Reference Architecture: Red Hat OpenStack Platform with ThinkSystem Servers

Last update: 26 September 2017

Version 1.0

Provides both economic and high performance options for cloud workloads

Describes Lenovo ThinkSystem servers, networking, and systems management software

Describes architecture for high availability and distributed functionality

Includes validated and tested deployment and sizing guide for quick start

Jiang Xiaotong
Xu Lin
Mike Perks
Yixuan Huang
Srihari Angaluri

LENOVO PRESS

Table of Contents

1 Introduction

OpenStack continues to gain significant traction in the industry because of the growing adoption of cloud usage and the flexibility OpenStack offers as an open source product. This document describes the reference architecture (RA) for deploying the Red Hat OpenStack Platform on industry leading servers, storage, networking, and systems management tools from Lenovo.

Lenovo and Red Hat have collaborated to promote best practices and validate reference architecture for deploying private cloud infrastructures by leveraging the Red Hat OpenStack Platform 11. This is an innovative, cost-effective cloud management solution that includes automation, metering, and security. The Red Hat OpenStack Platform provides a foundation to build a private or public Infrastructure as a Service (IaaS) cloud on top of Red Hat Enterprise Linux and offers a scalable, highly available platform for the management of cloud-enabled workloads. Red Hat OpenStack Platform version 11 is based on the OpenStack Ocata release that includes features such as the Red Hat OpenStack Platform 11 Director that can deploy the cloud environment to bare metal systems and support for high availability (HA). It benefits from the improved overall quality of the open source Ocata release.

The Lenovo hardware platform provides an ideal infrastructure solution for cloud deployments. These servers provide the full range of form factors, features and functions that are necessary to meet the needs of small businesses all the way up to large enterprises. Lenovo uses industry standards in systems management on all these platforms, which enables seamless integration into cloud management tools such as OpenStack. Lenovo also provides data center network switches that are designed specifically for robust, scale-out server configurations and converged storage interconnect fabrics.

The target environments for this reference architecture include Managed Service Providers (MSPs), Cloud Service Providers (CSPs), and enterprise private clouds that require a complete solution for cloud IaaS deployment and management based on OpenStack.

This document features planning, design considerations, performance, and workload density for implementing Red Hat OpenStack Platform on the Lenovo hardware platform. The RA enables organizations to easily adopt OpenStack to simplify the system architecture, reduce deployment and maintenance costs, and address the lack of convergence in infrastructure management. The reference architecture focuses on achieving optimal performance and workload density by using Lenovo hardware while minimizing procurement costs and operational overhead.

Deploying OpenStack and other Red Hat branded software such as Red Hat CloudForms on the Lenovo hardware platform to instantiate compute, network, storage, and management services. The Lenovo XClarity™ Administrator solution consolidates systems management across multiple Lenovo servers that span the data center. XClarity enables automation of firmware updates on servers via compliance policies, patterns for system configuration settings, hardware inventory, bare-metal OS and hypervisor provisioning, and continuous hardware monitoring. XClarity easily extends via the published REST API to integrate into other management tools, such as OpenStack.

The intended audience for this document is IT professionals, technical architects, sales engineers, and consultants. Readers should have a basic knowledge of Red Hat Enterprise Linux and OpenStack.

2 Business problem and business value

This chapter outlines the value proposition of the Red Hat OpenStack Platform with Lenovo hardware.

2.1 Business problem

Virtualization and cloud have achieved tremendous growth in recent years. Initially, virtualization provided immediate relief to server sprawl by enabling consolidation of multiple workloads onto a single server. As the hypervisor space evolved and cloud proved to be a more cost-effective means for deploying IT, independent software vendors (ISVs) moved up the software stack as a means of differentiating themselves while driving customer lock-in and profit.

MSPs and CSPs face intense challenges that drive them to look for economic but scalable infrastructure solutions that enable cost-effective IT services while easily expanding capacity to meet user demand. OpenStack provides an open source alternative for cloud management that is increasingly used to build and manage cloud infrastructures to support enterprise operations. The OpenStack free license reduces the initial acquisition and expansion costs. However, IT Administrators require robust skill sets to build an advanced cloud service based on OpenStack's flexibility and various deployment and configuration options. Moreover, scale out of such a cloud cluster introduces risk of an unbalanced infrastructure that may lead to performance issues, insufficient network bandwidth, or increased exposure to security vulnerabilities.

2.2 Business value

The Red Hat OpenStack Platform reference architecture solves the problems described in section 2.1 by providing a blueprint for accelerating the design, piloting, and deployment process of an enterprise-level OpenStack cloud environment on Lenovo hardware. The reference architecture reduces the complexity of OpenStack deployments by outlining a validated system configuration that scales and delivers an enterprise level of redundancy across servers, storage and networking to help enable HA. This document provides the following benefits:

- Consolidated and fully integrated hardware resources with balanced workloads for compute, network, and storage.
- An aggregation of compute and storage hardware, which delivers a single, virtualized resource pool customizable to different compute and storage ratios to meet the requirements of various solutions.
- Ease of scaling (vertical and horizontal) compute and storage resources at runtime based on business requirements.
- Elimination of single points of failure in every layer by delivering continuous access to virtual machines (VMs).
- Hardware redundancy and full utilization.
- Rapid OpenStack cloud deployment, including updates, patches, security, and usability enhancements with enterprise-level support from Red Hat and Lenovo.
- Unified management and monitoring for VMs.

3 Requirements

The requirements for a robust cloud implementation are described in this section.

3.1 Functional requirements

Table 1 lists the functional requirements for a cloud implementation.

Table 1. Functional requirements

Requirement	Description	Supported by
Mobility	Workload is not tied to any physical location	• Enabled VM is booted from distributed storage and runs on different hosts • Live migration of running VMs • Rescue mode support for host maintenance
Resource provisioning	Physical servers, virtual machines, virtual storage, and virtual network can be provisioned on demand	• OpenStack compute service • OpenStack block storage service • OpenStack network service • OpenStack bare-metal provisioning service
Management portal	Web-based dashboard for workloads management	OpenStack dashboard (Horizon) for most routine management operations
Multi-tenancy	Resources are segmented based on tenancy	Built-in segmentation and multi-tenancy in OpenStack
Metering	Collect measurements of used resources to allow billing	OpenStack metering service (Ceilometer)

3.2 Non-functional requirements

Table 2 lists the non-functional requirements for MSP or CSP cloud implementation.

Table 2. Non-functional requirements

Requirement	Description	Supported by
OpenStack environment	Supports the current OpenStack edition	OpenStack Ocata release through Red Hat OpenStack Platform 11
Scalability	Solution components can scale for growth	Compute nodes and storage nodes can be scaled independently within a rack or across racks without service downtime

Requirement	Description	Supported by
Load balancing	Workload is distributed evenly across servers	• Network interfaces are teamed and load balanced • Use of OpenStack scheduler for balancing compute and storage resources • Data blocks are distributed across storage nodes and can be rebalanced on node failure
High availability	Single component failure will not lead to whole system unavailability	• Hardware architecture ensures that computing service, storage service, and network service are automatically switched to remaining components • Controller node, compute node, and storage node are redundant • Data is stored on multiple servers and accessible from any one of them; therefore, no single server failure can cause loss of data • Virtual machines are persistent on shared storage service
Mobility	VM can be migrated or evacuated to different hosting server	• VM migration • VM evacuation
Physical footprint	Compact solution	• Lenovo System x server, network devices, and software are integrated into one rack with validated performance and reliability • Provides 1U compute node option
Ease of installation	Reduced complexity for solution deployment	• A dedicated deployment server with web-based deployment tool and rich command line provide greater flexibility and control over how you deploy OpenStack in your cloud • Optional deployment services

Requirement	Description	Supported by
Support	Available vendor support	• Hardware warranty and software support are included with component products • Standard or Premium support from Red Hat included with Red Hat OpenStack Platform subscription
Flexibility	Solution supports variable deployment methodologies	• Hardware and software components can be modified or customized to meet various unique customer requirements • Provides local and shared storage for workload
Robustness	Solution continuously works without routine supervision	• Red Hat OpenStack Platform 11 is integrated and validated on Red Hat Enterprise Linux 7.4 • Integration tests on hardware and software components
Security	Solution provides means to secure customer infrastructure	• Security is integrated in the Lenovo System x hardware with System x Trusted Platform Assurance, an exclusive set of industry-leading security features and practices • SELinux is enabled and in enforcing mode by default in Red Hat OpenStack Platform 11. • Network isolation using virtual LAN (VLAN) and virtual extensible LAN (VXLAN).
High performance	Solution components are high-performance	Provides 40 - 90 average workloads (2 vCPU, 8 GB vRAM, 80 GB disk) per host.

4 Architectural overview

The Red Hat OpenStack Platform 11 is based on the OpenStack Ocata release. This platform offers the innovation of the OpenStack community project and provides the security, stability, and enterprise-readiness of a platform built on Red Hat Enterprise Linux. Red Hat OpenStack Platform supports the configuration of data center physical hardware into private, public, or hybrid cloud platforms and has the following benefits:

- Fully distributed storage
- Persistent block-level storage
- VM provisioning engine and image storage
- Authentication and authorization mechanisms
- Integrated networking
- A web-based GUI for users and administration

Figure 1 shows the main components in Red Hat OpenStack Platform, which is a collection of interacting services that control the compute, storage, and networking resources.

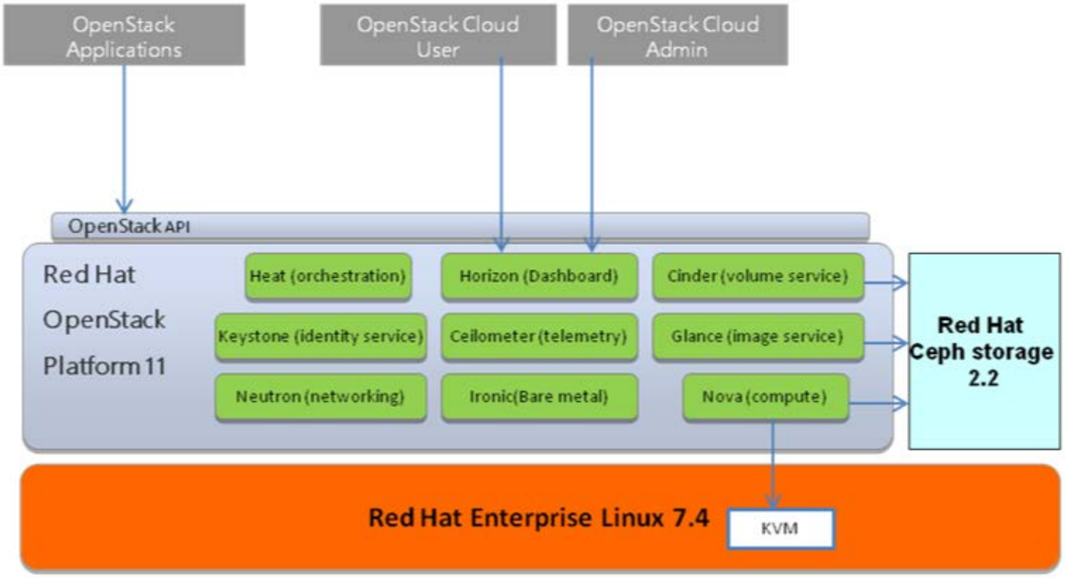


Figure 1. Overview of Red Hat OpenStack Platform

Administrators use a web-based interface to control, provision, and automate OpenStack resources. An extensive API available to end users of the cloud facilitates programmatic access to the OpenStack infrastructure.

5 Component model

The section describes the components that used in OpenStack, specifically the Red Hat OpenStack Platform distribution.

5.1 Core Red Hat OpenStack Platform components

Figure 2 shows the core components of Red Hat OpenStack Platform. It does not include some optional add-on components listed in the “Third-party components” section on page 9.

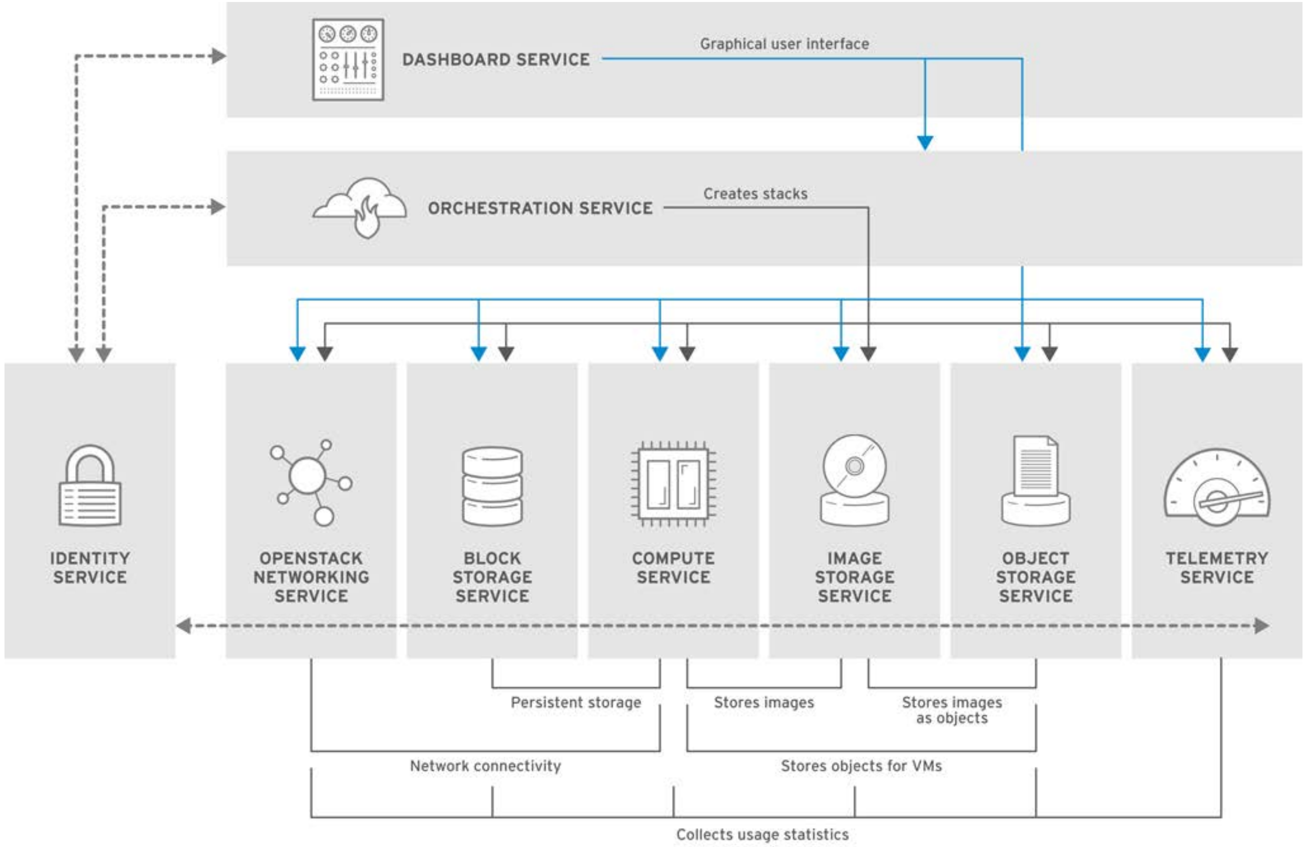


Figure 2. Components of Red Hat OpenStack Platform

Table 3 lists the core components of Red Hat OpenStack Platform as shown in Figure 2.

Table 3. Core components

Component	Code name	Description
Compute service	Nova	Provisions and manages VMs, which creates a redundant and horizontally scalable cloud-computing platform. It is hardware and hypervisor independent and has a distributed and asynchronous architecture that provides HA and tenant-based isolation.

Component	Code name	Description
Block storage service	Cinder	Provides persistent block storage for VM instances. The ephemeral storage of deployed instances is non-persistent; therefore, any data generated by the instance is destroyed after the instance terminates. Cinder uses persistent volumes attached to instances for data longevity, and instances can boot from a Cinder volume rather than from a local image.
Networking service	Neutron	OpenStack Networking is a pluggable “networking as a service” framework for managing networks and IP addresses. This framework supports several flexible network models, including Dynamic Host Configuration Protocol (DHCP) and VLAN.
Image service	Glance	Provides discovery, registration, and delivery services for virtual disk images. The images can be stored on multiple back-end storage units and cached locally to reduce image staging time.
Object storage service	Swift	Provides cloud storage software built for scale and optimized for durability, availability, and concurrency across the entire data set. It can store and retrieve data with a simple API, and is ideal for storing unstructured data that can grow without bound. (Red Hat Ceph Storage is used in this reference architecture, instead of Swift, to provide the object storage service)
Identity service	Keystone	Centralized service for authentication and authorization of OpenStack service and for managing users, projects and roles. Identity supports multiple authentication mechanisms, including user.
Telemetry service	Ceilometer	Provides infrastructure to collect measurements within OpenStack. Delivers a unique point of contact for billing systems to acquire all of the needed measurements to establish customer billing across all current OpenStack core components. An administrator can configure the type of collected data to meet operating requirements. Gnocchi is a multi-tenant, metrics and resource database that is as ceilometer backend.
Dashboard service	Horizon	Dashboard provides a graphical user interface for users and administrator to perform operations such as creating and launching instances, managing networking, and setting access control.
Orchestration	Heat	Heat is an orchestration engine to start multiple composite cloud applications based on templates in the form of text files. AWS CloudFormation templates are one example.
File Share Service	Manila	A file share service that presents the management of file shares (for example, NFS and CIFS) as a core service to OpenStack.

Table 4 lists the optional components in the Red Hat OpenStack Platform release. Actual deployment use cases will determine when and how these components are used.

Table 4. Optional components

Component	Code name	Description
Bare-metal provisioning service	Ironic	OpenStack Bare Metal Provisioning enables the user to provision physical, or bare metal machines, for a variety of hardware vendors with hardware-specific drivers.
Data Processing	Sahara	Provides the provisioning and management of Hadoop cluster on OpenStack. Hadoop stores and analyzes large amounts of unstructured and structured data in clusters.

Table 5 lists the OpenStack concepts to help the administrator further manage the tenancy or segmentation in a cloud environment.

Table 5. OpenStack tenancy concepts

Name	Description
Tenant	The OpenStack system is designed to have multi-tenancy on a shared system. Tenants (also called projects) are isolated resources that consist of separate networks, volumes, instances, images, keys, and users. These resources can have quota controls applied on a per-tenant basis.
Availability Zone	In OpenStack, an availability zone allows a user to allocate new resources with defined placement. The “instance availability zone” defines the placement for allocation of VMs, and the “volume availability zone” defines the placement for allocation of virtual block storage devices.
Host Aggregate	A host aggregate further partitions an availability zone. It consists of key-value pairs assigned to groups of machines and used by the scheduler to enable advanced scheduling.
Region	Regions segregate the cloud into multiple compute deployments. Administrators use regions to divide a shared-infrastructure cloud into multiple sites each with separate API endpoints and without coordination between sites. Regions share the Keystone identity service, but each has a different API endpoint and a full Nova compute installation.

5.2 Third-party components

The following third-party components can also be used:

- MariaDB is open source database software shipped with Red Hat Enterprise Linux as a replacement for MySQL. MariaDB Galera cluster is a synchronous multi-master cluster for MariaDB. It uses synchronous replication between every instance in the cluster to achieve an active-active multi-master topology, which means every instance can accept data retrieving and storing requests and the failed nodes do not affect the function of the cluster.

- RabbitMQ is a robust open source messaging system based on the AMQP standard, and it is the default and recommended message broker in Red Hat OpenStack Platform.
- Memcached and Redis offer persistence and shared storage and speed up dynamic web applications by reducing the database load.

5.3 Red Hat Ceph Storage Component

Red Hat Ceph Storage is open source software from Red Hat that provides Exabyte-level scalable object, block, and file storage from a completely distributed computer cluster with self-healing and self-managing capabilities. Red Hat Ceph Storage virtualizes the pool of the block storage devices and stripes the virtual storage as objects across the servers.

Red Hat Ceph Storage is integrated with Red Hat OpenStack Platform. The OpenStack Cinder storage component and Glance image services can be implemented on top of the Ceph distributed storage.

OpenStack users and administrators can use the Horizon dashboard or the OpenStack command-line interface to request and use the storage resources without requiring knowledge of where the storage is deployed or how the block storage volume is allocated in a Ceph cluster.

The Nova, Cinder, Swift, and Glance services on the controller and compute nodes use the Ceph driver as the underlying implementation for storing the actual VM or image data. Ceph divides the data into placement groups to balance the workload of each storage device. Data blocks within a placement group are further distributed to logical storage units called Object Storage Devices (OSDs), which often are physical disks or drive partitions on a storage node.

The OpenStack services can use a Ceph cluster in the following ways:

- VM Images: OpenStack Glance manages images for VMs. The Glance service treats VM images as immutable binary blobs and can be uploaded to or downloaded from a Ceph cluster accordingly.
- Volumes: OpenStack Cinder manages volumes (that is, virtual block devices) attached to running VMs or used to boot VMs. Ceph serves as the back-end volume provider for Cinder.
- Object Storage: OpenStack Swift manages the unstructured data that can be stored and retrieved with a simple API. Ceph serves as the back-end volume provider for Swift.
- VM Disks: By default, when a VM boots, its drive appears as a file on the file system of the hypervisor. Alternatively, the VM disk can be in Ceph and the VM started using the boot-from-volume functionality of Cinder or directly started without the use of Cinder. The latter option is advantageous because it enables maintenance operations to be easily performed by using the live-migration process, but only if the VM uses the RAW disk format.

If the hypervisor fails, it is convenient to trigger the Nova `evacuate` function and almost seamlessly run the VM machine on another server. When the Ceph back end is enabled for both Glance and Nova, there is no need to cache an image from Glance to a local file, which saves time and local disk space. In addition, Ceph can implement a copy-on-write feature ensuring the start-up of an instance from a Glance image does not actually use any disk space.

When deploying the hyper-converged platform, Red Hat OpenStack Platform 11 integrates Red Hat Ceph Storage OSD services and compute services into a hyper-converged node.

5.4 Red Hat OpenStack Platform specific benefits

Red Hat OpenStack Platform 11 offers a better and easier OpenStack based cloud management platform built on Red Hat Enterprise Linux that includes one year of production support. For more details, please see: Red Hat OpenStack Platform Life Cycle and Red Hat OpenStack Platform Director Life Cycle.

Compared with previous Red Hat OpenStack Platform edition, Red Hat OpenStack Platform 11 offers three kinds of deployment:

- Legacy deployment mode: The compute node provides compute services and the Ceph storage node provides storage services.
- Pure HCI (Hyper Converged Infrastructure) mode: Hyper-converged nodes provide compute services and storage services.
- Mixed HCI (Hyper Converged Infrastructure) mode: A mixture of hyper-converged nodes and normal compute nodes.

This document focuses on legacy deployment mode and pure HCI deployment mode.

Listed below are the most important new features in the Red Hat OpenStack Platform 11. For more details, please see: Red Hat OpenStack Platform 11 Release Notes.

- Composable services upgrades on deployment and management. Each composable service template now contains logic to upgrade the service across major. This provides a mechanism to accommodate upgrades through the custom role and composable service architecture
- NFS Snapshots. The NFS back end driver for the Block Storage service now supports snapshots.
- Composable high availability service. The Red Hat OpenStack Platform director now opens the composable service architecture to include high availability services. This means users can split high availability services from the Controller node or scale services with dedicated custom roles.
- Placement API Service. This service is a separate REST API stack and data model that tracks the inventory and usage of resource providers (Compute nodes).
- Improved Parity with Core OpenStack Services. This release now supports domain-scoped tokens (required for identity management in Keystone V3). This release adds support for launching Nova instances attached to an SR-IOV port.
- VLAN-Aware VMs. Instances can now send and receive VLAN-tagged traffic. This ability is particularly useful for network function virtualization (NFV), which expects 802.1q VLAN-tagged traffic. This allows multiple customers/services to be served. This implementation has full support with OVS-based and OVS-DPDK-based networks.
- Red Hat OpenStack Platform 11 includes full support for performance monitoring (collectd), log aggregation (fluentd), and availability monitoring (sensu). These agents called composable services by Red Hat OpenStack Platform director and are configured with Heat templates during installation.

6 Operational model

Because OpenStack provides a great deal of flexibility, the operational model for OpenStack normally depends upon a thorough understanding of the requirements and needs of the cloud users to design the best possible configuration to meet the requirements. For more information, see the OpenStack Operations Guide.

This section describes the Red Hat OpenStack Platform operational model using Lenovo hardware and software. It concludes with some example deployment models that use Lenovo servers and Lenovo RackSwitch™ network switches.

6.1 Hardware

Rack Servers:

- Lenovo ThinkSystem SR630
- Lenovo ThinkSystem SR650

Switches:

- Lenovo RackSwitch G8124E
- Lenovo RackSwitch G8272
- Lenovo RackSwitch G7028

6.1.1 Rack servers introduction

Following sections describe the server options for OpenStack.

Lenovo ThinkSystem SR650

The Lenovo ThinkSystem SR650 server (as shown in Figure 3 and Figure 4) is an enterprise class 2U two-socket versatile server that incorporates outstanding reliability, availability, and serviceability (RAS), security, and high efficiency for business-critical applications and cloud deployments. Unique Lenovo AnyBay technology provides the flexibility to mix-and-match SAS/SATA HDDs/SSDs and NVMe SSDs in the same drive bays. Four direct-connect NVMe ports on the motherboard provide ultra-fast read/writes with NVMe drives and reduce costs by eliminating PCIe switch adapters. Plus, storage can be tiered for greater application performance, to provide the most cost-effective solution. ThinkSystem SR650 server support up to 1.5TB of TruDDR4 Memory currently, and up to 3TB of TruDDR4 Memory in the near future. Its on-board Ethernet solution provides 2/4 standard embedded Gigabit Ethernet ports and 2/4 optional embedded 10 Gigabit Ethernet ports without occupying PCIe slots.

Combined with the Intel® Xeon® Scalable processors product family, the Lenovo ThinkSystem SR650 server offers a high density of workloads and performance that is targeted to lower the total cost of ownership (TCO) per VM. Its flexible, pay-as-you-grow design and great expansion capabilities solidify dependability for any kind of virtualized workload, with minimal downtime. Additionally, it supports two 300W high-performance GPUs and ML2 NIC adapters with shared management.

The Lenovo ThinkSystem SR650 server provides internal storage density of up to 100 TB (with up to 26 x 2.5-inch drives) in a 2U form factor with its impressive array of workload-optimized storage configurations. The ThinkSystem SR650 offers easy management and saves floor space and power consumption for the most demanding storage virtualization use cases by consolidating the storage and server into one system.

The key differences compared to the SR630 server are more expansion slots and chassis to support up to twenty-four 2.5-inch or fourteen 3.5-inch hot-swappable SAS/SATA HDDs or SSDs together with up to eight on-board NVMe PCIe ports that allow direct connections to the U.2 NVMe PCIe SSDs. The ThinkSystem SR650 server also supports up to two NVIDIA GRID cards for graphics acceleration.

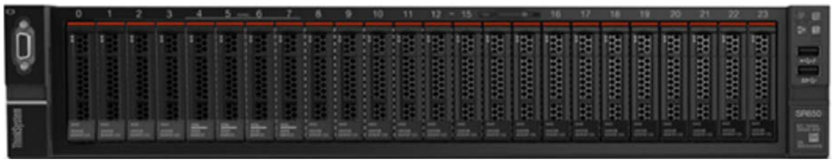


Figure 3. Lenovo ThinkSystem SR650 (with 24 x 2.5-inch disk bays)

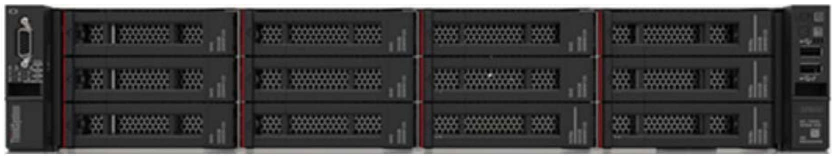


Figure 4. Lenovo ThinkSystem SR650 (with 12 x 3.5-inch disk bays)

For more information, see the following websites:

- [ThinkSystem SR650 Product Guide]

Lenovo ThinkSystem SR630

The Lenovo ThinkSystem SR630 server (as shown in Figure 5) is an ideal 2-socket 1U rack server for small businesses up to large enterprises that need industry-leading reliability, management, and security, as well as maximizing performance and flexibility for future growth. The SR630 server is designed to handle a wide range of workloads, such as databases, virtualization and cloud computing, virtual desktop infrastructure (VDI), infrastructure security, systems management, enterprise applications, collaboration/email, streaming media, web, and HPC. The ThinkSystem SR630 offers up to twelve 2.5-inch hot-swappable SAS/SATA HDDs or SSDs together with up to four on-board NVMe PCIe ports that allow direct connections to the U.2 NVMe PCIe SSDs.

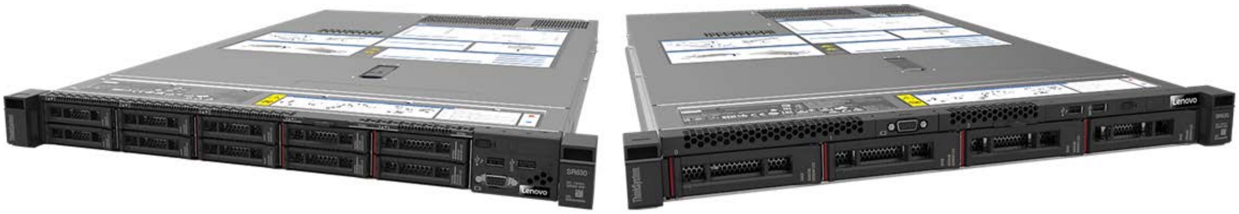


Figure 5. Lenovo ThinkSystem SR630

For more information, see the following websites:

- ThinkSystem SR630 Product Guide

6.1.2 Network switches introduction

Following sections describe the TOR switches used in this reference architecture. The 10 Gb switch is used for the internal and external network of Red Hat OpenStack Platform cluster, and 1 Gb switch is used for out-of-band server management.

Lenovo RackSwitch G8124E

The Lenovo RackSwitch™ G8124E (as shown in Figure 6) is a 10 GbE switch that is specifically designed for the data center and provides a virtualized, cooler, and easier network solution. The G8124E offers 24 10 GbE ports in a 1U footprint. Designed with top performance in mind, the RackSwitch G8124E provides line-rate, high-bandwidth switching, filtering, and traffic queuing without delaying data and large data-center grade buffers to keep traffic moving.

Figure 6. Lenovo RackSwitch G8124E

The G8124E switch is virtualized by providing rack-level virtualization of networking interfaces. The G8124E switch also supports Virtual Fabric, which allows for the distribution of a physical NIC into 2 to 8 vNICs and creates a virtual pipe between the adapter and the switch (for improved performance, availability and security, while reducing cost and complexity). The G8124E switch is easier to manage with server-oriented provisioning by using point-and-click management interfaces.

For more information, see the RackSwitch G8124E Product Guide.

Lenovo RackSwitch G8272

The Lenovo RackSwitch G8272 uses 10Gb SFP+ and 40Gb QSFP+ Ethernet technology and is specifically designed for the data center. It is an enterprise class Layer 2 and Layer 3 full featured switch that delivers line-rate, high-bandwidth switching, filtering, and traffic queuing without delaying data. Large data center-grade buffers help keep traffic moving, while the hot-swap redundant power supplies and fans (along with numerous high-availability features) help provide high availability for business sensitive traffic.

The RackSwitch G8272 (as shown in Figure 7) is ideal for latency sensitive applications, such as high-performance computing clusters and financial applications. In addition to the 10 Gb Ethernet (GbE) and 40 GbE connections, the G8272 can use 1 GbE connections. The G8272 supports the newest protocols, including Data Center Bridging/Converged Enhanced Ethernet (DCB/CEE) for Fibre Channel over Ethernet (FCoE), iSCSI and network-attached storage (NAS).

Figure 7. Lenovo RackSwitch G8272

The RackSwitch G8272 supports Lenovo Virtual Fabric, which helps clients significantly reduce cost and complexity that are related to I/O requirements of many virtualization deployments. Virtual Fabric helps reduce the number of multiple I/O adapters to a single dual-port 10 Gb adapter and the number of cables and required upstream switch ports.

By using Virtual Fabric, you can carve a dual-port 10 Gb server adapter into eight virtual network ports (vPorts) and create dedicated virtual pipes between the adapter and switch for optimal performance, higher availability, and improved security. With Virtual Fabric, you can make dynamic changes and allocate bandwidth per vPort so that you can adjust it over time without downtime.

For more information, see the RackSwitch G8272 Product Guide

Lenovo RackSwitch G7028

The Lenovo RackSwitch G7028 (as shown in Figure 8) is a 1 Gb top-of-rack switch that delivers line-rate Layer 2 performance at an attractive price. G7028 has 24 10/100/1000BASE-T RJ45 ports and four 10 Gb Ethernet SFP+ ports. It typically uses only 45 W of power, which helps improve energy efficiency.

Figure 8. Lenovo RackSwitch G7028

For more information, see the RackSwitch G7028 Product Guide.

6.2 Deployment of an OpenStack cluster

Figure 9 shows how the components and concepts described in “Component model” relate to each other and how they form an OpenStack cluster.

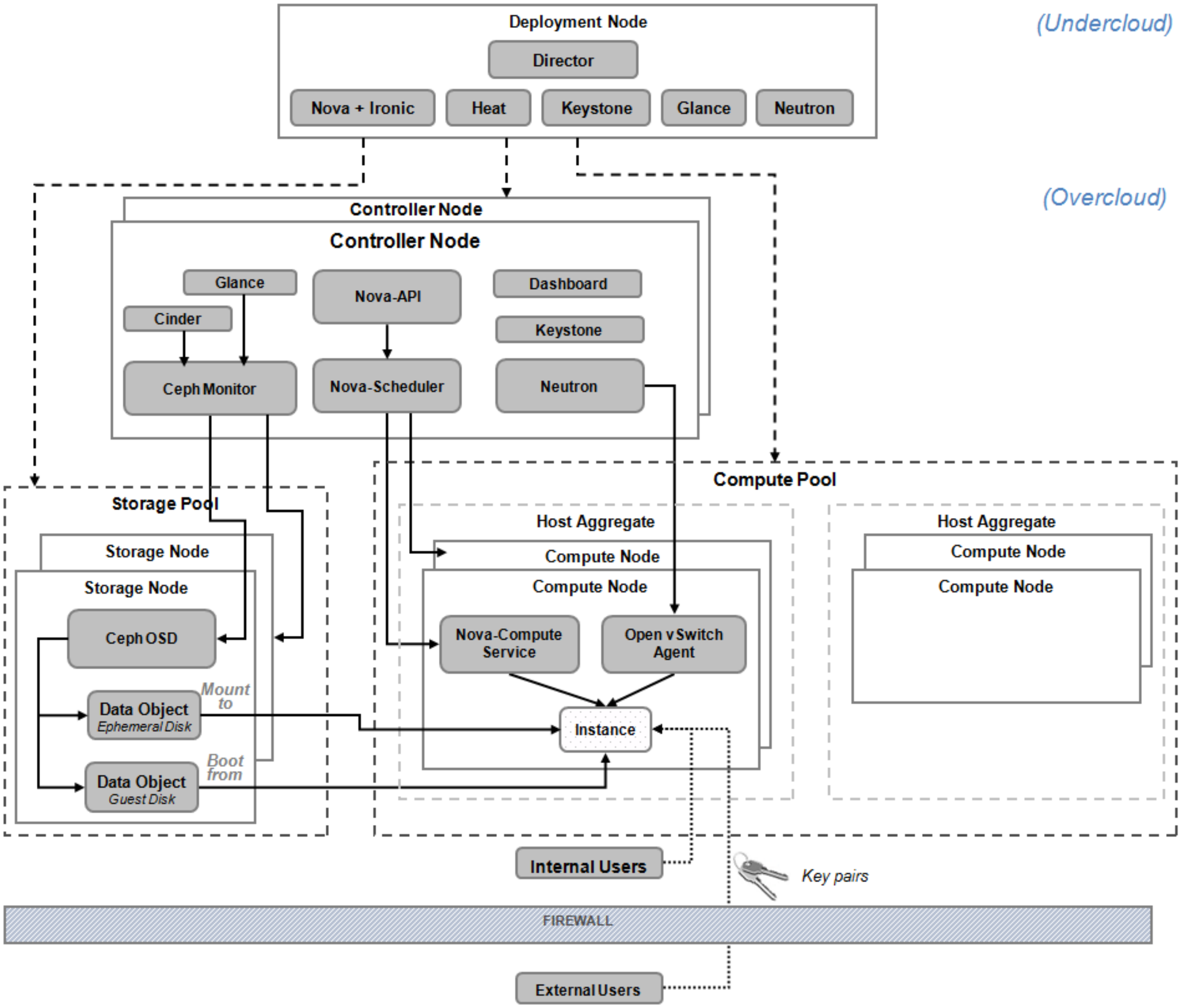


Figure 9. Legacy Deployment model for Red Hat OpenStack Platform

Figure 10 shows the components and concepts in pure HCI deployment mode.

The deployment node is responsible for the initial deployment of the controller node, compute node, storage node, and hyper-converged node leveraging the OpenStack bare metal provisioning service. It performs the service and network configuration between all deployed nodes. The deployment node adds additional nodes to the OpenStack cluster. Red Hat OpenStack Platform 11 is organized around two main concepts: an ‘Undercloud’ and an ‘Overcloud’. The Undercloud is the single-system OpenStack deployment node that provides environment planning, bare metal system control, installation and configuration of the Overcloud. While the Overcloud is resulting Red Hat OpenStack Platform Infrastructure as a Service (IaaS) environment created by the Undercloud.

The controller nodes in legacy deployment mode and pure HCI deployment mode are implemented as a cluster with three nodes. The nodes act as a central entrance for processing all internal and external cloud operation requests. The controller nodes manage the lifecycle of all VM instances that are running on the compute nodes and provide essential services, such as authentication and networking to the VM instances. The controller nodes rely on some support services, such as DHCP, DNS, and NTP.

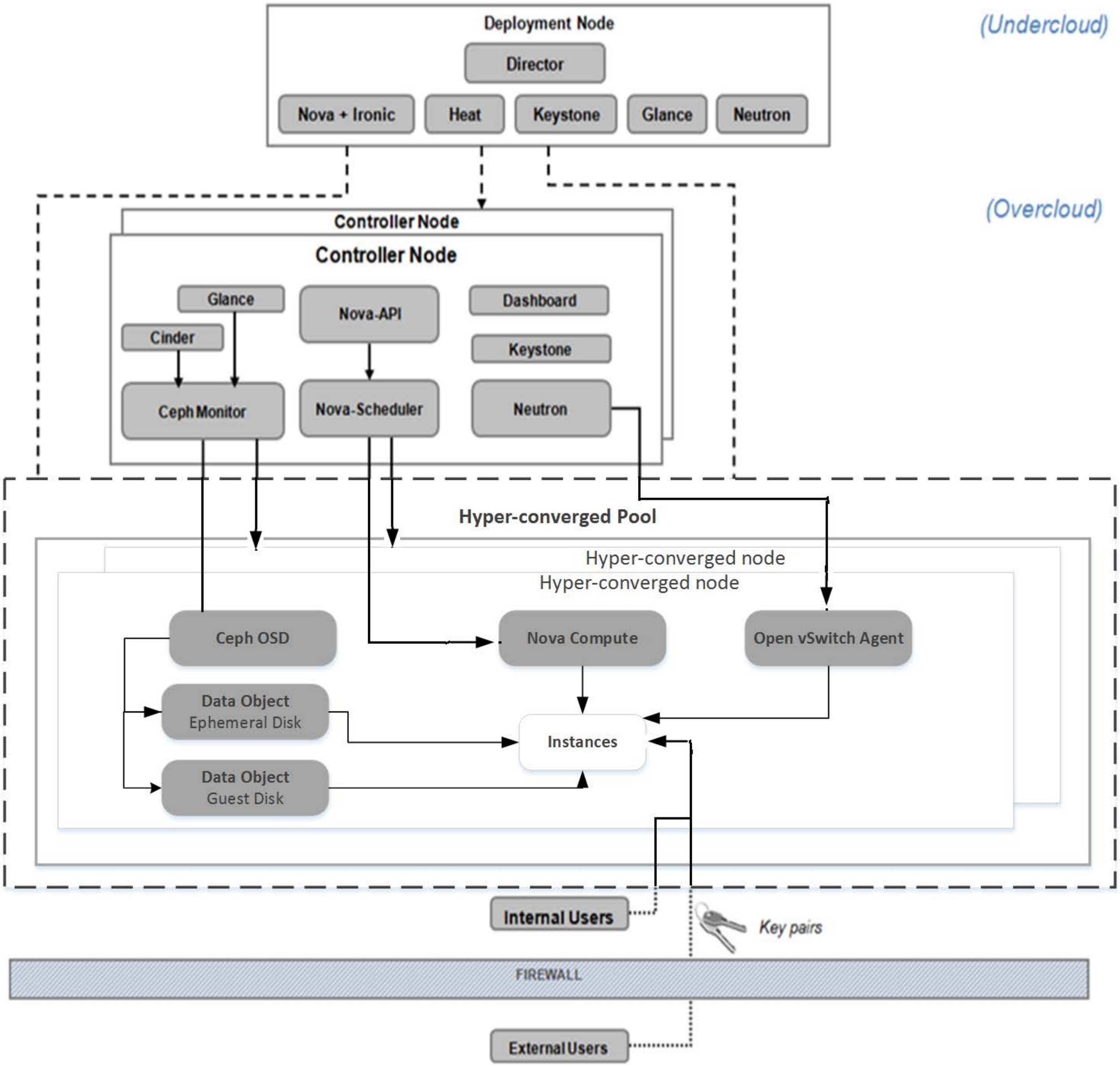


Figure 10. Pure HCI Deployment model for Red Hat OpenStack Platform

In legacy deployment mode, the Nova-Compute and Open vSwitch agents run on the compute nodes. The agents receive instrumentation requests from the controller node via RabbitMQ messages to manage the compute and network virtualization of instances that are running on the compute node. Compute nodes can be aggregated into pools of various sizes for better management, performance, or isolation. A Red Hat Ceph Storage cluster is created on the storage node. It is largely self-managed and supervised by the Ceph monitor installed on the controller node. The Red Hat Ceph Storage cluster provides block data storage for Glance image store and for VM instances via the Cinder service.

In pure HCI deployment mode, both compute node services and storage node services are co-located in hyper-converged node.

6.2.1 Compute pool in legacy deployment mode

The compute pool consists of multiple compute servers virtualized by OpenStack Nova to provide a redundant and scalable cloud-computing environment. Red Hat OpenStack Platform provides Nova drivers to enable virtualization on standard x86 servers (such as ThinkSystem SR650) and offers support for multiple hypervisors. Network addresses are automatically assigned to VM instances. Compute nodes inside the compute pool can be grouped into one or more “host aggregates” according to business need.

6.2.2 Storage pool in legacy deployment mode

There are two different types of storage pool. One uses local storage, and the other uses Red Hat Ceph Storage. Both have their own advantages and apply to different scenarios.

The local storage pool consists of local drives in a server. Therefore the configuration is easier and they provide high-speed data access. However, this approach lacks high-availability across servers.

The Red Hat Ceph storage pool consists of multiple storage servers that provide persistent storage resources from their local drives. In this reference architecture, all data are stored in a single Ceph cluster for simplicity and ease of management. Figure 11 shows the details of the integration between the Red Hat OpenStack Platform and Red Hat Ceph Storage.

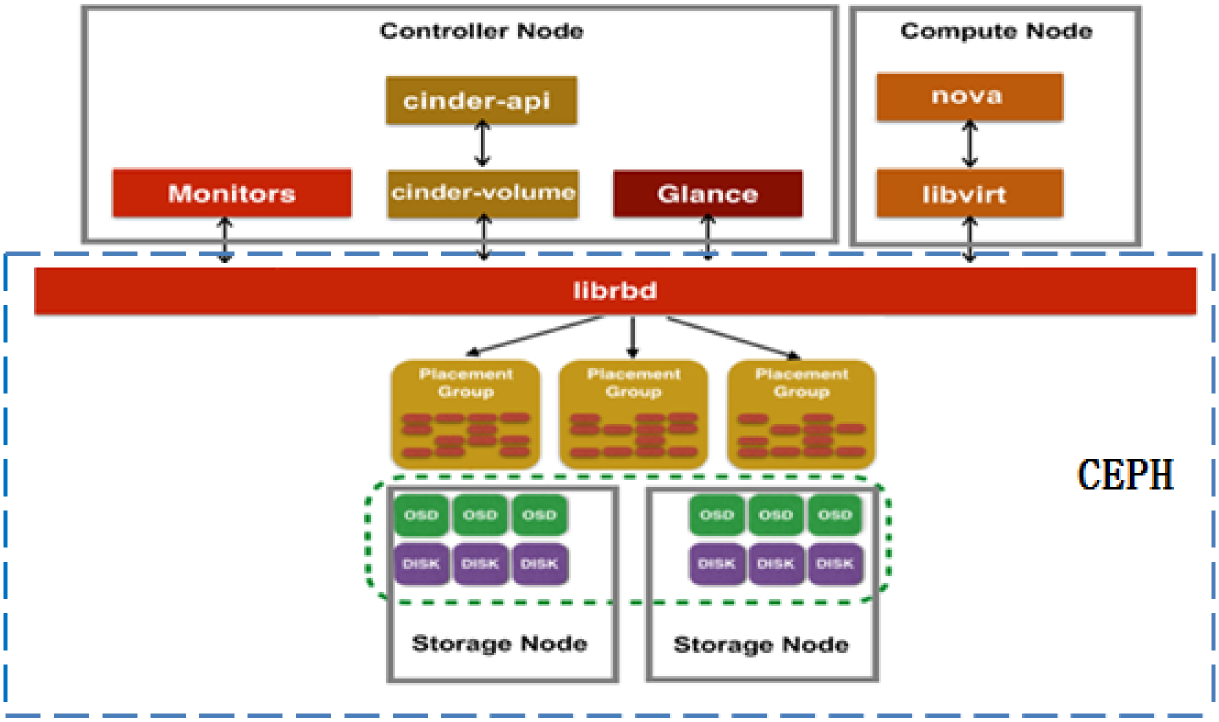


Figure 11. Red Hat Ceph Storage and OpenStack services integration

Ceph uses a write-ahead mode for local operations; a write operation hits the file system journal first and from there copied to the backing file store. To achieve optimal performance, two 2.5-inch SSDs have partitions for the operating system and the Ceph journal data. Refer to Red Hat Ceph Storage Installation Guide for Ceph OSD and SSD journal configuration details.

For better performance and security consideration, multiple Ceph clusters can be created for each OpenStack service or for different tenants. Because the complexity of Ceph I/O path and network bottleneck, the I/O speeds of Ceph storage is less than local storage. However, it has good scalability and high-availability, and it is very suitable for cloud computing.

Note: Red Hat Ceph Storage provides software-based data protection by using object replication and data striping and therefore there is no requirement for hardware-based RAID.

6.2.3 Hyper-converged pool in pure HCI mode

The Hyper-converged pool consists of multiple hyper-converged servers that provide virtualized computing environment and persistent Ceph storage resources. Lenovo recommends the high server configuration for both compute services and storage services in one server. Figure 12 shows the details of the hyper-converged pool.

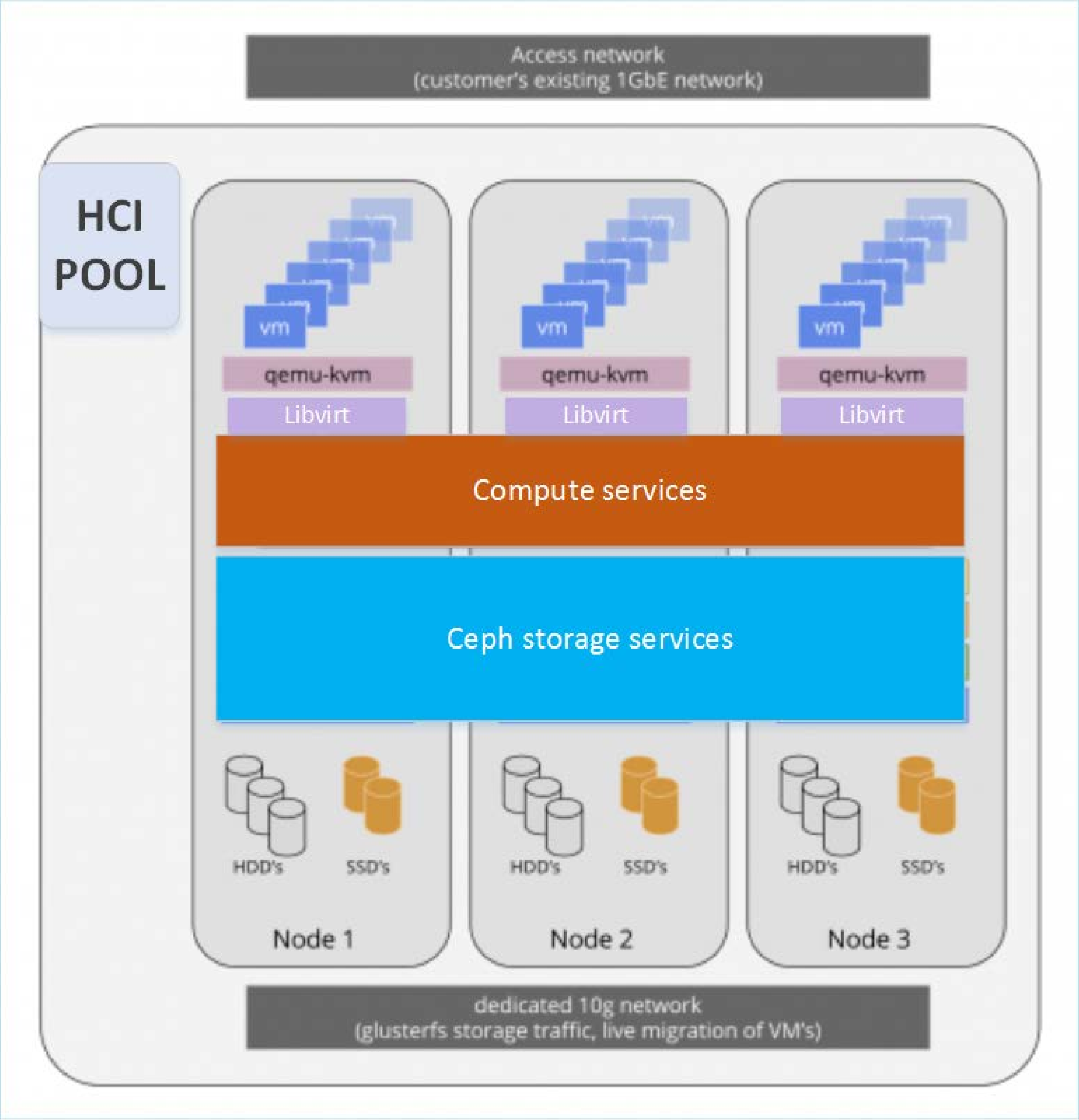


Figure 12. Hyper-converged pool

6.2.4 Controller cluster

The controller nodes are central administrative servers where users can access and provision cloud-based resources from a self-service portal. Controller node services are the same for both legacy deployment mode and pure HCI mode.

The HA of services that are running on controllers is primarily provided by using redundant controller nodes, as shown in Figure 13. Three types of services: *core*, *active-passive* and *systemd* are deployed in HA nodes. The Pacemaker component launches and manages core and active-passive services and systemd directly manages all the other services. An odd number of controller nodes are used (minimum of 3) because the HA capability is based on a quorum together with the Pacemaker and HAProxy technologies. HAProxy is configured on the controller nodes when using Red Hat OpenStack Platform Director to deploy more than one controller node. Pacemaker makes sure the cluster resource is running and available. If a service or a node fails, Pacemaker can restart the service, take the node out of the cluster, or reboot the failed node through HAProxy. For a three node controller cluster, if two controllers disagree, the third provides arbitration.

For more details, please visit：
https://access.redhat.com/documentation/en-us/red_hat_openstack_platform/11/html/understanding_red_hat_openstack_platform_high_availability/

For detailed service lists, please visit:

https://access.redhat.com/documentation/en-us/red_hat_openstack_platform/11/html/advanced_overcloud_customization/roles#Arch-Split_Controller

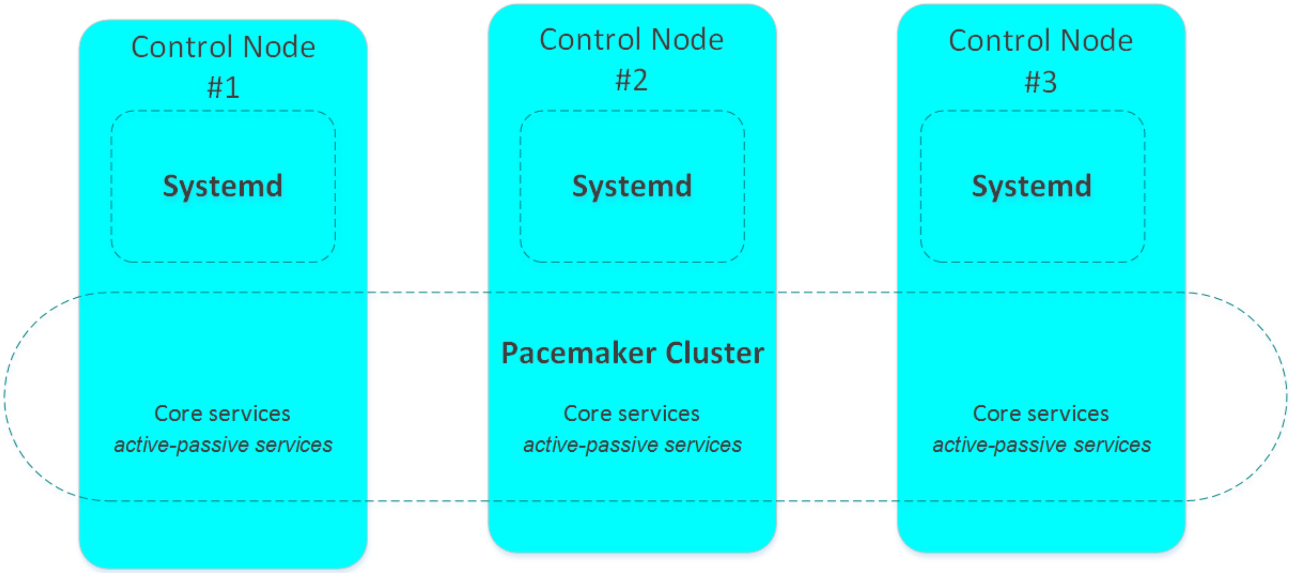


Figure 13. Controller cluster

The controller cluster hosts proxy and message broker services for scheduling compute and storage pool resources and provides the data store for cloud environment settings.

In addition, the controller servers act as network nodes that provide Software Defined Networking (SDN) functionality to the VMs. Network nodes are responsible for managing the virtual networking used by administrators to create public or private networks and uplink VMs into external networks. This forms the only ingress and egress points for instances that are running on top of OpenStack.

Each service location is referred to by using a Virtual IP (VIP) and Pacemaker automatically manages the mapping between the physical IP and VIP. Therefore switching a service provider between different controller nodes is transparent to service consumers. For a three-node controller cluster, each service needs three physical IP addresses and one virtual IP address.

Because all OpenStack API services are stateless, the user can choose an active/passive HA model or an active/active HA model with HAProxy to achieve load balancing. However, the Horizon web application needs to be enabled for session stickiness.

RabbitMQ and Red Hat Ceph include built-in clustering capabilities. MariaDB database uses the Galera library to achieve HA deployment.

6.2.5 Deployment server

The Red Hat OpenStack Platform 11 uses “Red Hat OpenStack Platform Director” as the toolset for installing and managing an OpenStack environment. The Red Hat OpenStack Platform Director is based primarily on the OpenStack TripleO project, which uses a minimal OpenStack installation to deploy an operational OpenStack environment, including controller nodes, compute nodes, and storage nodes as shown in the diagrams. The Ironic component enables bare metal server deployment and management. This tool simplifies the process of installing and configuring the Red Hat OpenStack Platform while providing a means to scale in the future.

6.2.6 Support services

Support services such as Domain Name System (DNS), Dynamic Host Configuration Protocol (DHCP), and Network Time Protocol (NTP) are needed for cloud operations. These services can be installed on the same deployment server or reused from the customer’s environment.

6.2.7 Deployment model

There are many models for deploying an OpenStack environment. This document will focus on legacy deployment mode and pure HCI deployment mode. To achieve the highest usability and flexibility, you can assign each node a specific role and place the corresponding OpenStack components on it.

The controller node has the following roles in supporting OpenStack:

- The controller node acts as an API layer and listens to all service requests (Nova, Glance, Cinder, and so on).The requests first land on the controller node. Then, the controller forwards requests to a compute node or storage node through messaging services for underlying compute workload or storage workload.
- The controller node is the messaging hub in which all messages follow and route through for cross-node communication.
- The controller node acts as the scheduler to determine the placement of a particular VM based on a specific scheduling driver.
- The controller node acts as the network node, which manipulates the virtual networks created in the cloud cluster.

However, it is not mandatory that all four roles be on the same physical server. OpenStack is a naturally distributed framework by design. Such all-in-one controller placement is used for simplicity and ease of deployment. In production environments, it might be preferable to move one or more of these roles and relevant services to another node to address security, performance, or manageability concerns.

On the compute nodes, the installation includes the essential OpenStack computing service (`openstack-nova-compute`), along with the `neutron-openvswitch-agent`, which enables software-defined networking. An optional metering component (Ceilometer) can be installed to collect resource usage information for billing or auditing.

The Ceph OSD service is installed on a storage node. It communicates with the Ceph monitor services that are running on controller nodes and contributes its local disks to the Ceph storage pool. The user of the Ceph storage resources exchanges data objects with the storage nodes through its internal protocol.

On a hyper-converged node, Compute services and Storage services are co-located and configured for optimized resource usage.

Table 6 lists the placement of major services on physical servers.

Table 6. Components placement on physical servers

Role	Host Name	Services

Red Hat OpenStack Platform 11 Director server (Undercloud)	deployer	`openstack-nova-* (Undercloud)`
		`openstack-glance-* (Undercloud)`
		`openstack-keystone(Undercloud)`
		`openstack-dashboard(Undercloud)`
		`neutron-* (Undercloud)`
		`openstack-ceilometer-*( Undercloud)`
		`openstack-heat-* (Undercloud)`
		`openstack-ironic-* (Undercloud)`
		`rabbitmq-server(Undercloud)`
		`opevswitch(Undercloud)`
		`mariadb(Undercloud)`

Compute node	compute01 – compute03	`neutron-openvswitch-agent`
		`openstack-ceilometer-compute`
		`openstack-nova-compute`

Storage node	storage01 - storage03	`ceph-osd`

Hyper-converged nodes	osdcompute01- osdcompute03	`neutron-openvswitch-agent`
		`openstack-ceilometer-compute`
		`openstack-nova-compute`
		`ceph-osd`

Role	Host Name	Services
Controller node	controller01 - controller03; osdcontroller01 - osdcontroller03	openstack-cinder-api openstack-cinder-backup openstack-cinder-scheduler openstack-cinder-volume openstack-glance-api openstack-glance-registry openstack-keystone openstack-nova-api openstack-nova-cert openstack-nova-conductor openstack-nova-consoleauth openstack-nova-novncproxy openstack-nova-scheduler neutron-dhcp-agent neutron-13-agent neutron-metadata-agent neutron-openvswitch-agent neutron-server openstack-ceilometer-alarm-evaluator openstack-ceilometer-alarm-notifier openstack-ceilometer-api openstack-ceilometer-central openstack-ceilometer-collector ceph-monitor rabbitmq, mariadb, mongodb

7 Deployment Example

This section describes the deployment of Red Hat OpenStack Platform 11 with Lenovo hardware. This document will focus on legacy deployment mode and pure HCI deployment mode.

7.1 Lenovo hardware preparation

The Lenovo x86 servers and switches previously described in “Hardware” on page 12 can be combined into a racked Red Hat OpenStack Platform cluster. Figure 14 and Figure 15 show a balanced configuration of compute, storage, networking, and power.

Components	Capacity	Rack layout
Rack	1 (42U)	
Monitor	1	
Controller nodes	3 (ThinkSystem SR630)	
Compute nodes	3 (ThinkSystem SR650)	
Storage nodes	3 (ThinkSystem SR650)	
Deployment server	1 (ThinkSystem SR630)	
Switch	2 (G8272), 1 (G7028)	

Figure 14. Deployment example 1: Full rack system legacy deployment

In rack of Figure 14, Legacy deployment platform (3 controller nodes, 3 compute nodes and 3 storage nodes) is setup for reference. Figure 15 shows the deployment of a pure HCI deployment platform (3 controller nodes and 6 hyper-converged nodes). The VM capacity is calculated on a per-host density basis. For more information about how to estimate the VM density, see “Resource isolation and sizing considerations” section.

Components	**Capacity**	**Rack layout**
Rack	1 (42U)	
Monitor	1	
Controller nodes (pure HCI deployment)	3 (ThinkSystem SR630)	
Hyper-converged nodes	6 (ThinkSystem SR650)	
Deployment server	1 (ThinkSystem SR630)	
Switch	2 (G8272), 1 (G7028)	

Figure 15. Deployment example 2: Full rack system pure HCI deployment

7.2 Networking

Combinations of physical and virtual isolated networks are configured at the host, switch, and storage layers to meet isolation, security, and quality of service (QoS) requirements. Different network data traffic is isolated into different switches according to the physical switch port and server port assignments.

There are two/four on-board 1 GbE Ethernet interfaces, plus one dedicated 1GbE Ethernet port for the XCC, and one/two installed dual-port 10 GbE Ethernet devices for each physical host (ThinkSystem SR630 or ThinkSystem SR650).

The 1 GbE management port (dedicated for the Integrated Management Module, or XCC) is connected to the 1GbE RackSwitch G7028 for out-of-band management, while the other 1 GbE ports are not connected. The corresponding Red Hat Enterprise Linux 7 Ethernet devices “eno[0-9]” are not configured.

Two G8272 10GbE Ethernet switches are used in pairs to provide redundant networking for the controller nodes, compute nodes, storage nodes, and deployment server. Administrators can create dedicated virtual pipes between the 10 GbE network adapters and the TOR switch for optimal performance and better security. Virtual LANs (VLANs) on the switch isolate Ethernet traffic by type.

Table 7 lists the five logical networks in a typical OpenStack deployment.

Table 7. OpenStack Logical Networks

Network	VLAN	Description
Management Network	2	The management network is used for OpenStack APIs and the message hub to connect various services inside OpenStack. The OpenStack Installer also uses this network for provisioning.
Tenant Network	3	This is the subnet for allocating the VM private IP addresses. Through this network, the VM instances can talk to each other.
External Network	4	This is the subnet for allocating the VM floating IP addresses. It is the only network where external users can access their VM instances.
Storage Network	5	The front-side storage network where Ceph clients (through Glance API, Cinder API, or Ceph CLI) access the Ceph cluster. Ceph Monitors operate on this network.
Data Cluster Network	6	The back-side storage network to which Ceph routes its heartbeat, object replication, and recovery traffic.
Internal API Network	7	The Internal API network is used for communication between the OpenStack services using API communication, RPC messages, and database communication.

Figure 16 shows the recommended network topology diagram with VLANs in legacy deployment mode.

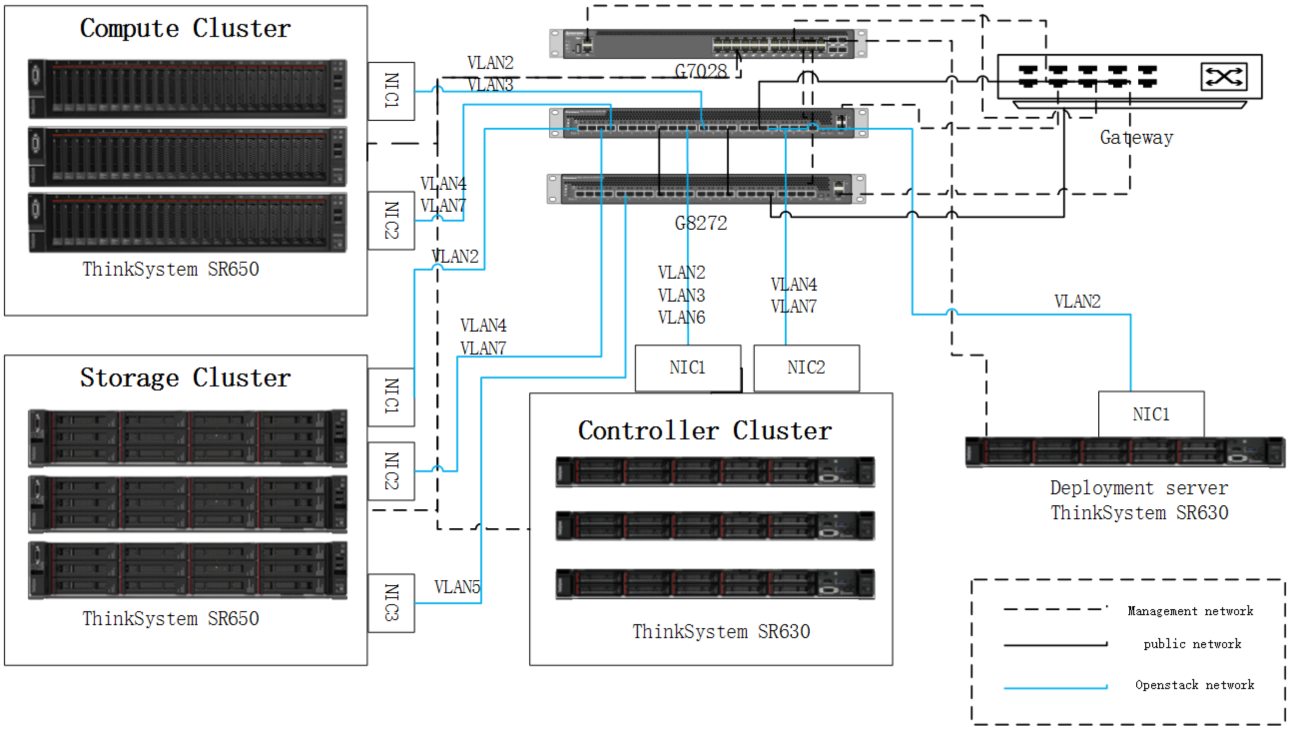


Figure 16. Network topology diagram in legacy deployment mode

Figure 17 shows the network topology diagram with VLANs in pure HCI deployment mode.

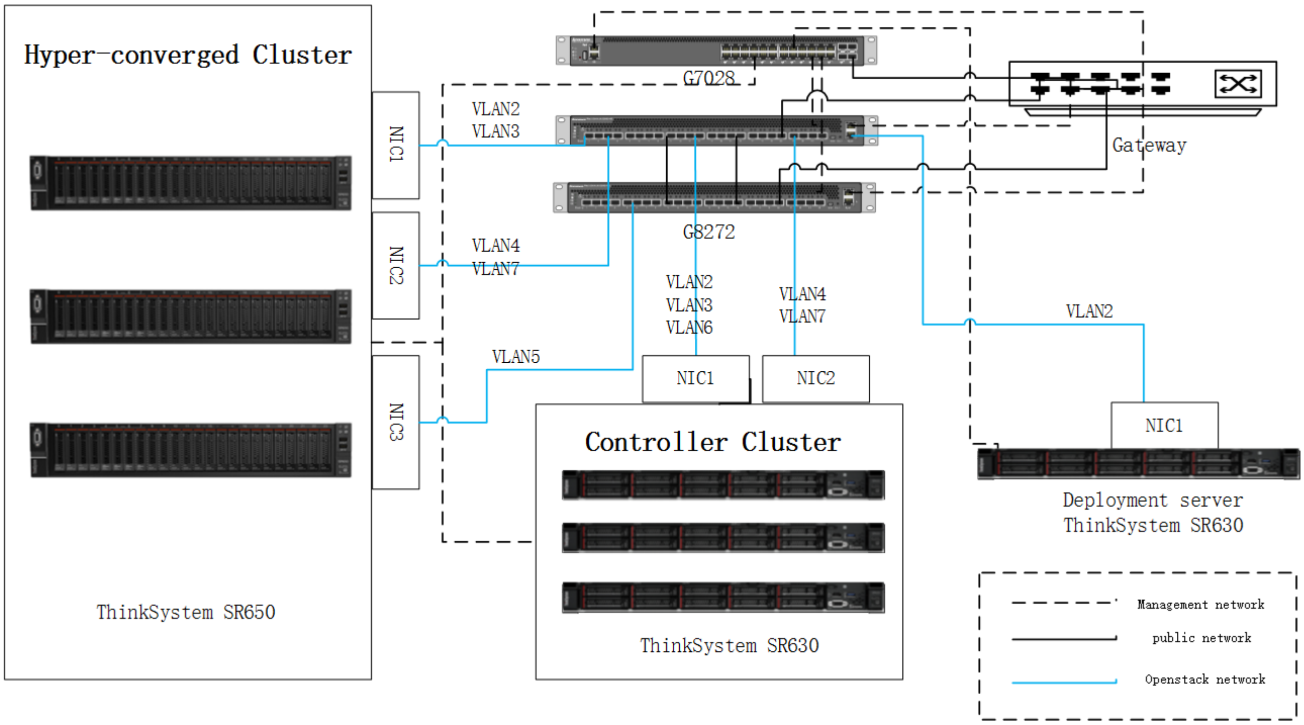


Figure 17. Network topology diagram in legacy deployment mode

Table 8 lists the recommended VLAN used by each type of node in an OpenStack deployment. These recommendations can be adjusted accordingly at runtime for the actual workload without service interruption.

Table 8. Recommended VLAN usage by OpenStack nodes

Role	Physical NIC	Usage	VLAN
	NIC1	Management Network	2
	NIC1	Tenant Network	3
	NIC2	Storage Network	4
	NIC1	External Network	6
	NIC2	Internal API Network	7
	NIC1	Management Network	2
	NIC1	Tenant Network	3
	NIC2	Storage Network	4
	NIC1	Reserved	N/A
	NIC2	Internal API Network	7
	NIC1	Management Network	2
	NIC1	Reserved	N/A
	NIC2	Storage Network	4
	NIC3	Data Cluster Network	5
	NIC2	Internal API Network	7
	NIC1	Management Network	2
	NIC1	Tenant Network	3
	NIC2	Storage Network	4
	NIC3	Data Cluster Network	5
	NIC2	Internal API Network	7
Deployment Server	NIC1	Management Network	2

The Emulex VFA5 adapter is used to offload packet encapsulation for VXLAN networking overlays, which results in higher server CPU effectiveness and higher server power efficiency.

Open vSwitch (OVS) is a fully featured, flow-based implementation of a virtual switch and can be used as a platform in software-defined networking. Open vSwitch supports VXLAN overlay networks, and serves as the default OpenStack network service in this reference architecture. Virtual Extensible LAN (VXLAN) is a network virtualization technology that addresses scalability problems that are associated with large cloud computing deployments by using a VLAN-like technique to encapsulate MAC-based OSI layer 2 Ethernet frames within layer 3 UDP packets.

7.3 Best practices of OpenStack Platform 11 deployment with Lenovo x86 Servers

This section describes best practices for OpenStack Platform 11 deployment using Lenovo x86 servers.

- For detailed deployment steps for OpenStack Platform 11 deployment, please see the Red Hat documentation "Director Installation and Usage"
- All nodes must be time synchronized by using NTP at all times.
- Lenovo XClarity Controller (XCC) is a unique management module of Lenovo x86 servers. Users can use web browsers or SSH to connect the management interface to configure and manage server. Firstly, configure an IP address for XCC (Boot the server → Enter "F1" → Select "RAID Setup " → Select "Manage Disk Drives" → Drop down to "Change all disk drives state from JBOD to UGood" → Select "Next" → Select "Advanced configuration" → Select "Next" → Create Virtual disks as your designed configuration.)
- RAID could be configured through XCC web console.(Login XCC web console → "Server Management" dropdown list → "Local Storage" → Select the RAID controller and click "Create Volume" → Complete the wizard)
- The information of NIC's MAC addresses need to be written into "instackenv.json" file. It is used by bare metal service (Ironic). These useful information can be obtained through XCC web console
    - Get MAC address (Login XCC web console → Select "Inventory" tab → Drop down to "PCI Adapters" → Launch the network adapter's properties and view "Physical Ports")
- Hyper-Converged compute nodes or Ceph storage nodes need multiple HDD and SSD disks. The SSDs are configured as journal disks (shared by all the OSDs), 2 HDDs are configured in RAID 1 for the boot device, and rest of HDDs are configured in RAID 0 for OSDs.
- Hyper-converged nodes deployed with same HW configurations, such as number of drivers and position of NICs, are requested for auto-deployment.
- The timeout set should be 60 sec in the ironic.conf, in order to avoid the XCC connection fail.
- UEFI mode configuration: Lenovo Servers' default boot mode is UEFI. If you want to deploy through UEFI mode, you need to change the properties of the nodes registered to ironic, and modify the configure file "undecloud.conf". Please see the Red Hat documentation "Director Installation and Usage".
- Legacy mode configuration: By default, Lenovo Servers' "PXE Boot" feature is "DISABLED". If you want to use the Legacy mode, you need to enable the PXE boot, use the Emulex PXESelect Utility and enable "PXE Boot".(press "Ctrl+p" to enter "Emulex PXESelect Utility" → Select "Personality" and press "F6" → Select one controller → "Boot Configuration" → "PXE Boot")

7.4 Resource isolation and sizing considerations

The expected performance of the OpenStack compute clusters relies on the actual workload that is running in the cloud and the correct sizing of each component. This section provides a sizing guide for different workloads and scenarios.

Table 9 lists the assumptions for the workload characteristics. The mixed workload consists of 45% small, 40% medium and 15% large VMs. This information serves as a baseline only to measure the user's real workload and does not necessarily mean to represent any specific application. In any deployment scenario, workloads are unlikely to have the same characteristics, and workloads may not be balanced between hosts. An appropriate scheduler with a real performance measurement of a workload should be used for compute and storage to determine if the target service level agreement (SLA) is being met.

Table 9. VM workloads characteristics

	vCPU	vRAM	Storage
Small	1	2 GB	20 GB
Medium	2	4 GB	60 GB
Large	4	8 GB	100 GB
xLarge	8	16 GB	200 GB

7.4.1 Resource considerations for legacy deployment mode

When sizing the solution, calculate the amount of required resources based on the amount of workload expected. Generally, 4 GB RAM is reserved for the system. For better resource utilization, consider putting similar workloads on the same guest OS type in the same host to use memory or storage de-duplication, as shown in the following calculations:

- Virtual CPU (vCPU) = Physical Cores * cpu_allocation_ratio
- Virtual Memory (vRAM) = (Physical Memory - OS reserved Memory - Instance overhead Memory) * (RAM_allocation_ratio)

The cpu_allocation_ratio indicates the number of virtual cores assigned to a node for each physical core. A ratio of 6:1 (6) provides a balanced choice for performance and cost effectiveness on models with Intel® Xeon® Scalable processors.

The RAM_allocation_ratio is used to allocate virtual resources in excess of what is physically available on a host through compression or de-duplication technology. The hypervisor uses it to improve infrastructure utilization of the RAM allocation ratio = (virtual resource/physical resource) formula. A ratio of 150% (1.5) provides a balanced choice for performance and cost effectiveness on models with Intel® Xeon® Scalable processors. Overhead per instance for the hypervisor is 0.1 GB.

By using these formulas, usable virtual resources can be calculated from the hardware configuration of compute nodes.

Table 10 shows two compute node configurations with different storage capabilities. The default configuration uses Red Hat Ceph storage nodes and there is no local storage apart from the Operating System. The standard configuration is used to provide different capacity and I/O performance for local storage.

Table 10 uses the same mixed workload compute capability for both configurations. This capability is generated from Table 9 in which each VM requires 2 vCPU, 8 GB vRAM, and 80 GB local disk.

Table 10. Resource for VMs in legacy deployment mode

	Default configuration (no local storage)	Standard configuration (local storage)
Server	ThinkSystem SR650	
vCPU	20 Cores*6 = 120 vCPU	
vRAM	(256 GB - 4GB – 4.6GB)*150% = 371.1GB	
Storage	Ceph storage	8*900GB with RAID-10 = 3.5 TB

7.4.2 Resource considerations for pure HCI deployment mode

Compute service and Ceph services are co-located on a hyper-converged node. Hyper-converged nodes need tuning in order to maintain stability and maximize the number of possible instances.

In hyper-converged node, the amount of vCPU and vRAM can be calculated as follows：

- Virtual CPU (vCPU) = Physical Cores * cpu_allocation_ratio
- Virtual Memory (vRAM) = (Physical Memory - OS reserved Memory - Instance overhead Memory - Memory allocated to Ceph) * RAM_allocation_ratio

The Compute scheduler uses cpu_allocation_ratio when choosing which Compute nodes on which to deploy an instance. A ratio of 4.5:1 (4.5) provides a balanced choice for performance and cost effectiveness on models with Intel® Xeon® Scalable processors.

To determine an appropriate value for hyper-converged nodes, assume that each OSD consumes 3GB of memory. Given a node with 768GB memory and 10 OSDs, you can allocate 30GB of memory for Ceph, leaving 738 GB for Compute. With 738 GB of system memory, a node can host, for example, 92 instances using 8GB of memory each.

However, you still need to consider additional overhead per instance for the hypervisor. Assuming this overhead is 0.5 GB, the same node can only host 86 instances, which accounts for the 738GB divided by 8.5GB.

The RAM_allocation_ratio is used to allocate virtual resources in excess of what is physically available on a host through compression or de-duplication technology. The hypervisor uses it to improve infrastructure utilization of the RAM allocation ratio = (virtual resource/physical resource) formula. For hyper-converged mode a ratio of 100% (1.0) provides a balanced choice for performance and cost effectiveness on models with Intel® Xeon® Scalable processors.

Table 11 shows one hyper-converged node configuration that uses 10 Ceph OSDs per node and VM capability generated from Table 10 in which each VM requires 2 vCPU, 8 GB vRAM, and 80 GB local disk.

Table 11. Resource for VMs in pure HCI mode

	Hyper-converged configuration (no local storage)
Server	ThinkSystem SR650
vCPU	40 Cores * 4.5 = 180 vCPU
vRAM	(768 GB - 4.5 - 43 - 30)*100% = 690.5GB
Storage	Ceph storage

7.5 Resource Management

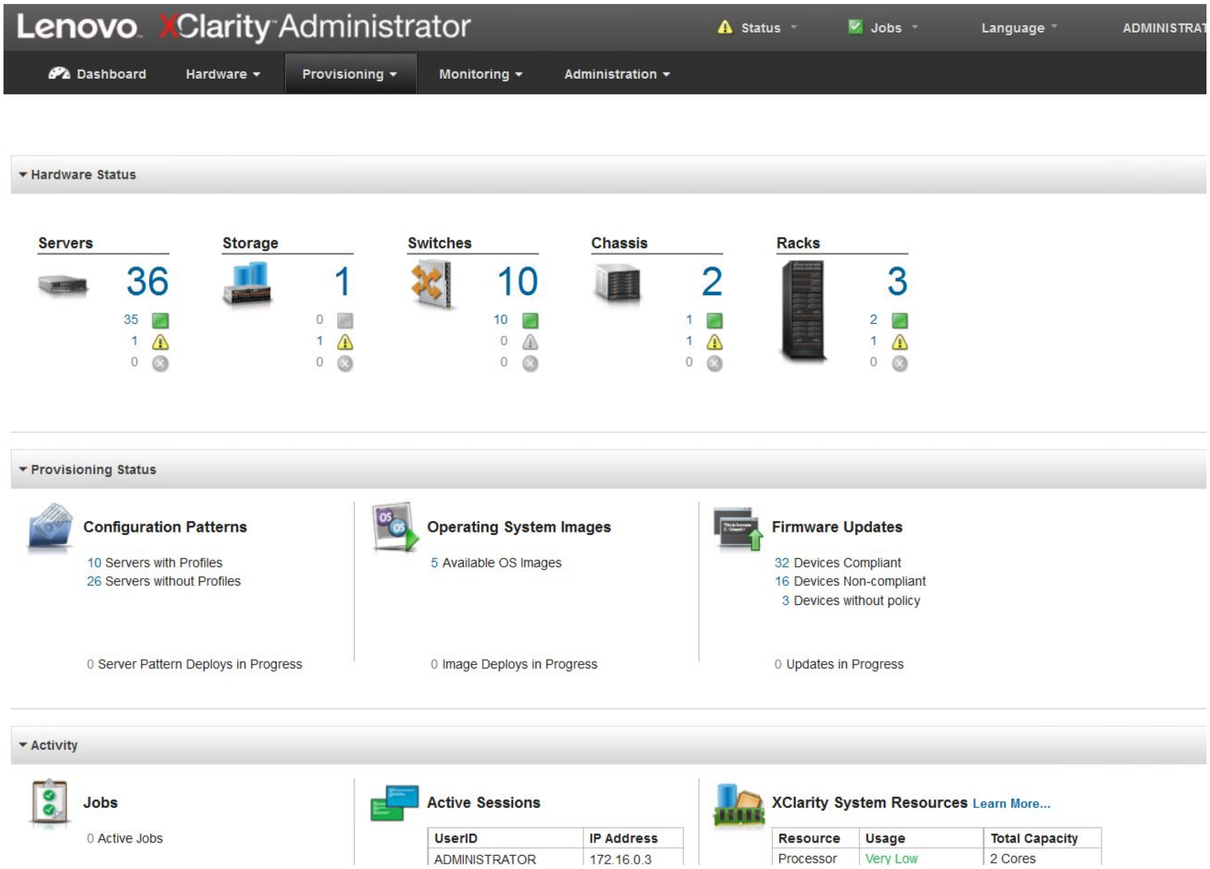
This section describes the software available for resource management.

7.5.1 Lenovo XClarity Administrator

Lenovo XClarity™ Administrator is a new centralized IT resource management solution that enables administrators to deploy infrastructure faster and with less effort. The solution seamlessly integrates into ThinkSystem servers. XClarity provides the following features:

- Intuitive, Graphical User Interface
- Auto-Discovery and Inventory
- Firmware Updates and Compliance
- Configuration Patterns
- Bare Metal Deployment
- Security Management
- Upward Integration
- REST API and PowerShell and Python Scripts
- SNMP, SYSLOG and Email Forwarding

Figure 18 shows the Lenovo XClarity Administrator interface, where Lenovo ThinkSystem servers are managed from the dashboard. For more information, please see "Lenovo XClarity".



Hardware Status

Servers	Storage	Switches	Chassis	Racks
36	1	10	2	3
35 🟩	0	10 🟩	1 🟩	2 🟩
1 ⚠	1 ⚠	0 ⚠	1 ⚠	1 ⚠
0 ⊗	0 ⊗	0 ⊗	0 ⊗	0 ⊗

Provisioning Status

Configuration Patterns
10 Servers with Profiles
26 Servers without Profiles
0 Server Pattern Deploys in Progress

Operating System Images
5 Available OS Images
0 Image Deploys in Progress

Firmware Updates
32 Devices Compliant
16 Devices Non-compliant
3 Devices without policy
0 Updates in Progress

Activity

Jobs
0 Active Jobs

Active Sessions

UserID	IP Address
ADMINISTRATOR	172.16.0.3

XClarity System Resources Learn More...

Resource	Usage	Total Capacity
Processor	Very Low	2 Cores

Figure 18. XClarity Administrator interface

7.5.2 Red Hat CloudForms 4.5

Red Hat CloudForms Management Engine delivers insight, control and automation that enterprises need to address the challenges of managing virtual environments. This technology enables enterprises with existing virtual infrastructures to improve visibility and control, and those starting virtualization deployments to build and operates a well-managed virtual infrastructure. Red Hat CloudForms has the following capabilities:

- Accelerate service delivery and reduce operational costs

- Improve operational visibility and control
- Ensure compliance and governance

Figure 19 shows the architecture and capabilities of Red Hat CloudForms. These features are designed to work together to provide robust management and maintenance of your virtual infrastructure.

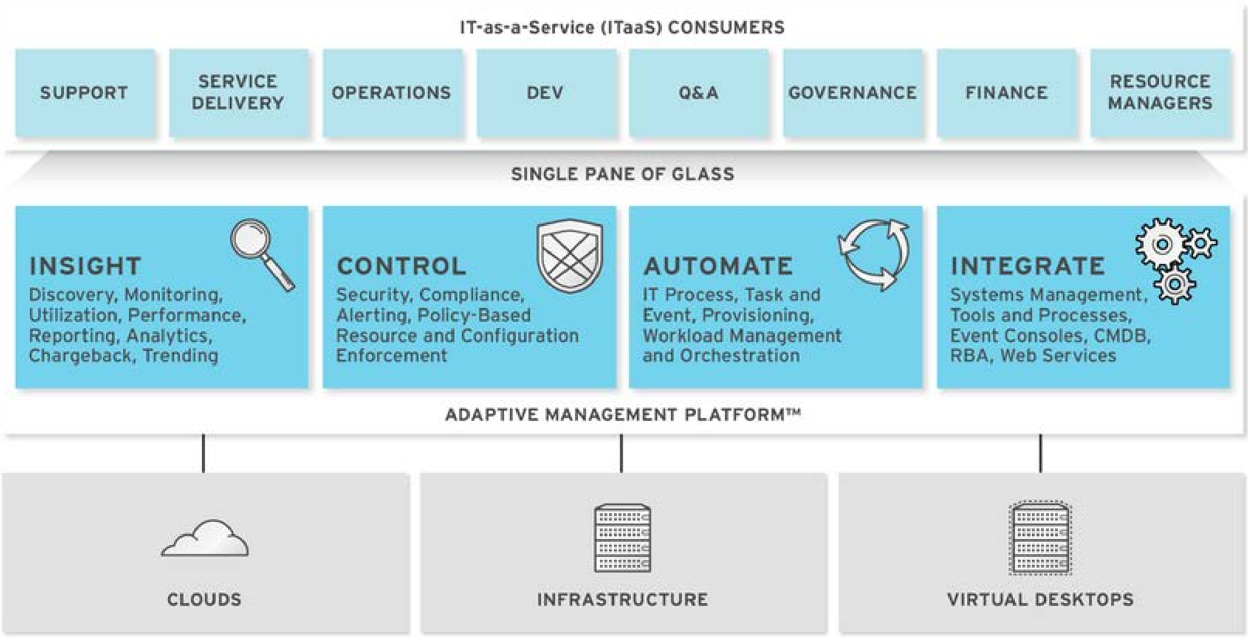
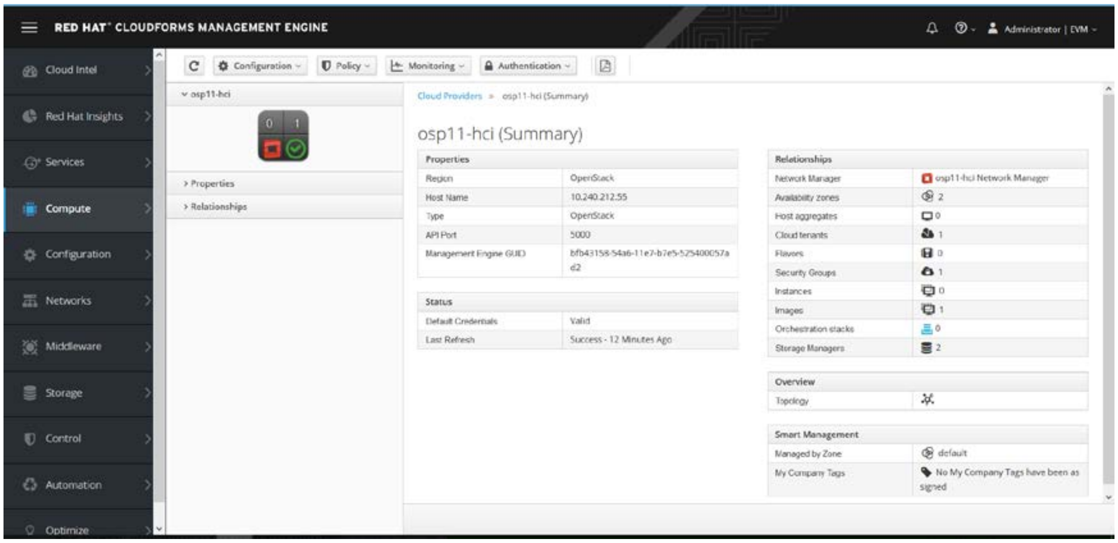

IT-as-a-Service (ITaaS) CONSUMERS
SUPPORT
SERVICE DELIVERY
OPERATIONS
DEV
Q&A
GOVERNANCE
FINANCE
RESOURCE MANAGERS
SINGLE PANE OF GLASS
INSIGHT
Discovery, Monitoring, Utilization, Performance, Reporting, Analytics, Chargeback, Trending
CONTROL
Security, Compliance, Alerting, Policy-Based Resource and Configuration Enforcement
AUTOMATE
IT Process, Task and Event, Provisioning, Workload Management and Orchestration
INTEGRATE
Systems Management, Tools and Processes, Event Consoles, CMDB, RBA, Web Services
ADAPTIVE MANAGEMENT PLATFORM™
CLOUDS
INFRASTRUCTURE
VIRTUAL DESKTOPS

Figure 19. Red Hat CloudForms Architecture

To install Red Hat CloudForms 4.5 on Red Hat OpenStack platform, please see Installing CloudForms.

After CloudForms installed, OpenStack cloud and infrastructure providers should be added, and then the CloudForms lifecycle can be managed. Figure 20 shows the CloudForms interface.



osp11-hci (Summary)

Properties	
Region	OpenStack
Host Name	10.240.212.55
Type	OpenStack
API Port	5000
Management Engine GUID	bfb43158-54a6-11e7-b7e5-525400057a d2

Status	
Default Credentials	Valid
Last Refresh	Success - 12 Minutes Ago

Relationships	
Network Manager	osp11-hci Network Manager
Availability zones	2
Host aggregates	0
Cloud tenants	1
Flavors	0
Security Groups	1
Instances	0
Images	1
Orchestration stacks	0
Storage Managers	2

Overview	
Topology	

Smart Management	
Managed by Zone	default
My Company Tags	No My Company Tags have been as signed

Figure 20. Red Hat CloudForms Interface

8 Appendix: Lenovo Bill of Materials

This appendix contains the Bill of Materials (BOMs) for different configurations of hardware for Red Hat OpenStack Platform deployments. There are sections for compute nodes, deployment server, controller nodes, storage nodes, networking, rack options, and software list.

8.1 Server BOM

The following section contains the BOM for the Red Hat OpenStack Platform 11 implementation using Lenovo System x Servers.

8.1.1 Director Server

Code	**Description**	**Quantity**
7X02CTO1WW	ThinkSystem SR630 - 3yr Warranty	1
AUW0	ThinkSystem SR630 2.5" Chassis with 8 bays	1
AWEP	Intel Xeon Gold 5118 12C 105W 2.3GHz Processor	2
AUNJ	ThinkSystem RAID 930-8i 2GB Flash PCIe 12Gb Adapter	1
AUM2	ThinkSystem 2.5" 1.8TB 10K SAS 12Gb Hot Swap 512e HDD	8
AUND	ThinkSystem 32GB TruDDR4 2666 MHz (2Rx4 1.2V) RDIMM	6
AT7S	Emulex VFA5.2 2x10 GbE SFP+ PCIe Adapter	1
AUKG	ThinkSystem 1Gb 2-port RJ45 LOM	1
A1PJ	3m Passive DAC SFP+ Cable	2
AVWB	ThinkSystem 1100W (230V/115V) Platinum Hot-Swap Power Supply	2

8.1.2 Controller Node

Code	**Description**	Quantity
7X02CTO1WW	ThinkSystem SR630 - 3yr Warranty	1
AWEL	Intel Xeon Gold 6126 12C 125W 2.6GHz Processor	2
AUND	ThinkSystem 32GB TruDDR4 2666 MHz (2Rx4 1.2V) RDIMM	3
AUNK	ThinkSystem RAID 930-16i 4GB Flash PCIe 12Gb Adapter	1
AUM2	ThinkSystem 2.5" 1.8TB 10K SAS 12Gb Hot Swap 512e HDD	8
AUMG	ThinkSystem 2.5" HUSMM32 400GB Performance SAS 12Gb Hot Swap SSD	2
AT7S	Emulex VFA5.2 2x10 GbE SFP+ PCIe Adapter	1
AUKG	ThinkSystem 1Gb 2-port RJ45 LOM	1
A1PJ	3m Passive DAC SFP+ Cable	2
AVWB	ThinkSystem 1100W (230V/115V) Platinum Hot-Swap Power Supply	2
6311	2.8m, 10A/100-250V, C13 to IEC 320-C14 Rack Power Cable	2

8.1.3 Compute Node (No Local Storage)

Code	Description	Quantity
7X06CTO1WW	ThinkSystem SR650 - 3yr Warranty	1
AUVV	ThinkSystem SR650 2.5" Chassis with 8, 16 or 24 bays	1
AWE1	Intel Xeon Gold 6140 18C 140W 2.3GHz Processor	1
AUND	ThinkSystem 32GB TruDDR4 2666 MHz (2Rx4 1.2V) RDIMM	8
AUNK	ThinkSystem RAID 930-16i 4GB Flash PCIe 12Gb Adapter	1
AUM2	ThinkSystem 2.5" 1.8TB 10K SAS 12Gb Hot Swap 512e HDD	2
AT7S	Emulex VFA5.2 2x10 GbE SFP+ PCIe Adapter	1
AUKG	ThinkSystem 1Gb 2-port RJ45 LOM	1
A1PJ	3m Passive DAC SFP+ Cable	2
AVWB	ThinkSystem 1100W (230V/115V) Platinum Hot-Swap Power Supply	2
6311	2.8m, 10A/100-250V, C13 to IEC 320-C14 Rack Power Cable	2

8.1.4 Compute Node (with Local Storage)

Code	Description	Quantity
7X06CTO1WW	ThinkSystem SR650 - 3yr Warranty	1
AUVV	ThinkSystem SR650 2.5" Chassis with 8, 16 or 24 bays	1
AWE1	Intel Xeon Gold 6140 18C 140W 2.3GHz Processor	1
AUND	ThinkSystem 32GB TruDDR4 2666 MHz (2Rx4 1.2V) RDIMM	8
AUV1	ThinkSystem RAID 930-24i 4GB Flash PCIe 12Gb Adapter	1
AUM2	ThinkSystem 2.5" 1.8TB 10K SAS 12Gb Hot Swap 512e HDD	4
AUMG	ThinkSystem 2.5" HUSMM32 400GB Performance SAS 12Gb Hot Swap SSD	2
AT7S	Emulex VFA5.2 2x10 GbE SFP+ PCIe Adapter	2
AUKG	ThinkSystem 1Gb 2-port RJ45 LOM	1
A51P	2m Passive DAC SFP+ Cable	2
AVWF	ThinkSystem 1100W (230V/115V) Platinum Hot-Swap Power Supply	2
6311	2.8m, 10A/100-250V, C13 to IEC 320-C14 Rack Power Cable	2

8.1.5 Storage Node

Code	Description	Quantity
7X06CTO1WW	ThinkSystem SR650 - 3yr Warranty	1
AUVV	ThinkSystem SR650 2.5" Chassis with 8, 16 or 24 bays	1
AWER	Intel Xeon Silver 4116 12C 85W 2.1GHz Processor	2
AUND	ThinkSystem 32GB TruDDR4 2666 MHz (2Rx4 1.2V) RDIMM	8
AUV1	ThinkSystem RAID 930-24i 4GB Flash PCIe 12Gb Adapter	1
AUM2	ThinkSystem 2.5" 1.8TB 10K SAS 12Gb Hot Swap 512e HDD	12
AUMG	ThinkSystem 2.5" HUSMM32 400GB Performance SAS 12Gb Hot Swap SSD	2
AT7S	Emulex VFA5.2 2x10 GbE SFP+ PCIe Adapter	2
AUKG	ThinkSystem 1Gb 2-port RJ45 LOM	1
A51P	2m Passive DAC SFP+ Cable	4
AVWF	ThinkSystem 1100W (230V/115V) Platinum Hot-Swap Power Supply	2
6311	2.8m, 10A/100-250V, C13 to IEC 320-C14 Rack Power Cable	2

8.1.6 Hyper-converged Node

Code	Description	Quantity
7X06CTO1WW	ThinkSystem SR650 - 3yr Warranty	1
AUVV	ThinkSystem SR650 2.5" Chassis with 8, 16 or 24 bays	1
AWE1	Intel Xeon Gold 6140 18C 140W 2.3GHz Processor	2
AUND	ThinkSystem 32GB TruDDR4 2666 MHz (2Rx4 1.2V) RDIMM	24
AUV1	ThinkSystem RAID 930-24i 4GB Flash PCIe 12Gb Adapter	1
AUM2	ThinkSystem 2.5" 1.8TB 10K SAS 12Gb Hot Swap 512e HDD	12
AUMG	ThinkSystem 2.5" HUSMM32 400GB Performance SAS 12Gb Hot Swap SSD	2
AT7S	Emulex VFA5.2 2x10 GbE SFP+ PCIe Adapter	2
AUKG	ThinkSystem 1Gb 2-port RJ45 LOM	1
A51P	2m Passive DAC SFP+ Cable	4
AVWF	ThinkSystem 1100W (230V/115V) Platinum Hot-Swap Power Supply	2
6311	2.8m, 10A/100-250V, C13 to IEC 320-C14 Rack Power Cable	2

8.2 Networking BOM

This section contains the BOM for different types of networking switches.

8.2.1 G7028 1GbE Switch

Code	Description	Quantity
7159BAX	Lenovo RackSwitch G7028 (Rear to Front)	1
39Y7938	2.8m, 10A/100-250V, C13 to IEC 320-C20 Rack Power Cable	2

8.2.2 G8124E 10GbE Switch

Code	Description	Quantity
7159BR6	Lenovo RackSwitch G8124E (Rear to Front)	2
39Y7938	2.8m, 10A/100-250V, C13 to IEC 320-C20 Rack Power Cable	4
90Y9427	1m Passive DAC SFP+ Cable	2

8.2.3 G8272 10 GbE (with 40 Gb uplink) Switch

Code	Description	Quantity
7159CRW	Lenovo RackSwitch G8272 (Rear to Front)	2
39Y7938	2.8m, 10A/100-250V, C13 to IEC 320-C20 Rack Power Cable	4
90Y9427	1m Passive DAC SFP+ Cable	2

8.3 Rack BOM

This section contains the BOM for the rack.

Code	Description	Quantity
93634PX	42U 1100mm Enterprise V2 Dynamic Rack	1
00YJ780	0U 20 C13/4 C19 Switched and Monitored 32A 1 Phase PDU	2

8.4 Red Hat Subscription Options

This section contains the BOM for the Red Hat Subscriptions. *See Lenovo Rep for final configuration.*

Code	Description	Quantity
00YH835	Red Hat OpenStack Platform, 2 socket, Premium RH Support, 3 yrs	Variable
00YH839	Red Hat OpenStack Platform Controller Node, 2 skt, Prem RH Support, 3 yrs	Variable
00YH849	Red Hat Ceph Storage, 12 Physical Nodes, to 256TB, Prem RH Support, 3 yrs	Variable

Resources

For more information about the topics in this document, see the following resources:

- OpenStack Project:
  www.openstack.org

- OpenStack Operations Guide:
  docs.openstack.org/ops/

- Red Hat OpenStack Platform:
  access.redhat.com/documentation/en/red-hat-openstack-platform/

- Red Hat Ceph Storage:
  www.redhat.com/en/technologies/storage/ceph

- Red Hat Hyper-converged Infrastructure:
  access.redhat.com/documentation/en-us/red_hat_openstack_platform/11/html/hyper-converged_infrastructure_guide/

Document History

Version 1.0	25 September 2017	• Initial version for Lenovo ThinkSystem server and OSP11

Trademarks and special notices

References in this document to Lenovo products or services do not imply that Lenovo intends to make them available in every country.

Lenovo, the Lenovo logo, AnyBay, AnyRAID, BladeCenter, NeXtScale, RackSwitch, Rescue and Recovery, ThinkSystem, System x, ThinkCentre, ThinkVision, ThinkVantage, ThinkPlus and XClarity are trademarks of Lenovo.

Red Hat, Red Hat Enterprise Linux and the Shadowman logo are trademarks of Red Hat, Inc., registered in the U.S. and other countries. Linux® is the registered trademark of Linus Torvalds in the U.S. and other countries. The OpenStack 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

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Intel, Intel Inside (logos), and Xeon are trademarks of Intel Corporation in the United States, other countries, or both.

Other company, product, or service names may be trademarks or service marks of others.

Information is provided "AS IS" without warranty of any kind.

All customer examples described are presented as illustrations of how those customers have used Lenovo products and the results they may have achieved. Actual environmental costs and performance characteristics may vary by customer.

Information concerning non-Lenovo products was obtained from a supplier of these products, published announcement material, or other publicly available sources and does not constitute an endorsement of such products by Lenovo. Sources for non-Lenovo list prices and performance numbers are taken from publicly available information, including vendor announcements and vendor worldwide homepages. Lenovo has not tested these products and cannot confirm the accuracy of performance, capability, or any other claims related to non-Lenovo products. Questions on the capability of non-Lenovo products should be addressed to the supplier of those products.

All statements regarding Lenovo future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only. Contact your local Lenovo office or Lenovo authorized reseller for the full text of the specific Statement of Direction.

Some information addresses anticipated future capabilities. Such information is not intended as a definitive statement of a commitment to specific levels of performance, function or delivery schedules with respect to any future products. Such commitments are only made in Lenovo product announcements. The information is presented here to communicate Lenovo’s current investment and development activities as a good faith effort to help with our customers' future planning.

Performance is based on measurements and projections using standard Lenovo benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput or performance improvements equivalent to the ratios stated here.

Photographs shown are of engineering prototypes. Changes may be incorporated in production models.

Any references in this information to non-Lenovo websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this Lenovo product and use of those websites is at your own risk.