CloudStack 2.2.13

Best Practices Guide

A guide capturing the answers to the questions that many system administrators have before deploying CloudStack.

Eric Christensen

CloudStack 2.2.13 Best Practices Guide
A guide capturing the answers to the questions that many system administrators have before deploying CloudStack.
Edition 1

Author Eric Christensen *eric.christensen@citrix.com*

Preface

1. Document Conventions

This manual uses several conventions to highlight certain words and phrases and draw attention to specific pieces of information.

In PDF and paper editions, this manual uses typefaces drawn from the *Liberation Fonts*[1] set. The Liberation Fonts set is also used in HTML editions if the set is installed on your system. If not, alternative but equivalent typefaces are displayed. Note: Red Hat Enterprise Linux 5 and later includes the Liberation Fonts set by default.

1.1. Typographic Conventions

Four typographic conventions are used to call attention to specific words and phrases. These conventions, and the circumstances they apply to, are as follows.

`Mono-spaced Bold`

Used to highlight system input, including shell commands, file names and paths. Also used to highlight keycaps and key combinations. For example:

> To see the contents of the file **`my_next_bestselling_novel`** in your current working directory, enter the **`cat my_next_bestselling_novel`** command at the shell prompt and press **`Enter`** to execute the command.

The above includes a file name, a shell command and a keycap, all presented in mono-spaced bold and all distinguishable thanks to context.

Key combinations can be distinguished from keycaps by the hyphen connecting each part of a key combination. For example:

> Press **`Enter`** to execute the command.
>
> Press **`Ctrl`**+**`Alt`**+**`F2`** to switch to the first virtual terminal. Press **`Ctrl`**+**`Alt`**+**`F1`** to return to your X-Windows session.

The first paragraph highlights the particular keycap to press. The second highlights two key combinations (each a set of three keycaps with each set pressed simultaneously).

If source code is discussed, class names, methods, functions, variable names and returned values mentioned within a paragraph will be presented as above, in **`mono-spaced bold`**. For example:

> File-related classes include **`filesystem`** for file systems, **`file`** for files, and **`dir`** for directories. Each class has its own associated set of permissions.

Proportional Bold

This denotes words or phrases encountered on a system, including application names; dialog box text; labeled buttons; check-box and radio button labels; menu titles and sub-menu titles. For example:

> Choose **System**□□□ **Preferences**□□□ **Mouse** from the main menu bar to launch **Mouse Preferences**. In the **Buttons** tab, click the **Left-handed mouse** check box

[1] https://fedorahosted.org/liberation-fonts/

> and click **Close** to switch the primary mouse button from the left to the right (making the mouse suitable for use in the left hand).
>
> To insert a special character into a **gedit** file, choose **Applications□□□ Accessories □□ Character Map** from the main menu bar. Next, choose **Search□□□ Find…** from the **Character Map** menu bar, type the name of the character in the **Search** field and click **Next**. The character you sought will be highlighted in the **Character Table**. Double-click this highlighted character to place it in the **Text to copy** field and then click the **Copy** button. Now switch back to your document and choose **Edit□□ Paste** from the **gedit** menu bar.

The above text includes application names; system-wide menu names and items; application-specific menu names; and buttons and text found within a GUI interface, all presented in proportional bold and all distinguishable by context.

Mono-spaced Bold Italic or ***Proportional Bold Italic***

Whether mono-spaced bold or proportional bold, the addition of italics indicates replaceable or variable text. Italics denotes text you do not input literally or displayed text that changes depending on circumstance. For example:

> To connect to a remote machine using ssh, type **ssh *username*@*domain.name*** at a shell prompt. If the remote machine is **example.com** and your username on that machine is john, type **ssh john@example.com**.
>
> The **mount -o remount *file-system*** command remounts the named file system. For example, to remount the **/home** file system, the command is **mount -o remount /home**.
>
> To see the version of a currently installed package, use the **rpm -q *package*** command. It will return a result as follows: ***package-version-release***.

Note the words in bold italics above — username, domain.name, file-system, package, version and release. Each word is a placeholder, either for text you enter when issuing a command or for text displayed by the system.

Aside from standard usage for presenting the title of a work, italics denotes the first use of a new and important term. For example:

> Publican is a *DocBook* publishing system.

1.2. Pull-quote Conventions

Terminal output and source code listings are set off visually from the surrounding text.

Output sent to a terminal is set in **mono-spaced roman** and presented thus:

```
books        Desktop   documentation  drafts  mss    photos   stuff  svn
books_tests  Desktop1  downloads      images  notes  scripts  svgs
```

Source-code listings are also set in **mono-spaced roman** but add syntax highlighting as follows:

```
package org.jboss.book.jca.ex1;

import javax.naming.InitialContext;

public class ExClient
```

```
{
   public static void main(String args[])
       throws Exception
   {
      InitialContext iniCtx = new InitialContext();
      Object         ref    = iniCtx.lookup("EchoBean");
      EchoHome       home   = (EchoHome) ref;
      Echo           echo   = home.create();

      System.out.println("Created Echo");

      System.out.println("Echo.echo('Hello') = " + echo.echo("Hello"));
   }
}
```

1.3. Notes and Warnings

Finally, we use three visual styles to draw attention to information that might otherwise be overlooked.

Note

Notes are tips, shortcuts or alternative approaches to the task at hand. Ignoring a note should have no negative consequences, but you might miss out on a trick that makes your life easier.

Important

Important boxes detail things that are easily missed: configuration changes that only apply to the current session, or services that need restarting before an update will apply. Ignoring a box labeled 'Important' will not cause data loss but may cause irritation and frustration.

Warning

Warnings should not be ignored. Ignoring warnings will most likely cause data loss.

2. We Need Feedback!

iIf you find a typographical error in this manual, or if you have thought of a way to make this manual better, we would love to hear from you! Please submit a report in Bugzilla: *http://bugs.cloud.com* against the component **Doc**

If you have a suggestion for improving the documentation, try to be as specific as possible when describing it. If you have found an error, please include the section number and some of the surrounding text so we can find it easily.


DRAFT

Building Your Cloud

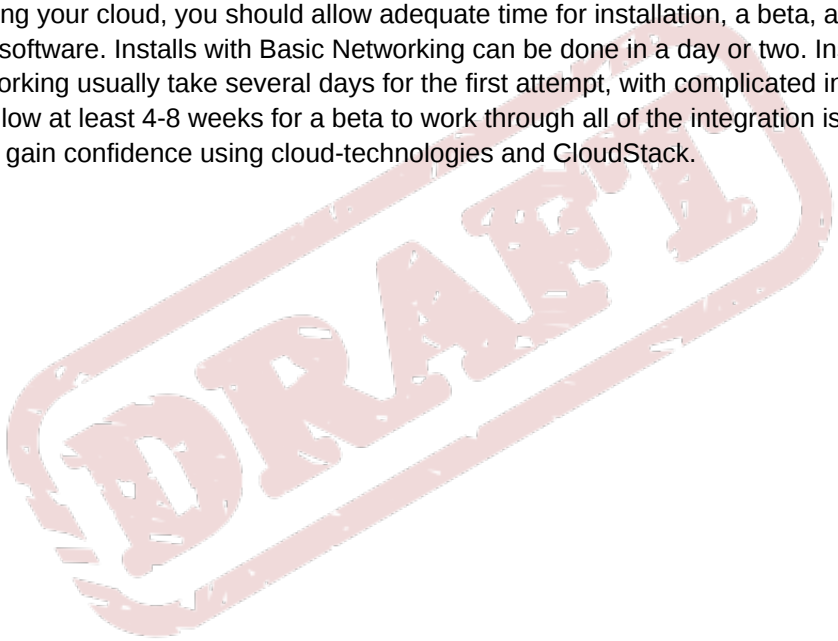
So you are ready to start building your cloud using the CloudStack management software. Here are a few tips and tricks to help you be successful in your rollout.

1.1. Staging Area

Having a staging system that models your production environment is strongly advised. It is critical if customizations have been applied to your CloudStack environment. A staging system allows you to make changes to your system so you can determine if there are any negative effects. Upgrades and patches all change your system's behavior. All changes should be tested before making changes to your production environment. Your staging system should be as identical to your production system as well. Small differences in hardware can be problematic when testing upgrades and patches. What affects one system may not affect the other.

1.2. Taking adequate time for planning.

When assembling your cloud, you should allow adequate time for installation, a beta, and learning the system and software. Installs with Basic Networking can be done in a day or two. Installs with Advanced Networking usually take several days for the first attempt, with complicated installations taking longer. Allow at least 4-8 weeks for a beta to work through all of the integration issues. It can takes months to gain confidence using cloud-technologies and CloudStack.




DRAFT

Clusters

In cloud organization, Clusters are the smallest organizational unit. Hosts, running a common hypervisor, join together to form a cluster and receive virtual machines.

2.1. Recommended Limits

In order to maintain an efficient cloud, Citrix recommends the following logical limits when building your cloud:

- >A maximum of six hosts per cluster. You can have an unlimited number of clusters per pod and the number of storage pools per cluster does not matter.
- No more than 200 virtual disks per LUN.
- Each hypervisor supports a maximum number of virtual machines per host. It is important to follow the manufacturer's recommendations.
- There are no practical limits on SRs per cluster.

It is also important to take into consideration host downtime. To maintain uptime whenever possible your cloud should be designed to have a host go down and have all VMs on that host migrate over to other hosts without violating the above recommendations.

2.2. Network Switch Density

Multiple clusters can be used per pod to achieve a certain switch density. In advance networking, too many VLANs can be problematic. To reduce the number of VLANs a cluster has to handle customers should be assigned to a particular cluster and not be scattered throughout the cloud. This can be accomplished by setting the value **`use.user.concentrated.pod.allocation`** to **`true`**.


DRAFT

Hypervisor Hosts

Hosts are where the action takes place. When CloudStack receives an instruction to create a virtual machine (VM) it polls all the hosts in a particular cluster in search of one that can handle the CPU and RAM demands of the VM. The hosts run the VMs using the hypervisor software.

3.1. Setting up your hosts

It is increadibly important that the host hardware is identical with respect to the processors, number of network interface cards (NICs), and the way the network is setup on all hosts within a cluster. VMs that are setup in one configuration will expect the same hardware to be present if it is migrated to another host.

3.2. Hypervisor Software

CloudStack currently supports the following hypervisor software versions:

- XenServer 5.6 SP2
- VMWare vSphere 4.1 with vCenter
- KVM (on RHEL/CentOS 5.6 (64-bit), RHEL6 (64-bit), Fedora 14 (64-bit), and Ubuntu 10.04 LTS (64-bit)
- Oracle VM (OVM) 2.3
- XCP 1.0 and 1.1

3.3. Host Capacity and Overprovisioning

Host capacity should generally be modeled in terms of RAM for the guests. Storage and CPU may be overprovisioned. RAM may not. RAM is usually the limiting factor in capacity designs.


DRAFT

KVM

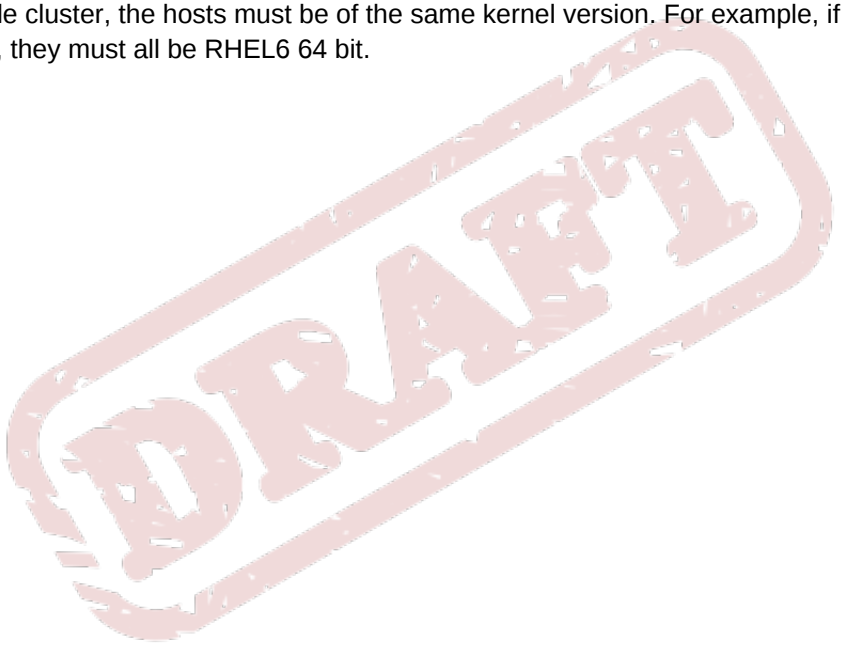
The KVM hypervisor is included in Red Hat® Enterprise Linux® and CentOS operating systems.

4.1. Installation

4.1.1. Prerequisites and Constraints

All KVM Hosts must follow these guidelines:

- Must be 64-bit.
- Must support HVM (Intel-VT or AMD-V enabled).
- Within a single cluster, the hosts must be homogenous. The CPUs must be of the same type, count, and feature flags.
- Within a single cluster, the hosts must be of the same kernel version. For example, if one Host is RHEL6 64 bit, they must all be RHEL6 64 bit.




DRAFT

XenServer

5.1. Hardware

All hosts must be 64-bit and must support HVM (Intel-VT or AMD-V enabled). All Hosts within a Cluster must be homogeneous. That means the CPUs must be of the same type, count, and feature flags. See http://docs.vmd.citrix.com/XenServer/4.0.1/reference/ch02.html for more information on homogeneous XenServer hosts.

5.2. VM Limits

All hypervisors have limits on the number of virtual machines they can handle due to memory and other constraints. In XenServer 5.x you should keep the number of virtual machines deployed on a host to 50 or less. Some improvements have been seen in version 6.0 and that number has been increased to 90.

5.3. Reusing Servers

You must re-install Citrix XenServer if you are going to re-use a host from a previous install.

5.4. Networking

Once XenServer has been installed you may need to do some additional network configuration. If you plan on using NIC bonding, the NICs on all hosts in the Cluster must be cabled exactly the same. For example, if eth0 is in the private bond on one host in a cluster, then eth0 must be in the private bond on all hosts in the cluster.

The IP address assigned for the private network interface must be static. It can be set on the host itself or obtained via static DHCP.

5.4.1. NIC Bonding

XenServer supports Source Level Balancing (SLB) NIC bonding. You must set bonds on the first host added to a cluster. Then you must use xe commands to establish the same bonds in the second and subsequent hosts added to a cluster. Slave hosts in a cluster must be cabled exactly the same as the master. For example, if eth0 is in the private bond on the master, it must be in the private network for added slave hosts.


DRAFT

VMWare vSphere

6.1. Installation

6.1.1. Prerequisites and Constraints

The following requirements must be met in order for the VMware vSphere installation to work properly:

- VMware vCenter Standard Edition 4.1 must be installed and available to manage the vSphere Hosts.
- vCenter must be configured to use the standard port 443 so that it can communicate with the CloudStack Management Server.
- You must re-install VMware ESXi if you are going to re-use a host from a previous install.
- The CloudStack requires VMware vSphere 4.1. VMware vSphere 4.0 is not supported.
- All hosts must be 64-bit and must support HVM (Intel-VT or AMD-V enabled). All Hosts within a Cluster must be homogenous. That means the CPUs must be of the same type, count, and feature flags.
- The CloudStack private network must not be configured as a separate tagged virtual network. The CloudStack private network is the same as the vCenter management network, and will inherit its configuration. See Configure vCenter Management Network on page 50.
- CloudStack requires ESXi. ESX is not supported.
- All resources used for CloudStack must be used for CloudStack only. CloudStack cannot shares instance of ESXi or storage with other management consoles. Do not share the same storage volumes that will be used by CloudStack with a different set of ESXi servers that are not managed by CloudStack.
- Put all target ESXi hypervisors in a Cluster in a separate Datacenter in vCenter.
- The cluster that will be managed by CloudStack product should not contain any VMs. Do not run the management server, vCenter or any other VMs on the cluster that is designated for CloudStack use. Create a separate cluster for use of CloudStack and make sure that they are no VMs in this cluster.
- All the required VLANs must be trunked into all the ESXi hypervisor hosts. These would include the VLANS for Management, Storage, vMotion, and guest VLANs. The guest VLAN (used in Advanced Networking; see Network Setup on page 18) is a contiguous range of VLANs that will be managed by the CloudStack Product. CloudStack does not support Distributed vSwitches in VMware.

6.2. VM Limits

All hypervisors have limits on the number of virtual machines they can handle due to memory and other constraints. VMware vSphere supports 100 to 128 virtual machines per host.

6.3. Networking

6.3.1. Increasing Ports

By default a virtual switch on ESXi hosts is created with 56 ports. We recommend setting it to 4096, the maximum number of ports allowed. To do that, click the "Properties..." link for virtual switch (note this is not the Properties link for Networking).

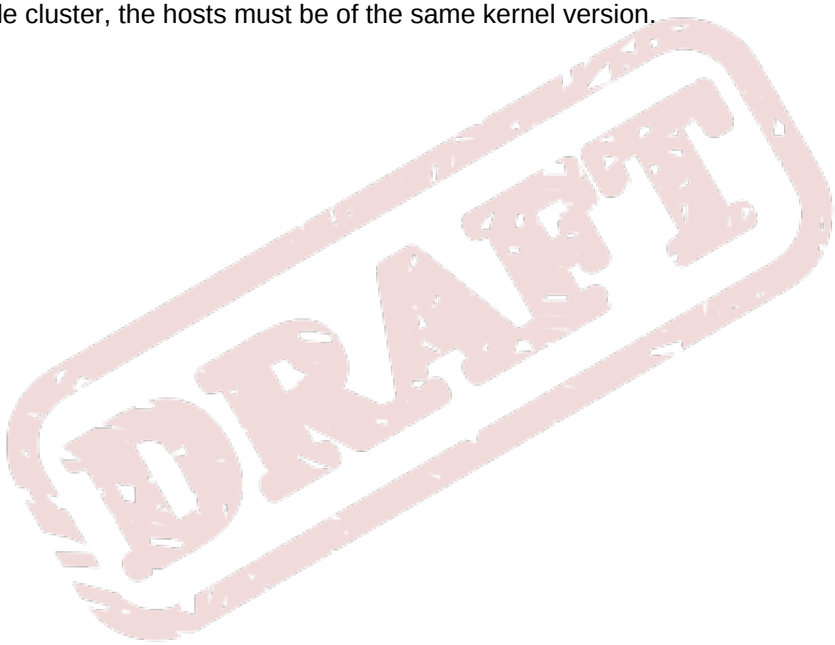
Oracle VM (OVM)

7.1. Installation

7.1.1. Prerequisites and Constraints

All OVM Hosts must follow these guidelines:

- Must be 64-bit.
- Must support HVM (Intel-VT or AMD-V enabled).
- Within a single cluster, the hosts must be homogenous. The CPUs must be of the same type, count, and feature flags.
- Within a single cluster, the hosts must be of the same kernel version.




DRAFT

Networking

Networking is as important a component to your cloud as other hardware and software. Without a good, redundant, high bandwidth network your cloud could have problems including long transfer times, virtual machine failures, and loss of data. Generally speaking, ten-gigabit networking is generally recommended for storage access when larger servers that can support relatively more VMs are used.

8.1. IP Address Planning

Several network subnets are required to support all the different networks that go into an operational cloud. Careful planning of IP addresses is necessary before deploying your cloud. The administrator should provide private IP addresses for the system in each pod and provision them in CloudStack.

8.2. NIC Bonding

Network Interface Card, or NIC, bonding is important to maintaining connectivity throughout your cloud when network problems loom. When done properly, NIC bonding provides a redundant path for data to flow in the event that one network fails. Because bonding is generally managed through the hypervisor, the procedure for creating bonds can differ depending on the software being used. Common among all hypervisors, however, is the requirement of network interfaces to be addressed identically on all hosts.

8.3. Networking in vSphere

For vSphere with advanced virtual networking, we recommend provisioning enough private IPs for your total number of customers, plus enough for the required CloudStack System VMs. Typically, about 10 additional IPs are required for the System VMs. For more information about System VMs, see Working with System Virtual Machines in the Administrator's Guide.

The Management Servers communicate with VMware vCenter servers on port 443 (HTTPS).

8.4. Networking in KVM and XenServer

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.

When Advanced Virtual 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 when Advanced Virtual networking is enabled, 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.

- The secondary storage VMs and console proxy VMs connect to the Management Server on port 8250. If you are using multiple Management Servers, the load balanced IP address of the Management Servers on port 8250 must be reachable. Note that you must not expose port 8250 to the public Internet. The secondary storage VM can be located on any Host in the Zone. It uses the private network for its communication.
- The Management Server and secondary storage VMs must be able to access vCenter and all ESXi hosts in the zone. To allow the necessary access through the firewall, keep port 443 open.
- The console proxy VMs connect to all hosts in the Zone over the private network. Therefore the private network of any given Pod in the Zone must have connectivity to the private network of all other Pods in the Zone.
- The secondary storage NFS export is mounted by the secondary storage VM. Secondary storage traffic goes over the private network even if there is a separate storage network. (Primary storage traffic goes over the storage network, if available.) If you choose to place secondary storage NFS servers on the storage network, you must make sure there is a route from the private network to the storage network.
- When external firewall integration is in place, the public IP VLAN must still be trunked to the Hosts. This is required to support the Secondary Storage and Console Proxy VMs.
- The Management Servers communicate with each other to coordinate tasks amongst themselves. This communication uses TCP on ports 8250 and 9090.
- With Advanced Networking, separate subnets must be used for private and public networks.
- The public internet must not be able to access port 8096 on the Management Server.
- The Management Servers communicate with the XenServers on ports 22 (ssh) and 80 (HTTP).
- The Management Servers communicate with the KVM servers on port 22 (ssh).

MySQL Database

The **MySQL** database houses all information about your cloud. It is imperative to protect the database against corruption, loss, unavailablilty.

9.1. Security

On RHEL and CentOS, MySQL does not set a root password by default. It is very strongly recommended that you set a root password as a security precaution. Run the following commands, and substitute your own desired root password for passwd.

```
# service mysqld start

# mysql -u root

mysql> SET PASSWORD = PASSWORD('passwd');
```

9.2. Replication

CloudStack supports database replication from one MySQL node to another. Creating a replica is not a backup solution. You should develop a backup procedure for the MySQL data that is distinct from replication.

Edit my.cnf on the master and add the following in the [mysqld] section below datadir.

```
log_bin=mysql-bin
server_id=1
```

The server_id must be unique with respect to other servers. The recommended way to achieve this is to give the master an ID of 1 and each slave a sequential number greater than 1, so that the servers are numbered 1, 2, 3, etc.


DRAFT

Primary Storage

Primary Storage is the storage used by hypervisors to store virtual machines. Depending on the hypervisor software, the primary storage could be an NFS share, fiber channel, or other type.

10.1. Primary Storage per Cluster

Primary storage mountpoints or LUNs should not exceed 6 TB in size. It is better to have multiple smaller primary storage elements per Cluster than one large one.

When exporting shares on primary storage, avoid data loss by restricting the range of IP addresses that can access the storage.

Monitor host disk space. Many host failures occur because the host's root disk fills up from logs that were not rotated adequately.

10.2. Utilizing RAID

It is important that the storage server contain many large disks and use a hardware RAID controller. Hardware RAID is faster than software RAID and has the added bonus of supporting hot plug functionality independent of the operating system so you can replace faulty disks without impacting the running operating system.

10.3. Limiting NFS Export

It is highly recommended that you limit the NFS export to a particular subnet by specifying a subnet mask (e.g.,"192.168.1.0/24"). By allowing access from only within the expected cluster, you avoid having non-pool member mount the storage. The limit you place must include the private network(s) and the storage network(s). If the two are the same network then one CIDR is sufficient. If you have a separate storage network you must provide separate CIDR's for both or one CIDR that is broad enough to span both.

```
# echo "/export CIDR(rw,async,no_root_squash)" >> /etc/exports
```

The following is an example with separate CIDRs:

- **`/export 192.168.1.0/24(rw,async,no_root_squash)`**
  `10.50.1.0/24(rw,async,no_root_squash)`

Removing the async flag. The async flag improves performance by allowing the NFS server to respond before writes are committed to the disk. Remove the async flag in your mission critical production deployment.

The volumes used for Primary and Secondary storage should be accessible from Management Server and the hypervisors. These volumes should allow root users to read/write data. These volumes must be for the exclusive use of CloudStack and should not contain any data.


DRAFT

Templates

Templates are virtual machines that have been customized for a specific purpose to allow easy implementation of new virtual machines of the same type. An example would be a RHEL® 6.2 server with Apache® installed and hardened. By creating a template from this virtual machine you can then make new installations of this virtual machine quickly.

11.1. Large Templates

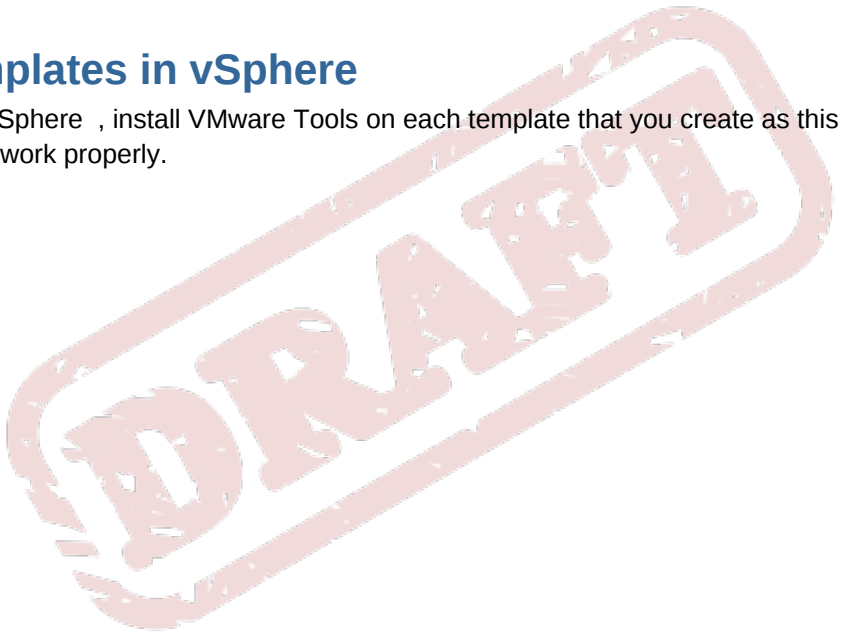
If you plan to use large templates (100GB or larger), be sure you have a 10-gigabit network to support the large templates. A slower network can lead to timeouts and other errors when large templates are used.

11.2. Templates in XenServer

When running XenServer , install PV drivers and Xen tools on each template you create as this will enable live migration and clean guest shutdown.

11.3. Templates in vSphere

When running vSphere , install VMware Tools on each template that you create as this will enable console view to work properly.




DRAFT

Appendix A. Revision History

Revision 2.2.13-1 **Sun Dec 18 2011** **Eric Christensen** *eric.christensen@citrix.com*

Added additional information.

Revision 2.2.13-0 **Tue Nov 29 2011** **Eric Christensen** *eric.christensen@citrix.com*

Initial creation of book by publican.
Updated default text to include CloudStack-specific text.
Added initial information.


DRAFT

Index


DRAFT