vmware | vCLOUD® ARCHITECTURE TOOLKIT™

VMware® vCloud® Architecture Toolkit
Consuming a VMware vCloud

Version 3.1

January 2013

VMware, Inc.
3401 Hillview Ave
Palo Alto, CA 94304
www.vmware.com

Contents

List of Figures

List of Tables

1. Overview

Cloud computing leverages the efficient pooling of on-demand, self-managed virtual infrastructure to provide resources that are consumable as a service. VMware® vCloud® is the VMware cloud solution.

Consumers consume vCloud resources. Understanding consumption requires an understanding of an organization's processes, constraints, and requirements. This applies both to enterprises and to service providers, with some variations depending upon use cases.

Consuming a VMware vCloud serves as a reference for infrastructure architects, managers, and end users who are considering the first steps on the journey to private, public, or hybrid vCloud computing. This document provides:

- An approach to consuming a vCloud from the consumer's perspective.
- A methodology for choosing consumption models.
- Special considerations for:
  - Developing service catalogs.
  - Working with VMware vCloud® vApps.
  - Interactions between enterprises and service providers.

Architecting a VMware vCloud, *Operating a VMware vCloud*, and *Consuming a VMware vCloud* are designed to be used together throughout the lifecycle of a VMware vCloud computing implementation. Using all three documents together, in combination with a private or public service definition, helps to develop a comprehensive view of VMware vCloud computing.

1.1 Audience

This document is designed for those who plan to consume vCloud resources, including architects and designers who have been trained on VMware vSphere and vCloud technologies.

There are two types of consumers:

- End users – People concerned with running applications in an environment regardless of the underlying virtualization infrastructure and vCloud layer.
- Administrators – IT administrators of enterprises and small businesses whose organizations might have purchased cloud computing resources from service providers to augment their in-house resources.

1.2 Scope

This document provides design considerations and patterns for consuming vCloud resources.

2. vCloud Consumption Approach

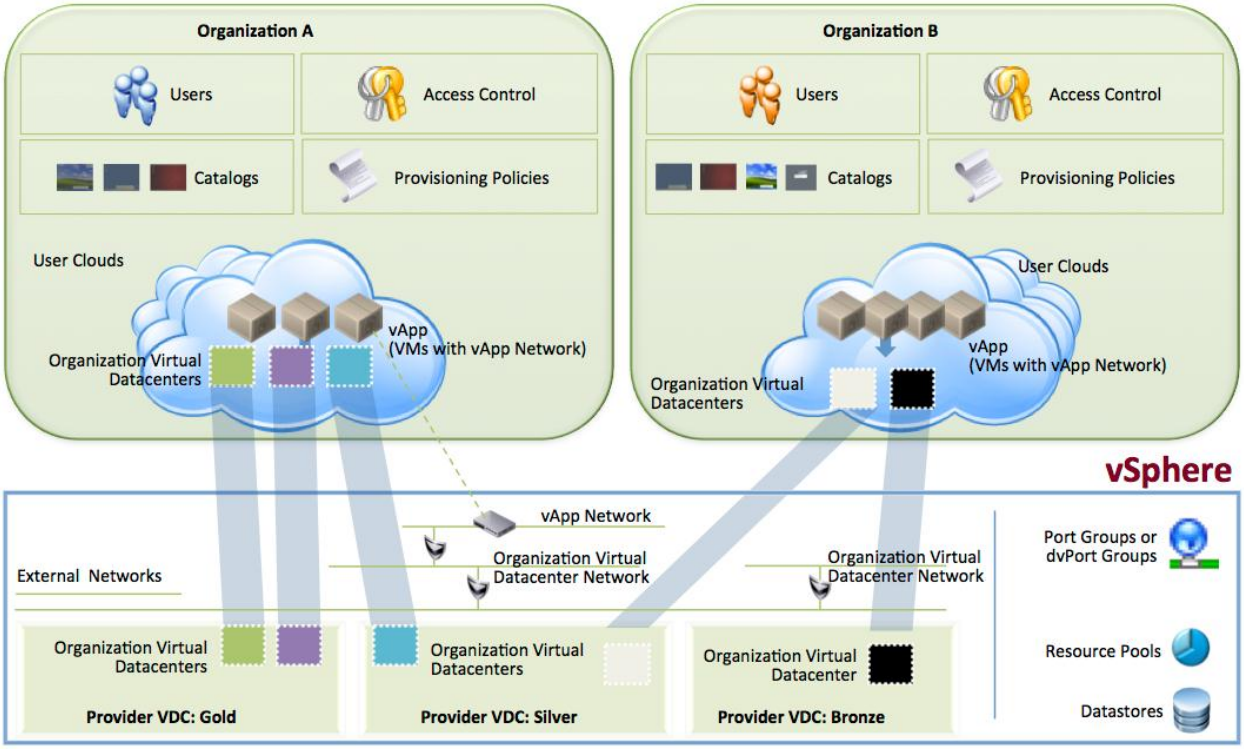
Adopting a consistent approach to vCloud resource consumption, and understanding the underlying vCloud and vSphere components can provide valuable insight into design guidelines for consumption. Such an approach is especially valuable for IT administrators who manage end-user resources that are hosted at a service provider.

2.1 vCloud Consumer Resources

vCloud consumer resources are provided by a VMware vSphere infrastructure dedicated to hosting vCloud workloads. VMware vCloud Director® builds on vSphere hardware abstraction capabilities and introduces logical constructs, such as *virtual datacenters*, *organizations*, and *organization virtual datacenter networks*, to facilitate multitenant resource consumption.

The following figure shows the logical constructs in vCloud Director and how an organization or end user views the vCloud environment and its related constructs.

Figure 1. Mapping vCloud Director Logical Constructs to vSphere



The following table describes the logical constructs in vCloud Director that abstract underlying vSphere resources.

Table 1. vCloud Director Logical Constructs, as Viewed by an Organization

vCloud Director Construct	Description
Organization	A unit of administration that represents a logical collection of users, groups, and computing resources. The organization also serves as a security boundary from which only users of a particular organization can deploy workloads and have visibility into deployed workloads in the vCloud.
Provider virtual datacenter	A collection of vSphere resources, such as CPU, memory, and storage, shared among tenants. This collection is usually based on business requirements.
Organization virtual datacenter	A subset of provider virtual datacenter resources assigned to an organization and backed by a VMware vCenter™ resource pool that is automatically created by vCloud Director. An organization virtual datacenter allocates resources using one of the following models: • Pay As You Go • Allocation pool • Reservation pool
vApp template and media catalogs	A collection of services available for consumption. Catalogs contain *vApp templates* (pre-configured containers of one or more virtual machines), media (ISO images of operating systems), or both.
Network pool	A set of pre-allocated networking resources that vCloud Director can draw from as needed to create virtual networks.
Internal and external organization virtual datacenter networks	Organization virtual datacenter networks are virtual networks that provide an organization with vApp network connectivity. Internal organization virtual datacenter networks are isolated networks used for connectivity between vApps within the organization virtual datacenter. External organization virtual datacenter networks provide connectivity outside the organization virtual datacentre by connecting to an existing external network, using either a direct connection or a connection routed through a vCloud Network and Security Edge Gateway. organization virtual datacenter networks can be shared with other virtual datacenters within the organization. Administrators can create and manage organization virtual datacenter networks, but there are limits to what an organization administrator is permitted to configure. Only system administrators can create external networks.
vApp network	Virtual network contained within a vApp that facilitates network connectivity between virtual machines in the vApp. vApp networks can be connected to an organization virtual datacenter network with a direct, NAT-routed, or fenced connection to enable communication with other vApps inside or outside the organization, if the organization virtual datacenter network is connected to an external network. vApp networks are backed by network pools. Most users with access to a vApp can create and manage their own vApp networks.

2.2 vCloud Consumer Resource Capacity

One of the key benefits of implementing a vCloud is the ability to rapidly provision vApps into the vCloud environment. Capacity management is designed so that the vCloud infrastructure has sufficient capacity to meet the current and future needs of consumers under normal circumstances. Maintaining sufficient reserve capacity in the vCloud infrastructure typically prevents vApps from contending for resources. This mitigates the risk of breaching a service level agreement (SLA).

Provisioning and consuming vApps reduces the capacity of the vCloud infrastructure. To provide additional capacity, vCloud providers typically implement robust management processes that make sufficient resources available to support the service level requirements associated with vApp provisioning and performance.

As vCloud resources are consumed, additional capacity must also be added to allow for anticipated future demand while preserving sufficient capacity for near-term needs. To predict future capacity needs, analyze current capacity usage and trends to determine growth rates, and then estimate future needs based on new consumers and projects.

3. Choosing a vCloud Consumption Model

vCloud providers offer consumer resources to support different SLAs, costing, and sharing models.

3.1 Consuming vCloud Services

The following vCloud service offerings apply to both private and public vCloud, but each has a different impact on consumption. Choose the service offering that best supports your use cases.

- *Basic service offering* (unreserved *pay-per-use* class) – Designed for quick-start pilot projects and workloads, such as software testing, that do not need reservations or guaranteed performance.
- *Committed service offering* (subscription model) – Provides reserved compute resources with the ability to burst above committed levels if additional capacity is available. It offers predictable performance by reserving resources for workloads within a multitenant infrastructure while enabling access to more resources as they become available.
- *Dedicated service offering* – Provides dedicated compute resources, sometimes known as a *virtual private cloud.* It offers predictable performance by reserving dedicated resources. This service offering can help to support SLAs for Tier 1 applications. Under this service offering, end users with appropriate privileges have flexibility in modifying compute allocations to the vApps and virtual machines.

Service classes are designed to help consumers move their workloads to a vCloud. Any existing VMware virtual machine or virtual application (vApp) can be run with little or no modification in a public vCloud. Compatibility with existing enterprise VMware deployments is a key design objective. There is no requirement to deploy a private vCloud to realize the benefits of vCloud computing because any VMware-based virtualized infrastructure is compatible.

All vCloud consumption models are fundamentally based on vCloud Director allocation models, regardless of the name or branding attached to a given service offering.

3.2 vCloud Director Allocation Models

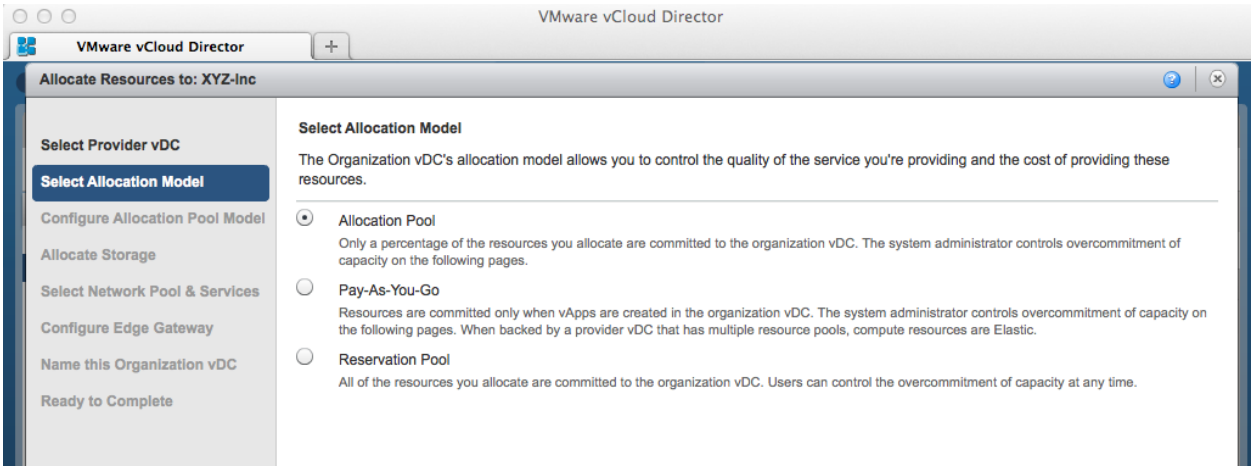
Allocation models define how resources are allocated to an organization's virtual datacenter and how resources are consumed when vApps are deployed.

Allocation models are differentiated by how resources are reserved or limited. A limit can be placed at the level of a virtual machine or a resource pool. To guarantee resources to a virtual machine or to an entire resource pool, vCloud Director sets a reservation or not depending upon the allocation models used. The following are vCloud allocation models:

- Allocation Pool.
- Pay As You Go.
- Reservation Pool.

The following figure shows the vCloud Director screen for selecting an allocation model.

Figure 2. Allocation Models



3.2.1 Allocation Pool

The customer is charged for the pre-allocation of resources to an organization virtual datacenter. A provider may charge the customer based upon the allocated resources, the guaranteed resources, or a combination of the guaranteed resources and the consumption of resources beyond the guarantee. The cloud provider manages resource overcommitments by defining a guaranteed percentage of allocated resources, such as CPU, memory resources, virtual CPU speed, and an optional, maximum virtual machine limit. Only the vCloud provider can expand or contract resources. The Allocation Pool model may consume resources from one or more resource pools, depending on the vCloud provider's configuration.

The Allocation Pool model enables an organization to procure resources under normal operating conditions but has the capability to *burst* for more resources when need arises. Because this model guarantees a specified percentage of the allocated resource, the remainder that is not guaranteed is shared with other tenants. This can result in resource contention.

The Allocation Pool model is usually a good fit for relatively steady state workloads that occasionally surge.

3.2.2 Pay As You Go

The customer is charged for each running virtual machine. As with the Allocation Pool model, the vCloud provider manages resource overcommitments by defining a guaranteed percentage of allocated resources, such as CPU, memory resources, and a maximum number of virtual machines. The cloud provider can specify a maximum virtual CPU speed, or limit. Unlike the allocation and Reservation Pool models, the Pay As You Go model facilitates an unlimited approach to resource consumption within the constraints of the vCloud provider's physical infrastructure. As with the Allocation Pool model, the Pay As You Go model may consume resources from one or more resource pools, depending on the vCloud provider's configuration.

Typically, the Pay As You Go model is a good fit for transient and training environments.

3.2.3 Reservation Pool

As with the Allocation Pool and Pay As You Go models, the Reservation Pool allocation model applies charges for the pre-allocation of resources to an organization virtual datacenter. The fundamental difference from the Allocation Pool model is that the vCloud provider cannot overcommit resources because all CPU and memory resources are 100% guaranteed. However, the vCloud provider can specify a maximum number of virtual machines. Only the vCloud provider can expand or contract resources. The Reservation Pool is unique in that it offers consumers full resource management controls in the form of shares, reservations, and limits. This is similar to *vSphere Resource Management*.

The Reservation Pool represents a good fit for steady state workloads that require guaranteed performance. To make best use of this model, you should know the customer's application profile well enough to optimize resource provisioning.

vCloud providers typically charge a premium for this type of allocation model because they cannot overcommit resources.

3.2.4 Storage Allocation

The allocation of storage resources is the same for all allocation models and is charged as capacity is allocated. The vCloud provider manages storage resource overcommitment by defining a maximum amount of storage that can be consumed within the organization virtual datacenter and by controlling whether storage is *thin-provisioned* (not pre-allocated) or *thick-provisioned* (pre-allocated in full).

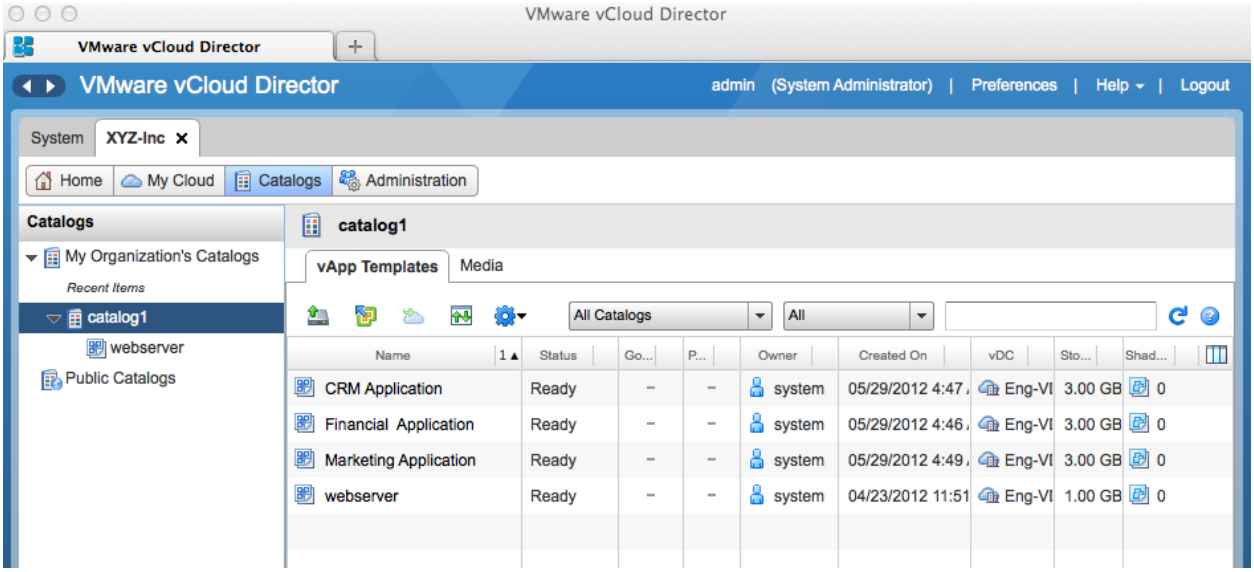
A vCloud Director 5.1 feature provides multiple classes or tiers of storage capacity within an organization virtual datacenter. This allows an application owner to deploy a multitier application with different classes of storage for different tiers.

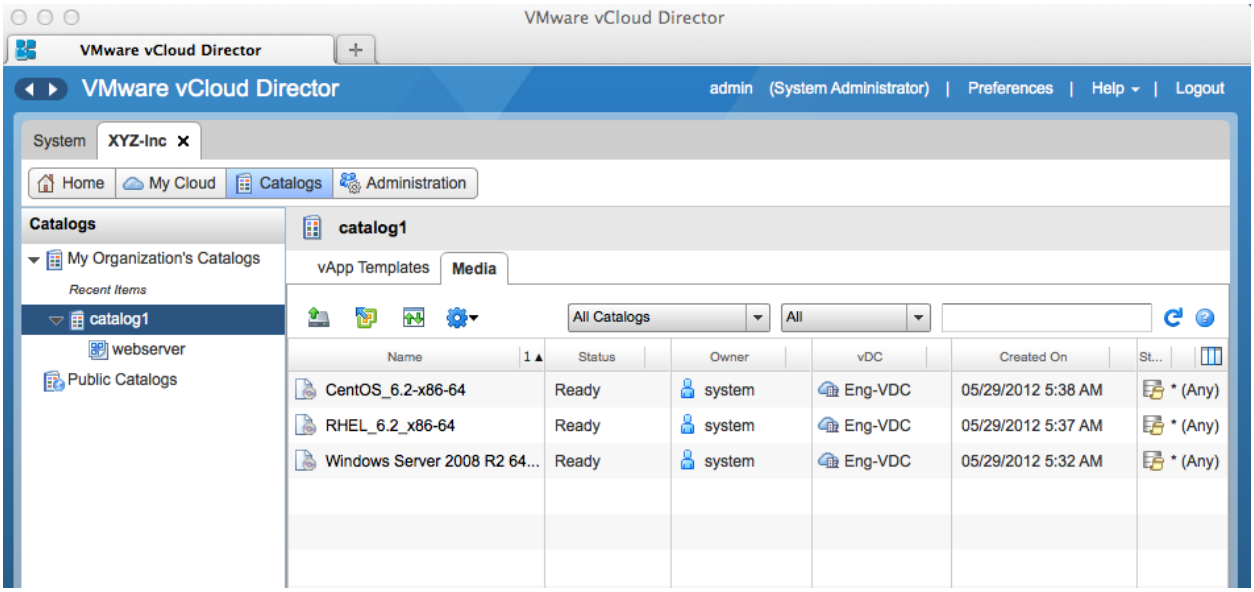
4. Organization Catalogs

An *organization catalog* is a container for vApp templates and media files. Detailed information about organization catalogs is available in the *VMware vCloud Director Documentation* (https://www.vmware.com/support/pubs/vcd_pubs.html).

The following figure shows two views of a sample published catalog. The vApp templates are visible from the **vApp Template** tab, and the media files are visible from the **Media** tab.

Figure 3. vApp Templates and Media Files in a Catalog





Organizations can offer the following types of service catalogs to their users or customers:

- A *vCloud service catalog* – Includes predefined vApps, virtual machines, and images (operating systems and applications) that users can deploy within an organization.
- An *operational service catalog* – Includes operational features such as development of a vCloud service catalog, backup and recovery services, archival services, managed services, and migration services.

The following sections focus on the vCloud service catalog and include design considerations for organizations and their virtual datacenters as well as their service catalogs. Catalogs can grow exponentially with the permutation and combination of guest operating system versions, service packs and patch levels, as well as any installed applications or configurations and their version levels. VMware vFabric™ Application Director™ can address the growing number of vApp templates in catalogs.

4.1 Understanding Catalogs

The vSphere approach to implementing catalog-like functionality helps to understand the contents of a vCloud catalog and how the catalog is used. Although vSphere does not use the concept of a catalog, it achieves similar functionality with virtual machine templates and media files (ISO/FLP). The following steps outline how to build a fairly common Linux, Apache, MySQL, and PHP (LAMP) stack configuration in vSphere.

1. Gather the relevant media. In some cases the media are readily available for download in the form of ISO files from operating system vendors. Applications might be available only as binaries, in the form of tar, zip, or RPM files. You can copy binaries directly to a virtual machine with tools such as FTP or SCP or by bundling the binaries into an ISO file. These ISO files are then typically available on a dedicated, shared datastore.
2. Create virtual machine templates with a bare guest operating system installed on each template, and clone them to create instances of virtual machines to install applications. You can use a single template to create a web server, application server, and database server. Users can mix and match templates to satisfy use case requirements.
3. After deploying virtual machines from templates, you can customize each virtual machine for its specific purpose by installing Apache/PHP, Tomcat, and MySQL software binaries, either directly or by mounting a custom ISO image.
4. After the individual virtual machines are configured, an administrator can clone them to templates for future use.

vSphere cannot transform a vApp into a template. Any modifications to a vApp require that templates be converted back to virtual machines. Only the virtual machines can be modified.

The following sections discuss how vCloud Director implements the same functionality.

4.1.1 Software Media Files

In a vCloud, ISO files are uploaded to a catalog, as shown in Figure 3. The catalog belongs to an organization that is backed by one or more organization virtual datacenters. These are then backed by provider virtual datacenters. A provider virtual datacenter comprises a collection of compute resources, including datastores. These media files might be private, public, or shared with named users.

Uploaded media files are located on the compute resource associated with the organization virtual datacenter. You can attach them to existing virtual machines or use them to install new virtual machines with a process almost identical to the vSphere procedure in Section 4.1.

A good catalog includes all commonly used operating systems, such as the various editions of Windows and common Linux distributions such as Red Hat, CentOS, or SUSE. Software media can consist of custom operating system builds, kickstart CDs, or software packages. There are no limits when including default operating system installation media.

4.1.2 vApp

In the LAMP stack example that follows, three virtual machines running a Linux guest operating system are deployed into a vSphere environment. The most efficient approach is to build and configure a single virtual machine, clone it to a template, and re-deploy the same virtual machine as many times as necessary to support application requirements. It might be necessary to customize the virtual machine's hardware configuration, depending on the application.

In vCloud Director, the smallest construct is a vApp. The vApp can contain one or more virtual machines.

To create a LAMP stack

1. Create a vApp.
2. Within the vApp, create a virtual machine with a specific guest operating system.
3. After installing the guest operating system, shut down the vApp and capture it to a catalog.
4. After capturing the vApp, create the LAMP stack by creating a new vApp.
5. Add the captured virtual machine repeatedly, as needed.
6. Install and configure the application as needed.
7. Shut down the vApp.
8. Capture the vApp again to the catalog.
9. Other users can now deploy the vApp from the catalog.

The foundation for a good catalog is a collection of commonly used configurations of virtual machines, operating systems, and applications that can be deployed as single vApps or used to build vApps that contain multiple virtual machines.

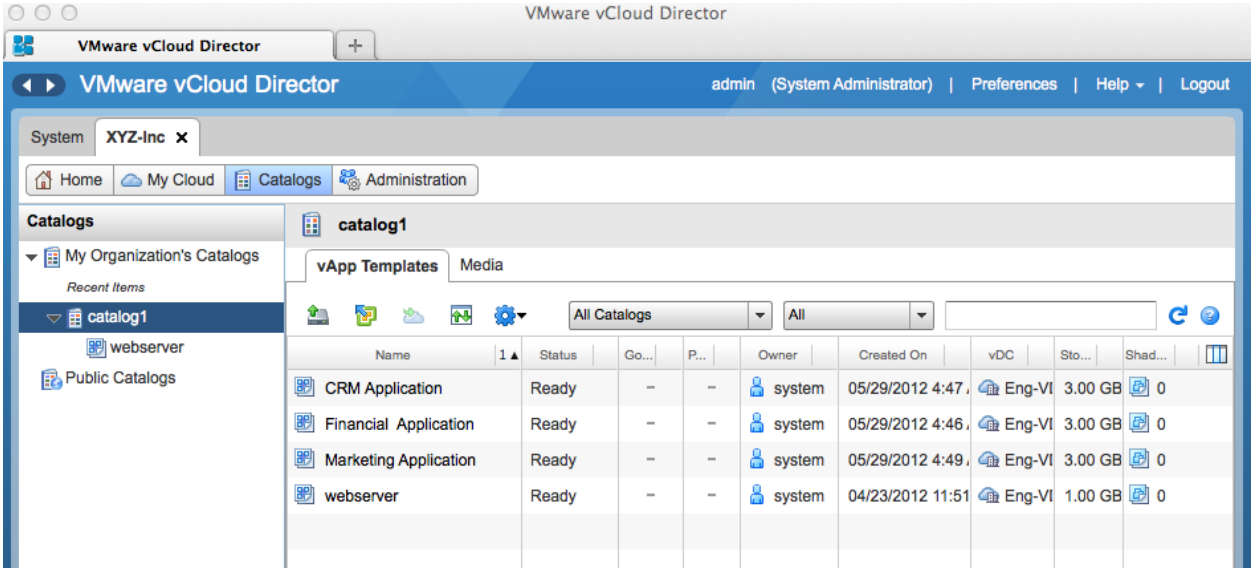
4.2 Populating a Catalog

Populating a catalog involves creating vApps and adding them to a catalog while considering costs associated with the catalog and vApps after they are deployed. Populating also involves determining whether the catalog should be placed in a global or organizational catalog.

4.2.1 Catalog Items

To populate catalog items such as vApp templates, use the vCloud Director **Catalogs** tab (see Figure 4), or add them from the **My Cloud** tab of a powered-off vApp.

Figure 4. vCloud Director Catalogs Tab



Consider the following when determining the number of vApp templates to add to the catalog:

- A need to enforce standardized virtual machine configurations.
- A requirement to provide applications pre-configured as vApps.
- User privileges needed to modify vApp and virtual machine configurations.

The size of the anticipated user base and the level of control granted to users determines how many vApp templates to create.

The following are examples of common use cases for catalog entities:

- A *vApp Author*, or application developer, who deploys a vApp as a development and testing platform.
- A *vApp User*, or business user, who deploys a vApp with the latest version of an internally developed application for user acceptance testing (UAT).

4.2.1.1. Application Developer Use Case

Application developers are skilled in IT and require many vApp configurations with different internal virtual machine configurations and software package installation configurations. Rather than trying to predict the requirements for each developer, the infrastructure team can provide a collection of media files and basic, predefined standard vApp templates based on an existing corporate standard. Developers can then construct their own individual vApps and modify the configuration (CPU, RAM, disk) of the contained virtual machines. You might not need to create many vApp templates in advance because a single vApp template for each major guest operating system and application might be enough. There is no requirement to define small, medium, and large hardware-based derivatives of the same vApp template, because the user can edit the virtual hardware after deployment.

4.2.1.2. Business User Use Case

Business users are not necessarily aware of IT requirements and often do not fully understand the differences and implications of virtual machine-level hardware changes. An infrastructure team would not delegate full control to business users and therefore could deploy a predefined vApp configuration that meets their requirements. Because the user cannot edit the hardware after deployment, there could be a requirement to offer a more extensive catalog, including small, medium, and large hardware-based derivatives of the same vApp template.

4.2.2 Cost

The cost of maintaining catalogs and vApps after deployment from a catalog is an additional consideration. Media files and vApp templates consume disk space and have associated costs, as does the actual configuration of virtual machines, which consume compute resources within a vApp. Oversized virtual machines can artificially reduce the overall capacity of an organization virtual datacenter and also have an inherent cost, depending on the allocation model and charging strategy. Given these costs, consider the following:

- Place a catalog in an organization virtual datacenter that uses lower-cost storage, such as NFS.
- Use shared and published catalogs to minimize the number of duplicated vApp templates and media files.
- Depending on the specific catalog requirements, use a catalog-only organization virtual datacenter, mapped to a lower-tier vSphere cluster that uses lower-cost datastores, or if using storage profiles, use a lower-tier storage profile. Using a lower-tier vSphere cluster or a lower-tier storage profile helps provide cost-effective storage for catalog items. Reserve more expensive storage tiers to run workloads.

4.2.3 Global Catalog

When consuming resources from a private vCloud, the provider often populates a *global catalog* with core operating system versions and hardware configurations that align with the organization's current physical and virtual hardware standards. Often these hardware standards are derived from legacy physical server configurations or from virtual machine configurations adopted in previous server consolidation activities. Requirements to maintain consistency with physical standards tend to use chargeback mechanisms based on capital hardware, depreciation, and recurring maintenance as opposed to actual resource allocation costing. This is the basis for using VMware vCenter Chargeback™ with vCloud Director. As confidence in self-service catalogs increases and additional control is delegated to users, a shift to a simplified catalog becomes possible.

A public vCloud provider, on the other hand, is less likely to offer a global catalog with core operating system versions, largely because of the variance in requirements from different consumer organizations and complications associated with licensing. In some cases, a public provider might offer standard media files and standard vApp templates that are not constrained by licensing restrictions. Over time, public vCloud providers can offer more services as they try to differentiate themselves from competitors. Similarly, software vendors might look to partner with public providers to offer their applications as a service.

4.2.4 vFabric Application Director

VMware vFabric Application Director helps to reduce the number of variations of vApp templates in a catalog. Because vFabric Application Director runs arbitrary scripts during vApp deployment, it can be scripted to install patches, service packs, applications, and third-party software. This helps to reduce the number of catalog items with different variations of guest operating system versions and patch levels. It also helps to reduce the number of applications and associated patch levels.

4.3 Working with Catalogs

When working with vCloud catalogs, consider creating, publishing, accessing, and searching catalogs, as well as how consumers interact with providers.

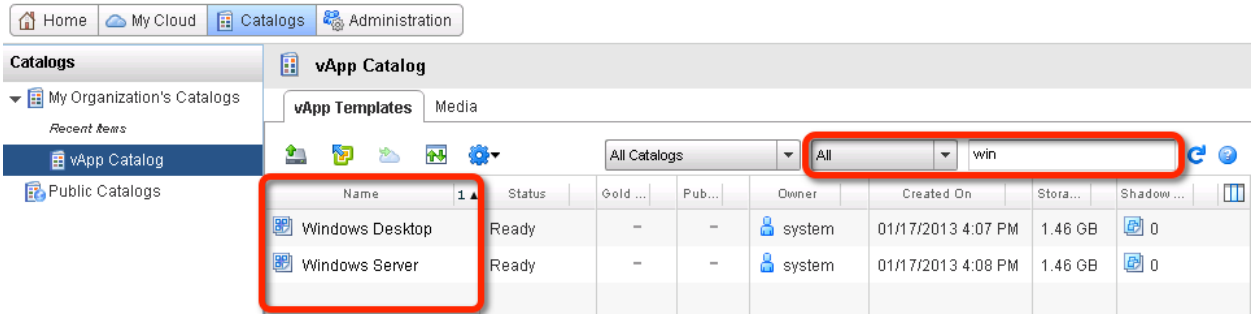
4.3.1 Browsing

To simplify catalog browsing, use a simple, logical, and methodical naming convention for media files and vApp templates. Also consider incorporating vApp template name versioning to identify patch updates and service pack updates. This enables users to quickly find templates that meet their requirements.

Figure 5 shows the use of the search drop-down menu to filter and narrow results. In the example that follows, users seek Windows vApp templates based on a keyword search for "Win." Because a logical naming convention is used, results are almost immediate.

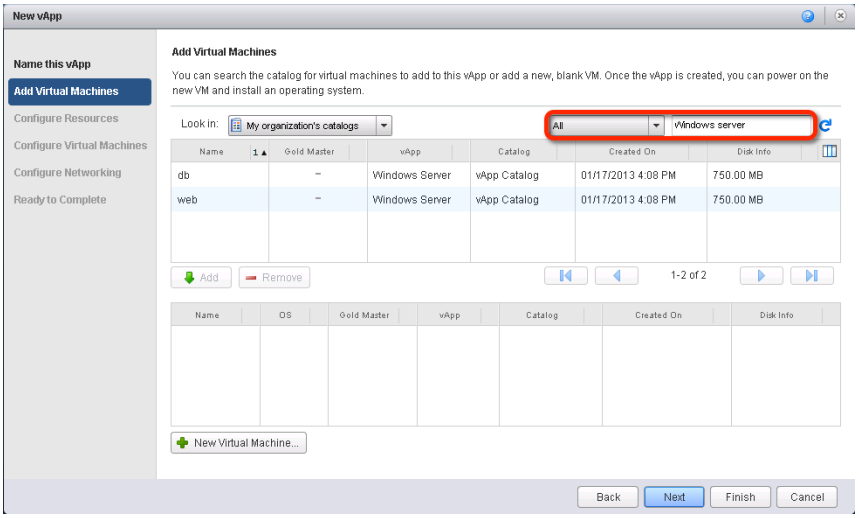
You can search for a vApp template or media file based on **Name**, **Status**, **Gold Master**, **Published, Owner**, **Created On**, or **Virtual Datacenter**. You can specify that the search be conducted on **All Catalogs**, **My Catalogs,** or **Catalogs Shared to Me**.

Figure 5. Browsing and Searching Catalogs



Use the search drop-down menu to filter and narrow your search when creating a new vApp. While adding virtual machines to a vApp, you can search for existing virtual machine configurations that are available from existing vApp templates, as shown in the following figure.

Figure 6. Browsing and Searching for Virtual Machines During vApp Creation



4.3.2 Catalog Access

Catalogs can be private, shared, or published. When a catalog is created, it is owned by the user who created it and is private by default. You can *share* a catalog with specific users and groups in an organization or *publish* it to all organizations within vCloud Director.

When deciding whether to share or publish a catalog, determine who should have access to the catalog and its contents. Publishing the catalog makes it accessible to all organizations, assuming that users have the *View Published Catalogs* permission. Sharing allows more granular access control based on users and groups within the organization where the catalog was created. The following are use case examples for a published catalog in enterprise and service provider environments.

4.3.3 Sharing a Catalog

The key question when deciding to share a catalog is whether you need to share catalog items with all or a subset of the users in your organization. For instance, sharing a catalog could make media files available to all users in your organization. A more restricted use case might involve a group of individuals in an organization who are working on a specific application or operating system and want to share a media file among themselves.

To access a shared catalog, users need the *View Private and Shared Catalogs* permission. This permission is not assigned to all roles by default.

4.3.4 Publishing a Catalog

You can publish a catalog during the creation process, or you can select an existing catalog, right-click, and select **Publish**.

4.3.4.1. Enterprise Environment Use Case

Within an enterprise organization, it is not uncommon for the IT team to create its own smaller organizations to build and update vApp templates. These templates are then shared with other business units through the global catalog and are aligned with corporate build standards to reduce variances and support increased security.

4.3.4.2. Service Provider Environment Use Case

Service providers often allow individual organizations to manage their own catalogs where the service provider offers a compute resource as opposed to a service. Some service providers, seeking to differentiate themselves or offer additional services, might publish common ISO files or basic virtual machine configurations consistent with licensing implications. Service providers might also offer published catalogs that contain applications developed or configured by software vendors.

ISO images cannot be shared with different organizations in a published catalog. For an ISO image to be mapped to a virtual machine, the ISO image must be available in the same organization virtual datacenter as the virtual machine.

4.3.5 Media File Limitations

vCloud Director does not permit media file sharing outside of an organization. This means that even if a catalog is published and vApp templates are available, the media files are not. To address this issue, build local organization catalogs and share those catalogs with the entire organization. All users should have the *View Private and Shared Catalogs* permission, but the vApp User role does not provide this permission by default.

4.3.6 Updating vApp Templates

A user's ability to update vApp templates and hardware and network configuration details depends on the permissions delegated to the role associated with the user's account. For example:

- vApp User – This role does not allow the user to reconfigure vApps or vApp templates. Users cannot modify the compute resource or disk, or view the catalog or existing vApps, unless they are shared with them. The user who owns the vApp can add and remove network interface controllers (NICs) and update the networks to which a vApp is connected.
- vApp Author – This role allows for the provisioning of vApps but does not permit access to published catalogs by default. The user must create virtual machines from scratch unless granted access to the published or shared catalogs. The user can update vApp and virtual machine configurations such as CPU, RAM, and networking.

The effect of these permissions can differ, depending on whether a vApp is shared or owned. For these permissions to apply, the user must *own* the vApp. When the vApp is shared, the user can be granted **Read Only**, **Read/Write**, or **Full Control** permissions.

4.3.7 Deploying vApps

A user's ability to deploy vApp templates depends on the user's role and the permissions delegated. For example:

- vApp User – This role can deploy vApps from a catalog and copy it to an alternative organization virtual datacenter.
- vApp Author – With the proper permissions, this role can deploy vApps, and select a destination organization virtual datacenter. Following deployment, the vApp Author can copy a vApp from one organization virtual datacenter to another in the same way as a vApp User, or the vApp Author can move a vApp from one organization virtual datacenter to another. A vApp Author can also add the vApp to the catalog.

4.3.8 Selecting Networks

The impact of using a catalog on network connectivity is subtly different depending on whether vApp network or organization virtual datacenter network types are in use.

4.3.8.1. Internal vApp Networks

When a vApp containing internal vApp networks is deployed, the configuration of the internal networks is consistent with the configuration at the time the vApp was added to the catalog. The configuration, including firewall rules, is predefined and does not need to be updated. A user with the appropriate permissions can change this configuration.

The exception to this rule is when you use network address translation (NAT) with the default one-to-one mapping configuration. All virtual machines connected to the given vApp network use this NAT configuration automatically, regardless of any change you made when creating the vApp. For example, if a vApp containing three virtual machines is created and mapped to an organization virtual datacenter network with NAT, three NAT-routed IP addresses are allocated. Although it is possible to disable these IP addresses manually as in the LAMP stack example, changes are lost when the vApp is deployed from the catalog. For the changes to remain in effect, use the port forwarding approach where all rules are retained as they would be for firewalls.

4.3.8.2. Organization Virtual Datacenter Networks

For a vApp and its associated virtual machines to have connectivity outside the vApp, the virtual machines must be connected to an internal or external organization virtual datacenter network. Available organization virtual datacenter networks are usually created and defined by an organization administrator. When the vApp is deployed, it is automatically attached to the organization virtual datacenter network defined when it was created and uploaded to the catalog. It is also possible for the user to select which network to connect to the virtual machine.

A vApp deployed from a published catalog created in a different organization might have a different or invalid network defined, in which case the user should select a valid organization virtual datacenter network and attach the vApp.

5. Creating and Managing vApps

The vApp is a virtual application container, and is the basic construct for workloads in a vCloud deployment. vCloud vApps are similar to vApps in vSphere, but vCloud extends the attributes of a vApp. A vSphere and vCloud Director vApp can contain one or more virtual machines. A multitiered, vCloud vApp includes the attributes of each component—SLAs, role-based access controls (RBAC), and lifecycle management.

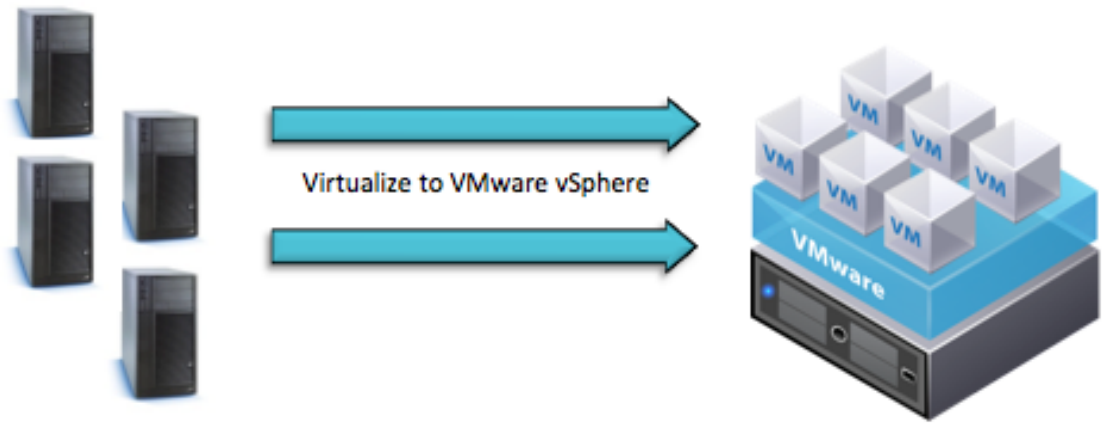
5.1 Migrating Workloads to a vCloud

Physical workloads cannot be directly migrated to a vCloud. However, they can be migrated after they have been initially migrated to vSphere.

5.1.1 Migrating Physical Workloads to vSphere

Physical machines cannot be directly migrated to a vCloud datacenter because of limitations such as device drivers, any hardware dependencies, and the static attributes of a physical system. vCloud requires virtualized workloads. Apart from the basic virtualization foundation, a fully implemented vCloud datacenter also requires availability, scalability, resource pooling, and VMware vSphere Distributed Resource Scheduler™ (DRS). The first stage of moving a physical machine to a vCloud datacenter is to virtualize it on vSphere.

Figure 7. Migrating from a Physical Machine to a vSphere Virtual Machine

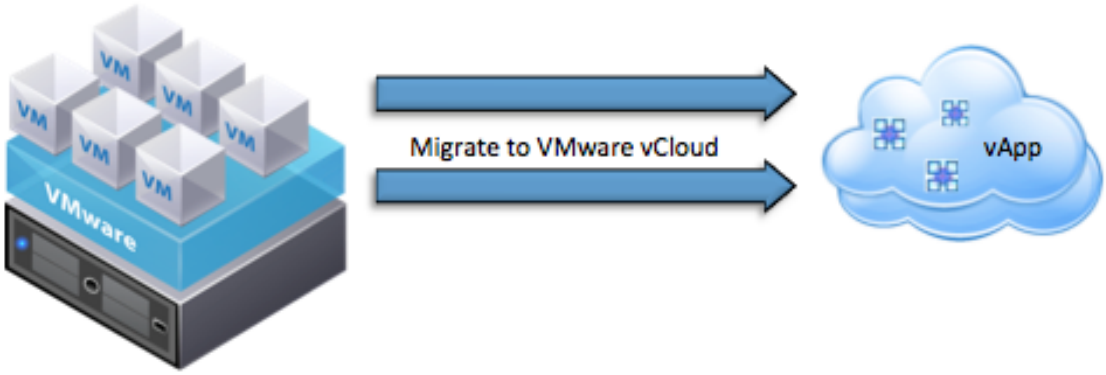


5.1.2 Migrating Virtual Workloads to vCloud Director

Before migrating the newly created virtual machine to the vCloud datacenter, VMware recommends that you identify a minimum period of time to monitor and optimize the virtual machine hardware for the guest operating system and guest application requirements. For many applications, an appropriate amount of time for this process is a typical full application business cycle, such as the first day through the last day of a business month. As an example, use VMware vCenter Operations Manager™ to determine whether the virtual machine is sized properly.

After the physical machine is virtualized, stabilized, and optimized to run in a virtual datacenter, you can move the virtualized workload to a VMware vCloud datacenter.

Figure 8. Migrating from a vSphere Virtual Machine to a vCloud vApp



In vCloud, a virtual machine is encapsulated by a vApp. A vCloud vApp is a logical entity comprising one or more virtual machines. It uses the Open Virtualization Format (OVF) to specify and encapsulate all components of a multitier application and the associated operational policies and service levels. A vCloud vApp might or might not be associated with vApp networks for inter-virtual machine communication in a vCloud vApp.

To move a virtual machine from a virtual datacenter to a vCloud datacenter, export the virtual machine and then import it in OVF file format.

Note If vCloud Director is attached to a vCenter server, you can import the virtual machines without having to export them.

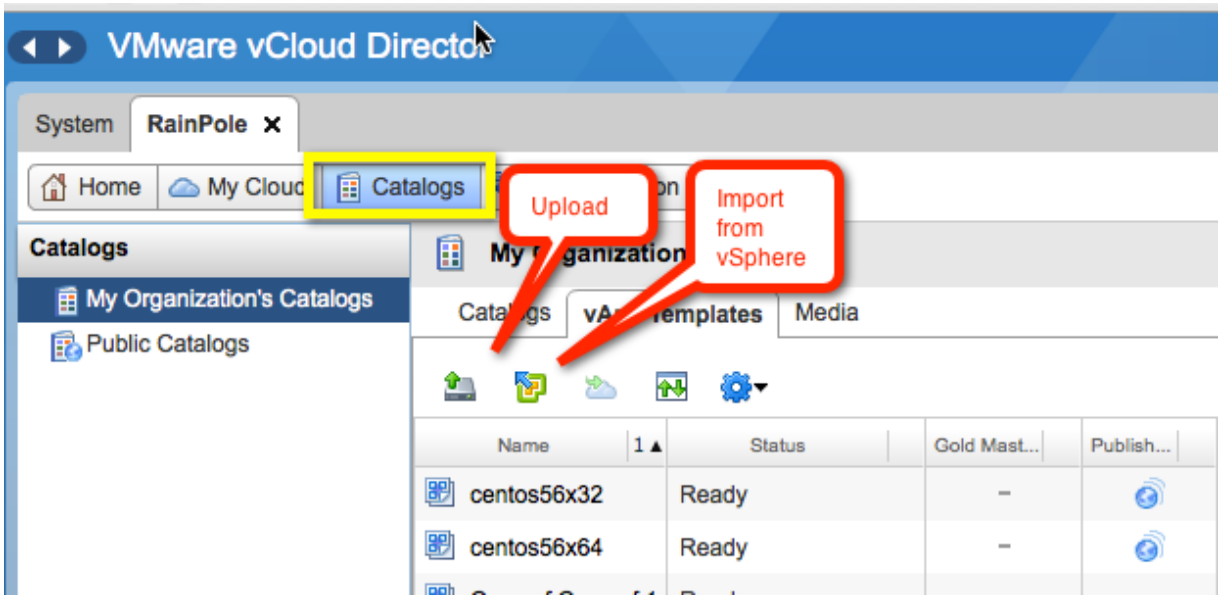
OVF is an industry standard approved and certified by the Distributed Management Task Force (DMTF). A standard OVF package consists of the following components:

- At least one OVF descriptor with extension `.ovf`.
- Zero or one OVF manifest with extension `.mf` (contains the SHA-1 digests of individual files in the package).
- Zero or one OVF certificate with extension `.cert` (contains digest of the manifest file and base64-encoded X.509 signed certificate).
- Zero or more disk image files with extension `.vmdk` (required for VMware vCloud).
- Zero or more additional resource files, such as ISO images.

There are several ways to convert a virtual machine or a vApp to an OVF package that can be consumed directly by the organizations and catalogs in the vCloud datacenter:

- Export and upload the OVF manually:
  - Use the vSphere Client™ to export the OVF file manually and select **Catalogs – Upload** in the vCloud UI to upload the OVF file. The user privilege required for this operation is at least *Catalog Author*, *Organization Administrator*, or *System Administrator*.
  - Manually import directly into the vCloud from a connected vCenter instance. The virtual machine must be powered off, without any snapshots or vSphere Fault Tolerance enabled. Select **Catalogs** > **Import from vSphere** in the vCloud UI to import the virtual machine. This operation requires the *System Administrator* user privilege.

Figure 9. Manually Import to vCloud



- Use a hybrid vCloud plug-in, such as VMware vCloud Connector™. The vCloud Connector virtual appliance must be installed and configured to be used with the source vCenter instance. You must also add a vCloud to vCloud Connector to manage it.

You can use any of these methods to move workloads into the vCloud, but first consider and understand the limitations imposed by the physical locations of the vCloud environments.

5.1.3 vApp Migration

In conjunction with vCloud Connector, a hybrid vCloud enables vApps to move between public and private vCloud services. Prior to vCloud Director 5.1, some configurable vApp elements are not transported with the migration process. Keep track of what is and is not transported as you might need to reconfigure the vApp after the migration so that it functions properly. You can view the configuration parameters from two perspectives—from the point of view of the vApp itself or of the virtual machines contained within the vApp. With vCloud Director 5.1, lossless OVF export is provided to enable greater portability between vCloud environments.

Note Metadata defined on vCloud entities is not transported during the migration process.

5.1.3.1. Migration Process

To move a vApp from vCloud A to vCloud B with the vCloud Connector:

1. Export the virtual machines in the vApp from vCloud A to an OVF.
5. Stage them on the vCloud Connector appliance with the vCloud API. The data is transferred to the vCloud Connector appliance over HTTPS.
2. Use the `HTTPS upload` command to upload the staged OVF to vCloud B.
3. Import the OVF into vCloud B's catalog or directly to an organization datacenter.

The migration process can be carried out manually with the upload and download capabilities of the vCloud Director user interface or by programming the vCloud API. vCloud Connector simplifies vApp migrations, but is not required.

5.1.3.2. vApp Power Action Configuration

When a vApp is migrated, the vApp itself defines the virtual machine *start* and *stop* settings within the vApp. The parameters encapsulated within the vApp are:

- Start/Stop Order.
- Start Action.
- Start Delay.
- Stop Action.
- Stop Delay.

All of these parameters are transferred with the vApp from one vCloud to another. The configuration in vCloud A remains consistent in vCloud B.

5.1.3.3. vApp Network Configuration

Configuration parameters are associated with virtual machines in the migrated vApp, including network segments (vApp networks) that are private to the vApp itself in the source vCloud. Some of these configuration items are maintained during a migration, while others require reconfiguration by the vApp user after migration.

vApp networks are not maintained across a vCloud migration—vApp networks that are defined in a vApp in vCloud A are *not* created and mapped to the virtual machines within the vApp in vCloud B. Assuming that vCloud A and vCloud B use the same network topology, after moving the application to vCloud B, the end user must complete the configuration process by recreating the vApp networks as they were defined in vCloud A.

5.1.3.4. Hardware Configuration

Hard disk configuration and disk bus type are maintained across a vCloud migration. If a specific disk bus type is selected for a virtual machine's disk, this setting is maintained when the virtual machine migrates from vCloud A to vCloud B.

Prior to vCloud Director 5.1, network interface assignments, configuration types, and network adapter types are not maintained across a vCloud migration. When the vApp is transferred, the virtual machine network interfaces revert to the default adapter type, **flexible**. With vCloud Director 5.1, the lossless OVF export capability preserves these settings.

IP addresses are not held across a vApp migration and must be reconfigured to the new vCloud IP addressing scheme. The end user must reconfigure all virtual machine network adapter settings for a migrated vApp.

5.1.3.5. Guest OS Customization

Even if **Enable Guest Customization** is not selected in vCloud A, it is enabled in vCloud B when the vApp is migrated, regardless of whether the virtual machine has an installed operating system and VMware Tools™.

Most password reset parameters are maintained. When the password reset parameters are configured for the virtual machine, the settings are transferred from vCloud A to vCloud B. However, if the virtual machine configuration specifies a default local administrator password, this information is not transferred to vCloud B. The end user must re-enter the default password for the virtual machine.

5.1.3.6. Resource Allocation

Resource allocation controls set on the individual virtual machines in a vApp are maintained across a migration from vCloud A to vCloud B. However, for the settings to be used, the vApp must be deployed to a virtual datacenter that uses a Reservation Pool allocation model. Datacenters that use the Allocation Pool or Pay As You Go model do not support resource allocation controls.

The following table summarizes vApp parameters and indicates whether they are maintained across a migration.

Table 2. vApp Parameters

Configuration Item	Maintained?	Notes
Start/Stop Order	Yes	
Start Action	Yes	
Stop Action	Yes	
Start/Stop Delay	Yes	
vApp Networks	No	Networks and associated configurations are lost. Networks must be redefined and NICs attached.
NIC Assignment (non-vApp network)	No	NICs must be reassigned.
Guest Customization Enabled	No	Checkbox is selected in target vCloud.
CPU Hot Add	No	
Memory Hot Add	No	
Synchronize Virtual Machine Time	No	
Password Reset	Yes	
Default Password	No	Defaults to blank field in UI.
Disk Bus Type	Yes	
Network Adapter Type	No	Defaults to flexible.
Resource Allocation	Yes	When vApp is deployed to Reservation Pool virtual datacenter.
OVF Properties	Yes	Accessed through VMware Tools and used for scripting, fetching, and for post-configuration and provisioning.

5.2 Using vCloud Workloads

The following sections provide examples of logging into vCloud Director, deploying a vApp, interacting with the vApp, and managing runtime and storage leases.

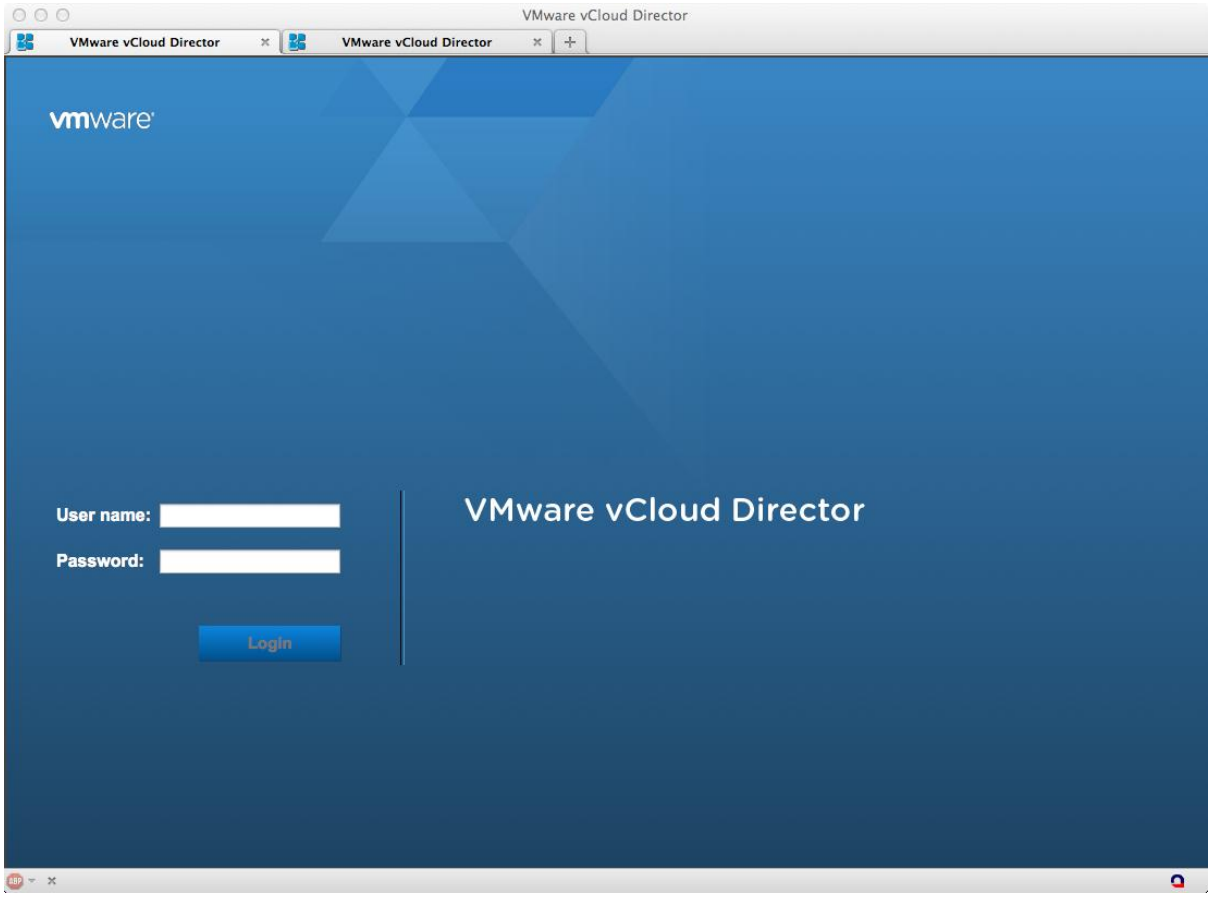
5.2.1 Logging in to the vCloud Director Portal

vCloud Director requires Microsoft Internet Explorer 7.0 or later, or Mozilla Firefox 3.x or later. For advanced browser configuration options, see the *vCloud Director User's Guide* (https://www.vmware.com/support/pubs/vcd_pubs.html).

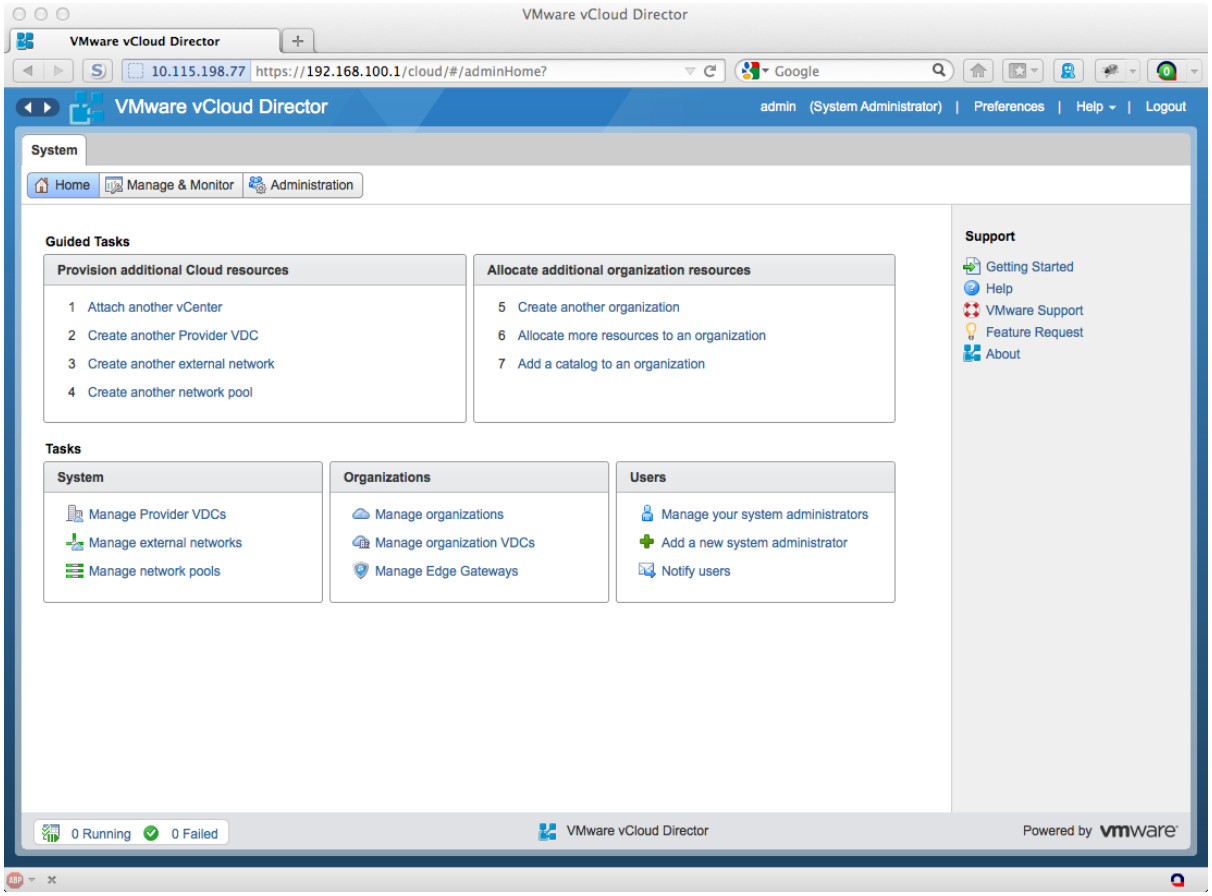
The following procedure provides an example of how a consumer logs in to the vCloud Director portal.

To log in to the vCloud Directory portal

1. Go to your organization URL (https://<vCloud-Director-IP-address>/cloud/org/<OrganizationName>/).
2. To log in, enter your username and password, as provided by your administrator.



3. After you have logged in, you can perform various tasks from the main vCloud Director **Home** view.

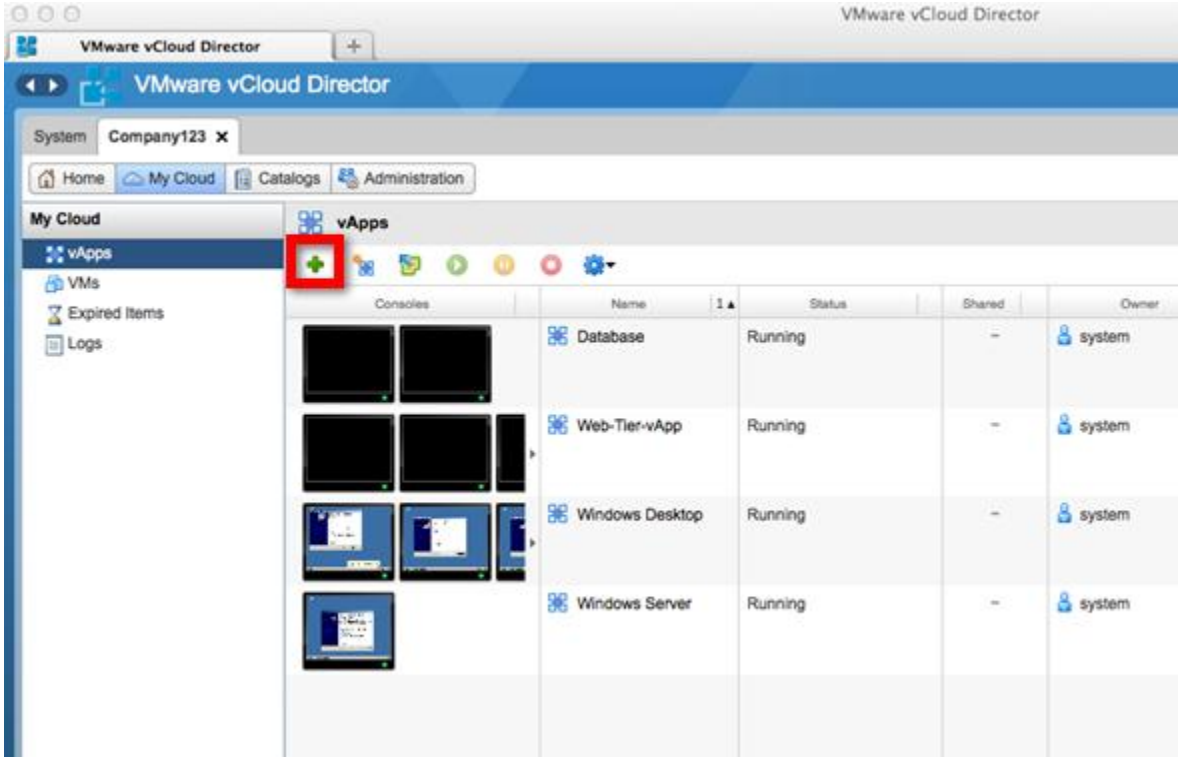


5.2.2 Deploying a vApp

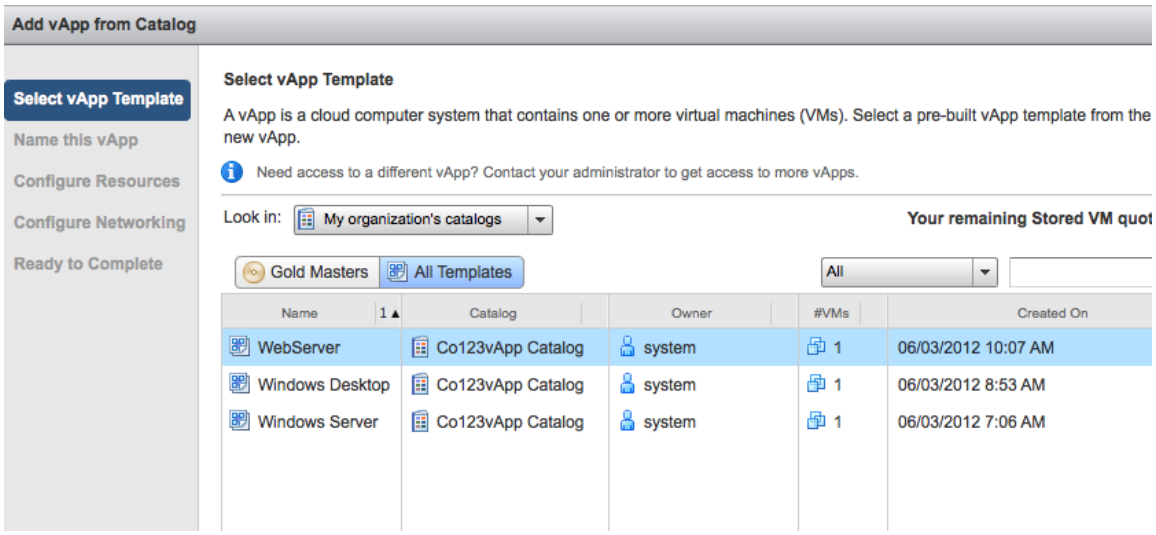
The following procedure provides an example of how a consumer can deploy a vCloud vApp from a catalog.

To deploy a vApp from a catalog

1. On the vCloud Director administration screen, click **Add vApp from Catalog**.
2. Review the available catalogs, and select a vApp.

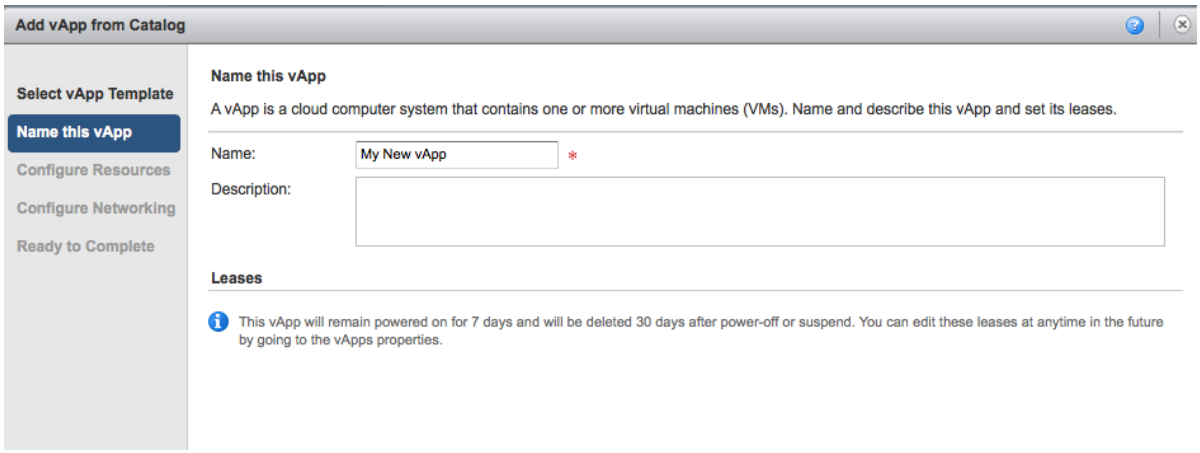


3. Click **Next** (not shown)



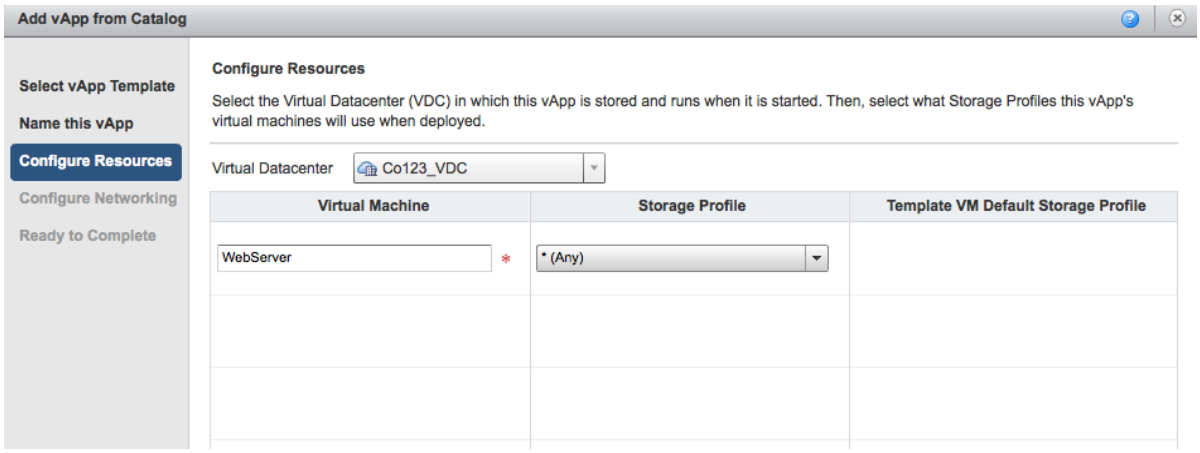
4. Select a name and enter a description for your new vApp.

   Usually, runtime and storage leases are predefined, but in some cases you might be permitted to specify a time period.

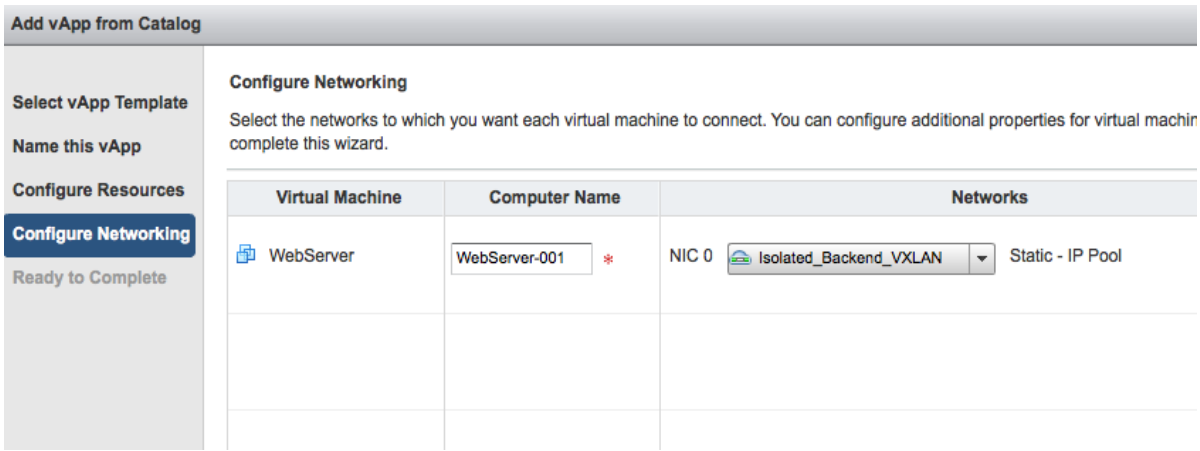


5. Select the destination organization virtual datacenter from the **Virtual Datacenter** drop-down menu.
6. Select the storage profile for this vApp. This is a new feature in vCloud Director 5.1.

7. Click **Next**.

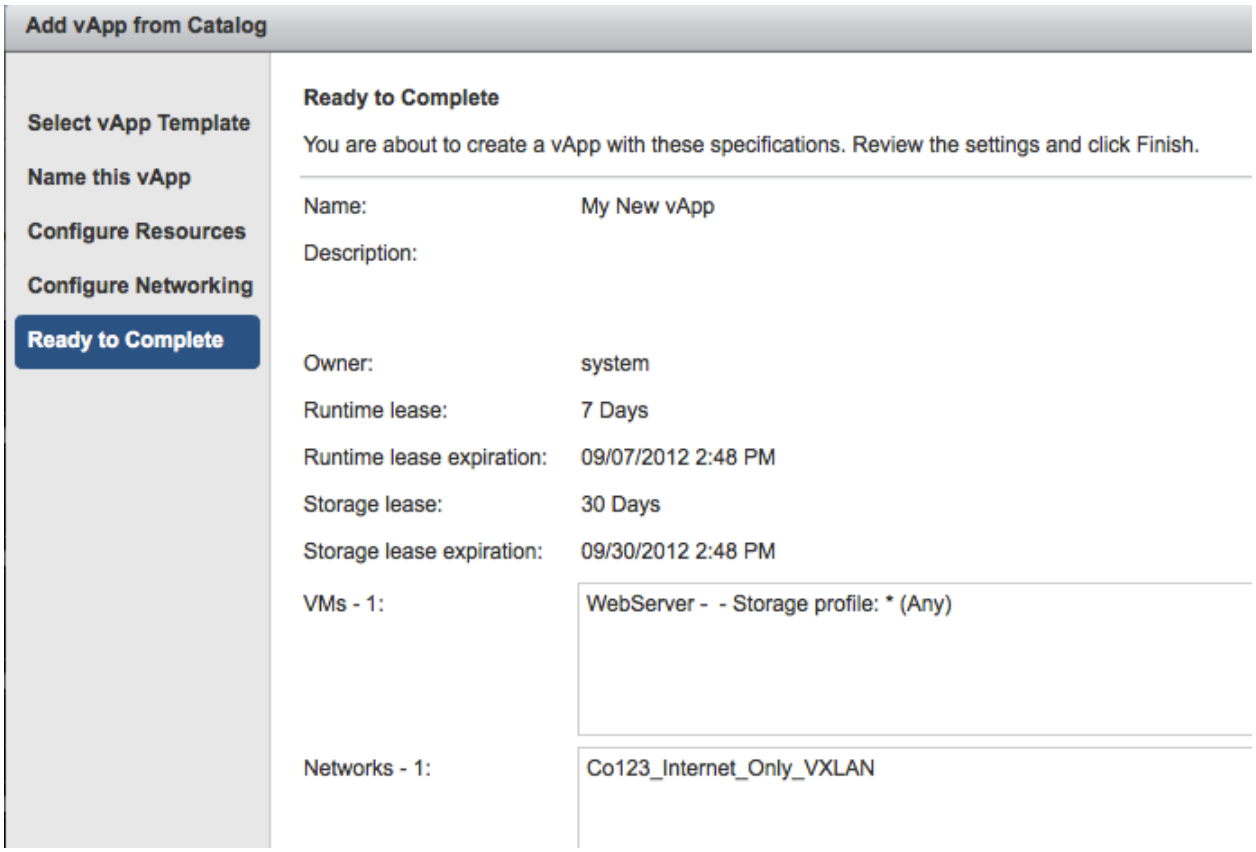


8. Specify the **Full Name** and **Computer Name** for each virtual machine in the vApp.
9. Select the network and IP address assignment. Typically, an external or internal network is available.
10. Click **Next**.



11. On the **Ready to Complete** screen, verify that everything is correct.
12. Click **Finish**.

13. The status for the requested vApp deployment is displayed on the main administration screen.

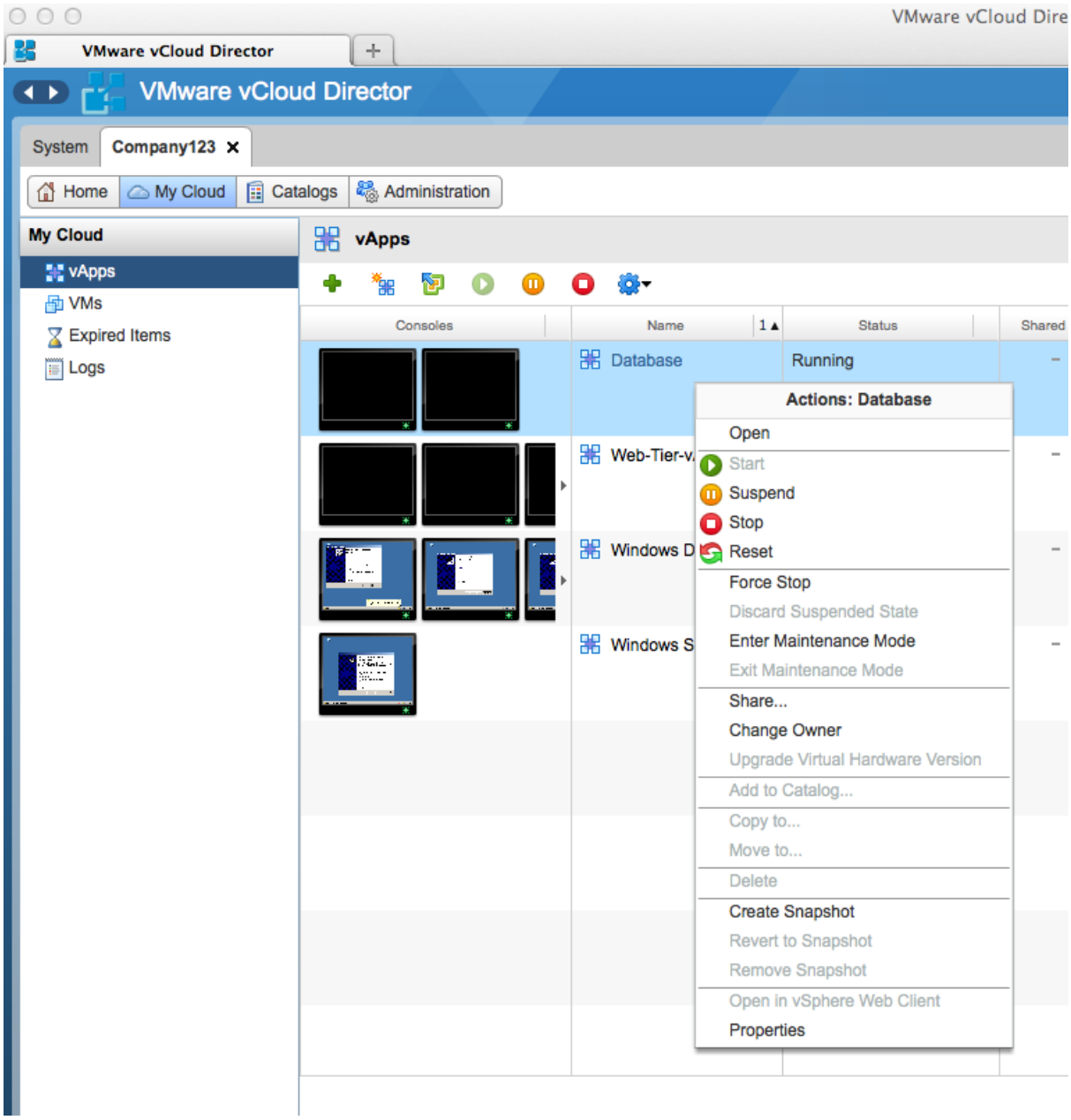


5.2.3 Interacting with the vApp

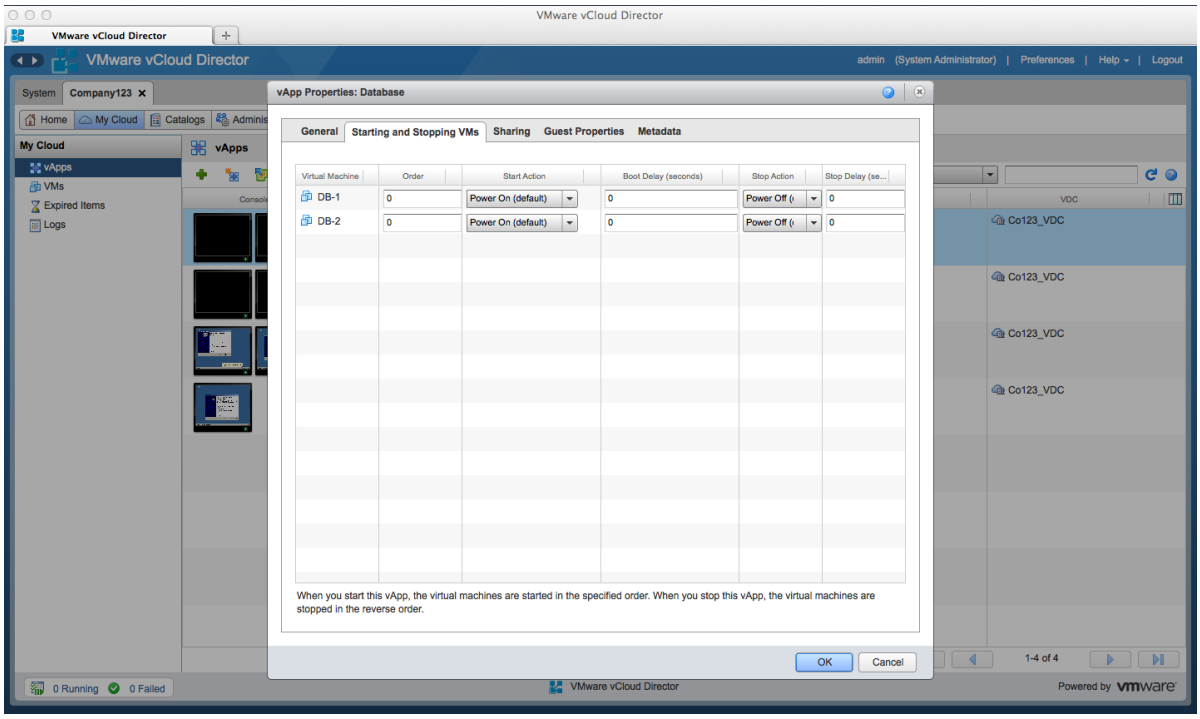
The following example illustrates the deployed vApp options available to a user and describes where and how to implement configuration changes.

To interact with the vApp

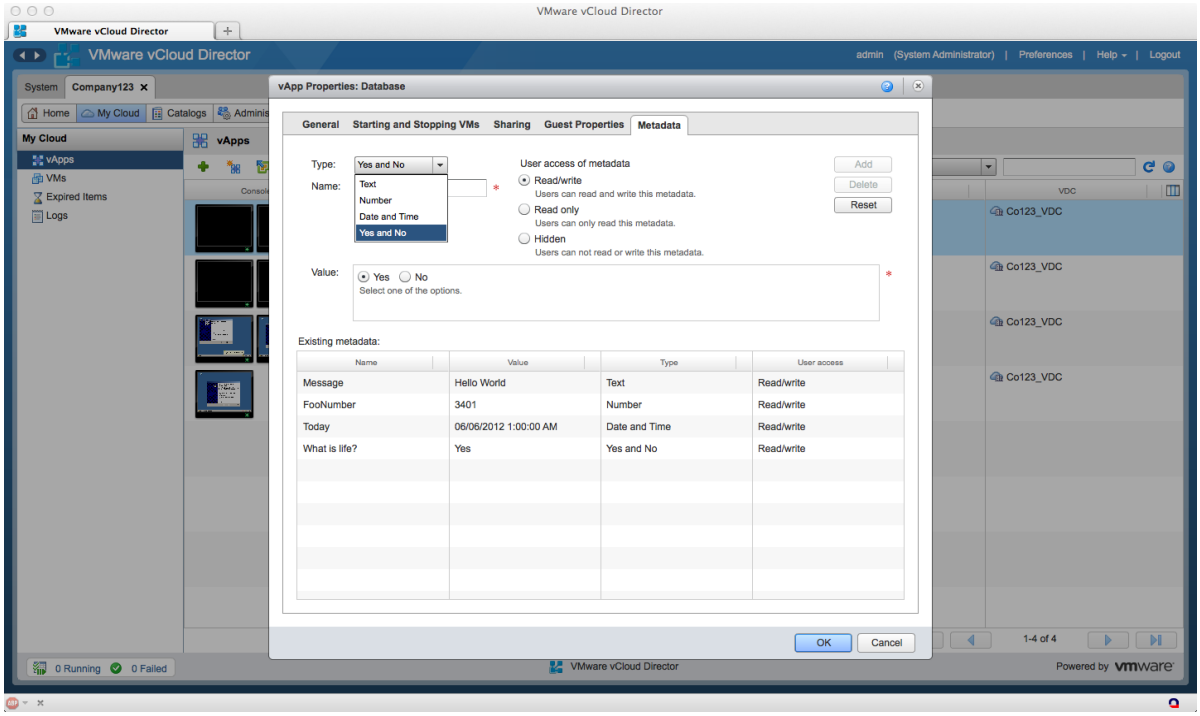
1. From the vCloud Director Organization Administration screen, click **My Cloud**.
2. From the list of vApps, right-click the desired vApp to display an **Actions** menu.



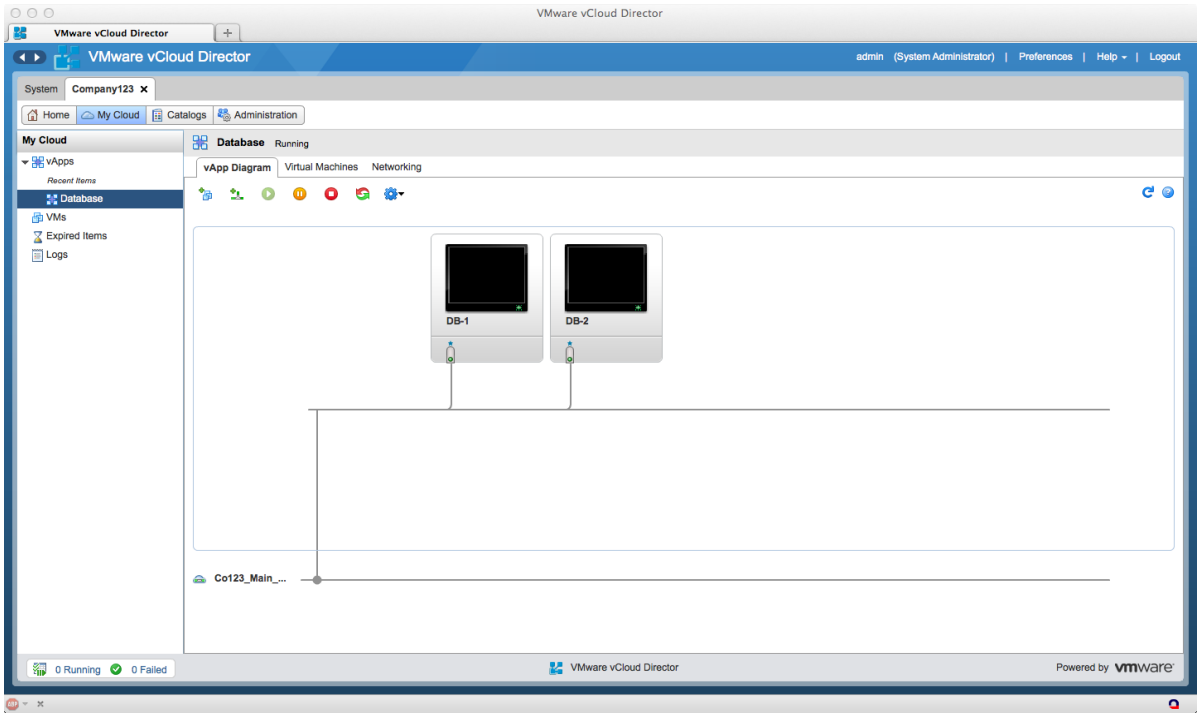
When a vApp is powered on, all virtual machines within that vApp are also powered on. You can modify this behavior by selecting **Properties** in the **Actions** menu, clicking the **Starting and Stopping VMs** tab, and using the drop-down menus to modify and re-order start and stop actions.



3. Under the **Metadata** tab, enter any arbitrary value.



4. To view a diagram of all virtual machines, see networking details and access each virtual machine individually.
5. Select a vApp from the vApp menu.
6. Click **Open**.



7. Click the **Virtual Machines** tab to gain console access to each machine.
8. Right-click the console to access virtual machine console controls.

5.2.4 Runtime and Storage Leases

Because vCloud is a shared environment, your administrator might specify *runtime leases* and *storage leases* to prevent you from indefinitely running your vCloud computing resources. Leases provide a level of control over an organization's storage and compute resources by specifying the maximum amount of time that vApps can run and that vApps and vApp templates can be stored. The computing and storage resources that represent your vCloud computing system expire at a point determined by your vCloud administrator and are automatically freed up for other uses at that time.

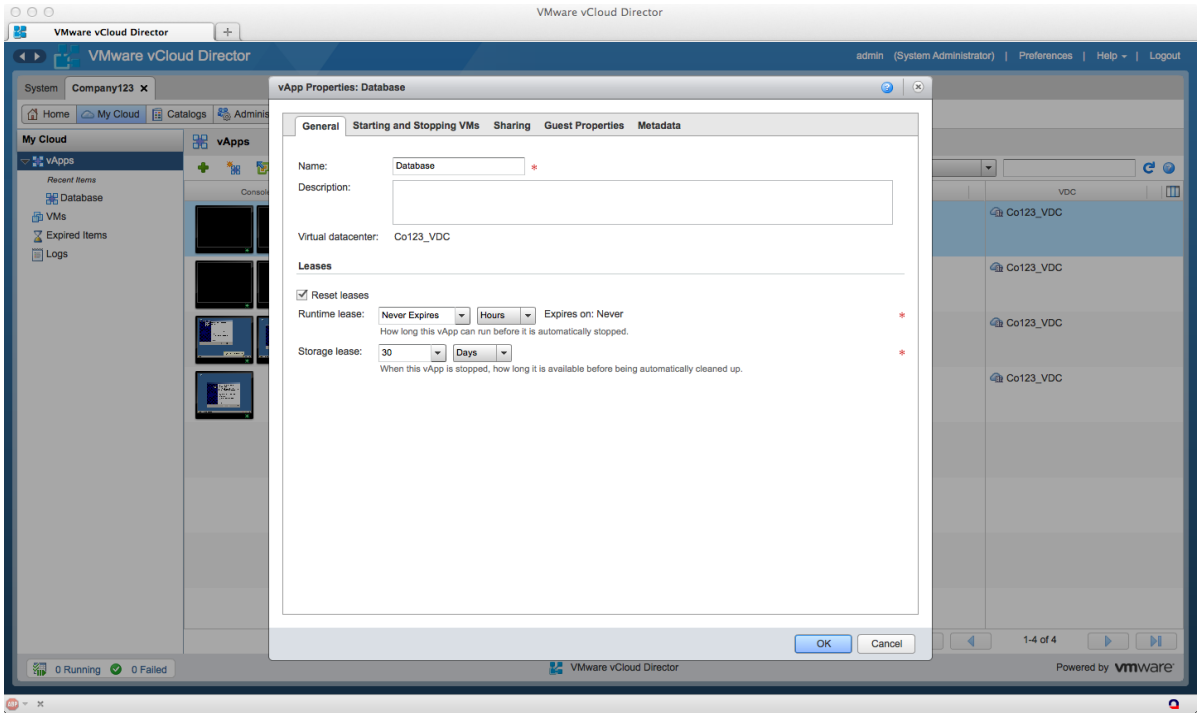
The runtime lease prevents inactive vApps from consuming compute resources. If a virtual machine is unused, it is powered down so that memory and compute resources can be used for other workloads. The storage lease functions the same way but also reclaims storage.

The vCloud administrator defines what happens when these leases expire. In the following example, the runtime lease is initially set to 14 days. The user can log in and reset to extend the lease. In this example, the user extends the lease for an additional 14 days.

As your vApp approaches the expiration date of its lease, you typically receive an expiration notice in email.

To extend the lease of a running vApp

1. Choose **Manage vApps** from the **Home** tab.
2. Right-click your vApp.
3. Select **Properties**.
4. Select **Reset Leases**.
5. Change the values of the leases as desired.
6. Click **OK**.

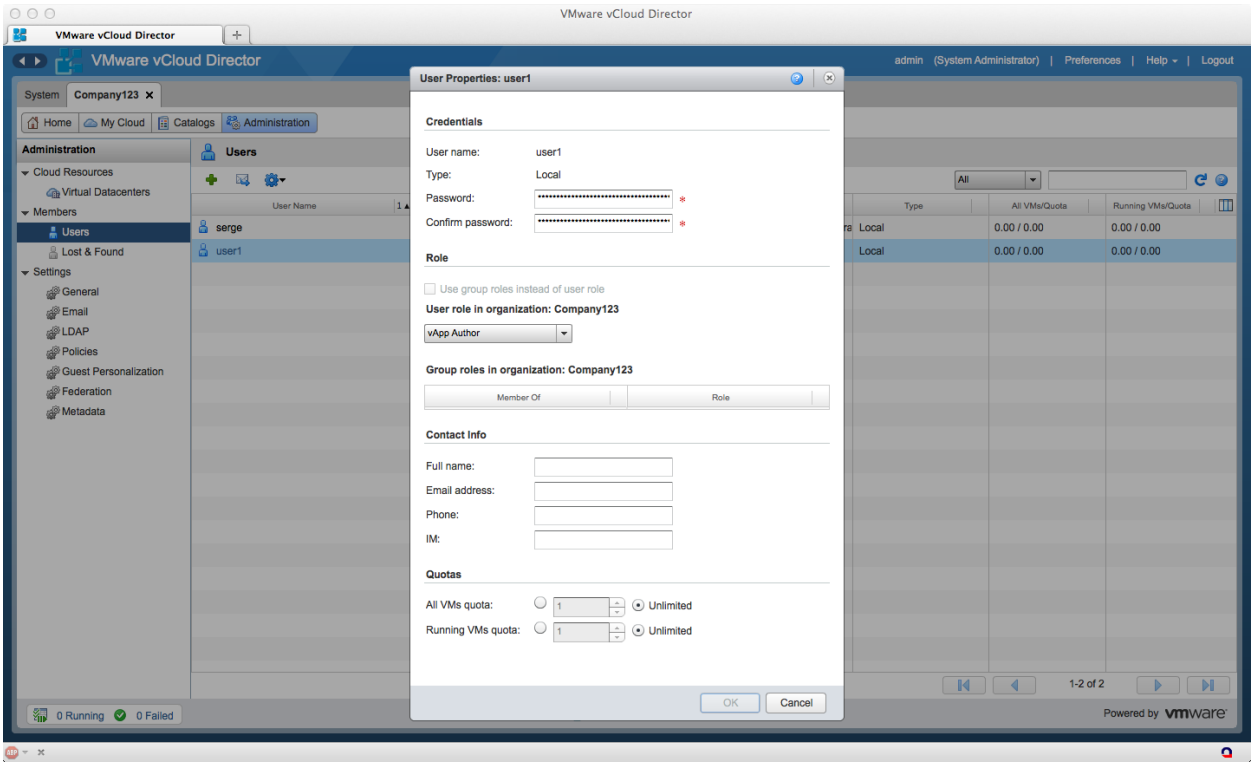


After the virtual machine is stopped when the runtime lease expires, the storage lease goes into effect. Depending on the configuration set by the administrator, the vApp might be moved to an expired state or deleted.

Ask your administrator what are the default options for your organization. The following quotas and limits can be applied:

- Running virtual machines quota.
- Stored virtual machines quota.
- Number of resource-intensive operations per user limit.
- Number of resource-intensive operations per organization limit.
- Number of simultaneous connections per virtual machine limit.

Similarly the running virtual machine and stored virtual machine quotas can be applied to individual users. As a user, you can find out what your individual limits are by clicking **Administration** > **Users**, right-clicking your user account, and selecting **Properties**. Scroll to the bottom to see the defined quotas.



If your vApp is missing, the storage lease of your vApp might have expired. Contact your administrator to determine whether your vApp has expired and might be recoverable or was deleted and might not be recoverable.

5.3 Directory Services in vCloud

Directory services serve several purposes in vCloud Director. This discussion refers to Microsoft Active Directory, but the same considerations apply for other directory services. Directory services used with vCloud Director include authentication services in the following areas:

- Infrastructure – vCloud Director, vCenter, and other supporting applications that are not managed by vCloud Director.
- vCloud Director Management – Within the vCloud Director portal for the system and organization realms.
- Organizational – Internal to an organization, including vApps contained in an organization.

Note For considerations that apply to running Active Directory or other clock-shift sensitive directory services within a virtual environment, see the white paper Virtualizing a Windows Active Directory Domain Infrastructure (http://www.vmware.com/files/pdf/Virtualizing_Windows_Active_Directory.pdf).

The best location for directory services depends on how the services are being used. Locations can be external to an organization, shared inside an organization, or encapsulated and distributed per vApp.

Placement can be guided based on a few key measures. Use the information in the following sections in conjunction with other reference materials to form the basis for an appropriate design.

5.3.1 Hosting Locations for Directory Services

When selecting the placement of directory servers in the vCloud, consider the strict availability and longevity requirements for directory service servers in balance with the applications that will be supported by the deployed instance.

5.3.1.1. External to the vCloud

VMware recommends that any services that support the vCloud Director instance and infrastructure be hosted externally to the vCloud Director managed environment. Specifically, directory services should be hosted external to the vCloud and can be configured using standard procedures for virtualizing the service. Follow design guidelines for virtualizing Active Directory and other directory platforms. Using vCloud does not change these practices.

In a private vCloud architecture, directory services can be hosted external to the vCloud environment if there is no geographical separation between host platforms.

5.3.1.2. Within the vCloud

vCloud-hosted applications are dependent on directory service. After you determine the level of dependency, performance gains are achieved by encapsulating and distributing directory servers with dependent vApps services that are offered in the vCloud.

The expiration of run-time and storage leases for vApp hosting directory services can lead to unexpected outages for dependent applications. A solution to vApp expiration is to create a separate vCloud Director organization that has indefinite leases to host services that should not expire.

Provide for isolation between any redundant directory servers hosted within vCloud. To avoid single points of failure it might be necessary to distribute directory servers over multiple provider virtual datacenters that do not share physical dependencies.

5.3.1.3. Single Sign-On

As IT systems proliferate to support business processes, users and system administrators face an increasingly complicated interface. Users typically have to sign on to multiple systems, with multiple sign-on dialogs that might involve different user names and authentication information. System administrators must coordinate and manage user accounts within each system to maintain the integrity of security policy enforcement.

The goal of the vCloud Director 5.1 Web Single Sign-On (SSO) feature is to simplify the sign-on process to provide an authentication service that can be used by service providers and enterprise customers.

Access control is a key security model component because it restricts unauthorized users. It is part of what is known as the *triple A process* of authentication, authorization, and accountability. Authentication systems have traditionally been based on passwords, and many organizations now use more advanced technologies such as tokens or biometrics. Some organizations enforce two-factor authentication.

Although knowing who should be authenticated serves as a basis of access control, authorization is also an issue. Authorization defines what access the user has and what capabilities are available. A vCloud administrator is normally authorized to perform more functions than an ordinary user. To control access to the end tenant's vCloud organization, limit authorization to only the required functions.

Single sign-on addresses a problem common to all service providers and enterprise customers. Various systems within the service provider and enterprise likely require the user to log on to each system with different credentials. Single sign-on addresses this problem by authenticating users once to a single authentication authority and then providing access to all other protected resources without re-authenticating. Kerberos and directory services are examples of authentication systems that can implement single sign-on. Before implementing single sign-on, consider security implications. For example, if an attacker can authenticate as a given individual, that attacker can then access multiple systems.

Compliance requires that identities be controlled. *Risk management* involves event identification, analysis, and response mechanisms faced by a service provider or enterprise. Risk management is not only a defensive operation to minimize risk effects, but it is also proactive, enabling the service provider or enterprise to take advantage of the triggering of a risk event. *Compliance* is the process of implementing procedures to meet the governance policy. Compliance requires a level of monitoring, analysis, and reporting. These elements are tied to identity management. Governance policies establish who has access to which functions in the service provider or enterprise and the conditions that are imposed on that access.

Service providers and enterprises typically request single sign-on functionality because they want end tenants to log in to their own portals and be redirected to the vCloud Director portal without re-authenticating. Service providers also encourage single sign-on because it significantly decreases administrative costs by reducing password-related tasks and support. Handling authentication can be done on a centralized basis rather than a per-application basis. Single sign-on additionally enhances security and compliance for service providers and enterprises by providing a central facility to log all system and application access.

vCloud Director and its single sign-on feature must be interoperable and work with the existing service provider and enterprise infrastructures. Providing interoperability greatly increases the use of the vCloud Director single-sign on functionality and reduces the inconvenience users experience when asked to re-authenticate.

Service providers and enterprises can offer a vCloud Director web-based portal application to enable vCloud end tenants to administer and troubleshoot identity information and perform self-service requests to add, remove, or change user roles.

5.4 vApp Deployment Readiness

vApp deployment readiness metrics should be assessed before deploying a vApp. These include vApp design considerations, vApp limitations inside vCloud, and vApp validation that should be performed before uploading to a vCloud service catalog.

5.4.1 vApp Design Considerations

A vCloud vApp differs from a vSphere vApp in the way it is instantiated and consumed in the vCloud. A vApp is a container for a distributed software solution and is the standard unit of deployment in vCloud Director. It allows power on and off operations to be defined and ordered, consists of one or more virtual machines, and can be imported or exported as an OVF package. A vCloud vApp can have additional vCloud-specific constructs such as networks and security definitions.

5.4.1.1. General Design Considerations

Some general recommendations for designing vApps include:

- Default to one vCPU unless requirements call for more (multithreaded application virtual machines).
- Always install the latest version of VMware Tools.
- Deploy virtual machines using default shares, reservations, and limit settings unless a clear requirement exists to avoid defaults.
- For virtual network adaptors, use VMXNET3 where supported.
- Secure virtual machines as you would physical machines.
- Use standard virtual machine naming conventions.

5.4.1.2. vApp and Virtual Machine Hardware Version Considerations

Virtual machine hardware version 9 is supported in vSphere 5.1. This support is carried over to vCloud Director. For maximum configuration values, see *VMware Configuration Maximums (VMware vSphere 5.1)* (http://www.vmware.com/pdf/vsphere5/r51/vsphere-51-configuration-maximums.pdf). The major use cases for using hardware version 9 are:

- Windows 8 XP mode – XP mode allows a virtualized Windows XP instance to run on Windows 8 for compatibility with older applications that do not run natively on Windows 8. Users running XP mode in Windows 8 must choose an organization virtual datacenter that is backed by a provider virtual datacenter with support for virtual hardware version 9. After adding support for virtual hardware version 9, you must also enable the *Nested HV* feature.
- 64-bit nested virtualization – Hyper-V and virtualized VMware vSphere ESXi™ nested virtualization can be helpful for non-production use cases, such as training and demonstration environments. Virtualized Hyper-V or virtualized ESXi running nested 64-bit virtual machines requires virtual hardware version 9 with the *Nested HV* feature enabled.
- CPU-intensive workloads – Running a CPU-intensive workload in a virtual machine requiring between 32 and 64 vCPUs requires virtual hardware version 9.

5.4.1.3. Network Design Considerations

A vApp network provides network connectivity to virtual machines within a vApp. Virtual machines in a vApp use an organization virtual datacenter network to connect to the outside world or to other vApps in the organization. A vApp network is backed by a network pool unless it is directly attached to an organization virtual datacenter network that is directly attached to an external network. vApp networks are created with one of the following methods:

- Dynamic – Created when a vApp is directly connected to an organization virtual datacenter network and deployed in fenced mode. There is no opportunity to use the DHCP, NAT, or firewall services at the vApp network level because this network is created automatically. It is not accessible from the vCloud UI.
- Manual – Created and either connected to an organization virtual datacenter network in NAT mode or left isolated. DHCP, NAT, or firewall service rules can be defined manually at the vApp network level as needed.

A vApp network can be directly connected to an organization virtual datacenter network, whether routed, isolated, or connected with NAT. The following are types of vApp networks:

- Direct – Virtual machines in a vApp are configured to connect directly to the organization virtual datacenter network port group and are assigned IP addresses from the organization's network range.

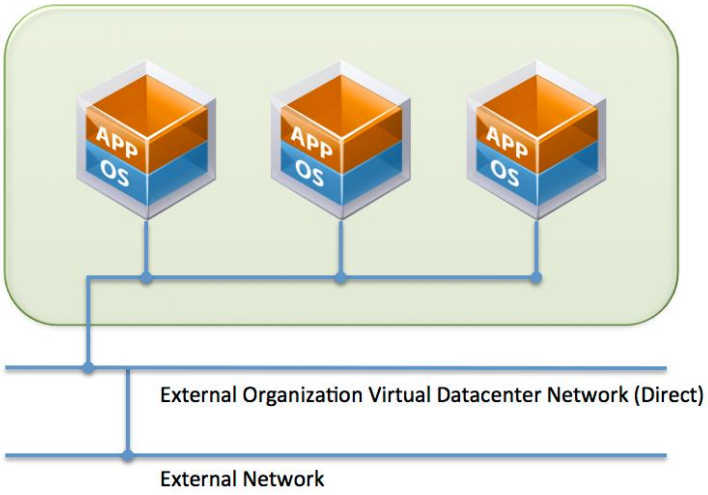
- NAT-routed – vApps are protected behind a VMware vCloud Networking and Security Edge (Edge) instance that provides NAT services for outbound and inbound access.
- Fenced – Allows identical virtual machines to exist in different vApps by isolating their MAC addresses. Fenced vApps are protected behind an Edge instance with proxy Address Resolution Protocol (ARP) capabilities.
- None – Isolated, with no external access to an organization virtual datacenter network or other vApps in the organization.

The most common vApp network configurations are described in the following sections.

5.4.1.4. Direct – External Organization Virtual Datacenter Network

Connecting a vApp to an organization virtual datacenter network that has a *direct* connection to an external network connects the vApp directly to the external network and deploys the vApp there with the external network's IP addressing. An example vApp with three virtual machines using this configuration is shown in the following figure.

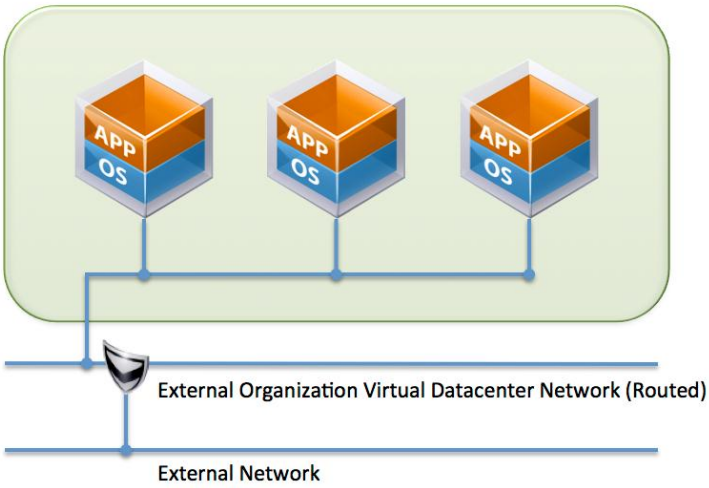
Figure 10. Direct Connection to a Directly-Connected External Organization Virtual Datacenter Network



5.4.1.5. Direct – External Organization Virtual Datacenter Network (Routed)

If the same example vApp with three virtual machines is connected to an organization virtual datacenter network that has a *routed* connection to an external network, the vApp is connected to an organization virtual datacenter network and is deployed there with the organization virtual datacenter network's IP addressing. The Edge Gateway device then provides a routed connection between the organization virtual datacenter network and the external network. This scenario is shown in the following figure.

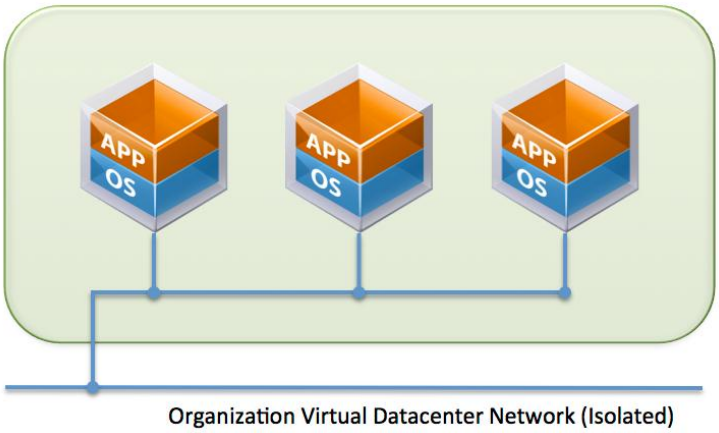
Figure 11. Direct Connection to a Routed External Organization Virtual Datacenter Network



5.4.1.6. Direct – Internal Organization Virtual Datacenter Network (Isolated)

As shown in the following figure, if the same vApp is connected directly to an *isolated* organization virtual datacenter network, the vApp is deployed there with the organization virtual datacenter network's IP addressing.

Figure 12. Direct Connection to an Isolated Internal Organization Virtual Datacenter Network



5.4.1.7. Fenced – Dynamically or Manually Created

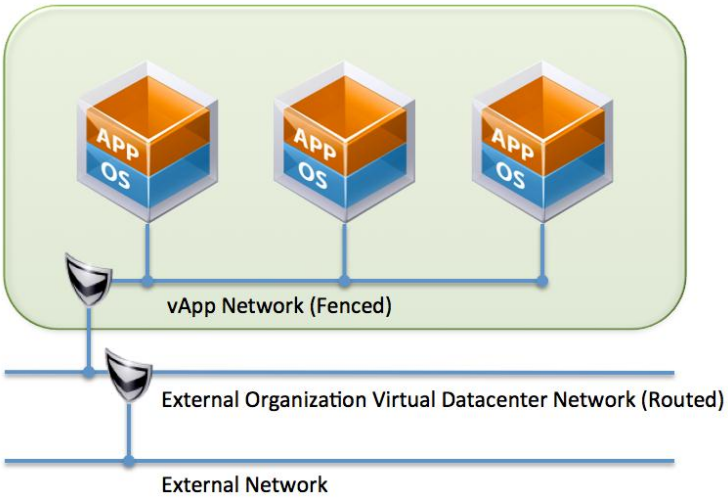
In vCloud Director, a network type is *fenced* when the virtual machines in the vApp share the same Layer 2 network as their organization virtual datacenter network. This is a special case of a NAT-routed network in which the inside and outside address of the Edge device are on the same Layer 2 network. In this mode the Edge device provides proxy ARP services to the virtual machines in the vApp.

From a vApp network perspective, depending on the type of connected organization virtual datacenter network, a NAT or double NAT might occur for incoming or outgoing traffic. The following scenarios describe a double and single NAT situation.

5.4.1.8. NAT-Routed – External Organization Virtual Datacenter Network (Routed)

If a vApp configured with a NAT-routed vApp network is connected to an external NAT-routed organization virtual datacenter network, the deployment results in a double NAT. In this scenario, the virtual machines are connected to a NAT-routed vApp network and are deployed there with the vApp network's IP addressing. The first Edge device provides NAT between the vApp network and the organization virtual datacenter network, and the second Edge device provides NAT between the organization virtual datacenter network and the external network. This scenario is shown in the following figure.

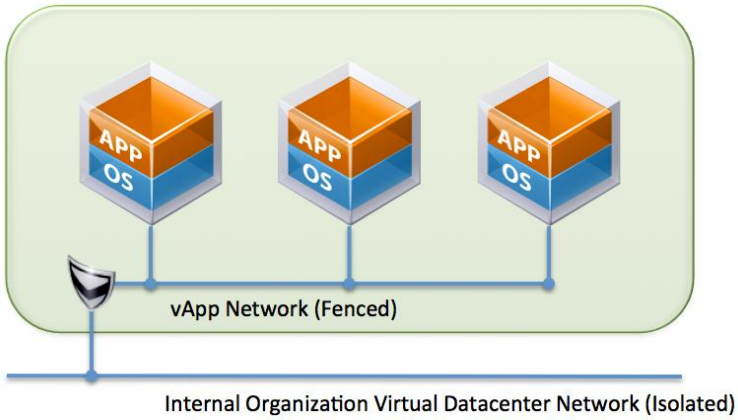
Figure 13. NAT-Routed – External Organization Virtual Datacenter Network (Routed)



5.4.1.9. NAT-Routed – Internal Organization Virtual Datacenter network (Isolated)

If the same vApp, configured with a NAT-routed vApp network, is connected to an isolated organization virtual datacenter network, the deployment results in a single NAT. The virtual machines are connected to the vApp network and deployed with the vApp network's IP addressing. The Edge device then provides NAT between the vApp network and the organization virtual datacenter network. This scenario is shown in the following figure.

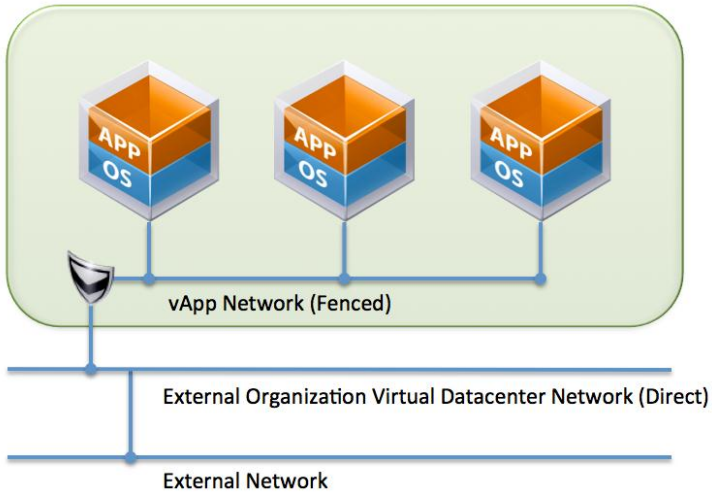
Figure 14. NAT-Routed – Internal Organization Virtual Datacenter Network (Isolated)



5.4.1.10. NAT-Routed – External Organization Virtual Datacenter Network (Direct)

The following figure shows a scenario where a vApp is configured with a NAT-routed vApp network that is connected to an external organization virtual datacenter network. The virtual machines are connected to a NAT-routed vApp network and deployed there with the vApp network IP addressing. The Edge device provides NAT between the vApp network and the external network.

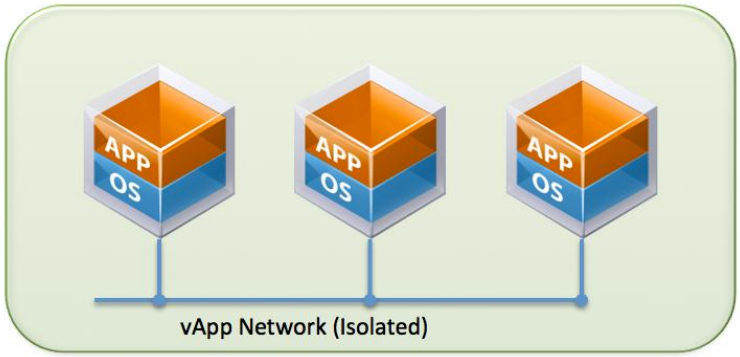
Figure 15. NAT-Routed – External Organization Virtual Datacenter Network (Direct)



5.4.1.11. Isolated vApp Network

A vApp network that is configured with no organization virtual datacenter network connectivity is completely isolated. The network is isolated at Layer 2, and no connectivity outside the vApp is possible. This configuration is usually used to build multi-tier applications.

Figure 16. Isolated vApp Network



5.4.2 vApp Limitations within vCloud

Some OVF sections are not supported by vCloud Director. Backup limitations and the nature of vCloud Director vApps render vApp backups a complex undertaking.

5.4.2.1. OVF Restrictions

Because vCloud Director does not currently support all of the OVF sections supported by vSphere, the following sections of the OVF representation of the vSphere vApp are not carried over to vCloud Director:

- `AnnotationSection`
- `DeploymentOptionSection`
- `InstallSection`
- `ProductSection`
- `ResourceAllocationSection`

All other vSphere OVF properties are parsed by vCloud Director. When a section is ignored by vCloud Director, its contents are not interpreted as they would be by vSphere, and there could be differences in behavior when you instantiate the imported vApp in a virtual datacenter. When a vApp is downloaded from vCloud Director, the OVF descriptor contains only vCloud Director-supported OVF semantics.

5.4.2.2. Backup Limitations

Using vCloud APIs for virtual machine backups poses some major limitations for backup of vApps, and the intricate nature of vApp constructs stored in the vCloud Director database requires additional overhead for backups and restores in case of a failure. vApp networks, whether fenced or NAT-routed, are backed by VMware vCloud Networking and Security Manager, making the situation even more complex.

A simpler approach might be to orchestrate vApp backups by using the *full clone* mechanism at scheduled intervals. An orchestration engine, such as VMware vCenter Orchestrator™, makes this a reasonable alternative until this capability becomes available through the native APIs or from third-party products or plug-ins.

To help customers and third-party backup vendors implement backup solutions, vCloud Director 5.1 supports snapshots of vApps and virtual machines. Snapshots provide the ability to roll back to an earlier point in time. However, snapshots are intended for use as a temporary resiliency method, not as a permanent backup, disaster recovery, or business continuity process.

One use case for enterprises and service providers is to allow users to take snapshots before they modify their virtual machines and vApps. This feature can be exposed to end users through the vCloud Director portal and through APIs in a custom portal. This can be a value-added feature for the providers, and it additionally benefits customers by reducing the administrator overhead.

Similarly, a third-party vendor can take a snapshot immediately before taking a backup of the vApp and then delete the snapshot after the backup is complete. Customers can schedule backup windows as done in a vSphere environment.

In vCloud Director, virtual machines and vApps can have only one snapshot. Any subsequent creation of snapshots overwrites the previous one. Snapshots are supported on virtual machines and vApps, both in a powered-on or powered-off state. Reverting the snapshot restores the previous state of the virtual machines in which the snapshot was taken. However, it does not restore the network mapping and OVF properties associated with the virtual machines and vApps. While a snapshot is active, the network configuration cannot be modified. An `Undeploy` operation on vApps preserves the snapshot, but cloning and capturing to the catalog operation does not. Snapshots at the vCenter layer are automatically consolidated.

The following considerations must be addressed by snapshot consumers:

- Storage consumed by snapshots.
- Storage consumed by snapshots in fast-provisioned environments.
- Snapshot management.
- Backup and recovery of vApps and virtual machines with snapshots.
- For a service provider, the costs of snapshots are readily made available to the end customer.
- For an enterprise IT provider, snapshot usage affects future storage needs.

5.4.2.3. Disaster Recovery in vCloud

In vCloud Director 5.1, disaster recovery is not integrated natively with VMware vCenter Site Recovery Manager™. VMware consultants have a workaround to help enterprises protect investments in vCloud. See the following whitepapers:

- Blog post – *Overview of Disaster Recovery in vCloud Director* (http://blogs.vmware.com/vcloud/2012/02/overview-of-disaster-recovery-in-vcloud-director.html).
- Whitepaper – *VMware vCloud Director Infrastructure Resiliency Case Study* (http://www.vmware.com/files/pdf/techpaper/vcloud-director-infrastructure-resiliency.pdf).

5.4.3 vApp Validations

vCloud APIs have no pre-built mechanism to automatically evaluate whether a vApp is ready to be uploaded to vCloud Director for consumption. One way to address this is to create a separate user, such as vApp Tester, who would be responsible for checking the validity and functionality of the vApp's readiness. The flow of events could be as follows:

1. The user, vApp Author, creates the vApp to solve a problem or business case.
2. The vApp is passed on to vApp Tester. vApp Author and vApp Tester might have access to a separate, dedicated catalog. vApp Author uploads the vApp to the catalog and vApp Tester takes it for testing and validation.
3. After vApp Tester completes the functionality testing of the vApp and obtains the expected results, someone with appropriate credentials can upload it to a public or private catalog.

This is similar to the *test* and *dev* roles and responsibilities in a standard vSphere deployment.

5.4.4 vApp Lifecycle Considerations

You must understand vApp dependencies and verify that business requirements can continue to be fulfilled throughout the lifecycle.

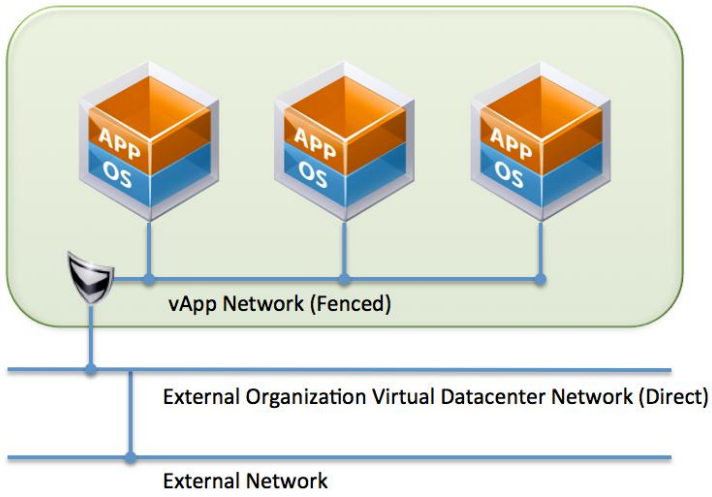
Providing Infrastructure as a Service (IaaS) requires new approaches to workload management, which can be disruptive, but there are many benefits (see *Operating a VMware vCloud* for information about proven approaches for managing vCloud environments). VMware vCloud Director supports delivery of Infrastructure as a Service.

When using vCloud Director, one common concern is the underlying network configuration implemented to support deployed vApps, in particular, how the configuration impacts network connectivity outside the vCloud and the associated impact on application dependencies.

To effectively design a new vApp, you must have a detailed understanding of application behavior and external dependencies. vCloud Director leverages the Edge networking appliance, which offers DHCP, firewall, NAT, static routing, and VPN capabilities that can be configured and managed from within vCloud Director. Various scenarios that illustrate vApp network configurations, with and without Edge, are described in Section 5.4.1.

The following figure shows an Edge appliance being used to fence a vApp so that it can be isolated, yet it uses NAT for connectivity to and from external resources. This configuration can present challenges for agent-based solutions, such as backup or antivirus updates.

Figure 17. Sample vApp Backed by a Fenced Network



In this example, private addressing is used behind the Edge appliance, and a single NAT address is used to map a single public IP address for the web-based application. Access to the vApp from outside the vCloud goes through the public IP address on port 80, which is redirected to the Web server using its internal private IP address. In some cases an application requires an agent installed on the virtual machines inside the vApp.

5.4.5 OVF Properties

When users migrate virtual machines from any virtualization environment to vCloud Director, the OVF format is required. vCloud Director imports virtual machine artifacts in the OVF format. Open Virtualization Application (OVA) format is not supported for importing vApps in vCloud Director. The difference between OVF and OVA, which is usually a tar file with the same contents as several OVF files, is that OVA can have multiple virtual machines with a single metadata file.

Users can pass custom properties when importing via OVF properties. This creates opportunities for further customization of virtual machines.

The vCloud API and the user interface support OVF properties. OVF properties can take any of the following forms:

- String
- Integer
- Boolean
- String Choice
- IP
- Custom Types

The guest operating system can use the following mechanisms to get these properties:

- ISO image mounted on the first available CD-ROM drive in the guest operating system. The properties can be read from an XML file named `ovf-env.xml` in the root directory of the mounted image. This method can be used even if VMware Tools is not installed.
- With VMware Tools installed, the guest operating system can access the properties by issuing a `vmtoolsd` query:

  ```
  vmtoolsd --cmd "info-get guestinfo.ovfEnv"
  ```

Programs or scripts executed within the guest OS can obtain these OVF properties. This provides the potential for:

- Passing in configuration parameters.
- Passing a message to a program or script, enabling dynamic behavior.
- Allowing users to select from an options list to pass in information (using the String Choice property type).

Programs or scripts running in the guest operating system can query OVF properties at any time. One use case runs a script when the guest operating system boots. This script obtains the OVF properties and follows a set of decision paths to configure the system based on these values.

Another use case executes an automatically scheduled script or program to query the OVF properties at known intervals. If OVF properties change between iterations, the scheduled script or program can alter its behavior as the message changes.

As with the vSphere API Guest Operations (formerly the VIX API), the guest operating system does not need to be connected to a network to enable OVF properties to be read. The properties are made available to the guest OS with VMware Tools or as a file contained in an ISO image mounted on a CD-ROM drive.

The vCloud API provides calls to get and set OVF properties. It can be used to query vApp templates and instantiated vApp objects to obtain `ProductSections` that contain the `ProductSectionList`. When using the vCloud API to add, update, or remove a property, you must supply a `ProductSectionList` object.

With the vCloud Director UI, you can set OVF property values when you add a new vApp. To set properties on an instantiated vApp, vCloud Director provides tabs to configure custom properties at the vApp and virtual machine levels.

For more information on using the properties, see *Leveraging vApp & VM Custom Properties In vCloud Director* (http://blogs.vmware.com/vsphere/2012/06/leveraging-vapp-vm-custom-properties-in-vcloud-director.html).

5.4.6 OVF Package Upload Latency Considerations

The transfer for OVF files from client devices to a vCloud Director cell typically occurs in a WAN environment where bandwidth and network speed are limited. Considerations for transferring OVF files are discussed in more detail in the *VMware vCloud Director 1.0 Performance and Best Practices* (http://www.vmware.com/files/pdf/techpaper/VMW-Performance-vCloud-Director-1-0.pdf).

5.4.7 Relocate Existing vApps

You can choose between two primary methods for migrating a vApp between datastores within an organization virtual datacenter:

- vCloud Director-initiated vApp relocate.
- VMware vSphere Storage vMotion®.

Choose the method that best fits your case after considering storage efficiency and provisioning time.

5.4.7.1. Prerequisites

The following are prerequisites for moving a vApp between virtual datacenters:

- The target datastore and vApp must exist within the same organization virtual datacenter.
- All virtual disks for an individual virtual machine must reside within a single datastore.

Note Capacity requirements to store a virtual machine might be affected by storage configuration policies, and this might impact fast provisioning.

5.4.7.2. Use Cases

There are several potential use cases for the migration of vApps between organization virtual datacenters:

- Planned maintenance of underlying storage.
- Rebalancing storage utilization for capacity or performance management.
- Upgrading or replacing a back end storage system.

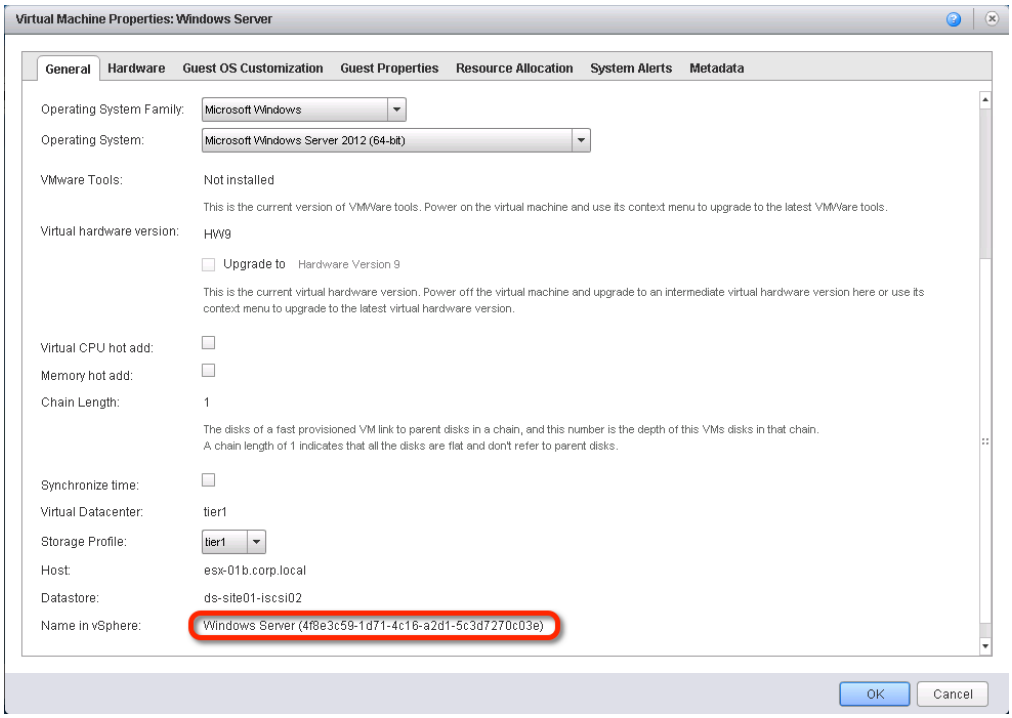
5.4.7.3. Storage vMotion Procedure

Storage vMotion can be performed from the vSphere Client or through scripts that leverage vCenter procedures to complete the task. The following is a procedure for the migration process as performed from the vSphere Client.

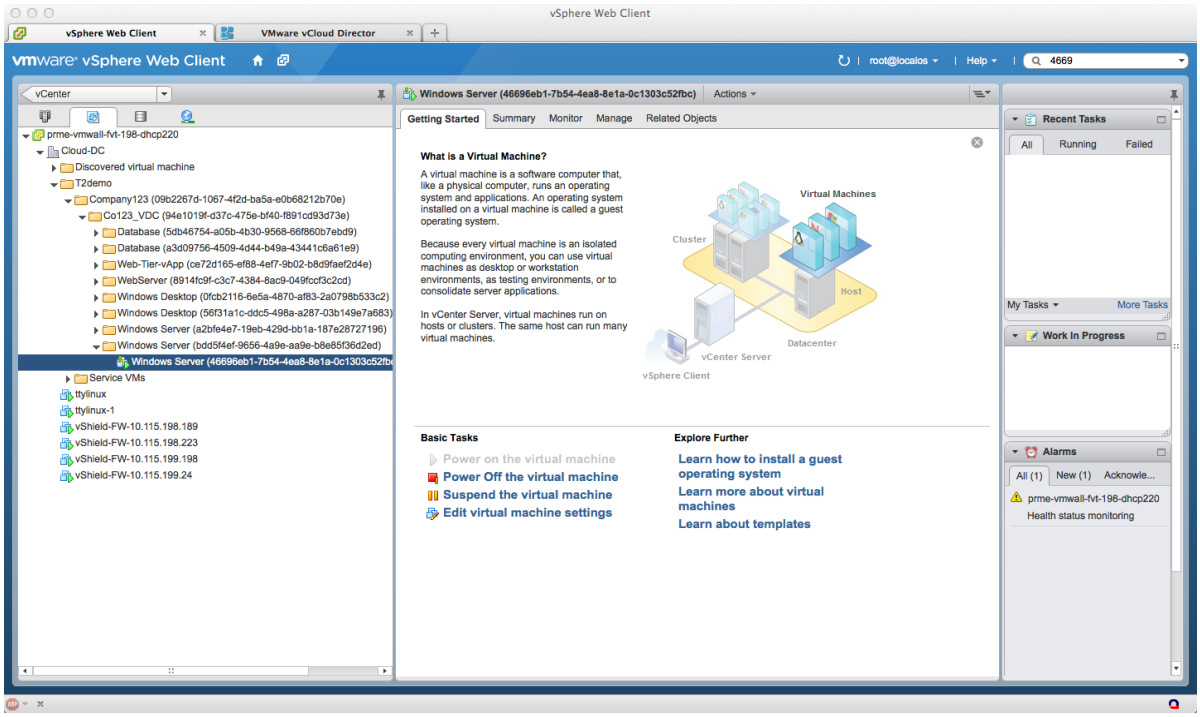
To migrate a vApp

1. Log in to the vCloud Director portal as a System Administrator.
2. Find the organization that contains the vApp.
3. Open the organization.
4. Find and open the vApp.
5. Click the **Virtual Machines** tab.
6. Right-click the virtual machine to display the **vApp VM Menu**.
7. Select **Properties**.
8. Note the value of the **Name in vSphere** property of the virtual machine. Use this value to find the virtual machine in vSphere. The UUID portion is unique, but names might be duplicated within an environment.

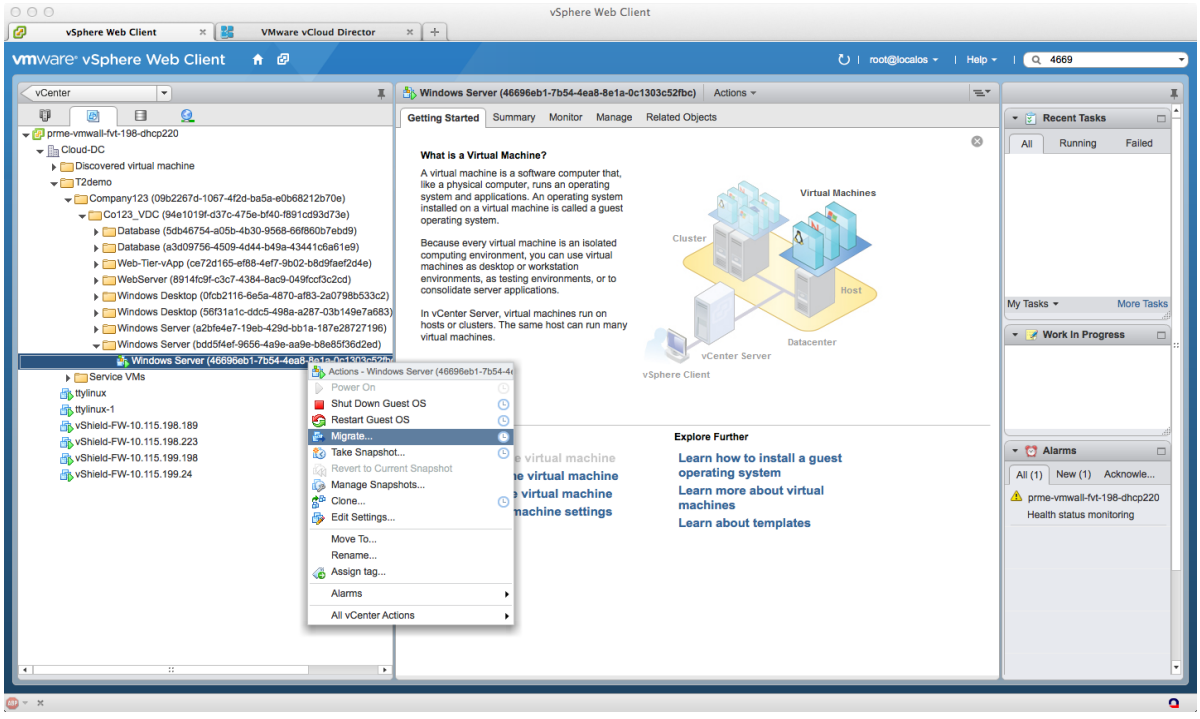
   Note If this section is not visible, the logged-in user does not have vCloud Director System Administrator privileges.



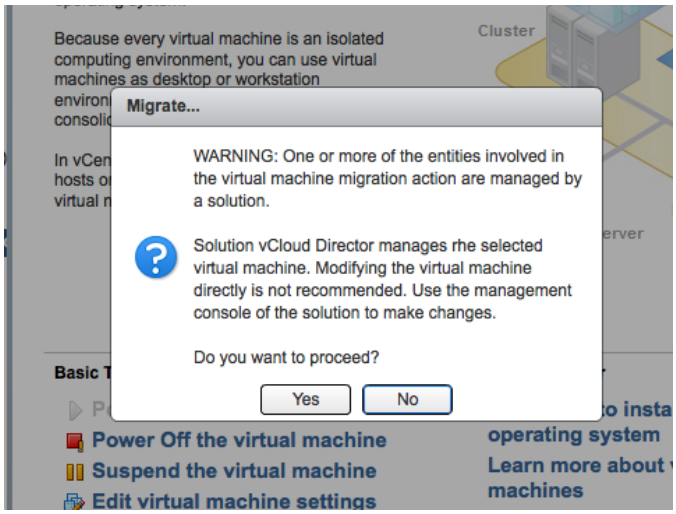
9. Using the vSphere Client, connect to the vCenter server that manages the host specified in the virtual machine properties.
10. Type the UUID of the target virtual machine in the inventory search panel.



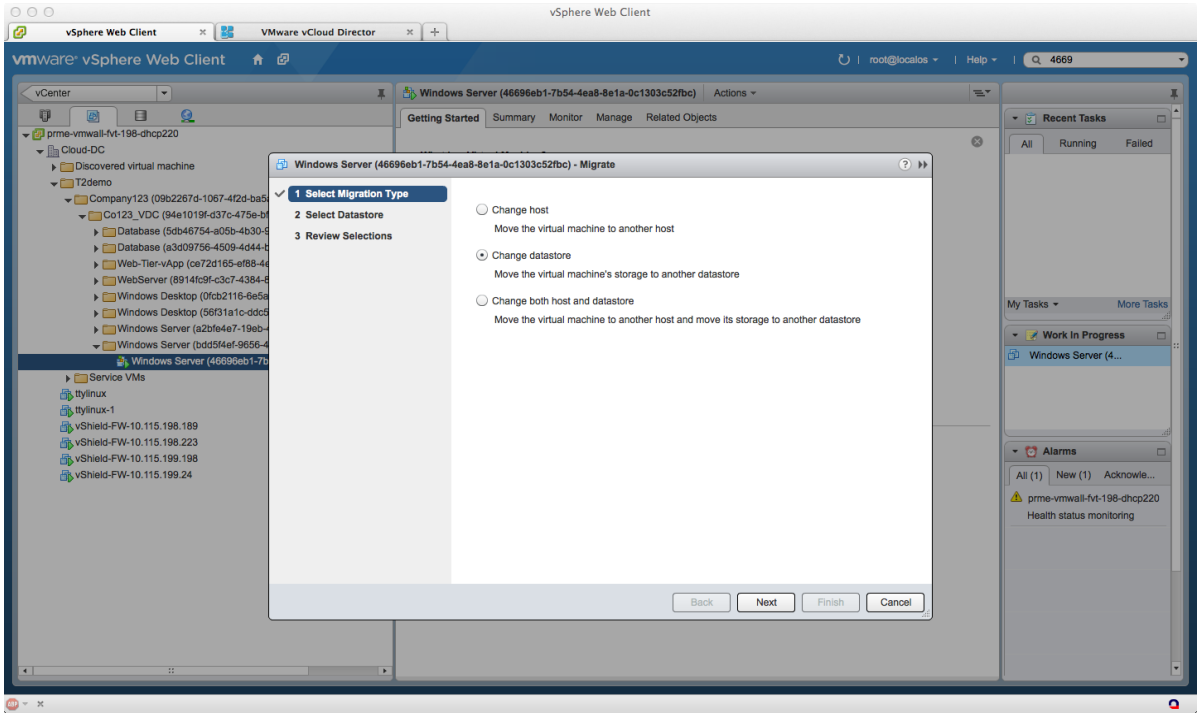
11. Locate the virtual machine in the results list.
12. Select the virtual machine.
13. Click **Inventory** > **Virtual Machine**.
14. Select **Migrate**.



15. If running vSphere 5 or later, read the Warning dialog. Click **Yes** to continue.



16. Select the **Change datastore** option.



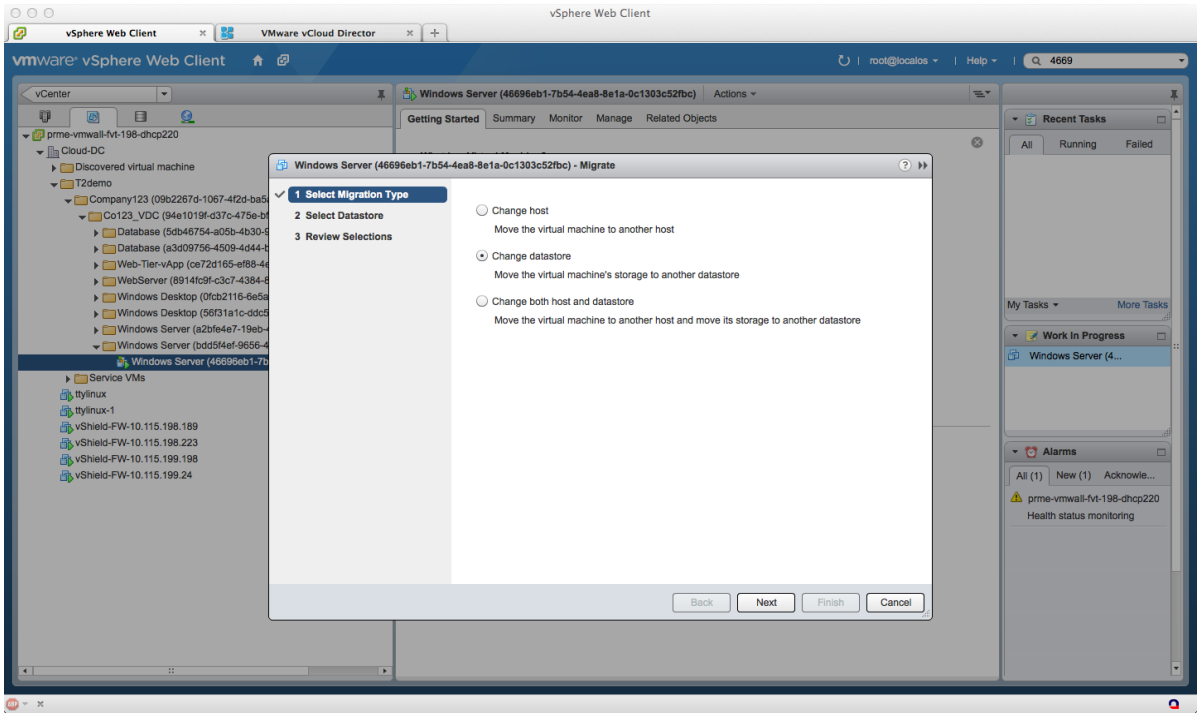
17. Click **Next** to continue.

18. Do not modify the currently assigned default resource pool selection.

    Caution Never use the vSphere Client to modify the virtual machine resource pool. If using a vSphere version earlier than vSphere 5, you must manually select the correct resource pool. The selected resource pool must match the current resource pool that contains the virtual machine.

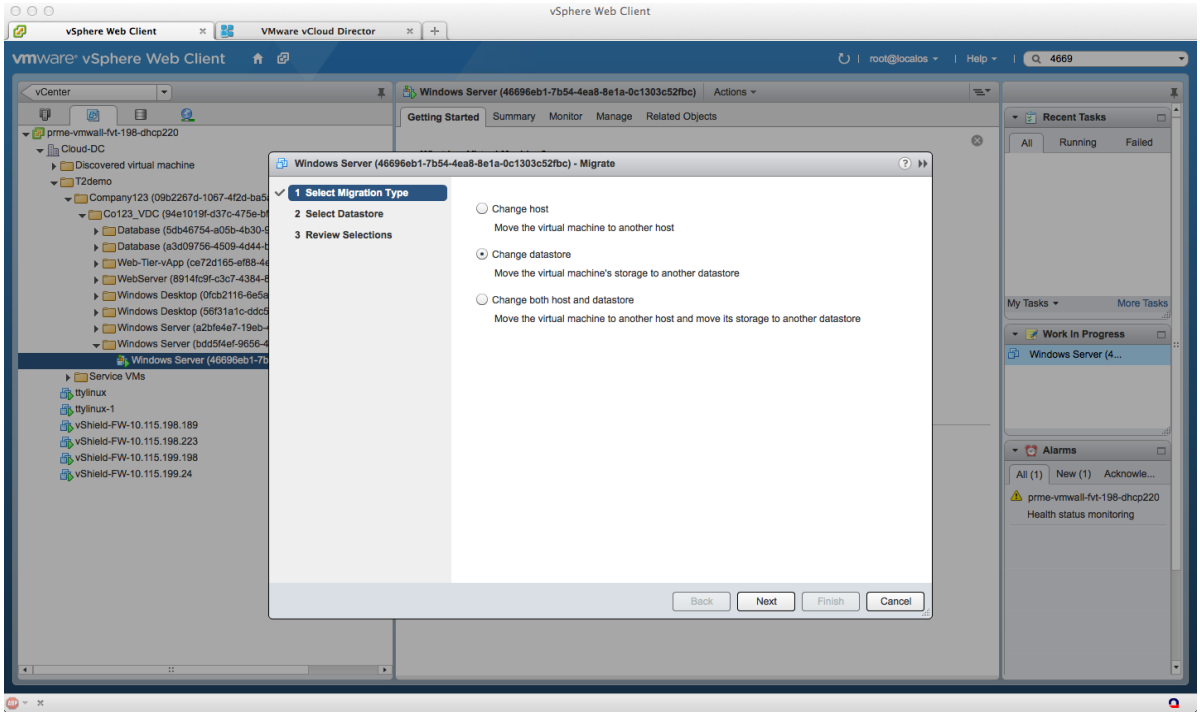
    Note Configurations that have a single virtual machine using multiple virtual disks distributed over more than one datastore are not supported. All virtual disks supporting any virtual machine must be contained within a single datastore.

19. Select the target datastore from the list, and click **Next** to continue.



Before proceeding, confirm that the target datastore is a member of the virtual datacenter that contains the vApp within vCloud Director.

20. Click **Finish** to start the Storage vMotion process.



21. Continue to monitor event status for progress. After the storage migration is complete, the new datastore name is displayed in the virtual machine properties.

5.4.7.4. Risks

Configuration changes at the vSphere layer do not propagate to vCloud Director. The following risks are associated with migrating vApps and virtual machines at the vSphere layer without vCloud Director involvement.

- If the virtual machine is migrated to datastores that are not assigned to the appropriate virtual datacenter management, anomalies might occur.
- If the resource pool assignment of the virtual machine is changed in the vSphere Client, abnormal behavior might result.

Changes made in vCloud Director are stored in the database. To mitigate the risks, use Representational State Transfer (REST) APIs with integrated fail-safes.

5.4.7.5. Impact

The virtual machine is relocated as requested between datastores. Any services provided by the virtual machine are unavailable while it is stopped or suspended during the relocate event. The relocate process is performed as a clone operation. To minimize risk, it is followed by a delete operation.

5.4.8 Moving a vApp Between Organization Virtual Datacenters

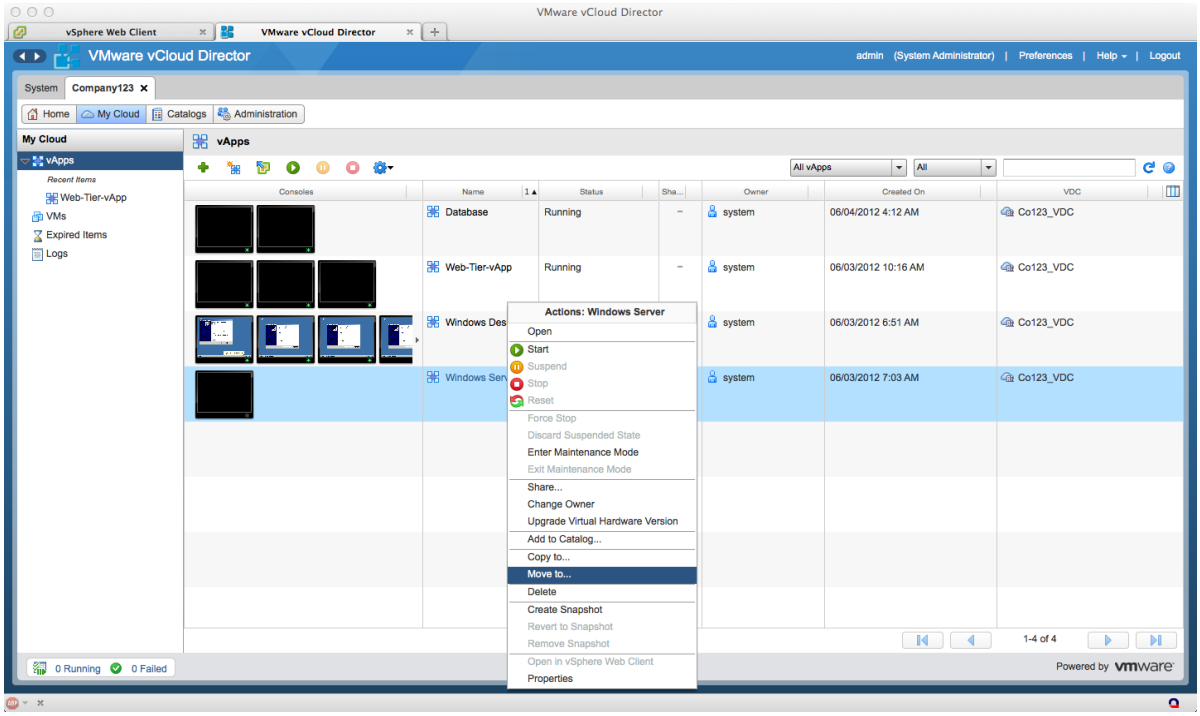
The following are potential use cases for vApp migration between organization virtual datacenters:

- vApp lifecycle management.
- vApp performance management.
- Virtual datacenter resource management.

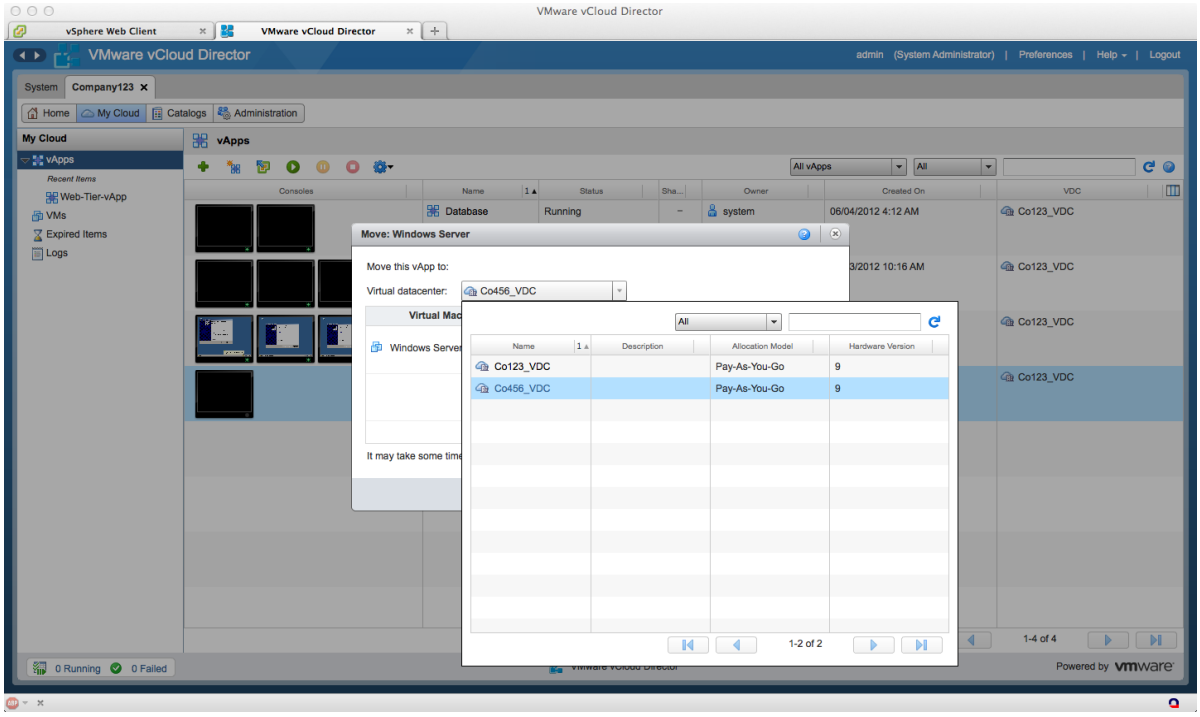
For lifecycle management, processes can be developed to migrate a vApp between virtual datacenters. These processes serve different purposes during vApp development and promotion cycles. For example, a vApp might be initially provisioned within a development virtual datacenter. At the end of the development cycle, the vApp can be migrated or copied to a quality assurance and testing virtual datacenter. In the final stage, the vApp can be copied to the catalog for publication to users.

To migrate a vApp between virtual datacenters

1. Log in as a user with the appropriate permissions.
2. Click the **My Cloud** tab.
3. Locate the vApp that requires migration.
4. If the vApp is running, stop or suspend the vApp by right-clicking and selecting **Stop** or **Suspend**.
5. After the vApp has been selected and stopped or suspended, right-click and select **Move to**.



6. When prompted, select the target virtual datacenter from the drop-down menu, and click **OK**.



7. Wait for the copy and delete operations to finish, and verify that the vApp is located in the target virtual datacenter.
8. Following migration, verify the networking requirements within the target virtual datacenter. Virtual datacenters are often defined based on physical vSphere resources, which can have networking configurations that differ from the source networks.

During the migration, virtual machines contained within the vApp are copied to new datastores. Copy time varies depending on whether the source or target organization virtual datacenter has been prepared for fast provisioning.

5.5 Updating vApps

You can change or reset the virtual machine MAC addresses, the CPU and memory values, and the number of vNICs. The following section describes how to change the MAC address.

5.5.1 Changing a Virtual Machine MAC Address

You can change the MAC address assigned to a virtual machine in a vApp. vCloud Director assigns a MAC address to all deployed vApps and virtual machines. Perform any modifications to assigned MAC addresses through the vCloud Director portal, API, or tools that use the API. If permitted on the virtual switch, you can alternatively edit the MAC address in the guest operating system.

5.5.1.1. Prerequisites

The following are requirements for modifying the MAC address of a virtual machine:

- The virtual machine must be in a powered-off state.
- The user performing the action must have full control rights to the virtual machine.

5.5.1.2. Use Cases

Use cases that might require changing the MAC address of a virtual machine include:

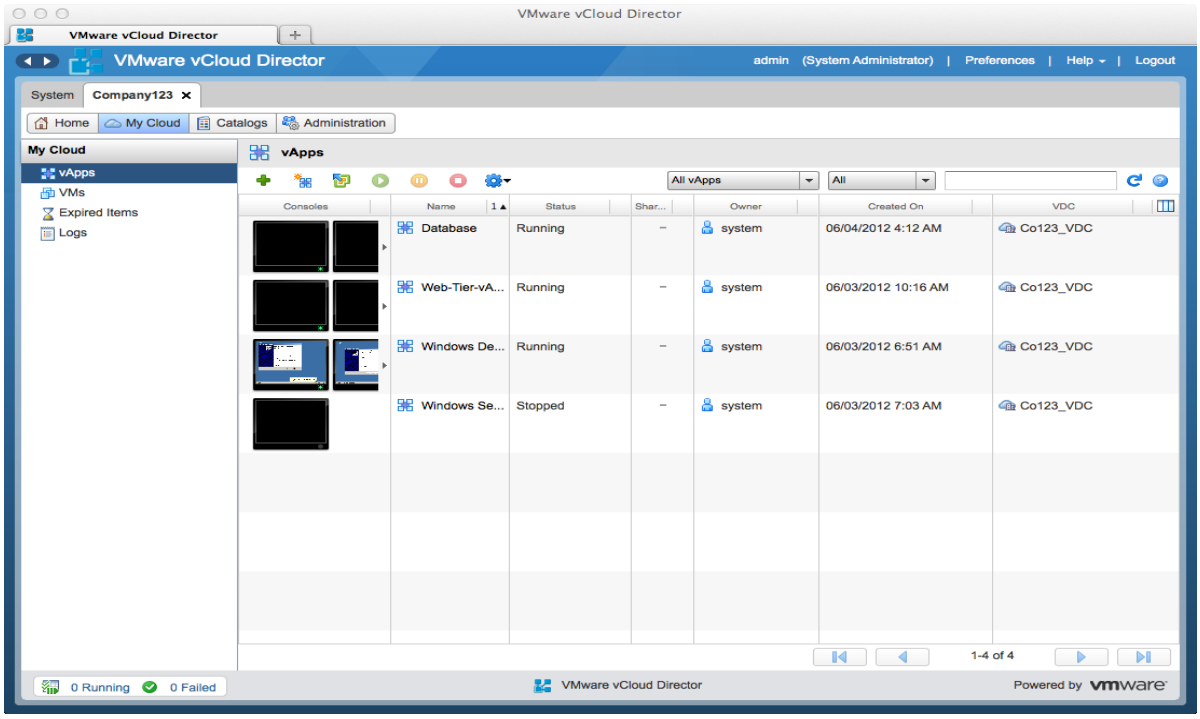
- Forcing new, unique MAC address creation for guest software licensing requirements.
- Preventing duplication of MAC addresses within a single network when virtual machine network connections are relocated during the lifecycle.
- Assigning a MAC address based on the current vCloud Director environment for imported virtual machines.

5.5.1.3. vCloud Director MAC Address Reset Procedure

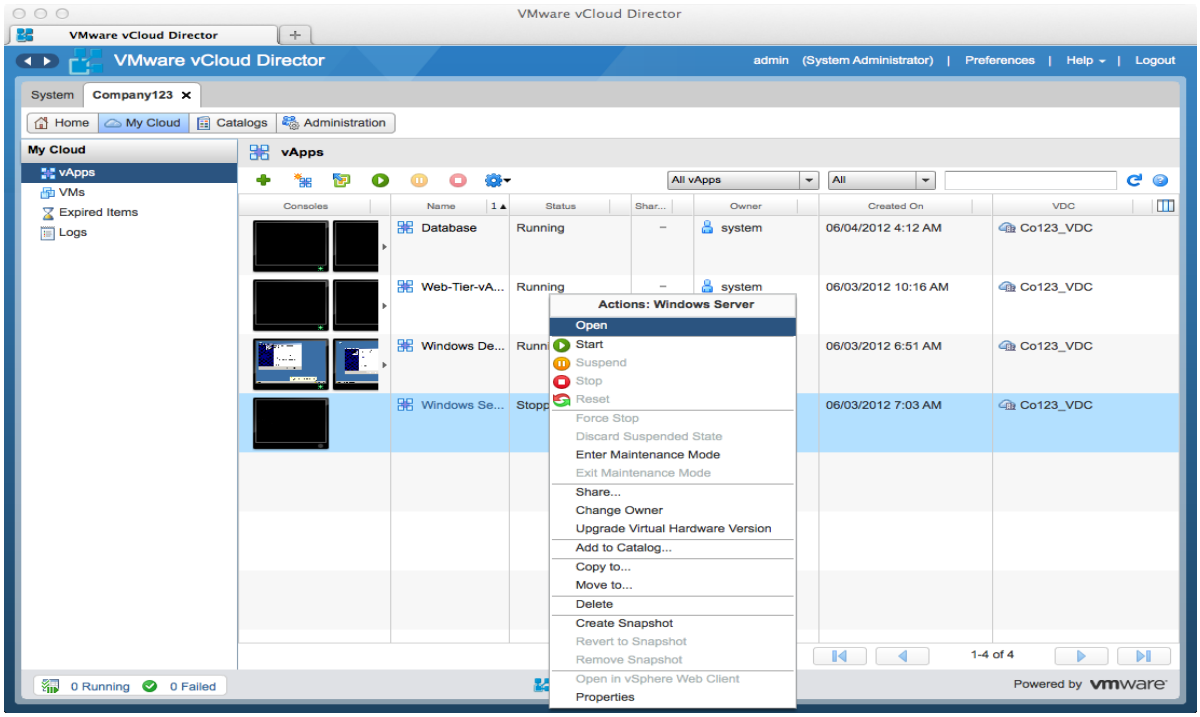
A new MAC address can be generated quickly for any virtual machine network interface within the vCloud Director portal.

To generate a new MAC address for a virtual machine network interface

1. Log in to the vCloud Director portal as a user with rights to target the vApp virtual machine.

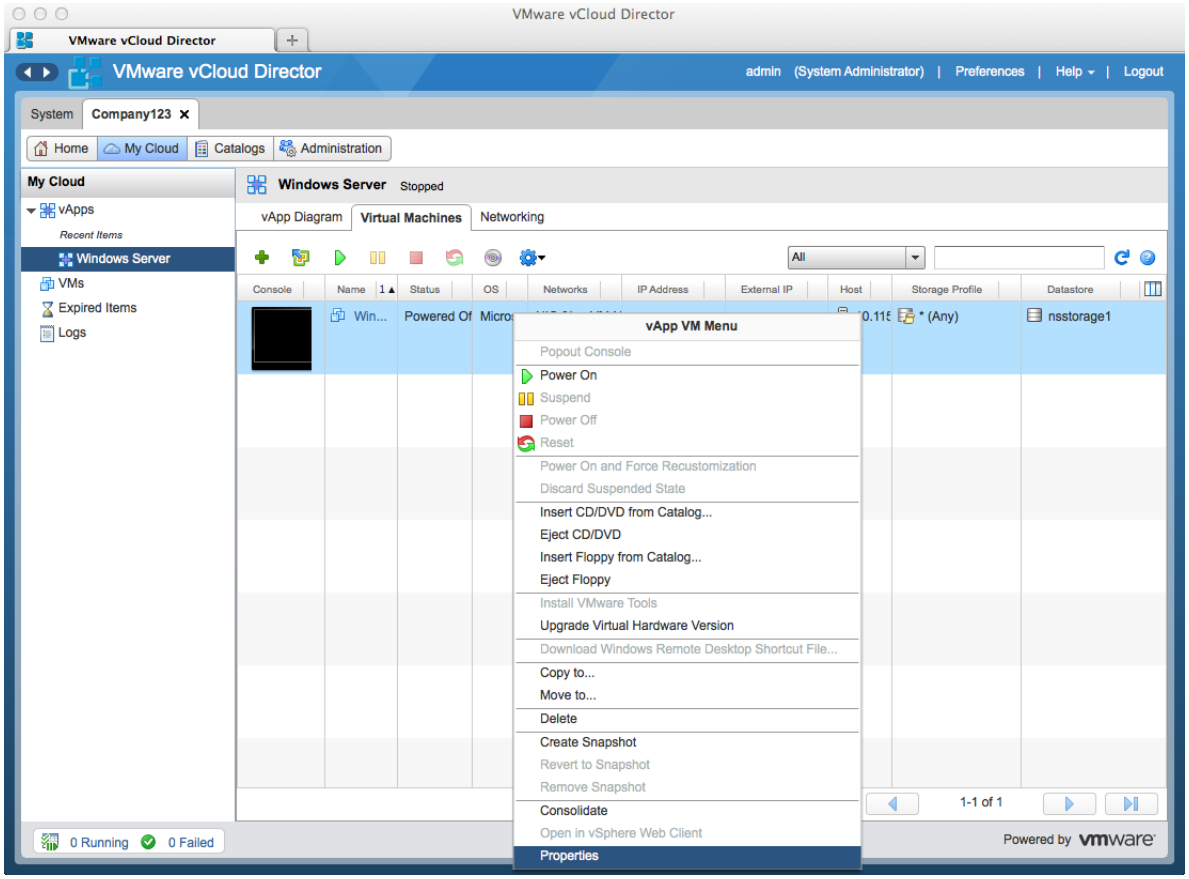


2. Navigate to the target vApp.
3. Right-click, and select **Open**.



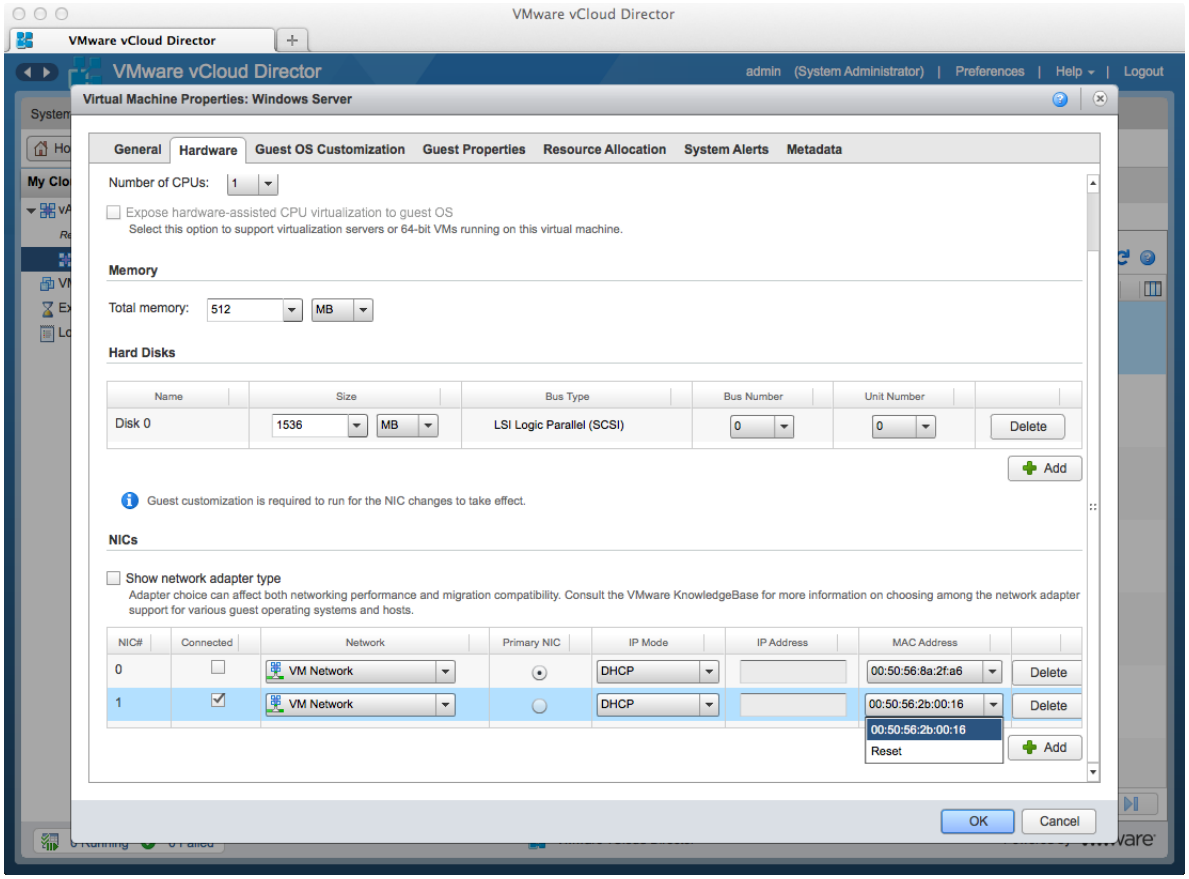
4. Open the vApp, and select the target virtual machine.

5. Right-click, and select **Properties**.



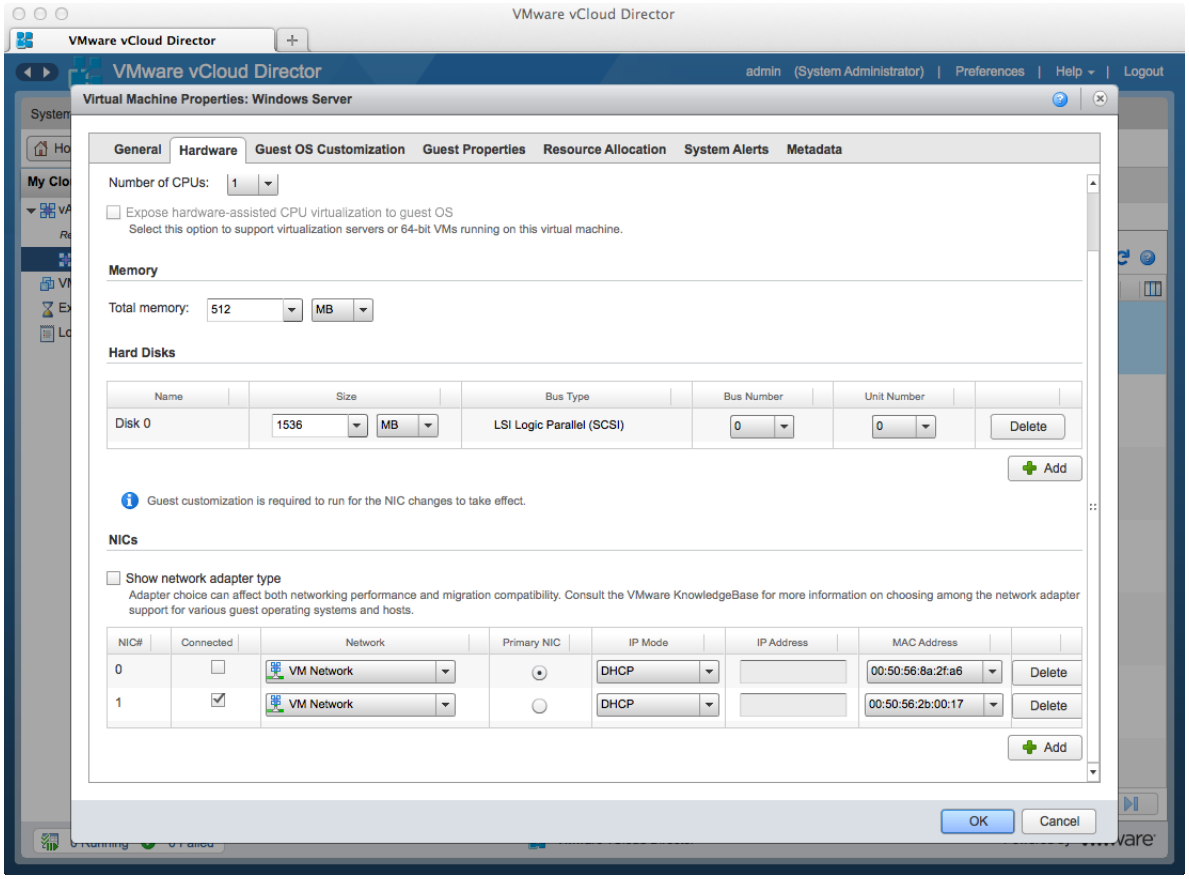
6. Switch to the **Hardware** tab, scrolling down to the **NICs** section, if necessary.
7. Determine which network interfaces require a new MAC address.
8. Click the drop-down menu with the MAC address.

9. Select **Reset** for each NIC requiring a new MAC address.



10. Select **OK** to activate the changes. vCloud Director assigns and configures a new MAC address within the virtual machine hardware.

11. Verify your changes by inspecting the **Virtual Machine Properties NICs** panel.



5.5.2 Impact and Risks

The MAC address of the virtual machine interface selected for MAC regeneration has a new MAC address allocated based on the identity of the vCloud Director instance.

Applications and services that depend on MAC address registration might be impacted. This includes DHCP reservations and MAC address-aware software licensing.

5.6 Establishing Service Levels

Service Level Agreements (SLAs) between consumers and providers are needed to align business requirements with offered services. This becomes increasingly important as business-critical applications move to a vCloud-based environment. Providers can be enterprises or public vCloud service providers. This section provides guidance for consumers.

5.6.1 Defining a Service Level Agreement

For a service provider to optimally provide vCloud consumers with services, technology services must be seamlessly integrated. Generally, a workflow engine called the *orchestration layer* is used for this purpose. Invoking a service can automatically trigger one or more technology services. The rules governing these workflows must be pre-configured and pre-approved for control, and they must provide an agreed-upon level of service to the consumer. This agreed-upon level of service is known as a *Service Level Agreement* and is typically defined as a pre-determined agreement between the service consumer and the service provider that measures the quality and performance of available services.

5.6.2 vCloud Layers and Service Level Agreements

A typical vCloud-computing environment consists of multiple layers, such as IaaS, Platform as a Service (PaaS), and Software as a Service (SaaS). Consumers, based on their business requirements, have a number of options for implementing the vCloud stack, including:

- Creating a private vCloud.
- Using a public vCloud provider.
- Adopting a hybrid vCloud model in which both private and public vCloud resources are used.

The ability of an organization to guarantee availability and performance at every vCloud layer enables this flexibility. Flexibility is achieved by establishing SLAs externally with service providers or by creating SLAs internally for a private vCloud.

For an organization with an IaaS layer that is hosted by a public vCloud provider but where PaaS and SaaS layers are maintained internally, the following SLAs might apply:

- IaaS layer:
  - Uptime and Availability SLA – Established with the external vCloud service provider.
  - Network Performance SLA – Established with the external vCloud service provider.
  - Request Fulfillment SLA – Measure of response time for provisioning and access configuration requests.
  - Restore Time SLA.
- PaaS layer:
  - Uptime and Availability SLA – For development environment.
  - Uptime and Availability SLA – For critical development environment components.
  - Restore Time SLA – For development environment.
- SaaS layer:
  - Uptime and Availability SLA – Specific to an application.
  - Application Response Time SLA – Measure of how the application performs for the business users.
  - Time to Resolution SLA – Time to recover an application in case of failure.

5.6.3 vCloud Considerations for Service Level Agreements

SLAs are required to provide efficiency and accountability at every layer for both external providers and internal IT groups. Managing SLAs in every layer helps to isolate systemic problems and eliminate delays.

There are inter-relationships between SLAs at different vCloud layers, and any change in quality of the service or breach of one of the lower vCloud layers might impact multiple SLAs at a higher vCloud layer. Using the previous example of an organization with an IaaS layer hosted by a public vCloud provider and PaaS and SaaS layers maintained internally, an SLA performance breach resulting in external vCloud providers unable to support operating system performance needs would propagate problems to the SaaS layer, decreasing application performance and response time for business users.

Business needs are continuously evolving, as are business requirements for a vCloud. SLAs must be continually updated to reflect current business requirements. In the previous example, adding a thousand new business users to a particular application is likely to increase the criticality and cause the application to be deemed mission-critical. This business change means that corresponding SLAs supporting the application might require revision for increased uptime and availability. This, in turn, leads to increased demands at the IaaS layer, requiring revisions to SLAs with an external IaaS provider as well.

6. Consuming vCloud with the API

The vCloud API is used to interact with vCloud Director from outside its native portal. The VMware vCloud API provides support for developers who build interactive clients of VMware vCloud Director using a RESTful application development style. The vCloud API clients and servers communicate over HTTP, exchanging representations of vCloud objects. These representations take the form of XML elements. HTTP GET requests are used to retrieve the current representation of an object, HTTP POST and PUT requests are used to create or modify an object, and HTTP DELETE requests are typically used to delete an object.

The VMware vCloud API was designed to be supported by the ecosystem of vCloud-ready service providers and to interact with what are known as *pure virtual* resources. Any IaaS cloud architecture consists of raw hardware resources abstracted and presented to end users in the form of vCloud resources. For example, in a VMware vCloud environment, networking and datastore resources are connected to vSphere hosts in the form of switches, port groups, and storage in datastores and datastore clusters. When you build a cloud infrastructure, network and storage configuration should be transparent to the end user.

6.1 Characteristics of the API

Depending upon desired operations and the assigned privileges, there are specific APIs available.

6.1.1 Self Service APIs

These are typical operations performed by an end user who consumes cloud resources:

- Creating vApps and virtual machines.
- Power operations on vApps and virtual machines.
- Listing of resources available to the user.
- Managing the vApp lifecycle.

Typically users never see the administration side of the vCloud API and do not have permission to use it. They use the vCloud API for automated tasks and the graphical vCloud interface for daily tasks and operations. Independent software vendor (ISV) partners might write workflow applications to take advantage of reduced privileges and provide an interface to their customers on their custom portal. This prevents breaches from escalating to administrator privileges. The roles defined for users in the vCloud Director GUI are the same as those in the vCloud API when the same logins are used.

6.1.2 Administrative APIs

Administration APIs have higher privileges and access to the following vCloud Director management functions:

- Creating, updating, and deleting virtual datacenters.
- Creating, updating, and deleting organizations.
- Creating, updating, and deleting networks.

6.2 API Functions

Some of the most important API functions include the following:

- Authentication Operations – These API functions provide basic authentication over TLS/SSL. They also provide information about supported versions of the API.
- Resource Navigation Operations – APIs are REST-based. Each resource is a named URI that is used by the HTTP or HTTPS protocol to access and act upon resources, such as a vApp.
- Long-Running Operations – A task that takes a longer time to run, such as deploying a vApp with many virtual machines. The status of these tasks is modeled as a resource, so the status of the task can be retrieved like other resources using a URI.
- Error Reporting – One of the key functions of the API is to report any errors that occur in vCloud Director.

6.3 What's New in the vCloud 5.1 API

vCloud Director 5.1 provides the following new and expanded features, and APIs have been updated accordingly:

- Query Service – Introduced in the vCloud 1.5 API. Useful when searching for resources. You do not have to traverse the resource tree to find specific children nodes. The query service has been expanded in the vCloud 5.1 API including the ability to query for metadata tags.
- Metadata Tagging – Expanded in the vCloud 5.1 API. Annotate vCloud resources with typed metadata that can be system defined and either hidden from or read-only for users, or user defined that is read writable. Consumers can interact with metadata from the UI in vCloud Director 5.1; in previous releases, this feature is available only from the API.
- API Extensions – New in the vCloud 5.1 API. Provides the ability to extend the vCloud API to modify existing functionality or to add new capabilities to the API.
- Block Tasks and Notifications – Introduced in the vCloud 1.5 API. This feature relates to the messages published and consumed by vCloud Director. In blocking tasks, the system waits for a user to take an action. For example, if a manager must provide approval when a developer requests vCloud resources, a blocking task might be created. When a developer deploys a vApp, the manager is notified and must approve before proceeding with deployment.
- In notifications, also called *non-blocking events*, a message is sent to the event's Advanced Message Queuing Protocol (AMQP) broker.

6.4 vCloud SDK

APIs involve creation of XML payloads to post, and parsing of an XML response. The SDK simplifies the development process. Currently, Java, .NET, and PHP have supported language bindings.

These language bindings can be downloaded from VMware vCloud API (http://www.vmware.com/go/vcloudapi).

7. Consuming vCloud with vFabric Application Director

Cloud computing is aggressively driving efficiencies in processes, compliance, and innovation in applications, and optimizing the infrastructure on which they run. Requirements vary according to who consumes the services. Application owners want the simplicity and agility to deploy platforms for development in the cloud. The platforms are fully configured environments that meet internal enterprise compliance requirements. Infrastructure owners want the ability to create self-service models for application owners.

Virtualization has matured to a point where virtual machines can be deployment in a matter of hours where it once took weeks to provision physical servers. When all levels of control are in place, vCloud Director provisions solutions in minutes with self-service portal capabilities and integration with third-party provisioning systems using vCloud APIs.

vFabric Application Director furthers the deployment of applications in the cloud. Major challenges that vFabric Application Director addresses are the following:

- Legacy deployments.
- Proliferation of virtual machine templates and customization scripts.
- Disconnected application operations.

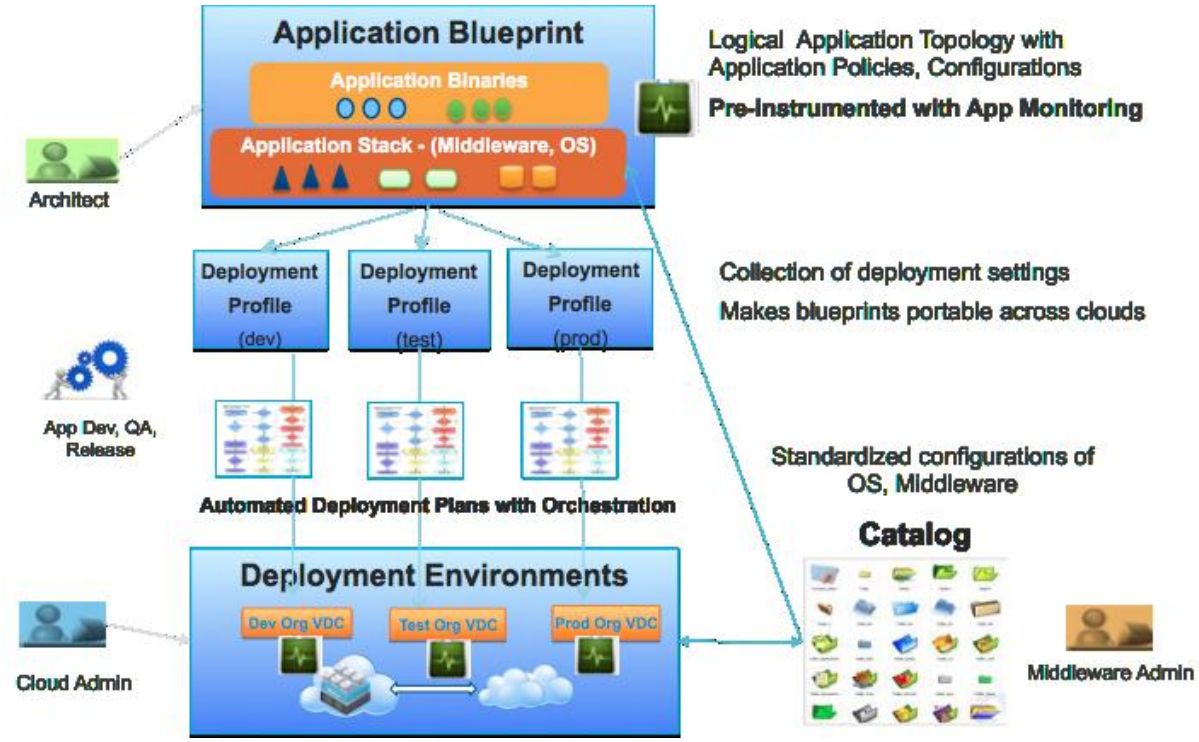
In legacy deployments, there is little or no automation. Complex dependencies might or might not be abstracted. Maintaining those abstractions can increase overhead. The deployments might not be cloud aware, might have tighter dependency on the cloud infrastructure, and might not be flexible enough to meet developer needs in timely manner.

Permutation and combination of application and guest operating system versions in vApp templates have increased. This introduces enormous challenges on administrators to standardize configurations, security, and compliance.

Applications after their deployments are not automatically discovered by monitoring systems. This leads to long troubleshooting cycles, and capacity planning processes with little or no auto-scaling capabilities for applications.

vFabric Application Director addresses many of these shortcomings. It has an architecture, model-based application deployment, integrated active application monitoring and management, and provisioning to public, private, and hybrid clouds**.**

Figure 18. vFabric Application Director



8. References

- *VMware vSphere Documentation*
  https://www.vmware.com/support/pubs/vsphere-esxi-vcenter-server-pubs.html
- *VMware vCloud Director Documentation*
  https://www.vmware.com/support/pubs/vcd_pubs.html
- *Virtualizing existing domain controllers*
  http://kb.vmware.com/kb/1006996
- *FSMO placement and optimization on Active Directory domain controllers*
  http://support.microsoft.com/kb/223346
- *Active Directory Replication over Firewalls*
  http://technet.microsoft.com/en-us/library/bb727063.aspx
- *How to configure a firewall for domains and trusts*
  http://support.microsoft.com/kb/179442
- *Virtualizing a Windows Active Directory Domain Infrastructure*
  http://www.vmware.com/files/pdf/Virtualizing_Windows_Active_Directory.pdf
- *VMware vCloud Director Security Hardening Guide*
  http://www.vmware.com/files/pdf/techpaper/VMW_10Q3_WP_vCloud_Director_Security.pdf
- *VMware vCloud Director 1.0 Performance and Best Practices*
  http://www.vmware.com/files/pdf/techpaper/VMW-Performance-vCloud-Director-1-0.pdf
- *Active Directory Branch Office Guide Series*
  http://technet.microsoft.com/en-us/library/cc749926.aspx
  http://www.microsoft.com/downloads/details.aspx?FamilyId=9A4C7AC3-185E-4644-9E98-4876B2A477E7&displaylang=en
- *Description of support boundaries for Active Directory over NAT*
  http://support.microsoft.com/kb/978772

Third-party URLs are subject to changes that VMware cannot control. You might be able to locate a third-party document by searching from the their home page.