Citrix CloudPlatform (powered by Apache CloudStack) Version 4.5 Concepts Guide

Revised January 30, 2015 06:00 pm IST

Citrix CloudPlatform

Citrix CloudPlatform (powered by Apache CloudStack) Version 4.5 Concepts Guide
Revised January 30, 2015 06:00 pm IST

Author Citrix CloudPlatform

If you want to learn the concepts about CloudPlatform before you learn the ongoing operation and maintenance of a CloudPlatform-powered cloud, read this document.

About this Guide

1.1. About the Audience for this Guide

This guide is meant for anyone responsible for configuring and administering the public cloud infrastructure and the private cloud infrastructure of enterprises using CloudPlatform such as cloud administrators and Information Technology (IT) administrators.

1.2. Using the Product Documentation

The following guides provide information about CloudPlatform:

- *Citrix CloudPlatform (powered by Apache CloudStack) Installation Guide*
- *Citrix CloudPlatform (powered by Apache CloudStack) Concepts Guide*
- *Citrix CloudPlatform (powered by Apache CloudStack) Getting Started Guide*
- *Citrix CloudPlatform (powered by Apache CloudStack) Administration Guide*
- *Citrix CloudPlatform (powered by Apache CloudStack) Hypervisor Configuration Guide*
- *Citrix CloudPlatform (powered by Apache CloudStack) Developer's Guide*

For complete information on any known limitations or issues in this release, see the *Citrix CloudPlatform (powered by Apache CloudStack) Release Notes*.

For information about the Application Programming Interfaces (APIs) that is used in this product, see the API documents that are available with CloudPlatform.

1.3. Experimental Features

CloudPlatform product releases include some experimental features for customers to test and experiment with in non-production environments, and share any feedback with Citrix. For any issues with these experimental features, customers can open a support ticket but Citrix cannot commit to debugging or providing fixes for them.

The following experimental featues are inluded in this release:

- Advanced Networking in Baremetal
- Linux Containers
- Supported Management Server OS and Supported Hypervisors: RHEL7/CentOS 7 for experimental use with Linux Containers

1.4. Additional Information and Help

Troubleshooting articles by the Citrix support team are available in the Citrix Knowledge Center at *support.citrix.com/product/cs/*.

1.5. Contacting Support

The support team is available to help customers plan and execute their installations. To contact the support team, log in to the support portal at *support.citrix.com/cloudsupport*[1] by using the account credentials you received when you purchased your support contract.

[1] http://support.citrix.com/cloudsupport

Concepts

2.1. About CloudPlatform

CloudPlatform is a software platform that pools computing resources to build public, private, and hybrid Infrastructure as a Service (IaaS) clouds. CloudPlatform manages the network, storage, and compute nodes that make up a cloud infrastructure. Use CloudPlatform to deploy, manage, and configure cloud computing environments.

Typical users are service providers and enterprises. With CloudPlatform, you can:

- Set up an on-demand, elastic cloud computing service. Service providers can sell self service virtual machine instances, storage volumes, and networking configurations over the Internet.
- Set up an on-premise private cloud for use by employees. Rather than managing virtual machines in the same way as physical machines, with CloudPlatform an enterprise can offer self-service virtual machines to users without involving IT departments.

2.2. CloudPlatform Features

Multiple Hypervisor Support

CloudPlatform works with a variety of hypervisors. A single cloud deployment can contain multiple hypervisor implementations. You have the complete freedom to choose the right hypervisor for your workload.

CloudPlatform is designed to work with open source XenServer and KVM hypervisors as well as enterprise-grade hypervisors such as Citrix XenServer, Hyper-V, and VMware vSphere.

Massively Scalable Infrastructure Management

CloudPlatform can manage tens of thousands of servers installed in multiple geographically distributed datacenters. The centralized management server scales linearly, eliminating the need for intermediate

cluster-level management servers. No single component failure can cause cloud-wide outage. Periodic maintenance of the management server can be performed without affecting the functioning of virtual machines running in the cloud.

Automatic Configuration Management

CloudPlatform automatically configures each guest virtual machine’s networking and storage settings.

CloudPlatform internally manages a pool of virtual appliances to support the cloud itself. These appliances offer services such as firewalling, routing, DHCP, VPN access, console proxy, storage access, and storage replication. The extensive use of virtual appliances simplifies the installation, configuration, and ongoing management of a cloud deployment.

Graphical User Interface

CloudPlatform offers an administrator's Web interface, used for provisioning and managing the cloud, as well as an end-user's Web interface, used for running VMs and managing VM templates. The UI can be customized to reflect the desired service provider or enterprise look and feel.

API and Extensibility

CloudPlatform provides an API that gives programmatic access to all the management features available in the UI. This API enables the creation of command line tools and new user interfaces to suit particular needs.

The CloudPlatform pluggable allocation architecture allows the creation of new types of allocators for the selection of storage and hosts.

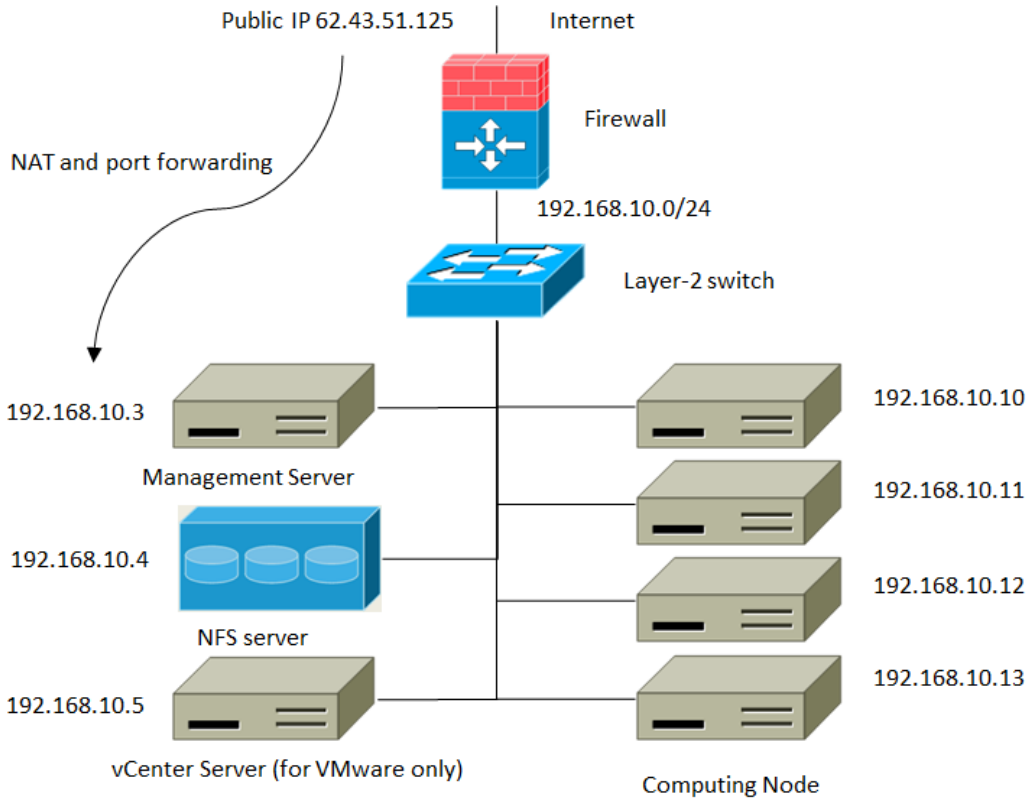
High Availability

CloudPlatform has a number of features to increase the availability of the system. The Management Server itself, which is the main controlling software at the heart of CloudPlatform, may be deployed in a multi-node installation where the servers are load balanced. MySQL may be configured to use replication to provide for a manual failover in the event of database loss. For the hosts, CloudPlatform supports NIC bonding and the use of separate networks for storage as well as iSCSI Multipath.

Choosing a Deployment Architecture

The architecture used in a deployment will vary depending on the size and purpose of the deployment. This section contains examples of deployment architecture, including a small-scale deployment useful for test and trial deployments and a fully-redundant large-scale setup for production deployments.

3.1. Small-Scale Deployment

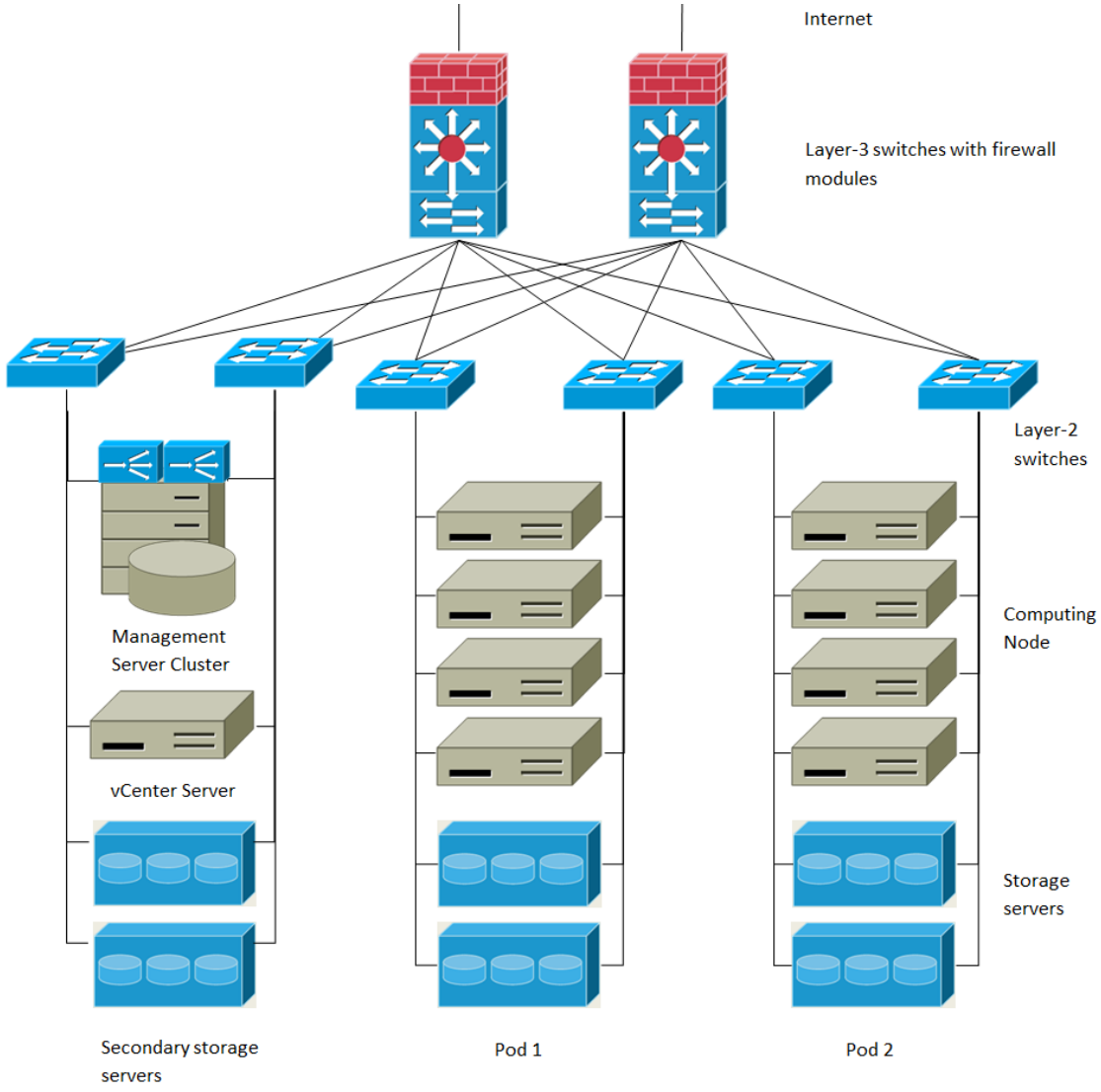


Small-Scale Deployment

This diagram illustrates the network architecture of a small-scale CloudPlatform deployment.

- A firewall provides a connection to the Internet. The firewall is configured in NAT mode. The firewall forwards HTTP requests and API calls from the Internet to the Management Server. The Management Server resides on the management network.
- A layer-2 switch connects all physical servers and storage.
- A single NFS server functions as both the primary and secondary storage.
- The Management Server is connected to the management network.

3.2. Large-Scale Redundant Setup



Large-Scale Redundant Deployment

This diagram illustrates the network architecture of a large-scale CloudPlatform deployment.

- A layer-3 switching layer is at the core of the data center. A router redundancy protocol like VRRP should be deployed. Typically high-end core switches also include firewall modules. Separate firewall appliances may also be used if the layer-3 switch does not have integrated firewall capabilities. The firewalls are configured in NAT mode. The firewalls provide the following functions:
  - Forwards HTTP requests and API calls from the Internet to the Management Server. The Management Server resides on the management network.
  - When the cloud spans multiple zones, the firewalls should enable site-to-site VPN such that servers in different zones can directly reach each other.
- A layer-2 access switch layer is established for each pod. Multiple switches can be stacked to increase port count. In either case, redundant pairs of layer-2 switches should be deployed.
- The Management Server cluster (including front-end load balancers, Management Server nodes, and the MySQL database) is connected to the management network through a pair of load balancers.

- Secondary storage servers are connected to the management network.
- Each pod contains storage and computing servers. Each storage and computing server should have redundant NICs connected to separate layer-2 access switches.

3.3. Separate Storage Network

In the large-scale redundant setup described in the previous section, storage traffic can overload the management network. A separate storage network is optional for deployments. Storage protocols such as iSCSI are sensitive to network delays. A separate storage network ensures guest network traffic contention does not impact storage performance.

3.4. Multi-Node Management Server

The CloudPlatform Management Server is deployed on one or more front-end servers connected to a single MySQL database. Optionally a pair of hardware load balancers distributes requests from the web. A backup management server set may be deployed using MySQL replication at a remote site to add DR capabilities.

The administrator must decide the following.

- Whether or not load balancers will be used.
- How many Management Servers will be deployed.
- Whether MySQL replication will be deployed to enable disaster recovery.

3.5. Multi-Site Deployment

The CloudPlatform platform scales well into multiple sites through the use of zones.

There are two ways to configure the storage network:

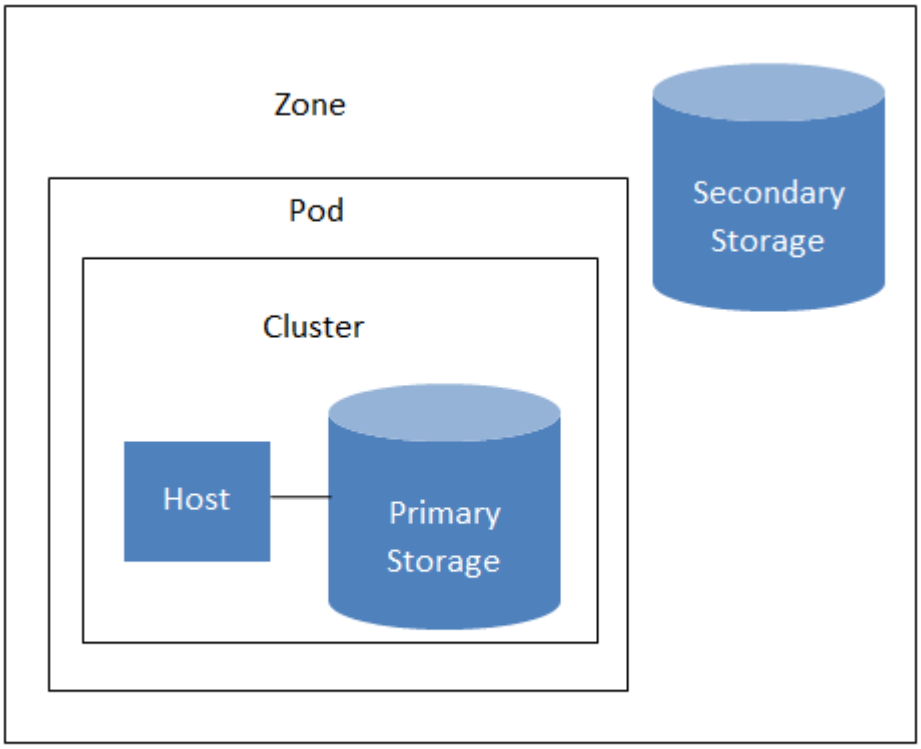
- Bonded NIC and redundant switches can be deployed for NFS. In NFS deployments, redundant switches and bonded NICs still result in one network (one CIDR block+ default gateway address).
- iSCSI can take advantage of two separate storage networks (two CIDR blocks each with its own default gateway). Multipath iSCSI client can failover and load balance between separate storage networks.

Cloud Infrastructure Concepts

4.1. Cloud Infrastructure Overview

The Management Server manages one or more zones (typically, datacenters) containing host computers where guest virtual machines will run. The cloud infrastructure is organized as follows:

- Region: To increase reliability of the cloud, you can optionally group resources into multiple geographic regions. A region consists of one or more zones.
- Zone: Typically, a zone is equivalent to a single datacenter. A zone consists of one or more pods and secondary storage.
- Pod: A pod is usually one rack of hardware that includes a layer-2 switch and one or more clusters.
- Cluster: A cluster consists of one or more hosts and primary storage.
- Host: A single compute node within a cluster. The hosts are where the actual cloud services run in the form of guest virtual machines.
- Primary storage is associated with a cluster, and it can also be provisioned on a zone-wide basis. It stores the disk volumes for all the VMs running on hosts in that cluster.
- Secondary storage is associated with a zone, and it can also be provisioned as object storage that is available throughout the cloud. It stores templates, ISO images, and disk volume snapshots.



Nested organization of a zone

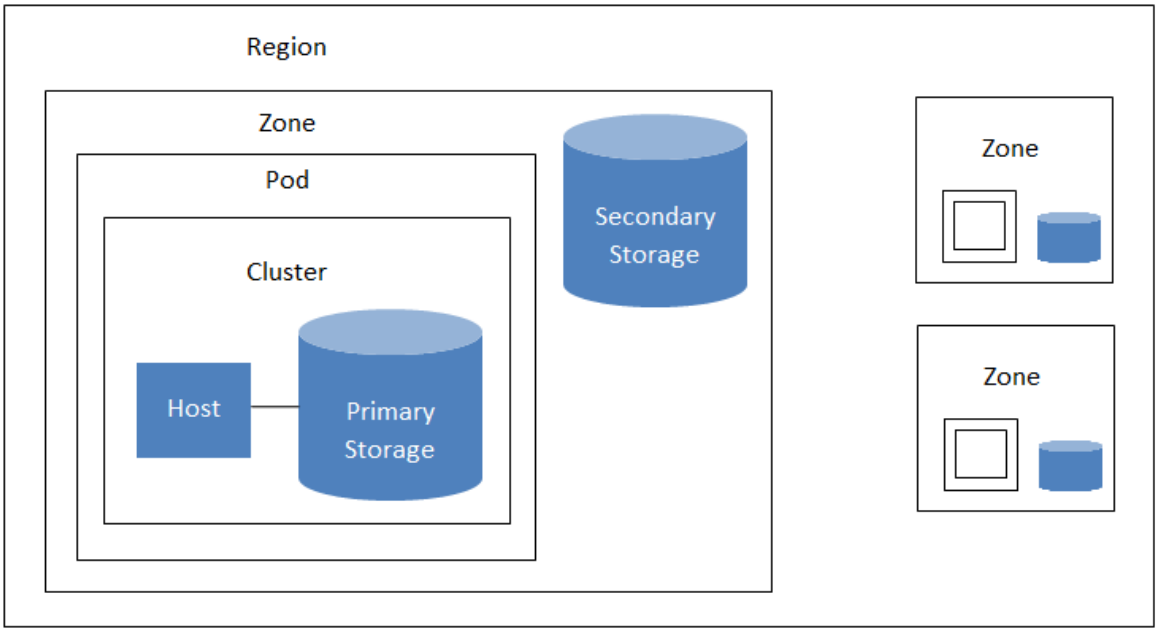
More Information

4.2. Regions

To increase reliability of the cloud, you can optionally group resources into multiple geographic regions. A region is the largest available organizational unit within a CloudPlatform deployment. A region is made up of several availability zones, where each zone is equivalent to a datacenter. Each region is controlled by its own cluster of Management Servers, running in one of the zones. The zones in a region are typically located in close geographical proximity. Regions are a useful technique for providing fault tolerance and disaster recovery.

By grouping zones into regions, the cloud can achieve higher availability and scalability. User accounts can span regions, so that users can deploy VMs in multiple, widely-dispersed regions. Even if one of the regions becomes unavailable, the services are still available to the end-user through VMs deployed in another region. And by grouping communities of zones under their own nearby Management Servers, the latency of communications within the cloud is reduced compared to managing widely-dispersed zones from a single central Management Server.

Usage records can also be consolidated and tracked at the region level, creating reports or invoices for each geographic region.



A region with multiple zones

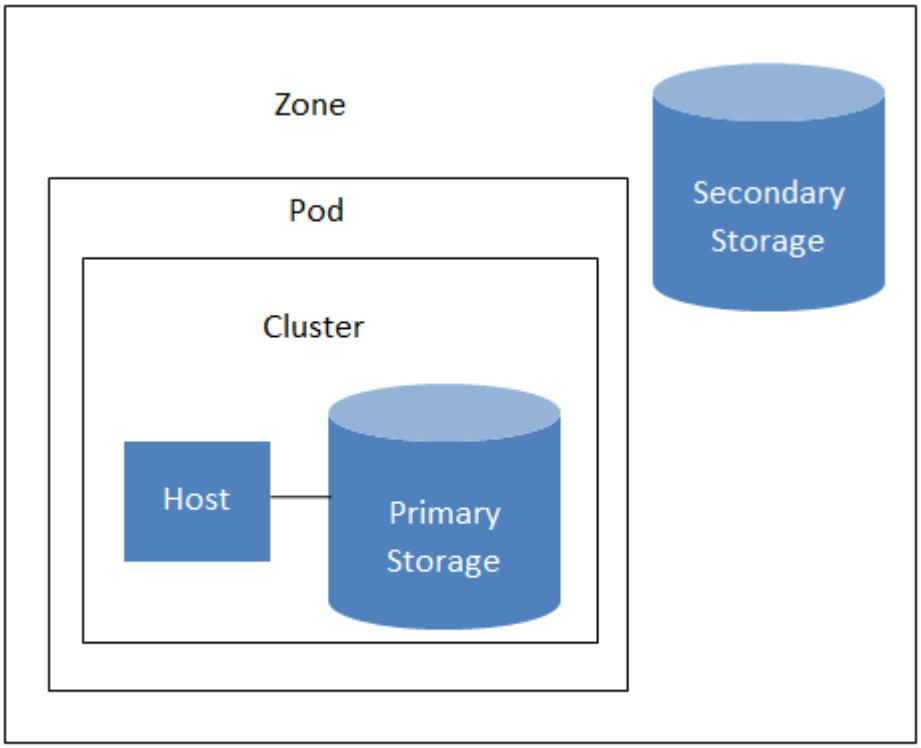
Regions are visible to the end user. When a user starts a guest VM on a particular CloudPlatform Management Server, the user is implicitly selecting that region for their guest. Users might also be required to copy their private templates to additional regions to enable creation of guest VMs using their templates in those regions.

4.3. Zones

A zone is the second largest organizational unit within a CloudPlatform deployment. A zone typically corresponds to a single datacenter, although it is permissible to have multiple zones in a datacenter. The benefit of organizing infrastructure into zones is to provide physical isolation and redundancy. For example, each zone can have its own power supply and network uplink, and the zones can be widely separated geographically (though this is not required).

A zone consists of:

- One or more pods. Each pod contains one or more clusters of hosts and one or more primary storage servers.
- (Optional) If zone-wide primary storage is desired, a zone may contain one or more primary storage servers, which are shared by all the pods in the zone. (Supported for KVM and VMware hosts)
- Secondary storage, which is shared by all the pods in the zone.



Nested organization of a zone

Zones are visible to the end user. When a user starts a guest VM, the user must select a zone for their guest. Users might also be required to copy their private templates to additional zones to enable creation of guest VMs using their templates in those zones.

Zones can be public or private. Public zones are visible to all users. This means that any user may create a guest in that zone. Private zones are reserved for a specific domain. Only users in that domain or its subdomains may create guests in that zone.

Hosts in the same zone are directly accessible to each other without having to go through a firewall. Hosts in different zones can access each other through statically configured VPN tunnels.

For each zone, the administrator must decide the following.

- How many pods to place in a zone.
- How many clusters to place in each pod.
- How many hosts to place in each cluster.
- (Optional) If zone-wide primary storage is being used, decide how many primary storage servers to place in each zone and total capacity for these storage servers. (Supported for KVM and VMware hosts)

- How many primary storage servers to place in each cluster and total capacity for these storage servers.
- How much secondary storage to deploy in a zone.

When you add a new zone, you will be prompted to configure the zone's physical network and add the first pod, cluster, host, primary storage, and secondary storage.

(VMware) In order to support zone-wide functions for VMware, CloudPlatform is aware of VMware Datacenters and can map each Datacenter to a CloudPlatform zone. To enable features like storage live migration and zone-wide primary storage for VMware hosts, CloudPlatform has to make sure that a zone contains only a single VMware Datacenter. Therefore, when you are creating a new CloudPlatform zone, you can select a VMware Datacenter for the zone. If you are provisioning multiple VMware Datacenters, each one will be set up as a single zone in CloudPlatform.

Note

If you are upgrading from a previous CloudPlatform version, and your existing deployment contains a zone with clusters from multiple VMware Datacenters, that zone will not be forcibly migrated to the new model. It will continue to function as before. However, any new zone-wide operations introduced in CloudPlatform 4.2, such as zone-wide primary storage and live storage migration, will not be available in that zone.

4.4. Physical Networks

Part of adding a zone is setting up the physical network. One or (in an advanced zone) more physical networks can be associated with each zone. The network corresponds to a NIC on the hypervisor host. Each physical network can carry one or more types of network traffic. The choices of traffic type for each network vary depending on whether you are creating a zone with basic networking or advanced networking.

A physical network is the actual network hardware and wiring in a zone. A zone can have multiple physical networks. An administrator can:

- Add/Remove/Update physical networks in a zone
- Configure VLANs on the physical network
- Configure a name so the network can be recognized by hypervisors
- Configure the service providers (firewalls, load balancers, etc.) available on a physical network
- Configure the IP addresses trunked to a physical network
- Specify what type of traffic is carried on the physical network, as well as other properties like network speed

4.4.1. Basic Zone Network Traffic Types

When basic networking is used, there can be only one physical network in the zone. That physical network carries the following traffic types:

- Guest. When end users run VMs, they generate guest traffic. The guest VMs communicate with each other over a network that can be referred to as the guest network. Each pod in a basic zone is a broadcast domain, and therefore each pod has a different IP range for the guest network. The administrator must configure the IP range for each pod.
- Management. When CloudPlatform's internal resources communicate with each other, they generate management traffic. This includes communication between hosts, system VMs (VMs used by CloudPlatform to perform various tasks in the cloud), and any other component that communicates directly with the CloudPlatform Management Server. You must configure the IP range for the system VMs to use.

Note

We strongly recommend the use of separate NICs for management traffic and guest traffic.

- Public. Public traffic is generated when VMs in the cloud access the Internet. Publicly accessible IPs must be allocated for this purpose. End users can use the CloudPlatform UI to acquire these IPs to implement NAT between their guest network and the public network, as described in the **9.14 Acquiring a New IP Address** section of the *CloudPlatform (powered by Apache CloudStack) Version 4.5 Administration Guide*. Public traffic is generated only in EIP-enabled basic zones. For more information, refer to the **9.20 About Elastic IP** section of the *CloudPlatform (powered by Apache CloudStack) Version 4.5 Administration Guide*.
- Storage. Traffic such as VM templates and snapshots, which is sent between the secondary storage VM and secondary storage servers. CloudPlatform uses a separate Network Interface Controller (NIC) named storage NIC for storage network traffic. Use of a storage NIC that always operates on a high bandwidth network allows fast template and snapshot copying. You must configure the IP range to use for the storage network.

In a basic network, configuring the physical network is fairly straightforward. In most cases, you only need to configure one guest network to carry traffic that is generated by guest VMs. If you use a NetScaler load balancer and enable its elastic IP and elastic load balancing (EIP and ELB) features, you must also configure a network to carry public traffic. CloudPlatform takes care of presenting the necessary network configuration steps to you in the UI when you add a new zone.

4.4.2. Basic Zone Guest IP Addresses

When basic networking is used, CloudPlatform will assign IP addresses in the CIDR of the pod to the guests in that pod. The administrator must add a direct IP range on the pod for this purpose. These IPs are in the same VLAN as the hosts.

4.4.3. Advanced Zone Network Traffic Types

When advanced networking is used, there can be multiple physical networks in the zone. Each physical network can carry one or more traffic types, and you need to let CloudPlatform know which type of network traffic you want each network to carry. The traffic types in an advanced zone are:

- Guest. When end users run VMs, they generate guest traffic. The guest VMs communicate with each other over a network that can be referred to as the guest network. This network can be isolated or shared. In an isolated guest network, the administrator needs to reserve VLAN ranges to

provide isolation for each CloudPlatform account's network (potentially a large number of VLANs). In a shared guest network, all guest VMs share a single network.

- Management. When CloudPlatform's internal resources communicate with each other, they generate management traffic. This includes communication between hosts, system VMs (VMs used by CloudPlatform to perform various tasks in the cloud), and any other component that communicates directly with the CloudPlatform Management Server. You must configure the IP range for the system VMs to use.
- Public. Public traffic is generated when VMs in the cloud access the Internet. Publicly accessible IPs must be allocated for this purpose. End users can use the CloudPlatform UI to acquire these IPs to implement NAT between their guest network and the public network, as described in the **9.14 Acquiring a New IP Address** section of the *CloudPlatform (powered by Apache CloudStack) Version 4.5 Administration Guide*.
- Storage. Traffic such as VM templates and snapshots, which is sent between the secondary storage VM and secondary storage servers. CloudPlatform uses a separate Network Interface Controller (NIC) named storage NIC for storage network traffic. Use of a storage NIC that always operates on a high bandwidth network allows fast template and snapshot copying. You must configure the IP range to use for the storage network.

These traffic types can each be on a separate physical network, or they can be combined with certain restrictions. When you use the Add Zone wizard in the UI to create a new zone, you are guided into making only valid choices.

4.4.4. Advanced Zone Guest IP Addresses

Guest IP addresses are private IP addresses that are used for internal communication. When advanced networking is used, the administrator can create additional networks for use by the guests. These networks can either be made available to all accounts in the zone, or they can be scoped to a single account. If they are scoped to a single account, only the named account may create guests that attach to these networks. The networks are defined by a VLAN ID, IP range, and gateway. The administrator may provision thousands of these networks if desired. Additionally, the administrator can reserve a part of the IP address space for non-CloudPlatform VMs and servers (For more information, refer to the **9.17 IP Reservation in Isolated Guest Networks** section of the *CloudPlatform (powered by Apache CloudStack) Version 4.5 Administration Guide*.).

4.4.5. Advanced Zone Public IP Addresses

Public IP addresses are used for communicating with the Internet. When advanced networking is used, the administrator can create additional networks for use by the guests. The networks are defined by a VLAN ID, IP range, and gateway. The administrator may provision thousands of these networks if desired. Network Address Translation (NAT) protocol is used for converting the private IP address to a public IP address.

4.4.6. System Reserved IP Addresses

In each zone, you need to configure a range of reserved IP addresses for the management network. This network carries communication between the CloudPlatform Management Server and various system VMs, such as Secondary Storage VMs, Console Proxy VMs, and Virtual Router VM.

The reserved IP addresses must be unique across the cloud. You cannot, for example, have a host in one zone which has the same private IP address as a host in another zone.

The hosts in a pod are assigned private IP addresses. These are typically RFC1918 addresses. The Console Proxy and Secondary Storage system VMs are also allocated private IP addresses in the CIDR of the pod that they are created in.

Make sure computing servers and Management Servers use IP addresses outside of the System Reserved IP range. For example, suppose the System Reserved IP range starts at 192.168.154.2 and ends at 192.168.154.7. CloudPlatform can use .2 to .7 for System VMs. This leaves the rest of the pod CIDR, from .8 to .254, for the Management Server and hypervisor hosts.

In all zones:

Provide private IPs for the system in each pod and provision them in CloudPlatform.

For KVM and XenServer, the recommended number of private IPs per pod is one per host. If you expect a pod to grow, add enough private IPs now to accommodate the growth.

In a zone that uses advanced networking:

When advanced networking is being used, the number of private IP addresses available in each pod varies depending on which hypervisor is running on the nodes in that pod. Citrix XenServer and KVM use link-local addresses, which in theory provide more than 65,000 private IP addresses within the address block. As the pod grows over time, this should be more than enough for any reasonable number of hosts as well as IP addresses for guest virtual routers. VMWare ESXi, by contrast uses any administrator-specified subnetting scheme, and the typical administrator provides only 255 IPs per pod. Since these are shared by physical machines, the guest virtual router, and other entities, it is possible to run out of private IPs when scaling up a pod whose nodes are running ESXi.

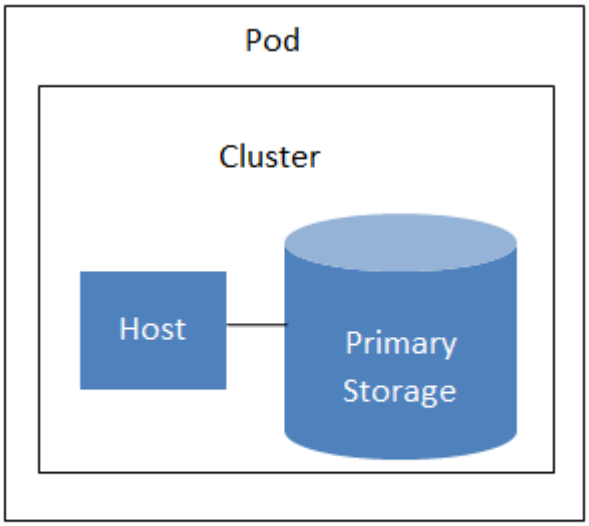
To ensure adequate headroom to scale private IP space in an ESXi pod that uses advanced networking, use one or more of the following techniques:

- Specify a larger CIDR block for the subnet. A subnet mask with a /20 suffix will provide more than 4,000 IP addresses.
- Create multiple pods, each with its own subnet. For example, if you create 10 pods and each pod has 255 IPs, this will provide 2,550 IP addresses.

For vSphere with advanced networking, we recommend provisioning enough private IPs for your total number of customers, plus enough for the required CloudPlatform System VMs. Typically, about 10 additional IPs are required for the System VMs. For more information about System VMs, refer to **Chapter 11 Working with System Virtual Machines** in the *CloudPlatform (powered by Apache CloudStack) Version 4.5 Administration Guide*.

4.5. Pods

A pod often represents a single rack. Hosts in the same pod are in the same subnet. A pod is the third-largest organizational unit within a CloudPlatform deployment. Pods are contained within zones, and zones can be contained within regions. Each zone can contain one or more pods. A pod consists of one or more clusters of hosts and one or more primary storage servers. Pods are not visible to the end user.



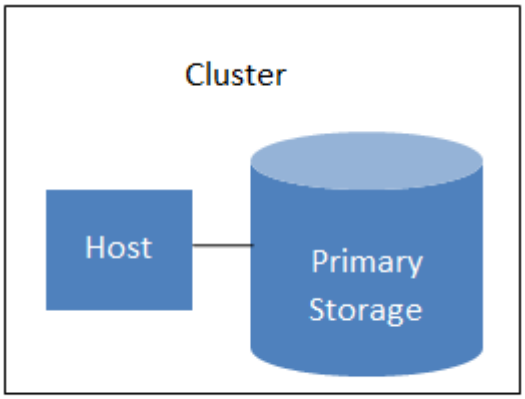
A simple pod

4.6. Clusters

A cluster provides a way to group hosts. To be precise, a cluster is a XenServer server pool, a set of KVM servers or a VMware cluster preconfigured in vCenter. The hosts in a cluster all have identical hardware, run the same hypervisor, are on the same subnet, and access the same shared primary storage. Virtual machine instances (VMs) can be live-migrated from one host to another within the same cluster without interrupting service to the user.

A cluster is the fourth-largest organizational unit within a CloudPlatform deployment. Clusters are contained within pods, pods are contained within zones, and zones can be contained within regions. Size of the cluster is only limited by the underlying hypervisor, although the CloudPlatform recommends you stay below the theoretically allowed maximum cluster size in most cases.

A cluster consists of one or more hosts and one or more primary storage servers.



A simple cluster

Even when local storage is used, clusters are still required. In this case, there is just one host per cluster.

(VMware) If you use VMware hypervisor hosts in your CloudPlatform deployment, each VMware cluster is managed by a vCenter server. The CloudPlatform administrator must register the vCenter

server with CloudPlatform. There may be multiple vCenter servers per zone. Each vCenter server may manage multiple VMware clusters.

4.7. Hosts

A host is a single computer. Hosts provide the computing resources that run guest virtual machines. Each host has hypervisor software installed on it to manage the guest VMs. For example, a host can be a Citrix XenServer server, a Linux KVM-enabled server, an ESXi server, or a Windows Hyper-V server.

The host is the smallest organizational unit within a CloudPlatform deployment. Hosts are contained within clusters, clusters are contained within pods, pods are contained within zones, and zones can be contained within regions.

Hosts in a CloudPlatform deployment:

- Provide the CPU, memory, storage, and networking resources needed to host the virtual machines
- Interconnect using a high bandwidth TCP/IP network and connect to the Internet
- May reside in multiple data centers across different geographic locations
- May have different capacities (different CPU speeds, different amounts of RAM, etc.), although the hosts within a cluster must all be homogeneous

Additional hosts can be added at any time to provide more capacity for guest VMs.

CloudPlatform automatically detects the amount of CPU and memory resources provided by the hosts.

Hosts are not visible to the end user. An end user cannot determine which host their guest has been assigned to.

For a host to function in CloudPlatform, you must do the following:

- Install hypervisor software on the host
- Assign an IP address to the host
- Ensure the host is connected to the CloudPlatform Management Server.

User Services Overview

In addition to the physical and logical infrastructure of your cloud, and the CloudPlatform software and servers, you also need a layer of user services so that people can actually make use of the cloud. This means not just a user UI, but a set of options and resources that users can choose from, such as templates for creating virtual machines, disk storage, and more. If you are running a commercial service, you will be keeping track of what services and resources users are consuming and charging them for that usage. Even if you do not charge anything for people to use your cloud – say, if the users are strictly internal to your organization, or just friends who are sharing your cloud – you can still keep track of what services they use and how much of them.

5.1. Service Offerings, Disk Offerings, Network Offerings, and Templates

A user creating a new instance can make a variety of choices about its characteristics and capabilities. CloudPlatform provides several ways to present users with choices when creating a new instance:

- Service Offerings, defined by the CloudPlatform administrator, provide a choice of CPU speed, number of CPUs, RAM size, tags on the root disk, and other choices. See Creating a New Compute Offering.
- Disk Offerings, defined by the CloudPlatform administrator, provide a choice of disk size for primary data storage. See Creating a New Disk Offering.
- Network Offerings, defined by the CloudPlatform administrator, describe the feature set that is available to end users from the virtual router or external networking devices on a given guest network. See Network Offerings.
- Templates, defined by the CloudPlatform administrator or by any CloudPlatform user, are the base OS images that the user can choose from when creating a new instance. For example, CloudPlatform includes CentOS as a template. See Working with Templates.

In addition to these choices that are provided for users, there is another type of service offering which is available only to the CloudPlatform root administrator, and is used for configuring virtual infrastructure resources. For more information, see Upgrading a Virtual Router with System Service Offerings.

Service Offerings

In this chapter we discuss compute, disk, and system service offerings. Network offerings are discussed in the section on setting up networking for users.

6.1. Compute and Disk Service Offerings

A service offering is a set of virtual hardware features such as CPU core count and speed, memory, and disk size. The CloudPlatform administrator can set up various offerings, and then end users choose from the available offerings when they create a new VM. Based on the user's selected offering, CloudPlatform emits usage records that can be integrated with billing systems.

Some characteristics of service offerings must be defined by the CloudPlatform administrator, and others can be left undefined so that the end-user can enter their own desired values. This is useful to reduce the number of offerings the CloudPlatform administrator has to define. Instead of defining a compute offering for every imaginable combination of values that a user might want, the administrator can define offerings that provide some flexibility to the users and can serve as the basis for several different VM configurations.

A service offering includes the following elements:

- CPU, memory, and network resource guarantees
- How resources are metered
- How the resource usage is charged
- How often the charges are generated

For example, one service offering might allow users to create a virtual machine instance that is equivalent to a 1 GHz Intel® Core™ 2 CPU, with 1 GB memory at $0.20/hour, with network traffic metered at $0.10/GB.

CloudPlatform separates service offerings into compute offerings and disk offerings. The compute service offering specifies:

- Guest CPU (optional). If not defined by the CloudPlatform administrator, users can pick the CPU attributes.
- Guest RAM (optional). If not defined by the CloudPlatform administrator, users can pick the RAM.
- Guest Networking type (virtual or direct)
- Tags on the root disk

The disk offering specifies:

- Disk size (optional). If not defined by the CloudPlatform administrator, users can pick the disk size.
- Tags on the data disk

6.1.1. Custom Compute Offering

CloudPlatform provides you the flexibility to specify the desired values for the number of CPU, CPU speed, and memory while deploying a VM. As an admin, you create a Compute Offering by marking it as custom, and the users will be able to customize this dynamic Compute Offering by specifying the

memory, and CPU at the time of VM creation or upgrade. Use this offering to deploy VM by specifying custom values for the dynamic parameters.

Dynamic Compute Offerings can be used in following cases: deploying a VM, changing the compute offering of a stopped VM and running VMs, which is nothing but scaling up. To support this feature a new field, Custom, has been added to the Create Compute Offering page in the UI. If the Custom field is checked, the user will be able to create a custom Compute Offering. During VM deployment you can specify desired values for number of CPU, CPU speed, and memory.

To support this feature, usage events has been enhanced to register events for dynamically assigned resources. Usage events are registered when a VM is created from a custom compute offering, and upon changing the compute offering of a stopped or running VM. The values of the parameters, such as CPU, speed, RAM are recorded.

6.2. System Service Offerings

System service offerings provide a choice of CPU speed, number of CPUs, tags, and RAM size, just as other service offerings do. But rather than being used for virtual machine instances and exposed to users, system service offerings are used to change the default properties of virtual routers, console proxies, and other system VMs. System service offerings are visible only to the CloudPlatform root administrator. CloudPlatform provides default system service offerings. The CloudPlatform root administrator can create additional custom system service offerings.

When CloudPlatform creates a virtual router for a guest network, it uses default settings which are defined in the system service offering associated with the network offering. You can upgrade the capabilities of the virtual router by applying a new network offering that contains a different system service offering. All virtual routers in that network will begin using the settings from the new service offering.

Storage Concepts Used in CloudPlatform

7.1. Storage Overview

CloudPlatform defines two types of storage: primary and secondary. Primary storage can be accessed by either iSCSI or NFS. Additionally, direct attached storage may be used for primary storage. Secondary storage is always accessed using NFS or a combination of NFS and object storage.

There is no ephemeral storage in CloudPlatform. All volumes on all nodes are persistent.

7.2. About Primary Storage

Primary storage is associated with a cluster or (in KVM and VMware) a zone, and it stores the disk volumes for all the VMs running on hosts.

You can add multiple primary storage servers to a cluster or zone. At least one is required. It is typically located close to the hosts for increased performance. CloudPlatform manages the allocation of guest virtual disks to particular primary storage devices.

It is useful to set up zone-wide primary storage when you want to avoid extra data copy operations. With cluster-based primary storage, data in the primary storage is directly available only to VMs within that cluster. If a VM in a different cluster needs some of the data, it must be copied from one cluster to another, using the zone's secondary storage as an intermediate step. This operation can be unnecessarily time-consuming.

For Hyper-V, SMB/CIFS storage is supported. Note that Zone-wide Primary Storage is not supported in Hyper-V.

CloudPlatform is designed to work with all standards-compliant iSCSI and NFS servers that are supported by the underlying hypervisor, including, for example:

- Dell EqualLogic™ for iSCSI
- Network Appliances filers for NFS and iSCSI
- Scale Computing for NFS

If you intend to use only local disk for your installation, you can skip adding separate primary storage.

7.2.1. Runtime Behavior of Primary Storage

Root volumes are created automatically when a virtual machine is created. Root volumes are deleted when the VM is destroyed. Data volumes can be created and dynamically attached to VMs (although, when the Oracle VM hypervisor is used, the VM must be stopped before an additional volume can be attached). Data volumes are not deleted when VMs are destroyed.

Administrators should monitor the capacity of primary storage devices and add additional primary storage as needed. See the Advanced Installation Guide.

Administrators add primary storage to the system by creating a CloudPlatform storage pool. Each storage pool is associated with a cluster.

7.2.2. Hypervisor Support for Primary Storage

The following table shows storage options and parameters for different hypervisors.

	VMware vSphere	Citrix XenServer	KVM	Hyper-V
Format for Disks, Templates, and Snapshots	VMDK	VHD	QCOW2	VHD Snapshorts are not supported.
iSCSI support	VMFS	Clustered LVM	Yes, via Shared Mountpoint	No
Fiber Channel support	VMFS	Yes, via Existing SR	Yes, via Shared Mountpoint	No
NFS support	Y	Y	Y	No
Local storage support	Y	Y	Y	Y
Storage over-provisioning	NFS and iSCSI	NFS	NFS	No
SMB/CIFS	No	No	No	Yes

XenServer uses a clustered LVM system to store VM images on iSCSI and Fiber Channel volumes and does not support over-provisioning in the hypervisor. The storage server itself, however, can support thin-provisioning. As a result the CloudPlatform can still support storage over-provisioning by running on thin-provisioned storage volumes.

KVM supports "Shared Mountpoint" storage. A shared mountpoint is a file system path local to each server in a given cluster. The path must be the same across all Hosts in the cluster, for example /mnt/primary1. This shared mountpoint is assumed to be a clustered filesystem such as OCFS2. In this case the CloudPlatform does not attempt to mount or unmount the storage as is done with NFS. The CloudPlatform requires that the administrator insure that the storage is available

With other hypervisors, CloudPlatform takes care of mounting the iSCSI target on the host whenever it discovers a connection with an iSCSI server and unmounting the target when it discovers the connection is down.

With NFS storage, CloudPlatform manages the overprovisioning. In this case the global configuration parameter storage.overprovisioning.factor controls the degree of overprovisioning. This is independent of hypervisor type.

Local storage is an option for primary storage for vSphere, XenServer, and KVM. When the local disk option is enabled, a local disk storage pool is automatically created on each host. To use local storage for the System Virtual Machines (such as the Virtual Router), set system.vm.use.local.storage to true in global configuration.

CloudPlatform supports multiple primary storage pools in a Cluster. For example, you could provision 2 NFS servers in primary storage. Or you could provision 1 iSCSI LUN initially and then add a second iSCSI LUN when the first approaches capacity.

7.2.3. Storage Tags

Storage may be "tagged". A tag is a text string attribute associated with primary storage, a Disk Offering, or a Service Offering. Tags allow administrators to provide additional information about the

storage. For example, that is a "SSD" or it is "slow". Tags are not interpreted by CloudPlatform. They are matched against tags placed on service and disk offerings. CloudPlatform requires all tags on service and disk offerings to exist on the primary storage before it allocates root or data disks on the primary storage. Service and disk offering tags are used to identify the requirements of the storage that those offerings have. For example, the high end service offering may require "fast" for its root disk volume.

The interaction between tags, allocation, and volume copying across clusters and pods can be complex. To simplify the situation, use the same set of tags on the primary storage for all clusters in a pod. Even if different devices are used to present those tags, the set of exposed tags can be the same.

7.2.4. Maintenance Mode for Primary Storage

Primary storage may be placed into maintenance mode. This is useful, for example, to replace faulty RAM in a storage device. Maintenance mode for a storage device will first stop any new guests from being provisioned on the storage device. Then it will stop all guests that have any volume on that storage device. When all such guests are stopped the storage device is in maintenance mode and may be shut down. When the storage device is online again you may cancel maintenance mode for the device. The CloudPlatform will bring the device back online and attempt to start all guests that were running at the time of the entry into maintenance mode.

7.3. About Secondary Storage

Secondary storage stores the following:

- Templates — OS images that can be used to boot VMs and can include additional configuration information, such as installed applications
- ISO images — disc images containing data or bootable media for operating systems
- Disk volume snapshots — saved copies of VM data which can be used for data recovery or to create new templates

The items in secondary storage are available to all hosts in the scope of the secondary storage, which may be defined as per zone or per region.

To make items in secondary storage available to all hosts throughout the cloud, you can add object storage in addition to the zone-based NFS Secondary Staging Store. It is not necessary to copy templates and snapshots from one zone to another, as would be required when using zone NFS alone. Everything is available everywhere.

For Hyper-V hosts, SMB storage is supported.

Note

Object storage is not supported on Hyper-V.

Warning

Heterogeneous Secondary Storage is not supported in Regions. For example, you cannot set up multiple zones, one using NFS secondary and the other using S3 secondary.

7.4. About Storage Volumes

A volume provides storage to a guest VM. The volume can provide for a root disk or an additional data disk. CloudPlatform supports additional volumes for guest VMs.

Volumes are created for a specific hypervisor type. A volume that has been attached to guest using one hypervisor type (e.g, XenServer) may not be attached to a guest that is using another hypervisor type (e.g. vSphere, Oracle VM, KVM). This is because the different hypervisors use different disk image formats.

CloudPlatform defines a volume as a unit of storage available to a guest VM. Volumes are either root disks or data disks. The root disk has “/” in the file system and is usually the boot device. Data disks provide for additional storage (e.g. As “/opt” or “D:”). Every guest VM has a root disk, and VMs can also optionally have a data disk. End users can mount multiple data disks to guest VMs. Users choose data disks from the disk offerings created by administrators. The user can create a template from a volume as well; this is the standard procedure for private template creation. Volumes are hypervisor-specific: a volume from one hypervisor type may not be used on a guest of another hypervisor type.

Note

CloudPlatform supports attaching up to 13 data disks to a VM on XenServer hypervisor versions 6.0 and above. For the VMs on other hypervisor types, the data disk limit is 6.

7.5. About Volume Snapshots

CloudPlatform supports snapshots of disk volumes. Snapshots are a point-in-time capture of virtual machine disks. Memory and CPU states are not captured.

Snapshots may be taken for volumes, including both root and data disks. The administrator places a limit on the number of stored snapshots per user. Users can create new volumes from the snapshot for recovery of particular files and they can create templates from snapshots to boot from a restored disk.

Users can create snapshots manually or by setting up automatic recurring snapshot policies. Users can also create disk volumes from snapshots, which may be attached to a VM like any other disk volume. Snapshots of both root disks and data disks are supported. However, CloudPlatform does not currently support booting a VM from a recovered root disk. A disk recovered from snapshot of a root disk is treated as a regular data disk; the data on recovered disk can be accessed by attaching the disk to a VM.

A completed snapshot is copied from primary storage to secondary storage, where it is stored until deleted or purged by newer snapshot.

This feature is supported for the following hypervisors: XenServer VMware, and KVM.

7.5.1. Automatic Snapshot Creation and Retention

(Supported for the following hypervisors: **XenServer**, **VMware vSphere**, and **KVM**)

Users can set up a recurring snapshot policy to automatically create multiple snapshots of a disk at regular intervals. Snapshots can be created on an hourly, daily, weekly, or monthly interval. One snapshot policy can be set up per disk volume. For example, a user can set up a daily snapshot at 02:30.

With each snapshot schedule, users can also specify the number of scheduled snapshots to be retained. Older snapshots that exceed the retention limit are automatically deleted. This user-defined limit must be equal to or lower than the global limit set by the CloudPlatform administrator. The limit applies only to those snapshots that are taken as part of an automatic recurring snapshot policy. Additional manual snapshots can be created and retained.

7.5.2. Incremental Snapshots and Backup

Snapshots are created on primary storage where a disk resides. After a snapshot is created, it is immediately backed up to secondary storage and removed from primary storage for optimal utilization of space on primary storage.

CloudPlatform does incremental backups for some hypervisors. When incremental backups are supported, every backup is a full backup.

	VMware vSphere	Citrix XenServer	KVM	Hyper-V
Support for Incremental Backup	N	Y	N	N

7.5.3. Volume Status

When a snapshot operation is triggered by means of a recurring snapshot policy, a snapshot is skipped if a volume has remained inactive since its last snapshot was taken. A volume is considered to be inactive if it is either detached or attached to a VM that is not running. CloudPlatform ensures that at least one snapshot is taken since the volume last became inactive.

When a snapshot is taken manually, a snapshot is always created regardless of whether a volume has been active or not.

7.5.4. Snapshot Restore

There are two paths to restoring snapshots. Users can create a volume from the snapshot. The volume can then be mounted to a VM and files recovered as needed. Alternatively, a template may be created from the snapshot of a root disk. The user can then boot a VM from this template to effect recovery of the root disk.

7.5.5. Snapshot Job Throttling

When a snapshot of a virtual machine is requested, the snapshot job runs on the same host where the VM is running or, in the case of a stopped VM, the host where it ran last. If many snapshots are requested for VMs on a single host, this can lead to problems with too many snapshot jobs overwhelming the resources of the host.

To address this situation, the cloud's root administrator can throttle how many snapshot jobs are executed simultaneously on the hosts in the cloud by using the new global configuration setting concurrent.snapshots.threshold.perhost. By using this setting, the administrator can better ensure that snapshot jobs do not time out and hypervisor hosts do not experience performance issues due to hosts being overloaded with too many snapshot requests.

Set concurrent.snapshots.threshold.perhost to a value that represents a best guess about how many snapshot jobs the hypervisor hosts can execute at one time, given the current resources of the hosts and the number of VMs running on the hosts. If a given host has more snapshot requests, the additional requests are placed in a waiting queue. No new snapshot jobs will start until the number of currently executing snapshot jobs falls below the configured limit.

The admin can also set job.expire.minutes to place a maximum on how long a snapshot request will wait in the queue. If this limit is reached, the snapshot request fails and returns an error message.

7.5.6. VMware Volume Snapshot Performance

When you take a snapshot of a data or root volume on VMware, CloudPlatform uses an efficient storage technique to improve performance.

A snapshot is not immediately exported from vCenter to a mounted NFS share and packaged into an OVA file format. This operation would consume time and resources. Instead, the original file formats (e.g., VMDK) provided by vCenter are retained. An OVA file will only be created as needed, on demand. To generate the OVA, CloudPlatform uses information in a properties file (*.ova.meta) which it stored along with the original snapshot data.

Note

For upgrading customers: This process applies only to newly created snapshots after upgrade to CloudPlatform 4.2. Snapshots that have already been taken and stored in OVA format will continue to exist in that format, and will continue to work as expected.

Networking for Users

8.1. Overview of Setting Up Networking for Users

People using cloud infrastructure have a variety of needs and preferences when it comes to the networking services provided by the cloud. As a CloudPlatform administrator, you can do the following things to set up networking for your users:

- Set up physical networks in zones
- Set up several different providers for the same service on a single physical network (for example, both Cisco and Juniper firewalls)
- Bundle different types of network services into network offerings, so users can choose the desired network services for any given virtual machine
- Add new network offerings as time goes on so end users can upgrade to a better class of service on their network
- Provide more ways for a network to be accessed by a user, such as through a project of which the user is a member

8.2. About Virtual Networks

A virtual network is a logical construct that enables multi-tenancy on a single physical network. In CloudPlatform a virtual network can be shared or isolated.

8.2.1. Isolated Networks

An isolated network can be accessed only by virtual machines of a single account. Isolated networks have the following properties.

- Resources such as VLAN are allocated and garbage collected dynamically
- There is one network offering for the entire network
- The network offering can be upgraded or downgraded but it is for the entire network

8.2.2. Shared Networks

A shared network can be accessed by virtual machines that belong to many different accounts. Network Isolation on shared networks is accomplished by using techniques such as security groups, which is supported only in Basic zones.

- Shared Networks are created by the administrator
- Shared Networks can be designated to a certain domain
- Shared Network resources such as VLAN and physical network that it maps to are designated by the administrator
- Shared Networks can be isolated by security groups
- Public Network is a shared network that is not shown to the end users

- Source NAT per zone is not supported when the service provider is virtual router. However, Source NAT per account is supported with virtual router in a Shared Network.

8.2.3. Runtime Allocation of Virtual Network Resources

When you define a new virtual network, all your settings for that network are stored in CloudPlatform. The actual network resources are activated only when the first virtual machine starts in the network. When all virtual machines have left the virtual network, the network resources are garbage collected so they can be allocated again. This helps to conserve network resources.

8.3. About Redundant Virtual Routers

Each CloudPlatform account uses a virtual router to provide network services to the resources that are part of that account. Typically in CloudPlatform, all the user VMs are in a private network and they communicate with the external network using a virtual router. They use virtual router as a gateway. If the virtual router fails, the network connection to the guest VMs gets disabled and the users face problems in accessing the guest VMs.

To address such exigencies, CloudPlatform provides the Redundant Virtual Router (RVR) feature. This feature enables the guest network to recover and resume operations if the virtual router that connects it to the external network fails. After you enable this feature, each guest network will have two virtual routers. The virtual router that receives and responds to the guest VM network is known as Master. The other virtual router is known as Backup. The Backup virtual router remains in the passive mode.

If the Master virtual router is down, the Backup virtual router takes the Master's role. The Backup router takes only a few seconds to get activated. This ensures minimum downtime. Another virtual router will be launched as a new Backup router. CloudPlatform uses the Virtual Redundant Router Protocol (VRRP) for the communication between the Master and the Backup virtual routers.

CloudPlatform uses the **`keepalived`** process and the **`conntrackd`** process that run on the Master router. The keepalived process implements the VRRP protocol. This process sends broadcast message every second to indicate that the master router is up and running. The **`conntrackd`** process tracks the TCP connections on the Master router.

If the Backup router does not receive three consecutive broadcast messages from the Master, it gets activated and starts functioning as the Master router. The Backup router uses the information on the TCP connection of the Master that the **`conntrackd`** process tracked for restoring the TCP connections.

The IP address of the gateway is independent and does not belong to the NICs of the Master and the Backup virtual routers. After the Backup router is activated as Master, the **`keepalived`** process send the gratuitous ARP to bind the gateway IP address to the MAC associated with the Guest VMs.

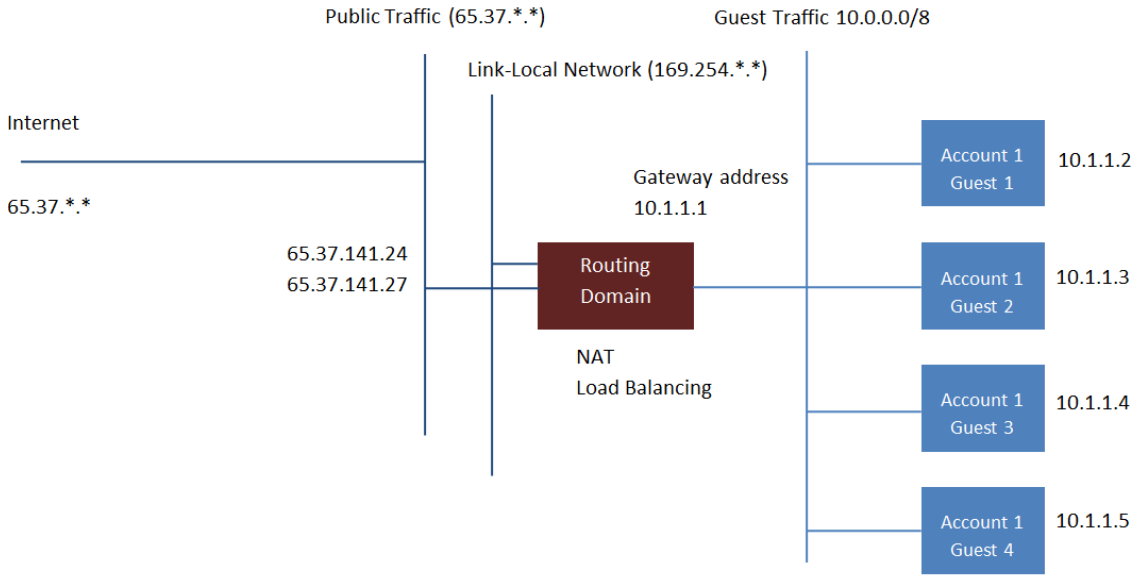
When a virtual router switches to Master, it does the following:

- Enables all the public interfaces.
- Sends out gratuitous ARP to the public gateway to update ARP cache.
- Starts password, dnsmasq, and VPN services.
- Updates the state of the conntrackd process to "primary".

8.4. Guest Traffic

A network can carry guest traffic only between VMs within one zone. Virtual machines in different zones cannot communicate with each other using their IP addresses; they must communicate with each other by routing through a public IP address.

See a typical guest traffic setup given below:



Guest Traffic Setup

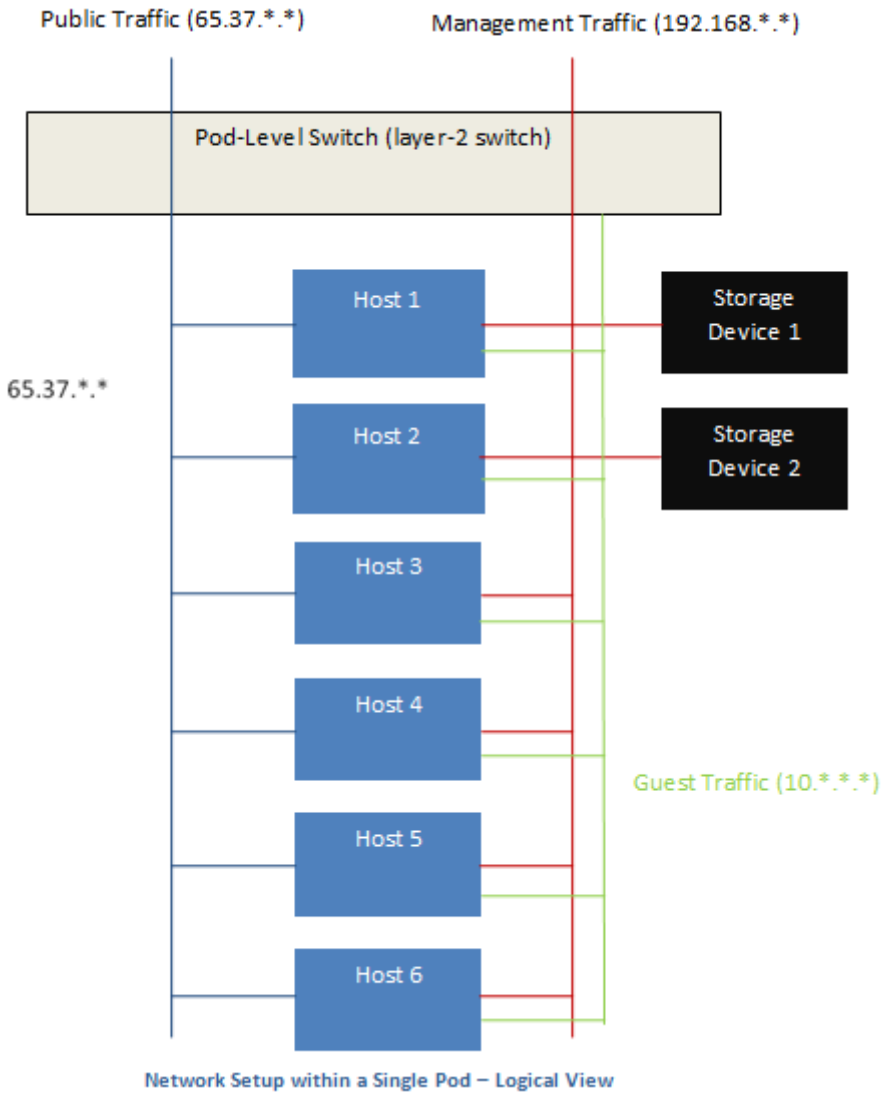
Typically, the Management Server automatically creates a virtual router for each network. A virtual router is a special virtual machine that runs on the hosts. Each virtual router in an isolated network has three network interfaces. If multiple public VLAN is used, the router will have multiple public interfaces. Its eth0 interface serves as the gateway for the guest traffic and has the IP address of 10.1.1.1. Its eth1 interface is used by the system to configure the virtual router. Its eth2 interface is assigned a public IP address for public traffic. If multiple public VLAN is used, the router will have multiple public interfaces.

The virtual router provides DHCP and will automatically assign an IP address for each guest VM within the IP range assigned for the network. The user can manually reconfigure guest VMs to assume different IP addresses.

Source NAT is automatically configured in the virtual router to forward outbound traffic for all guest VMs

8.5. Networking in a Pod

The figure below illustrates network setup within a single pod. The hosts are connected to a pod-level switch. At a minimum, the hosts should have one physical uplink to each switch. Bonded NICs are supported as well. The pod-level switch is a pair of redundant gigabit switches with 10 G uplinks.



Network Setup within a Single Pod – Logical View

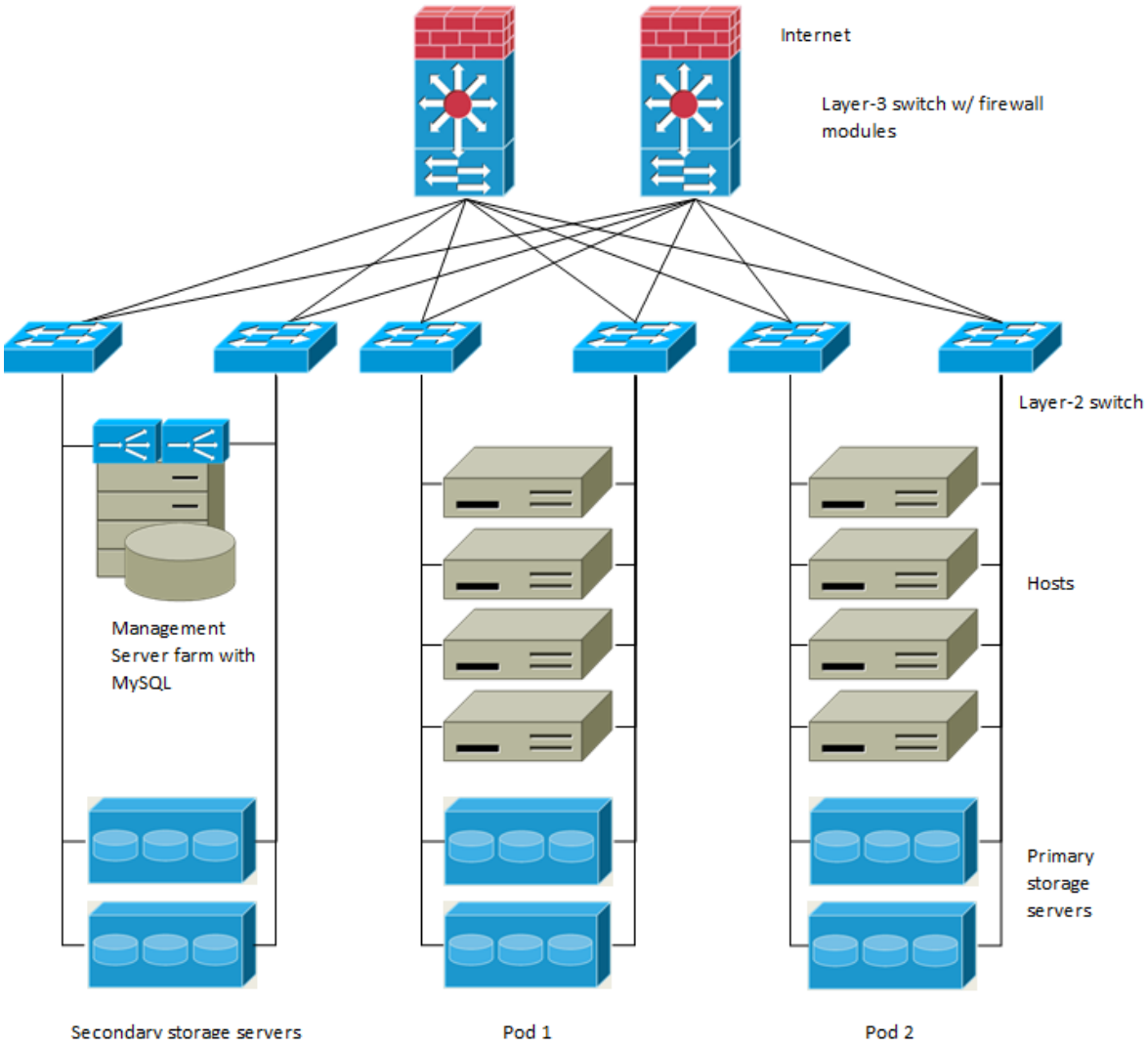
Servers are connected as follows:

- Storage devices are connected to only the network that carries management traffic.
- Hosts are connected to networks for both management traffic and public traffic.
- Hosts are also connected to one or more networks carrying guest traffic.

We recommend the use of multiple physical Ethernet cards to implement each network interface as well as redundant switch fabric in order to maximize throughput and improve reliability.

8.6. Networking in a Zone

The following figure illustrates the network setup within a single zone.



A firewall for management traffic operates in the NAT mode. The network typically is assigned IP addresses in the 192.168.0.0/16 Class B private address space. Each pod is assigned IP addresses in the 192.168.*.0/24 Class C private address space.

Each zone has its own set of public IP addresses. Public IP addresses from different zones do not overlap.

8.7. About Using a NetScaler Load Balancer

Citrix NetScaler is supported as an external network element for load balancing in zones that use isolated networking in advanced zones. Set up an external load balancer when you want to provide load balancing through means other than CloudPlatform's provided virtual router.

Note

In a Basic zone, load balancing service is only supported if Elastic IP or Elastic LB services are enabled.

When NetScaler load balancer is used to provide EIP or ELB services in a Basic zone, ensure that all guest VM traffic must enter and exit through the NetScaler device. When inbound traffic goes through the NetScaler device, traffic is routed by using the NAT protocol depending on the EIP/ELB configured on the public IP to the private IP. The traffic that is originated from the guest VMs usually goes through the layer 3 router. To ensure that outbound traffic goes through NetScaler device providing EIP/ELB, layer 3 router must have a policy-based routing. A policy-based route must be set up so that all traffic originated from the guest VM's are directed to NetScaler device. This is required to ensure that the outbound traffic from the guest VM's is routed to a public IP by using NAT. For more information on Elastic IP, see *Section 8.8, “About Elastic IP”*.

The NetScaler can be set up in direct (outside the firewall) mode. It must be added before any load balancing rules are deployed on guest VMs in the zone.

The functional behavior of the NetScaler with CloudPlatform is the same as described in the CloudPlatform documentation for using an F5 external load balancer. The only exception is that the F5 supports routing domains, and NetScaler does not. NetScaler can not yet be used as a firewall.

To install and enable an external load balancer for CloudPlatform management, see

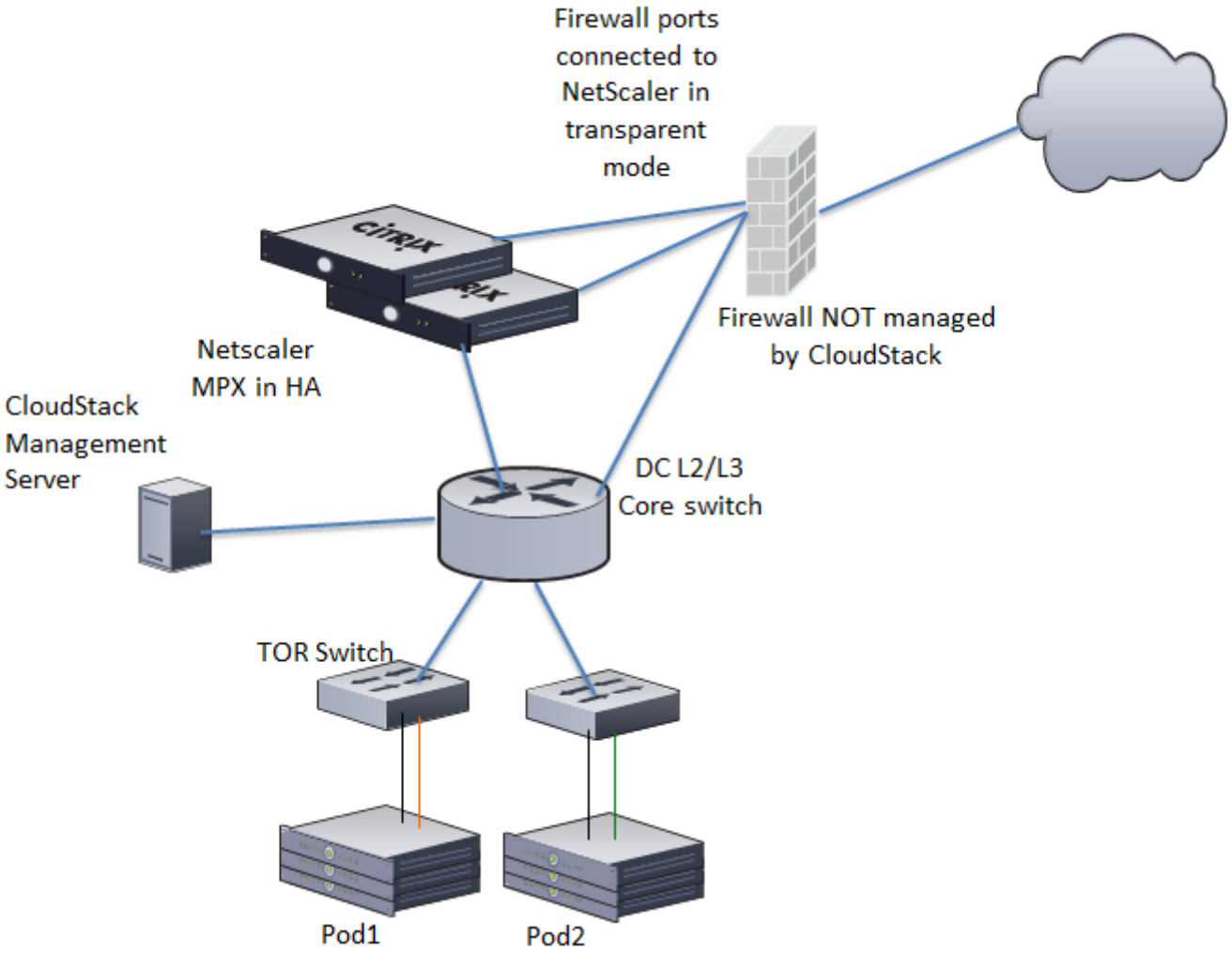
The Citrix NetScaler comes in three varieties. The following table summarizes how these variants are treated in CloudPlatform.

NetScaler ADC Type	Description of Capabilities	CloudPlatform Supported Features
MPX	Physical appliance. Capable of deep packet inspection. Can act as application firewall and load balancer	In advanced zones, load balancer functionality fully supported without limitation. In basic zones, static NAT, elastic IP (EIP), and elastic load balancing (ELB) are also provided.
VPX	Virtual appliance. Can run as VM on XenServer, ESXi, and KVM hypervisors. Same functionality as MPX	Supported on ESXi, XenServer, and KVM. Same functional support as for MPX. CloudPlatform will treat VPX and MPX as the same device type.
SDX	Physical appliance. Can create multiple fully isolated VPX instances on a single appliance to support multi-tenant usage	CloudPlatform will dynamically provision, configure, and manage the lifecycle of VPX instances on the SDX. Provisioned instances are added into CloudPlatform automatically – no manual configuration by the administrator is required. Once a VPX instance is added into CloudPlatform, it is treated the same as a VPX on an ESXi host.

8.8. About Elastic IP

Elastic IP (EIP) addresses are the IP addresses that are associated with an account, and act as static IP addresses. The account owner has the complete control over the Elastic IP addresses that belong to the account. As an account owner, you can allocate an Elastic IP to a VM of your choice from the EIP pool of your account. Later if required you can reassign the IP address to a different VM. This feature is extremely helpful during VM failure. Instead of replacing the VM which is down, the IP address can be reassigned to a new VM in your account.

Similar to the public IP address, Elastic IP addresses are mapped to their associated private IP addresses by using StaticNAT. The EIP service is equipped with StaticNAT (1:1) service in an EIP-enabled basic zone. The default network offering, DefaultSharedNetscalerEIPandELBNetworkOffering, provides your network with EIP and ELB network services if a NetScaler device is deployed in your zone. Consider the following illustration for more details.



In the illustration, a NetScaler appliance is the default entry or exit point for the CloudPlatform instances, and firewall is the default entry or exit point for the rest of the data center. Netscaler provides LB services and staticNAT service to the guest networks. The guest traffic in the pods and the Management Server are on different subnets / VLANs. The policy-based routing in the data center core switch sends the public traffic through the NetScaler, whereas the rest of the data center goes through the firewall.

The EIP work flow is as follows:

- When a user VM is deployed, a public IP is automatically acquired from the pool of public IPs configured in the zone. This IP is owned by the VM's account.
- Each VM will have its own private IP. When the user VM starts, Static NAT is provisioned on the NetScaler device by using the Inbound Network Address Translation (INAT) and Reverse NAT (RNAT) rules between the public IP and the private IP.

Note

Inbound NAT (INAT) is a type of NAT supported by NetScaler, in which the destination IP address is replaced in the packets from the public network, such as the Internet, with the private IP address of a VM in the private network. Reverse NAT (RNAT) is a type of NAT supported by NetScaler, in which the source IP address is replaced in the packets generated by a VM in the private network with the public IP address.

- This default public IP will be released in two cases:
  - When the VM is stopped. When the VM starts, it again receives a new public IP, not necessarily the same one allocated initially, from the pool of Public IPs.
  - The user acquires a public IP (Elastic IP). This public IP is associated with the account, but will not be mapped to any private IP. However, the user can enable Static NAT to associate this IP to the private IP of a VM in the account. The Static NAT rule for the public IP can be disabled at any time. When Static NAT is disabled, a new public IP is allocated from the pool, which is not necessarily be the same one allocated initially.

For the deployments where public IPs are limited resources, you have the flexibility to choose not to allocate a public IP by default. You can use the Associate Public IP option to turn on or off the automatic public IP assignment in the EIP-enabled Basic zones. If you turn off the automatic public IP assignment while creating a network offering, only a private IP is assigned to a VM when the VM is deployed with that network offering. Later, the user can acquire an IP for the VM and enable static NAT.

Note

The Associate Public IP feature is designed only for use with user VMs. The System VMs continue to get both public IP and private by default, irrespective of the network offering configuration.

New deployments which use the default shared network offering with EIP and ELB services to create a shared network in the Basic zone will continue allocating public IPs to each user VM.

8.9. About Global Server Load Balancing

Global Server Load Balancing (GSLB) is an extension of load balancing functionality, which is highly efficient in avoiding downtime. Based on the nature of deployment, GSLB represents a set of technologies that is used for various purposes, such as load sharing, disaster recovery, performance, and legal obligations. With GSLB, workloads can be distributed across multiple data centers situated at geographically separated locations. GSLB can also provide an alternate location for accessing a resource in the event of a failure, or to provide a means of shifting traffic easily to simplify maintenance, or both.

8.9.1. Components of GSLB

A typical GSLB environment is comprised of the following components:

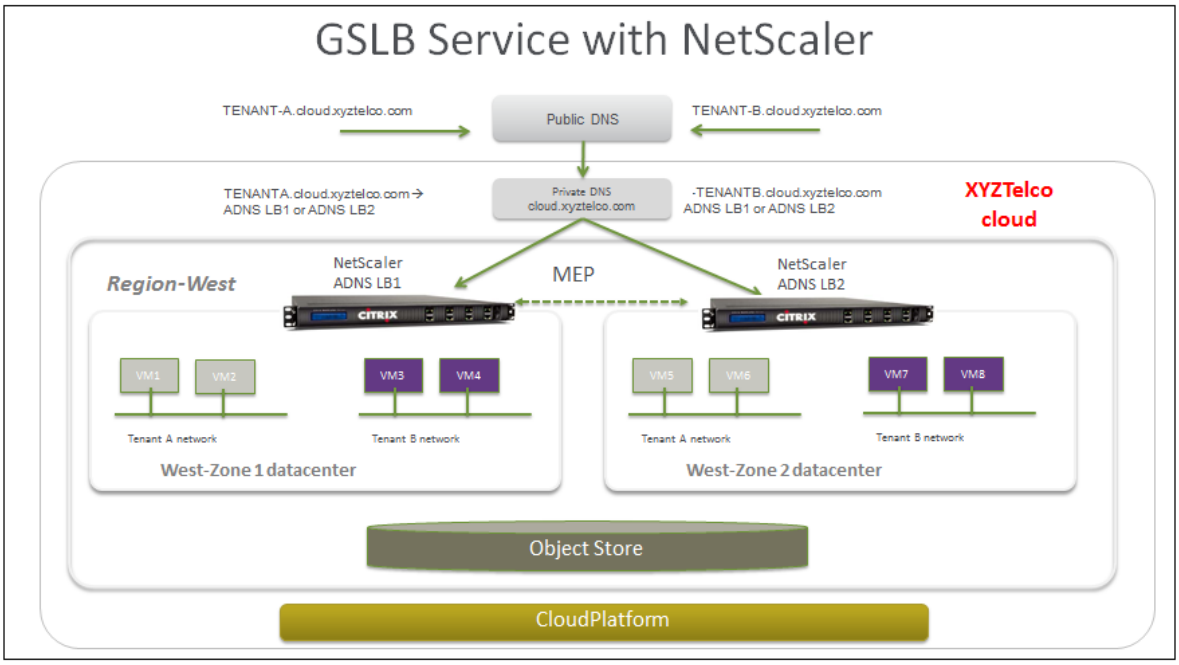
- **GSLB Site**: In CloudPlatformterminology, GSLB sites are represented by zones that are mapped to data centers, each of which has various network appliances. Each GSLB site is managed by a NetScaler appliance that is local to that site. Each of these appliances treats its own site as the local site and all other sites, managed by other appliances, as remote sites. It is the central entity in a GSLB deployment, and is represented by a name and an IP address.
- **GSLB Services**: A GSLB service is typically represented by a load balancing virtual server in a zone. In a GSLB environment, you can have a local as well as remote GSLB services. A local GSLB service represents a local load balancing or content switching virtual server. A remote GSLB service is the one configured at one of the other sites in the GSLB setup. At each site in the GSLB setup, you can create one local GSLB service and any number of remote GSLB services.
- **GSLB Virtual Servers**: A GSLB virtual server refers to a logical grouping of one or more GSLB services. CloudPlatform GSLB functionality ensures that traffic is load balanced across VMs in multiple zones. It evaluates the configured GSLB methods or algorithms to select a GSLB service to which to send the client requests. One or more virtual servers from different zones are bound to the GSLB virtual server. GSLB virtual server does not have a public IP associated with it, instead it will have a FQDN DNS name.
- **Load Balancing or Content Switching Virtual Servers**: According to Citrix NetScaler terminology, a load balancing or content switching virtual server represents one or many servers on the local network. Clients send their requests to the load balancing or content switching virtual server's virtual IP (VIP) address, and the virtual server balances the load across the local servers. After a GSLB virtual server selects a GSLB service representing either a local or a remote load balancing or content switching virtual server, the client sends the request to that virtual server's VIP address.
- **DNS VIPs**: DNS virtual IP represents a load balancing DNS virtual server on the GSLB service provider. The DNS requests for domains for which the GSLB service provider is authoritative can be sent to a DNS VIP.
- **Authoritative DNS**: ADNS (Authoritative Domain Name Server) is a service that provides actual answer to DNS queries, such as web site IP address. In a GSLB environment, an ADNS service responds only to DNS requests for domains for which the GSLB service provider is authoritative. When an ADNS service is configured, the service provider owns that IP address and advertises it. When you create an ADNS service, the NetScaler responds to DNS queries on the configured ADNS service IP and port.

8.9.2. How GSLB Works in CloudPlatform

Global server load balancing is used to manage the traffic flow to a web site hosted on two separate zones that ideally are in different geographic locations. The following is an illustration of how GLSB functionality is provided in CloudPlatform: An organization, xyztelco, has set up a public cloud

that spans two zones, Zone-1 and Zone-2, across geographically separated data centers that are managed by CloudPlatform. Tenant-A of the cloud launches a highly available solution by using xyztelco cloud. For that purpose, they launch two instances each in both the zones: VM1 and VM2 in Zone-1 and VM5 and VM6 in Zone-2. Tenant-A acquires a public IP, IP-1 in Zone-1, and configures a load balancer rule to load balance the traffic between VM1 and VM2 instances. CloudPlatform orchestrates setting up a virtual server on the LB service provider in Zone-1. Virtual server 1 that is set up on the LB service provider in Zone-1 represents a publicly accessible virtual server that client reaches at IP-1. The client traffic to virtual server 1 at IP-1 will be load balanced across VM1 and VM2 instances.

Tenant-A acquires another public IP, IP-2 in Zone-2 and sets up a load balancer rule to load balance the traffic between VM5 and VM6 instances. Similarly in Zone-2, CloudPlatform orchestrates setting up a virtual server on the LB service provider. Virtual server 2 that is setup on the LB service provider in Zone-2 represents a publicly accessible virtual server that client reaches at IP-2. The client traffic that reaches virtual server 2 at IP-2 is load balanced across VM5 and VM6 instances. At this point Tenant-A has the service enabled in both the zones, but has no means to set up a disaster recovery plan if one of the zone fails. Additionally, there is no way for Tenant-A to load balance the traffic intelligently to one of the zones based on load, proximity and so on. The cloud administrator of xyztelco provisions a GSLB service provider to both the zones. A GSLB provider is typically an ADC that has the ability to act as an ADNS (Authoritative Domain Name Server) and has the mechanism to monitor health of virtual servers both at local and remote sites. The cloud admin enables GSLB as a service to the tenants that use zones 1 and 2.



Tenant-A wishes to leverage the GSLB service provided by the xyztelco cloud. Tenant-A configures a GSLB rule to load balance traffic across virtual server 1 at Zone-1 and virtual server 2 at Zone-2. The domain name is provided as A.xyztelco.com. CloudPlatform orchestrates setting up GSLB virtual server 1 on the GSLB service provider at Zone-1. CloudPlatform binds virtual server 1 of Zone-1 and virtual server 2 of Zone-2 to GLSB virtual server 1. GSLB virtual server 1 is configured to start monitoring the health of virtual server 1 and 2 in Zone-1. CloudPlatform will also orchestrate setting up GSLB virtual server 2 on GSLB service provider at Zone-2. CloudPlatform will bind virtual server 1 of Zone-1 and virtual server 2 of Zone-2 to GLSB virtual server 2. GSLB virtual server 2 is configured to start monitoring the health of virtual server 1 and 2. CloudPlatform will bind the domain A.xyztelco.com to both the GSLB virtual server 1 and 2. At this point, Tenant-A service will be globally reachable at A.xyztelco.com. The private DNS server for the domain xyztelcom.com is configured by the admin out-of-band to resolve the domain A.xyztelco.com to the GSLB providers at both the zones, which are

configured as ADNS for the domain A.xyztelco.com. A client when sends a DNS request to resolve A.xyztelcom.com, will eventually get DNS delegation to the address of GSLB providers at zone 1 and 2. A client DNS request will be received by the GSLB provider. The GSLB provider, depending on the domain for which it needs to resolve, will pick up the GSLB virtual server associated with the domain. Depending on the health of the virtual servers being load balanced, DNS request for the domain will be resolved to the public IP associated with the selected virtual server.

8.10. Network Service Providers

Note

For the most up-to-date list of supported network service providers, see the CloudPlatform UI or call listNetworkServiceProviders.

A service provider (also called a network element) is hardware or virtual appliance that makes a network service possible; for example, a firewall appliance can be installed in the cloud to provide firewall service. On a single network, multiple providers can provide the same network service. For example, a firewall service may be provided by Cisco or Juniper devices in the same physical network.

You can have multiple instances of the same service provider in a network, for example, more than one Juniper SRX device.

If different providers are set up to provide the same service on the network, the administrator can create network offerings so users can specify which network service provider they prefer (along with the other choices offered in network offerings). Otherwise, CloudPlatform will choose which provider to use whenever the service is called for.

8.11. Network Service Providers Support Matrix

8.11.1. Individual

- Y = Supported
- N = Not Supported

	Virtual Router	VPC Virtual Router	BigIP F5	Juniper SRX	Citrix NetScaler	Cisco ASA
DHCP	Y	Y	N	N	N	N
DNS	Y	Y	N	N	N	N
User Data	Y	Y	N	N	N	N
Source NAT	Y	Y	N	Y	N	Y
Static NAT	Y	Y	N	Y	N	Y
Port Forwarding	Y	Y	N	Y	N	Y
Load Balancing	Y	Y	Y	N	Y	N

	Virtual Router	VPC Virtual Router	BigIP F5	Juniper SRX	Citrix NetScaler	Cisco ASA
Remote VPN	Y	N	N	N	N	N
Network ACL	N	Y	N	N	N	N
Usage Monitoring	Y	Y	Y	Y	Y	N
Security Group	N	N	N	N	N	N
Firewall	Y	N	N	Y	N	Y

8.11.2. Support Matrix for an Isolated Network (Combination)

- Y = Supported
- N = Not Supported

NW Devices	DHCP	DNS	User Data	Source NAT	Static NAT	Port Forwarding	Load Balanc--ing	Remote VPN	NW ACL	Usage Monitor-ing	Secur--ity Group	Firewall
Virtual Router (VR)	VR	VR	VR	VR	VR	VR	VR (TCP)	VR	N	VR	N	VR
VPC Virtual Router	VPC VR	VPC VR	VPC VR	VPC VR	VPC VR	VPC VR	VPC VR	N	VPC VR	Y	N	N
VR and F5 Side by side	VR	VR	VR	VR	VR	VR	F5	VR	N	Y	N	Static NAT / PF - Yes LB - No
VR and NetSca--ler Side by Side	VR	VR	VR	VR	VR	VR	NetSca--ler	VR	N	Y	N	Static NAT / PF - Yes LB - No
SRX and F5 Side by Side	VR	VR	VR	SRX	SRX	SRX	F5	N	N	Y	N	Static NAT / PF - Yes LB - No
SRX and NetSca--ler Side by Side	VR	VR	VR	SRX	SRX	SRX	NetSca--ler	N	N	Y	N	Static NAT / PF - Yes LB - No
SRX and F5 Inline	VR	VR	VR	SRX	SRX	SRX	F5	N	N	Y	N	Static NAT / PF - Yes LB - Yes

8.11.3. Support Matrix for Shared Network (Combination)

- Y = Supported
- N = Not Supported

NW Devices	DHCP	DNS	User Data	Source NAT	Static NAT	Port Forwarding	Load Balanc--ing	Remote VPN	NW ACL	Usage Monitor-ing	Secur--ity Group	Firewall
Virtual Router	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	N	Y
VR and F5 Side by side	VR	VR	VR	VR	VR	VR	F5	VR	N	Y	N	Static NAT / PF - Yes LB - No
VR and NetSca--ler Side by Side	VR	VR	VR	VR	VR	VR	NetSca--ler	VR	N	Y	N	Static NAT / PF - Yes LB - No
SRX and F5 Side by Side	VR	VR	VR	SRX	SRX	SRX	F5	N	N	Y	N	Static NAT / PF - Yes LB - No
SRX and NetSca--ler Side by Side	VR	VR	VR	SRX	SRX	SRX	NetSca--ler	N	N	Y	N	Static NAT / PF - Yes LB - No
SRX and F5 Inline	VR	VR	VR	SRX	SRX	SRX	F5	N	N	Y	N	Static NAT / PF - Yes LB - Yes

8.11.4. Support Matrix for Basic Zone

- Y = Supported
- N = Not Supported

NW Devices	DHCP	DNS	User Data	Source NAT	Static NAT	Port Forwarding	Load Balanc-ing	Remote VPN	NW ACL	Usage Monitor-ing	Secur-ity Group	Firewall
Virtual Router	VR	VR	VR	N	N	N	N	N	N	Y	Y	N
VR and NetSca-ler (EIP/ ELB)	VR	VR	VR	N	NetSca-ler	N	NetSca-ler	N	N	Y	Y	N

8.12. Network Offerings

Note

For the most up-to-date list of supported network services, see the CloudPlatform UI or call listNetworkServices.

A network offering is a named set of network services, such as:

- DHCP
- DNS
- Source NAT
- Static NAT
- Port Forwarding
- Load Balancing
- Firewall
- VPN
- Optional) Name one of several available providers to use for a given service, such as Juniper for the firewall
- (Optional) Network tag to specify which physical network to use

When creating a new VM, the user chooses one of the available network offerings, and that determines which network services the VM can use.

The CloudPlatform administrator can create any number of custom network offerings, in addition to the default network offerings provided by CloudPlatform. By creating multiple custom network offerings, you can set up your cloud to offer different classes of service on a single multi-tenant physical network. For example, while the underlying physical wiring may be the same for two tenants, tenant A may only need simple firewall protection for their website, while tenant B may be running a web server farm and require a scalable firewall solution, load balancing solution, and alternate networks for accessing the database backend.

Note

If you create load balancing rules while using a network service offering that includes an external load balancer device such as NetScaler, and later change the network service offering to one that uses the CloudPlatform virtual router, you must create a firewall rule on the virtual router for each of your existing load balancing rules so that they continue to function.

When creating a new virtual network, the CloudPlatform administrator chooses which network offering to enable for that network. Each virtual network is associated with one network offering. A virtual network can be upgraded or downgraded by changing its associated network offering. If you do this, be sure to reprogram the physical network to match.

CloudPlatform also has internal network offerings for use by CloudPlatform system VMs. These network offerings are not visible to users but can be modified by administrators.

About Virtual Machines in CloudPlatform

9.1. About Working with Virtual Machines

CloudPlatform provides administrators with complete control over the life cycle of all guest VMs executing in the cloud. CloudPlatform provides several guest management operations for end users and administrators. VMs may be stopped, started, rebooted, and destroyed.

Guest VMs have a name and group. VM names and groups are opaque to CloudPlatform and are available for end users to organize their VMs. Each VM can have three names for use in different contexts. Only two of these names can be controlled by the user:

- Instance name – a unique, immutable ID that is generated by CloudPlatform, and can not be modified by the user. This name conforms to the requirements in IETF RFC 1123.
- Display name – the name displayed in the CloudPlatform web UI. Can be set by the user. Defaults to instance name.
- Name – host name that the DHCP server assigns to the VM. Can be set by the user. Defaults to instance name.

Note

You can append the display name of a guest VM to its internal name.

For more information, refer to the **5.3. Appending a Display Name to the Internal Name of the Guest Virtual Machines** section of the *CloudPlatform (powered by Apache CloudStack) Version 4.5 Administration Guide*.

Guest VMs can be configured to be Highly Available (HA). A VM where HA is enabled is monitored by the system. If the system detects that the VM is down, it will attempt to restart the VM, possibly on a different host.

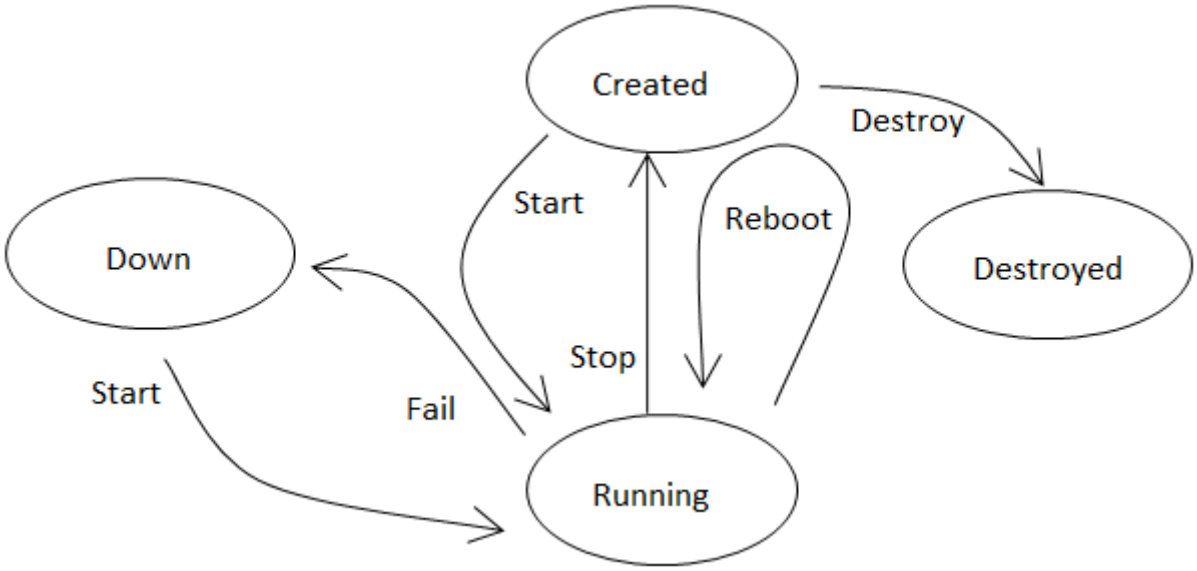
In a zone that uses basic networking with EIP enabled, each new VM is by default allocated one public IP address. When the VM is started, CloudPlatform automatically creates a static NAT between this public IP address and the private IP address of the VM.

If Elastic IP is in use (with the NetScaler load balancer), the IP address initially allocated to the new VM is not marked as elastic. The user must replace the automatically configured IP with a specifically acquired elastic IP, and set up the static NAT mapping between this new IP and the guest VM's private IP. The VM's original IP address is then released and returned to the pool of available public IPs. Optionally, you can also decide not to allocate a public IP to a VM in an EIP-enabled Basic zone.

CloudPlatform cannot distinguish a guest VM that was shut down by the user (such as with the "shutdown" command in Linux) from a VM that shut down unexpectedly. If an HA-enabled VM is shut down from inside the VM, CloudPlatform will restart it. To shut down an HA-enabled VM, you must go through the CloudPlatform UI or API.

9.2. VM Lifecycle

Virtual machines can be in the following states:



Once a virtual machine is destroyed, it cannot be recovered. All the resources used by the virtual machine will be reclaimed by the system. This includes the virtual machine's IP address.

A stop will attempt to gracefully shut down the operating system, which typically involves terminating all the running applications. If the operation system cannot be stopped, it will be forcefully terminated. This has the same effect as pulling the power cord to a physical machine.

A reboot is a stop followed by a start.

CloudPlatform preserves the state of the virtual machine hard disk until the machine is destroyed.

A running virtual machine may fail because of hardware or network issues. A failed virtual machine is in the down state.

The system places the virtual machine into the down state if it does not receive the heartbeat from the hypervisor for three minutes.

The user can manually restart the virtual machine from the down state.

The system will start the virtual machine from the down state automatically if the virtual machine is marked as HA-enabled.

9.3. Determining the Host for a VM

CloudPlatform uses the following methods to determine the host to place a VM on:

- Automatic default host allocation: CloudPlatform can automatically pick the most appropriate host to run each virtual machine.
- Instance type preferences: CloudPlatform administrators can specify certain hosts for particular types of guest instances. For example, an administrator can state that a host should have a preference to run Windows guests. The default host allocator will attempt to place guests of that OS type on such hosts first. If no such host is available, the allocator will place the instance wherever there is sufficient physical capacity.

- Vertical and horizontal allocation: Vertical allocation consumes all the resources of a given host before allocating any guests on another host. This reduces power consumption in the cloud. Horizontal allocation places a guest on each host in a round-robin fashion. This may yield better performance to the guests.
- End user preferences: Users cannot decide the host to run a given VM instance. However, they can specify a zone for the VM. Then, CloudPlatform allocates the VM only to one of the hosts in that zone.
- Host tags: The administrator can assign tags to hosts. These tags can be used to specify the host a VM should use. The CloudPlatform administrators decide whether to define host tags, Then, they create a service offering using those tags and offer it to the user.
- Affinity groups: By defining affinity groups and assigning VMs to them, you can influence (but not dictate) the VMs that must run on separate hosts. This feature lets users specify that certain VMs will not be placed on the same host.
- CloudPlatform also provides a pluggable interface for adding new allocators. These custom allocators can provide any policy the administrator desires.

9.4. Virtual Machine Snapshots

(Supported on VMware and XenServer)

In addition to the existing ability of CloudPlatform to snapshot individual VM volumes, you can take a VM snapshot to preserve all the VM's data volumes as well as (optionally) its CPU/memory state. This is useful for quick restore of a VM. For example, you can snapshot a VM, then make changes such as software upgrades. If anything goes wrong, simply restore the VM to its previous state using the previously saved VM snapshot.

The snapshot is created using the hypervisor's native snapshot facility. The VM snapshot includes not only the data volumes, but optionally also whether the VM is running or turned off (CPU state) and the memory contents. The snapshot is stored in CloudPlatform's primary storage.

VM snapshots will have a parent/child relationship. Each successive snapshot of the same VM is the child of the snapshot that came before it. Each time you take an additional snapshot of the same VM, it saves only the differences between the current state of the VM and the state stored in the most recent previous snapshot. The previous snapshot becomes a parent, and the new snapshot is its child. It is possible to create a long chain of these parent/child snapshots, which amount to a "redo" record leading from the current state of the VM back to the original.

For more information about VM snapshots on VMware, check out the VMware documentation and the VMware Knowledge Base: *Understanding virtual machine snapshots*[1].

9.5. Working with ISOs

CloudPlatform supports ISOs and their attachment to guest VMs. An ISO is a read-only file that has an ISO/CD-ROM style file system. Users can upload their own ISOs and mount them on their guest VMs.

ISOs are uploaded based on a URL. HTTP is the supported protocol. Once the ISO is available via HTTP specify an upload URL such as http://my.web.server/filename.iso.

[1] http://kb.vmware.com/selfservice/microsites/search.do?cmd=displayKC&externalId=1015180

ISOs may be public or private, like templates. ISOs are not hypervisor-specific. That is, a guest on vSphere can mount the exact same image that a guest on KVM can mount.

ISO images may be stored in the system and made available with a privacy level similar to templates. ISO images are classified as bootable and not bootable. A bootable ISO image is one that contains an OS image. CloudPlatform allows a user to boot a guest VM off of an ISO image. Users can also attach ISO images to guest VMs. For example, this enables installing PV drivers into Windows. ISO images are not hypervisor-specific.

About Templates in CloudPlatform

A template is a reusable configuration for virtual machines. When users launch VMs, they can choose from a list of templates in CloudPlatform.

Specifically, a template is a virtual disk image that includes one of a variety of operating systems, optional additional software such as office applications, and settings such as access control to determine who can use the template. Each template is associated with a particular type of hypervisor, which is specified when the template is added to CloudPlatform.

CloudPlatform ships with a default template. In order to present more choices to users, CloudPlatform administrators and users can create templates and add them to CloudPlatform.

10.1. The Default Template

CloudPlatform includes a CentOS template. This template is downloaded by the Secondary Storage VM after the primary and secondary storage are configured. You can use this template in your production deployment or you can delete it and use custom templates.

The root password for the default template is "password".

A default template is provided for each of XenServer, KVM, and vSphere. The templates that are downloaded depend on the hypervisor type that is available in your cloud. Each template is approximately 2.5 GB physical size.

The default template includes the standard iptables rules, which will block most access to the template excluding ssh.

```
# iptables --list
Chain INPUT (policy ACCEPT)
target     prot opt source               destination
RH-Firewall-1-INPUT  all  --  anywhere             anywhere

Chain FORWARD (policy ACCEPT)
target     prot opt source               destination
RH-Firewall-1-INPUT  all  --  anywhere             anywhere

Chain OUTPUT (policy ACCEPT)
target     prot opt source               destination

Chain RH-Firewall-1-INPUT (2 references)
target     prot opt source               destination
ACCEPT     all  --  anywhere             anywhere
ACCEPT     icmp --  anywhere        anywhere       icmp any
ACCEPT     esp  --  anywhere        anywhere
ACCEPT     ah   --  anywhere        anywhere
ACCEPT     udp  --  anywhere        224.0.0.251    udp dpt:mdns
ACCEPT     udp  --  anywhere        anywhere       udp dpt:ipp
ACCEPT     tcp  --  anywhere        anywhere       tcp dpt:ipp
ACCEPT     all  --  anywhere        anywhere       state RELATED,ESTABLISHED
ACCEPT     tcp  --  anywhere        anywhere       state NEW tcp dpt:ssh
REJECT     all  --  anywhere        anywhere       reject-with icmp-host-
```

10.2. Private and Public Templates

When a user creates a template, it can be designated private or public.

Private templates are only available to the user who created them. By default, an uploaded template is private.

When a user marks a template as “public,” the template becomes available to all users in all accounts in the user's domain, as well as users in any other domains that have access to the Zone where the template is stored. This depends on whether the Zone, in turn, was defined as private or public. A private Zone is assigned to a single domain, and a public Zone is accessible to any domain. If a public template is created in a private Zone, it is available only to users in the domain assigned to that Zone. If a public template is created in a public Zone, it is available to all users in all domains.

10.3. The System VM Template

The System VMs come from a single template. The System VM has the following characteristics:

- Debian 7.0
- Has a minimal set of packages installed, thereby reducing the attack surface
- 64-bit templates for all the hypervisors

  From the version 4.3 onwards, CloudPlatform supports only 64-bit templates for all the hypervisors.
- pvops kernel with Xen PV drivers, KVM virtio drivers, and VMware tools for optimum performance on all hypervisors
- Xen tools inclusion allows performance monitoring
- Latest versions of HAProxy, iptables, IPsec, and Apache from debian repository ensures improved security and speed
- Latest version of JRE from Sun/Oracle ensures improved security and speed

10.4. Managing the Number of System VM Templates

CloudPlatform uses the following procedure to manage the number of SSVMs:

1. Management Service scans the number of commands on all the running SSVMs, periodically.
2. If the global configuration setting, *`system.vm.auto.reserve.capacity`*, is enabled by default, the Management Service uses the auto loading logic.
3. Checks the value of global configuration setting, *`secstorage.capacity.standby`*. The default value is 10. This is the minimum number of command execution sessions that system is able to serve immediately (standby capacity), named as standbyCapacity.
4. Checks the value of global configuration setting, *`secstorage.session.max`*. The default value is 50, named as capacityPerSSVM.
5. Checks the number of active commands currently being processed by all SSVMs, named as activeCmds.

   To find activeCmds, use the following query:

   ```
   # select count(*) from cmd_exec_log cel, host h where h.id=cel.instance_id and
    h.status='Up'
   and h.data_center_id=<data center id> and cel.created > (now() - INTERVAL 30 MINUTE);
   ```
6. Checks the number of currently running SSVMs in zone, named as alreadyRunningSSVM.

7. The following logic decides when a new SSVM is required: (alreadyRunningSSVM * capacityPerSSVM - activeCmds) > standbyCapacity.

8. If equation is true, the Management Server creates a new SSVM.

   The values must be changed depending on the type of host workload on certain environment. It is a good practice to change the default system offering for SSVM. The default offering is not meant to be used in enterprise environments.

10.5. Multiple System VM Support for VMware

Every CloudPlatform zone has single System VM for template processing tasks such as downloading templates, uploading templates, and uploading ISOs. In a zone where VMware is being used, additional System VMs can be launched to process VMware-specific tasks such as taking snapshots and creating private templates. The CloudPlatform management server launches additional System VMs for VMware-specific tasks as the load increases. The management server monitors and weights all commands sent to these System VMs and performs dynamic load balancing and scaling-up of more System VMs.

10.6. Console Proxy

The Console Proxy is a type of System Virtual Machine that has a role in presenting a console view via the web UI. It connects the user's browser to the VNC port made available via the hypervisor for the console of the guest. Both the administrator and end user web UIs offer a console connection.

Clicking on a console icon brings up a new window. The AJAX code downloaded into that window refers to the public IP address of a console proxy VM. There is exactly one public IP address allocated per console proxy VM. The AJAX application connects to this IP. The console proxy then proxies the connection to the VNC port for the requested VM on the Host hosting the guest. .

Note

The hypervisors will have many ports assigned to VNC usage so that multiple VNC sessions can occur simultaneously.

The VNC traffic never goes through the guest virtual IP, and there is no need to enable VNC within the guest.

The console proxy VM will periodically report its active session count to the Management Server. The default reporting interval is five seconds. This can be changed through standard Management Server configuration with the parameter consoleproxy.loadscan.interval.

Assignment of guest VM to console proxy is determined by first determining if the guest VM has a previous session associated with a console proxy. If it does, the Management Server will assign the guest VM to the target Console Proxy VM regardless of the load on the proxy VM. Failing that, the first available running Console Proxy VM that has the capacity to handle new sessions is used.

Console proxies can be restarted by administrators but this will interrupt existing console sessions for users.

Note

Before the version 4.3, CloudPlatform was using a dynamic DNS service named **`realhostip.com`** for providing SSL security to console sessions. This domain name has been depreciated. As an alternate, CloudPlatform provides a new mechanism based on global settings to help administrators set up secure connections across various deployment environments. Using this mechanism, customers can use own domain.

10.7. Virtual Router

The virtual router is a type of System Virtual Machine. The virtual router is one of the most frequently used service providers in CloudPlatform. The end user has no direct access to the virtual router. Users can ping the virtual router and take actions that affect it (such as setting up port forwarding), but users do not have SSH access into the virtual router.

There is no mechanism for the administrator to log in to the virtual router. Virtual routers can be restarted by administrators, but this will interrupt public network access and other services for end users. A basic test in debugging networking issues is to attempt to ping the virtual router from a guest VM. Some of the characteristics of the virtual router are determined by its associated system service offering.

Securing Passwords in CloudPlatform

This section describes the password encryption in CloudPlatform and how you can modify the default password encryption to enhance your password secutrity.

11.1. About Password and Key Encryption

CloudPlatform stores several sensitive passwords and secret keys that are used to provide security. These values are always automatically encrypted:

- Database secret key
- Database password
- SSH keys
- Compute node root password
- VPN password
- User API secret key
- VNC password

CloudPlatform uses the Java Simplified Encryption (JASYPT) library. The data values are encrypted and decrypted using a database secret key, which is stored in one of CloudPlatform's internal properties files along with the database password. The other encrypted values listed above, such as SSH keys, are in the CloudPlatform internal database.

Of course, the database secret key itself can not be stored in the open – it must be encrypted. How then does CloudPlatform read it? A second secret key must be provided from an external source during Management Server startup. This key can be provided in one of two ways: loaded from a file or provided by the CloudPlatform administrator. The CloudPlatform database has a configuration setting that lets it know which of these methods will be used. If the encryption type is set to "file," the key must be in a file in a known location. If the encryption type is set to "web," the administrator runs the utility com.cloud.utils.crypt.EncryptionSecretKeySender, which relays the key to the Management Server over a known port.

The encryption type, database secret key, and Management Server secret key are set during CloudPlatform installation. They are all parameters to the CloudPlatform database setup script (cloudstack-setup-databases). The default values are file, password, and password. It is, of course, highly recommended that you change these to more secure keys.

11.2. Changing the Default Password Encryption

Passwords are encoded when creating or updating users. The default preferred encoder is SHA256. It is more secure than MD5 hashing, which was used in CloudPlatform 3.x. If you take no action to customize password encryption and authentication, SHA256 Salt will be used.

If you prefer a different authentication mechanism, CloudPlatform provides a way for you to determine the default encoding and authentication mechanism for admin and user logins. Two configurable lists are provided: userPasswordEncoders and userAuthenticators. userPasswordEncoders allow you to configure the order of preference for encoding passwords, and userAuthenticator allows you to configure the order in which authentication schemes are invoked to validate user passwords.

The following method determines what encoding scheme is used to encode the password supplied during user creation or modification.

When a new user is created, the user password is encoded by using the first valid encoder loaded as per the sequence specified in the **UserPasswordEncoders** property in the **ComponentContext.xml** or **nonossComponentContext.xml** files. The order of authentication schemes is determined by the **UserAuthenticators** property in the same files. If Non-OSS components, such as VMware environments, are to be deployed, modify the **UserPasswordEncoders** and **UserAuthenticators** lists in the **nonossComponentContext.xml** file. For OSS environments, such as XenServer or KVM, modify the **ComponentContext.xml** file. It is recommended to make uniform changes across both the files.

When a new authenticator or encoder is added, you can add them to this list. While doing so, ensure that the new authenticator or encoder is specified as a bean in both the files. The administrator can change the ordering of both these properties as desired to change the order of schemes. Modify the following list properties available in **client/tomcatconf/nonossComponentContext.xml.in** or **client/tomcatconf/componentContext.xml.in** as applicable, to the desired order:

```
<property name="UserAuthenticators">
         <list>
            <ref bean="SHA256SaltedUserAuthenticator"/>
            <ref bean="MD5UserAuthenticator"/>
            <ref bean="LDAPUserAuthenticator"/>
            <ref bean="PlainTextUserAuthenticator"/>
        </list>
    </property>
    <property name="UserPasswordEncoders">
        <list>
            <ref bean="SHA256SaltedUserAuthenticator"/>
             <ref bean="MD5UserAuthenticator"/>
             <ref bean="LDAPUserAuthenticator"/>
            <ref bean="PlainTextUserAuthenticator"/>
            </list>
```

In the above default ordering, SHA256Salt is used first for **UserPasswordEncoders**. If the module is found and encoding returns a valid value, the encoded password is stored in the user table's password column. If it fails for any reason, the MD5UserAuthenticator will be tried next, and the order continues. For **UserAuthenticators**, SHA256Salt authentication is tried first. If it succeeds, the user is logged into the Management server. If it fails, md5 is tried next, and attempts continues until any of them succeeds and the user logs in . If none of them works, the user is returned an invalid credential message.