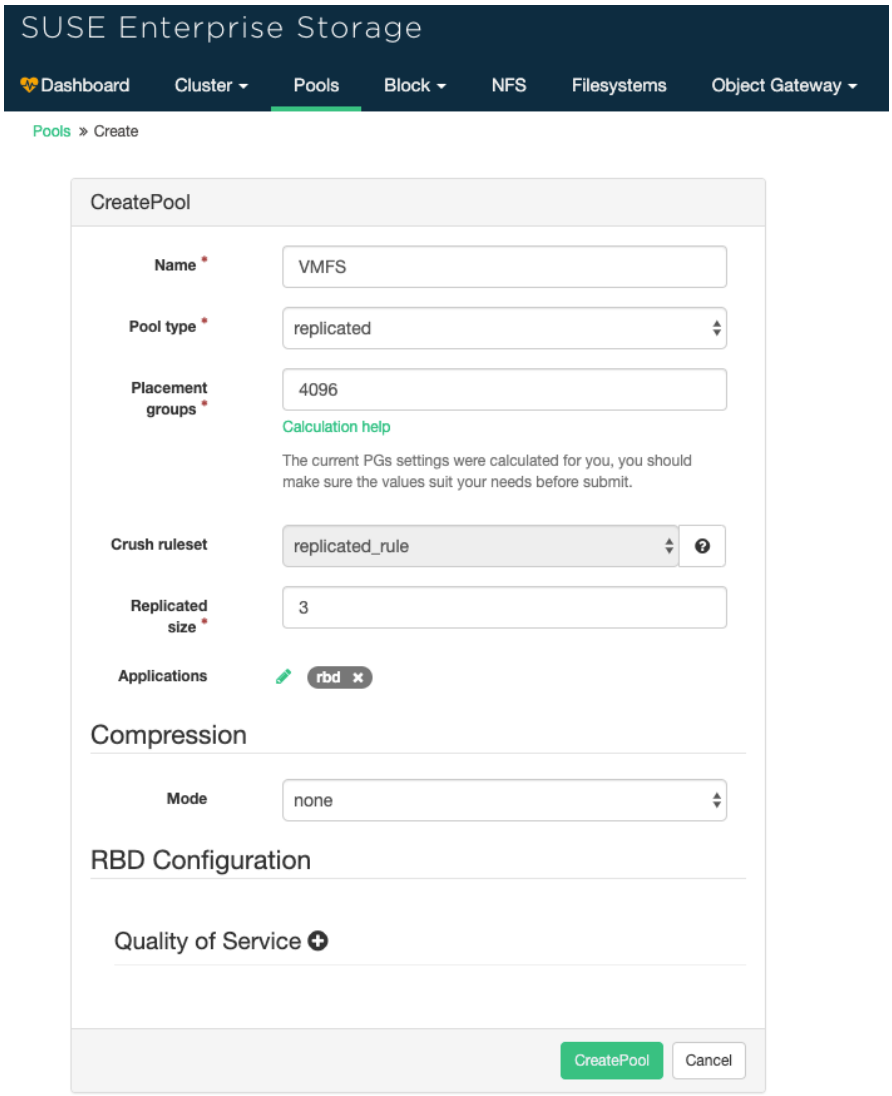
Build and deploy a SUSE Enterprise Storage Cluster as described in the SUSE Enterprise Storage Deployment Guide. During the deployment process, be sure to deploy two or more iSCSI gateway roles.

It is advisable to tune the cluster, as can be found in the SUSE Enterprise Storage Tuning Guide (https://documentation.suse.com/ses/6/html/ses-all/book-storage-tuning.html).

9.1.1 Create Pool(s) for VMFS Datastores

1. Create at least one pool for each protection scheme (replication and EC) being supported. To create a pool, open the Dashboard, select Pools, then click Create.

   - It is strongly recommended that the Pool Type is replicated.
   - Be sure to set the Application to RBD.

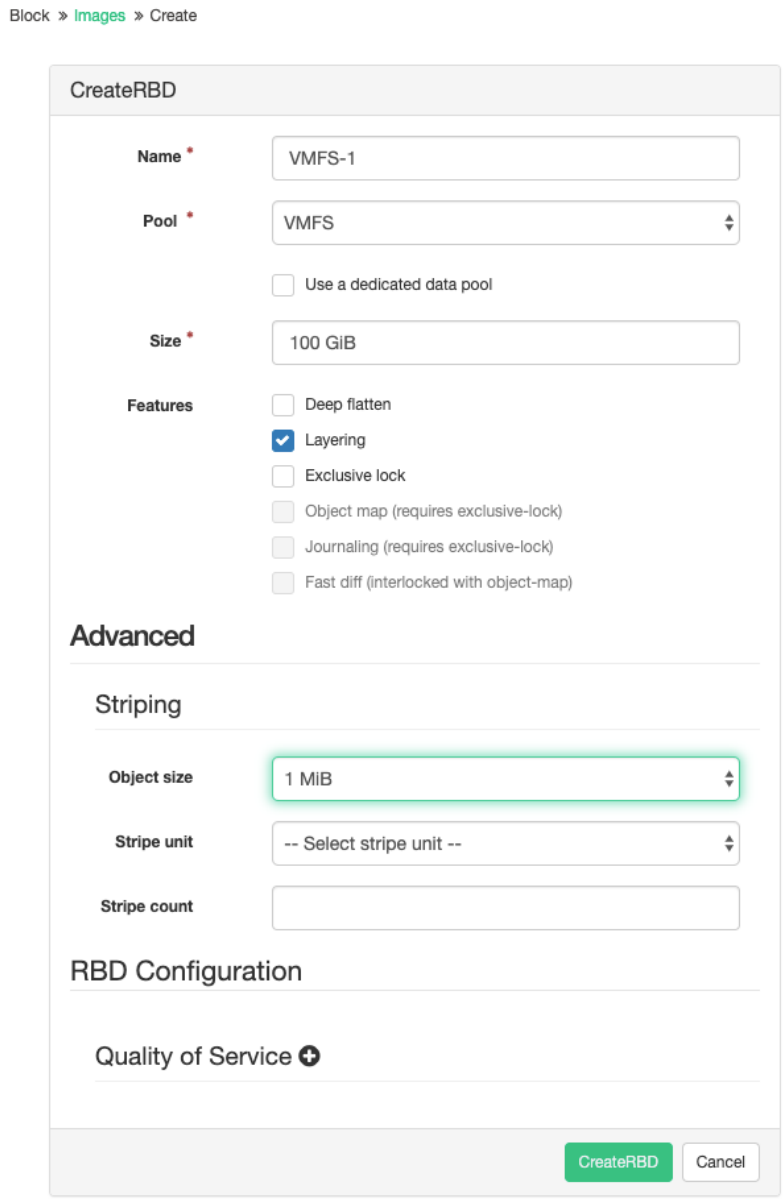


2. Create the RBD image
   To create the RBD that backs the iSCSI LUN, use the Dashboard interface under Block → Images.

- Select the pool created in the prior step.
- Click Advanced.
- Set the Object size to match the block size of the VMFS file system. The default for RBD is 4MiB, while VMFS defaults to 1MiB

  WARNING

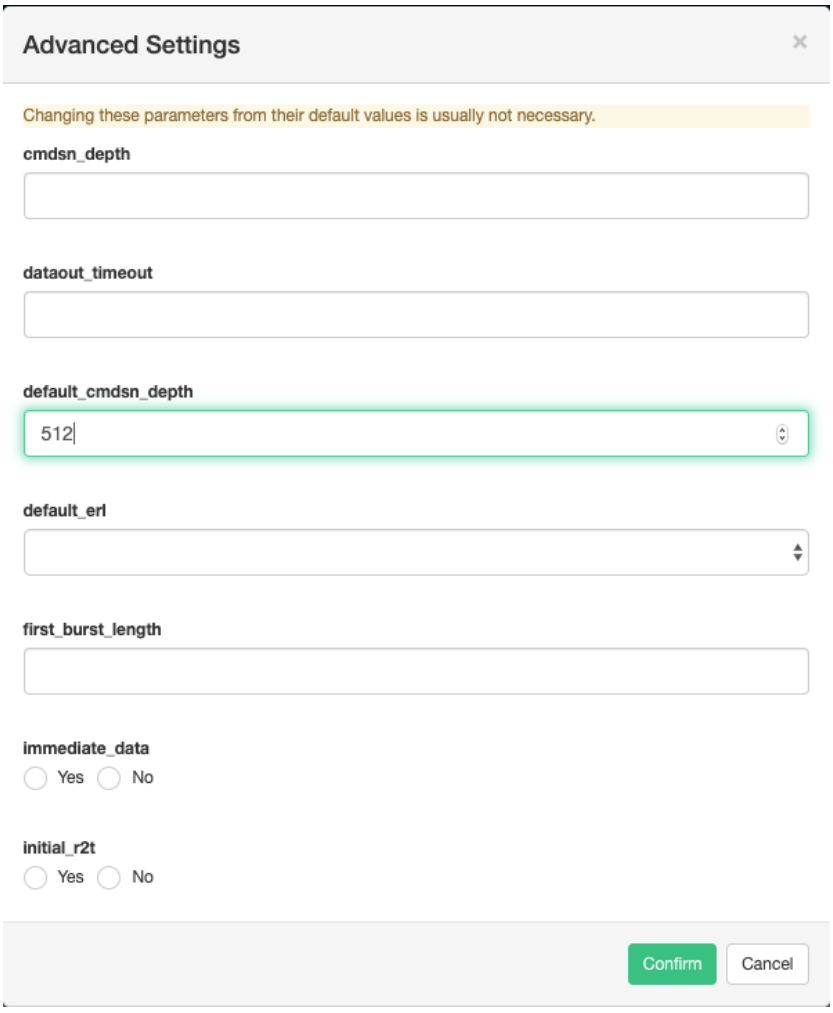
  Do not modify the stripe unit or stripe count; they should both be left blank.



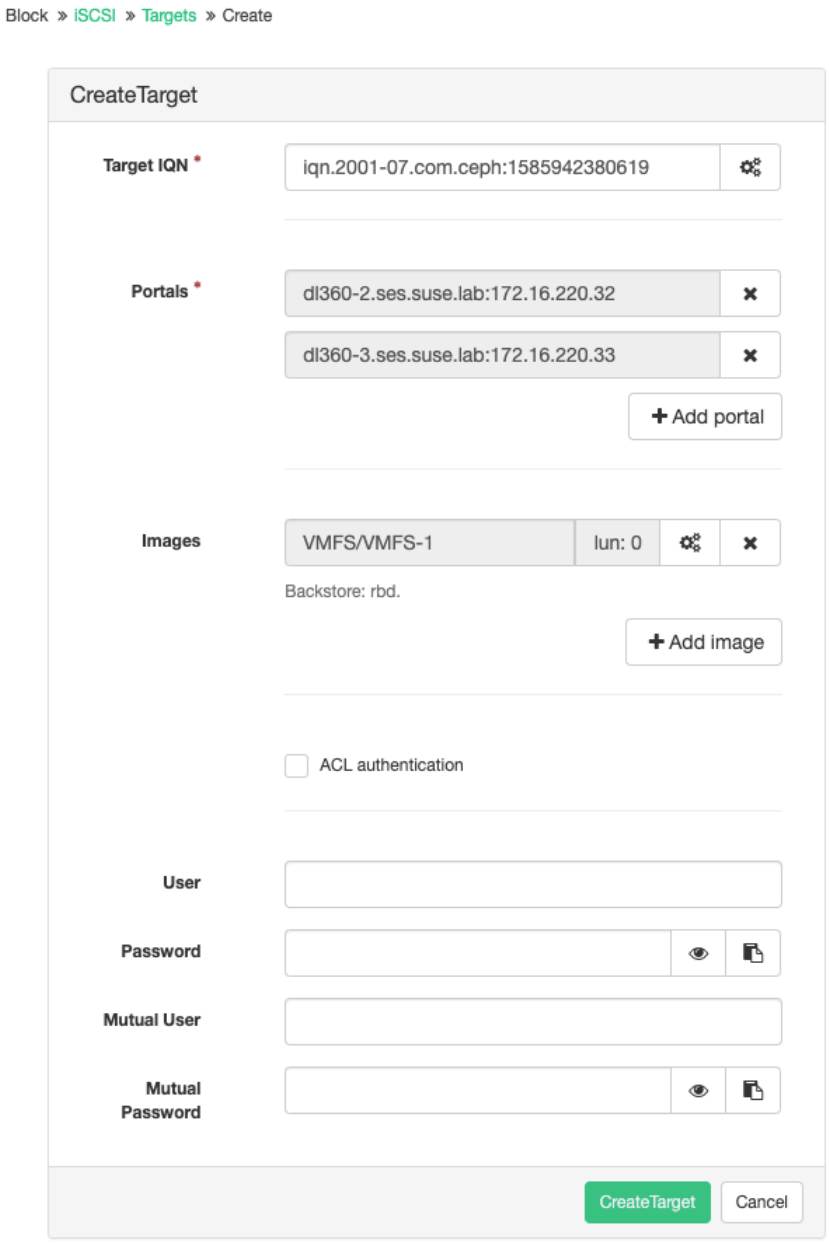
3. Create the iSCSI LUN

Correct creation of the iSCSI LUN is necessary to support functionality utilized by VMware.

- Under Bock→iSCSI→Targets, click Create.
- Select Use the existing IQN or create a custom IQN for easy identification.
- Modify the advanced settings for the target IQN by clicking the gear icon to the right. Set the following
  - default_cmdsn_depth:512



- Add the portals (iSCSI Gateways).
- Select the image created in the prior step.



- Click the gear next to the image name. This opens the advanced options dialogue. In this screen, set the following values:
  - lun:{Desired LUN ID number}
  - queue_depth:512
  - emulate_pr:0

- emulate_tpu:1
- emulate_tpws:1

NOTE

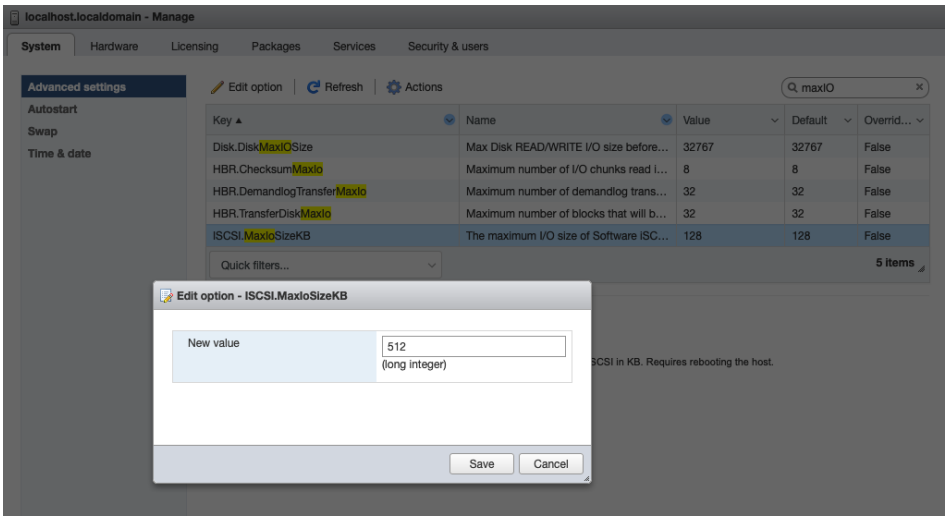
The emulate_tpu and emulate_tpws settings toggle target support for space reclamation via SCSI UNMAP and WRITE SAME, respectively. These have some performance implications. It is up to the user to determine whether to optimize storage utilization by enabling these or to optimize performance, by leaving them disabled. See VMware's documentation on space reclamation behavior (https://docs.vmware.com/en/VMware-vSphere/6.7/com.vmware.vsphere.storage.doc/GUID-BC1A172C-E649-4812-B8B2-A9E45AC97051.html) for more details.

4. Authentication
   If desired, CHAP authentication can be configured at this time.

9.2 VMware ESXi Configuration

There are a number of settings needed to ensure optimal configuration and operation of VMware with iSCSI targets. These settings are covered below.

- Set the iSCSI.MaxioSizeKB to 512



- Ensure ATS only Locking From the ESX CLI, issue this command:

```
esxcli storage vmfs lockmode list
```

  Information on changing a single LUN to ATS only can be found in the VMware documentation (https://docs.vmware.com/en/VMware-vSphere/6.7/com.vmware.vsphere.storage.doc/GUID-48E07447-5A9D-4DF1-BA70-713E07C054BC.html).

- Set multi-pathing policy
  VMware should always be multi-pathed to SUSE Enterprise Storage with at least two iSCSI gateways in the target portal. Traffic should be balanced across the gateways using the round-robin policy. In 6.7 and above the round-robin path selection now uses observed latencies. This results in the use of optimal pathing across the available gateways. Information on setting VMware multipath policy can be found in the VMware knowledgebase article #1017760 (https://kb.vmware.com/s/article/1017760).
  An additional item would be to set the IOPS to send down a particular path before changing paths. This can improve load-balancing and shorten failover times. This adjustment can be accomplished by using the command line on the ESXi server. Information on adjusting this setting can be found in the VMware knowledge base article 2069356 (https://kb.vmware.com/s/article/2069356).

10 Resources

- SUSE Enterprise Storage 6 – Deployment Guide https://documentation.suse.com/ses/6/html/ses-all/book-storage-deployment.html
- SUSE Enterprise Storage 6 – Administration Guide https://documentation.suse.com/ses/6/html/ses-all/book-storage-admin.html
- SUSE Enterprise Storage 6 - Tuning Guide https://documentation.suse.com/ses/6/html/ses-all/book-storage-tuning.html
- VMware Hardware Compatibility Guide - SUSE Enterprise Storage https://www.vmware.com/resources/compatibility/search.php?deviceCategory=san&details=1&partner=465&isSVA=0&page=1&display_interval=10&sortColumn=Partner&sortOrder=Asc