



Hewlett Packard
Enterprise

HPE best practices for deploying SUSE Enterprise Storage on HPE Synergy

Contents

Executive summary.....	3
Solution overview.....	3
HPE Synergy.....	4
SUSE Enterprise Storage.....	5
Solution components.....	5
Hardware.....	6
SUSE Enterprise Storage.....	6
Considerations for building this software-defined storage solution.....	7
Solution details.....	8
Best practices.....	8
Summary.....	10
Resources.....	11

Executive summary

For most enterprise-level businesses, the demand for data storage is growing much faster than the rate at which the price for storage is shrinking. As a result, you could be forced to increase your budget dramatically to keep up with data demands. This intelligent, software-defined storage solution powered by [SUSE Enterprise Storage](#) technology and composable HPE Synergy system hardware enables you to transform your [enterprise storage infrastructure](#) to reduce costs while providing unlimited scalability to keep up with your future demands. With this completely tested and certified approach, you will have the confidence to deploy a working solution in an agile manner and be able to maintain and scale it over time without capacity-based increases in software subscriptions.

Target audience

This document is intended for IT decision makers as well as architects, system administrators, and field technicians who implement the [HPE Synergy](#) platform and need a flexible, software-defined storage solution that can provide multiple protocol access. The reader should have a solid understanding of their storage use cases along with sizing and characterization concepts and limitations within their environments.

Solution overview

The coupling of a [composable infrastructure platform](#), such as HPE Synergy, and a software-defined storage solution, such as SUSE Enterprise Storage, provides an incredibly powerful and flexible combination. Nodes and storage capacity can be quickly added, replaced, or substituted over time as demands dictate. By utilizing the composable aspect provided by [HPE OneView](#) server profile templates, confidence in the deployment's consistency is guaranteed along with an increase in overall simplicity. Given the management paradigm can extend across multiple HPE Synergy frames, this solution can be easily scaled, both from the storage service and customer consumption access models. Use cases for such a flexible solution range from dynamically allocated storage pools for a virtual environment, or a custom testing and development solution, to more specific use cases such as SUSE OpenStack® Cloud and many other software systems.

In order to provide some focus, this solution focuses on a single use case of providing guest virtual machines, running on a Linux®-based KVM hypervisor host with access to block devices for their root filesystem. This is accomplished through the libvirt interaction to KVM and librbd to the storage cluster.

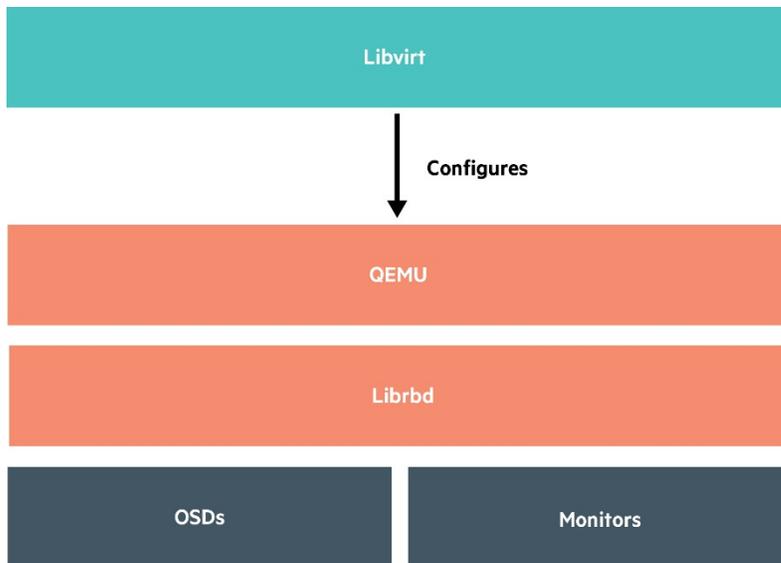


Figure 1. Libvirt overview

HPE Synergy

HPE Synergy, the first composable infrastructure, empowers IT to create and deliver new value easily and continuously. This single infrastructure reduces operational complexity for traditional workloads and increases operational velocity for the new breed of applications and services. Through a single interface, HPE Synergy composes compute, storage, and fabric pools into any configuration for any application. It also enables a broad range of workloads from bare metal to virtual machines to containers, and operational models like hybrid cloud and DevOps. HPE Synergy enables IT to rapidly react to new business demands.

HPE Synergy frames contain a management appliance called the HPE Synergy Composer, which hosts HPE OneView. HPE Synergy Composer manages the composable infrastructure and delivers

- Fluid pools of resources, where a single infrastructure of compute, storage, and fabric boots up ready for workloads
- Software-defined intelligence, with a single interface that precisely composes logical infrastructure at near-instant speeds, and demonstrates template-driven, frictionless operations
- Access to the infrastructure via a Unified API, which enables simple line-of-code programming of every infrastructure element, easily automates IT operational processes, and effortlessly automates applications through infrastructure deployment

HPE Synergy Composer

HPE Synergy Composer provides the enterprise-level management to compose and deploy system resources to your application needs. This management appliance uses software-defined intelligence with embedded HPE OneView to aggregate compute, storage, and fabric resources in a manner that scales to your application needs, instead of being restricted to the fixed ratios of traditional resource offerings. HPE Synergy template-based provisioning enables fast time to service with a single point for defining compute module state, pooled storage, network connectivity, and boot image.

HPE OneView is a comprehensive unifying management interface designed for converged infrastructure management. A unifying platform increases the productivity of every member of the internal IT team across servers, storage, and networking. By streamlining processes, incorporating best practices, and creating a new holistic way to work, HPE OneView provides organizations with a more efficient way to work. It is designed for open integration with existing tools and processes to extend these efficiencies.

HPE OneView is instrumental for the deployment and management of HPE servers and enclosure networking. It collapses infrastructure management tools into a single resource-oriented architecture that provides direct access to all logical and physical resources of the solution. Logical resources include server profiles and server profile templates, enclosures and enclosure groups, and logical interconnects and logical interconnect groups. Physical resources include server hardware blades and rack servers, networking interconnects, and computing resources.

HPE OneView converged infrastructure management console offers a uniform way for administrators to interact with resources by providing a RESTful API foundation. RESTful APIs enable administrators to utilize a growing ecosystem of integrations to further expand the advantages of the integrated resource model. It removes the need for the administrator to enter and maintain the same configuration data more than once and keep all versions up to date. It encapsulates and abstracts many underlying tools behind the integrated resource model, so the administrator can operate with new levels of simplicity, speed, and agility to provision, monitor, and maintain the solution.

HPE Synergy 12000 Frame

HPE Synergy 12000 Frame is a key element of HPE Synergy, providing the base for an intelligent infrastructure with embedded management and scalable links for expansion as business demands require. HPE Synergy 12000 Frame is the base infrastructure that pools resources of compute, storage, fabric, cooling, power, and scalability. With an embedded management solution combining HPE Synergy Composer and HPE Synergy Frame Link Modules, IT can manage, assemble, and scale resources on demand. The HPE Synergy 12000 Frame is designed for needs now and in the future with expanded compute and fabric bandwidths. HPE Synergy 12000 Frame specifications can be found [here](#). For more information on HPE Synergy architecture and components, visit the [HPE Synergy website](#).

HPE Synergy 480 Compute Module

HPE Synergy 480 Gen9 Compute Module delivers superior capacity, efficiency, and flexibility in a 2-socket, half-height form factor to support demanding workloads. HPE Synergy 480 Gen9 Compute Module provides a composable compute resource that can be auto-discovered, quickly provisioned, easily managed, and seamlessly redeployed to deliver the right compute capacity for changing workload needs. Powered by Intel® Xeon® E5-2600 v4 processors, HPE DDR4 SmartMemory supports up to 3TB flexible storage controller options, three I/O connectors, and GPU support. Designed to create a pool of flexible compute capacity within a composable infrastructure, HPE Synergy 480 Gen9 Compute Module is the ideal platform for general-purpose enterprise workload performance now and in the future. More information on HPE Synergy 480 Compute Modules can be found at the [HPE Synergy 480 Compute Module webpage](#).

HPE Synergy D3940 Storage Module

[HPE Synergy D3940 Storage Module](#) provides a fluid pool of storage resources for composable infrastructure. Additional capacity for compute modules is easily provisioned and intelligently managed with integrated data services for availability and protection. The 40 SFF drive bays per storage module can be populated with SSD or HDD storage in both SAS and SATA options to achieve the right cost per performance.

HPE Synergy D3940 Storage Module provides local storage to compute resources and can meet the demands of a wide range of data workloads. Pooled storage resources provide the flexibility and performance needed to accommodate a wide range of workloads. Changes such as updating firmware are automatically implemented, with the infrastructure online significantly reducing errors and delivering real-time compliance.

SUSE Enterprise Storage

SUSE Enterprise Storage solutions offer a single, unified software-defined storage cluster that provides unified object, block, and file storage. Designed with unlimited scalability from terabytes to petabytes and no single point of failure, SUSE Enterprise Storage maximizes system resiliency and application availability in the event of unexpected hardware issues or planned maintenance. A SUSE Enterprise Storage solution is

- Simple to set up and deploy, even more so within the documented guidelines for HPE Synergy's system hardware, networking, and management capabilities
- Resilient and adaptable to the physical and logical constraints needed by the business, both initially and as needed over time for performance, security, or scalability concerns
- Capable of providing optimized object, block, and file services to client access nodes, either directly or through gateway services

Solution components



Figure 2. HPE Synergy frame—front view

The configuration illustrated in this paper use HPE hardware, HPE software, and third-party components. Figure 2 shows the front view of HPE Synergy platform just for reference.

1. HPE Synergy 12000 Frame: 10U size
2. HPE Synergy Storage Module: Half-height doublewide
3. HPE Synergy 480 Compute Modules: Half-height singlewide
4. HPE Synergy 660, 620, and 680 Compute Modules: Full height and doublewide (not used in this solution)
5. Redundant management appliance modules: For HPE Synergy Composer
6. Front panel / HPE Synergy Console access ports

See the following software and hardware components lists used during testing of this solution.

Software

Table 1 lists the software used during testing of this solution.

Table 1. Software list

Component	Version
SUSE Enterprise Storage Base Configuration	4

Hardware

Table 2 lists the hardware used during testing of this solution.

Table 2. HPE Synergy components including storage options

Component	Description
HPE Synergy D3940 Storage Module	2X with 8 x HPE 800GB 6G SATA drives per storage module
HPE Synergy 480 Compute Module	8X (1 module for Administration Server, 3 modules for Monitor Nodes, and 4 modules for OSD Nodes)
HPE Synergy 12Gb SAS Connection Module	2 (redundant)
HPE Virtual Connect SE 40Gb F8 Module for Synergy	2 (redundant)
HPE Synergy 12000 Frame	1
HPE Synergy Composer	2 (redundant)

SUSE Enterprise Storage

SUSE Enterprise Storage provides unified object, block, and file access based on Ceph, an open source distributed storage software designed for scalability, reliability, and performance. As opposed to conventional systems, which have allocation tables to store and fetch data, Ceph uses a pseudo-random data distribution function, which reduces the number of lookups required. In addition to the required network interfaces, switches, and desired topology, the minimum Ceph storage cluster includes one Administration Server, a minimum of four object storage device (OSD) Nodes, three Monitor Nodes, and potentially one or more gateway systems.

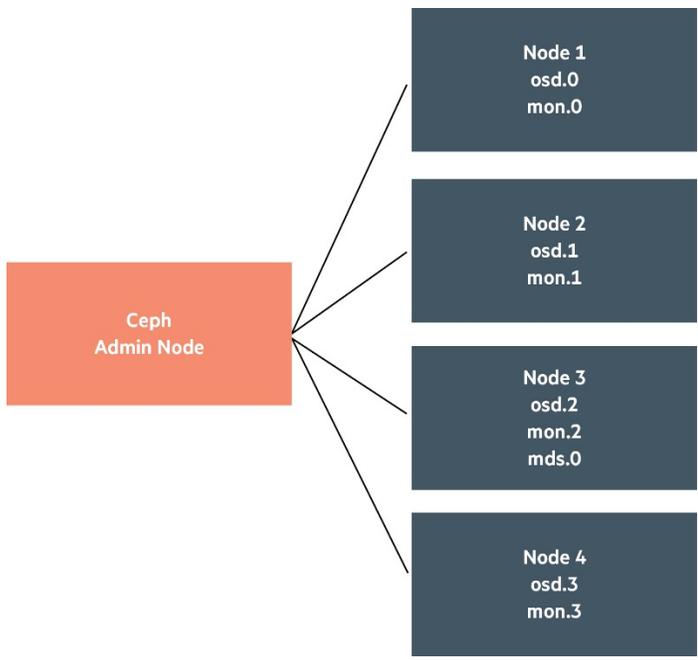


Figure 3. Simple Ceph Node organization

- The Administration Server is used to deploy and configure SUSE Enterprise Storage on the other nodes that will be OSDs, Monitor Nodes, and Object Gateways. Storage management services, like Calamari or openATTIC, which are web-based interfaces for Ceph, can also be hosted on the Administration Server although this is beyond the scope and requirements of this best practice guide. In this implementation, an HPE Synergy 480 Gen9 Compute Module was utilized as the Administration Server.
- Data is stored on hosts containing one or more intelligent object storage devices (OSDs). OSDs automate data management tasks such as data distribution, data replication, failure detection, and recovery using the CRUSH algorithm. The CRUSH algorithm determines how to store and retrieve data by computing data storage locations. CRUSH empowers Ceph clients to communicate with OSDs directly, rather than through a centralized server or broker. For each of the four OSD Nodes, HPE Synergy 480 Gen9 Compute Modules were utilized along with journal and data disk drives allocated from HPE Synergy D3940 Storage Modules.
- The monitors (MONs) maintain maps of the cluster state, including the monitor map, the OSD map, the Placement Group (PG) map, and the CRUSH map. Ceph maintains a history (called an epoch) of each state change in the Ceph Monitors, Ceph OSD Daemons, and Placement Groups. For high availability, at least three monitors should be run on a Ceph cluster, with HPE Synergy 480 Gen9 Compute Modules utilized for this role.
- A Ceph Object Gateway is an object storage interface that supports S3-compatible and Swift-compatible object storage functionality, using either the Amazon S3 or the OpenStack Swift RESTful APIs. For this best practice with the noted focus of providing block access to other virtual machine guests, no gateways were implemented or needed.

Considerations for building this software-defined storage solution

Before you start building your storage service, there are several recommended considerations that you must take into account to make your deployment successful.

Evaluation plan

The success of any storage service deployment greatly depends on a robust and fully thought out evaluation plan. You must determine the goals and access protocols and expected outcomes for your desired use cases. This will significantly influence how your hardware and software components will be procured, configured, and tuned.

A successful evaluation plan starts by assessing workloads, applications, and user experience requirements in your current environment. Storage deployments in general are application centric. Hardware and software configurations are largely dependent on the applications being used as well as user experience requirements. For the existing, performance data can be readily available and should be analyzed prior to any deployments. For greenfield deployments, a proof of concept is the recommended method of evaluation. Performing this analysis up front exposes system limitations and optimizes system performance and user experience. It also ensures smooth transition to production environments and eliminates surprises. Hewlett Packard Enterprise offers a host of business services to assist customers with developing and deploying cost-effective, software-defined storage environments.

Infrastructure requirements

Choosing what hardware to use for your storage deployment determines performance as well as ROI. In general, hardware requirements are an output of your evaluation plan and analysis results. This best practice is intended to complement the [SUSE Enterprise Storage™ Architectural Overview and Recommendations](#), which presents the concepts behind software-defined storage and the underlying technology of Ceph along with guidelines on hardware requirements and ratios. You are encouraged to read and understand that overview.

Deployment considerations

In addition, the following details are important to understand and taken into account during the planning and deployment cycles:

- While this best practice guide describes the use of a single HPE Synergy frame as a starting point to increase the tolerance and resiliency against outages, it is better to distribute the node types among multiple frames to reduce the impact of any single point of failure.
- The Best Practices section of the [SUSE Enterprise Storage 4 Administration and Deployment Guide](#) contains a wealth of general knowledge and recommendations on diagnosis and tuning aspects.
 - With the default replication setting of three, remember that the client-facing network will have about half or less of the traffic of the back end network. This is especially true when component failures occur or rebalancing happens on the OSD Nodes. For this reason, it is important not to underprovision this critical cluster and service resource.
 - It is important to maintain the minimum number of Monitor Nodes at three. As the cluster increases in size, it is best to increment it in pairs, keeping the total number of Monitor Nodes as an odd number. However, only large or very distributed clusters would likely need beyond the three Monitor Nodes cited in this reference implementation. For performance reasons, it is recommended to use distinct nodes for the monitor roles so that the OSD Nodes can be scaled as capacity requirements dictate.

Solution details

As mentioned previously, the focus of this best practice document is to create a software-defined storage service to provide block device access to virtual machine guests running on Linux KVM hypervisor hosts. In this example, a single HPE Synergy frame can provide this entire storage infrastructure and the hypervisor host and virtual machines would reside on other systems or frames.

Best practices

- Ensure all nodes within the respective HPE Synergy frame or frames have been discovered and accurately show the expected resources:
 - Validate the necessary CPU, memory, and interconnect quantity and type are present for each node and intended role
 - Ensure there are a pair of local, direct attached disk drives present, with a preference for SSDs here, on each node, which will later become the target for the operating system installation
 - Verify the HPE Synergy D3940 Storage Module has the expected number of drives and technologies present
- Prepare those attributes from the [HPE OneView Composer](#) interface, which are framewide:
 - **Network:** Prepare an IP addressing scheme and create both a storage cluster public and private network, along with the desired subnets and VLAN designations. Optionally, create a **Network Set** including both networks to make it easy to apply to each node
 - **Interconnects:** Ensure the necessary uplink ports are present and configured to access other frames and shared network infrastructure
- Prepare a **server profile template** for each of the three types of nodes—Administration Server, Monitor, OSD—with the following notable attributes:
 - **Connections:** Associate both the public and private networks, or the designated **Network Set** to the higher numbered Mezzanine Ports (leaving the “a” port unassociated for possible future usages)
 - **Local storage:**
 - For Administration Server and Monitor, configure the integrated local storage to create a RAID 1 LUN across the two local disk drives.
 - For the OSD Nodes, leave as configure manually and interactively use the Intelligent Provisioning boot option to create a RAID 1 LUN across the two local drives, prior to applying this template or installing the operating system. For the Mezzanine 1 storage controller, use HBA mode and add the desired number and type of drives for later use as journal, cache, and data drives. Follow the recommended ratios for journal:data from the [SUSE Enterprise Storage documentation](#).
 - **Boot settings:** Manage the boot node and select UEFI mode, with the primary device being hard disk
- Apply the respective **server profile template** to each node, address any issues until each has successfully been applied
- Following the [SUSE Enterprise Storage 4 Administration and Deployment Guide](#), begin installation of the operating system on each node type, starting with Administration Server, then Monitor Nodes, and finally OSD Nodes. Using this best practice document as a complement to that guide, only notable design decisions, material choices, and specific configuration changes that might differ from the published defaults are highlighted in the remainder of this document.
 - Ensure that you have access to a valid, reliable NTP service. This is a critical requirement in order for all nodes in the SUSE Enterprise Storage cluster to be maintained in time sync.
 - Set up and reserve DNS A records for the storage nodes.
 - Set up or use an existing SUSE Subscription Management Tool (SMT) system. This service provides a local mirror of the SUSE product package repositories from the SUSE Customer Center, allowing for rapid software deployment and later updating of packages.
 - For larger environments, SUSE Manager enables administrators to manage many Linux systems and keep them up to date, so SUSE Manager could be used in place of the SMT solution.
 - Include only the minimal pattern and components according to the procedure from deployment. This can be accomplished in any number of ways: with the virtual media option through HPE iLO or from a PXE network-boot environment.

Note

Currently HPE Synergy Image Streamer does not support the deployment of the SUSE Linux Enterprise Server operating system with the default filesystem of Btrfs.

- Use the suggested, default partitioning scheme on each node and validate that the target LUN for the operating system installation corresponds to the RAID 1 pair of local disk drives.

Note

Ceph does not currently support multipath access to disk drives provided by HPE Synergy D3940 Storage Modules. During the operating system installation, avoid enabling multipath disk support.

- After the operating system installation is complete across all the nodes, preform the following checks:
 - Ensure that each node has access to the necessary software repositories for later operations and updates. It is suggested that you apply all software updates, via **zypper up**.
 - NTP is configured and operational, synchronizing with a source outside the cluster via **ntpq -pn**.
 - If necessary, adjust the udev rules to ensure that network interfaces are identified (as needed) in the same logical order across the systems to make later steps easier. Ensure that the respective network interfaces are bonded together with the associated VLANs configured. While configuring these interfaces, it is also convenient to disable IPv6 functionality and the firewall on each node.

Note

For environments that require firewalls to be in place, see the Firewall Settings for Ceph section of the [SUSE Enterprise Storage 4 Administration and Deployment Guide](#) for detailed configuration settings.

- Begin the deployment of the storage cluster
 - To be more explicit in the access of the drives, which may show multiple access paths, for the OSD Nodes, the **ceph-deploy** command-line process will be used throughout the deployment, so see only those procedures.
 - A cephadm user is configured with passwordless SSH access from the Administration Server.
 - The Running ceph-deploy section of the [SUSE Enterprise Storage 4 Administration and Deployment Guide](#), and other portions noted in the following, will guide you through the steps to
 - Ensure that ceph-deploy is present on each node and distribute the necessary keyring and configuration files.
 - Set up each of the three Monitor Nodes and ensure that they are in a healthy, functional state via **ceph mon status**.

Note

An additional recommendation is to set up the Monitor Nodes as NTP peers.

- Set up each of the four OSD Nodes. Because SSD devices are present to act as journals, ensure that you invoke the **ceph-deploy** command with the appropriate syntax as you iterate across the devices to include this functionality: **ceph-deploy osd prepare HOST:DISK[:JOURNAL]**

Notes

1. Select a device designation via either access path to a drive or by alternating between the two paths, but avoid configuring both access paths
 2. As a reminder, there is no reservation system that explicitly ties any given drive to any specific node. So drives may not be reserved, and it is best to assume that any drive had a previous usage. Utilize the ceph-deploy disk zap function prior to including it into an active configuration
-

- As you progress, check the cluster status via `ceph health`.

Note

Upon completion of an OSD Node, verify that only the desired `ceph-osd` processes running are those associated with the desired disk paths. Remove any possible duplicates associated with the alternative paths, via

```
systemctl stop ceph-osd@###.service
systemctl disable ceph-osd@###.service
```

- Before proceeding on to utilize the cluster's service from other nodes, validate the health and functionality from any of the cluster nodes via:
 - `ceph health`
 - `ceph status`
 - `ceph osd pool create test 4096`
 - `rados bench -p test 300 write --no-cleanup`
 - `rados bench -p test 300 seq`

Note

After validation is complete, remove the test pool via `ceph osd pool delete test -yes-i-really-really-mean-it`

- At this point, you are ready to begin creating the block device and the virtual machines that use this storage service. See the Integration with Virtualization Tools section of the [SUSE Enterprise Storage 4 Administration and Deployment Guide](#). On the Linux KVM hypervisor host, assuming it is running SUSE Linux Enterprise Server 12 SP2
 - Install the Ceph package and copy the `/etc/ceph` configuration directory from the storage cluster to this virtualization host. Validate with **ceph health** to ensure you are interacting properly with the cluster
 - Configure the necessary block devices by following the Using libvirt with Ceph section of the [SUSE Enterprise Storage 4 Administration and Deployment Guide](#)
 - Create new virtual machines and configure them to utilize the newly accessible block device provided by the storage cluster and begin their installation

Summary

With the continued exponential growth of storage requirements in the world today, the object-based software-defined storage architecture provided by SUSE Enterprise Storage is a powerful complement to the HPE Synergy platform. This combined solution expands on the flexibility of a composable infrastructure and allows for an unmatched scalability model with regard to both storage capacity and performance customization. Customers are able to enjoy virtually unlimited capacity, while also reducing CAPEX and OPEX for your storage infrastructure with a truly unified block, object, and file solution.

For many businesses, the performance and customization available will mean that an HPE Synergy environment with SUSE Enterprise Storage will be able to solve a multitude of mission-critical data center issues. The HPE-SUSE partnership also brings a support infrastructure similar to what one would expect only from a purpose-built appliance model. By sharing the engineering resources between both products, we expect even greater things to come as both sides continue to work together to create the products of tomorrow.

Resources

[HPE Synergy](#)

[HPE OneView](#)

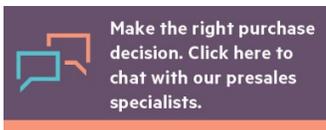
[HPE Software Defined Storage](#)

[HPE Storage](#)

[SUSE Enterprise Storage documentation](#)

To help us improve our documents, please provide feedback at hpe.com/contact/feedback.

Learn more at
hpe.com/info/synergy



Sign up for updates

© Copyright 2017 Hewlett Packard Enterprise Development LP. The information contained herein is subject to change without notice. The only warranties for Hewlett Packard Enterprise products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. Hewlett Packard Enterprise shall not be liable for technical or editorial errors or omissions contained herein.

Intel Xeon is a trademark of Intel Corporation in the U.S. and other countries. The OpenStack Word Mark is either a registered trademark/service mark or trademark/service mark of the OpenStack Foundation, in the United States and other countries and is used with the OpenStack Foundation's permission. We are not affiliated with, endorsed or sponsored by the OpenStack Foundation or the OpenStack community. Linux is the registered trademark of Linus Torvalds in the U.S. and other countries. All other third-party trademark(s) is/are property of their respective owner(s).

a00022567ENW, August 2017