VirtualCenter 2.x

Running VirtualCenter in a Virtual Machine

Running VirtualCenter in a virtual machine is fully supported by VMware to the same degree as if it were installed on a physical server. There are several reasons why deploying VirtualCenter in a virtual machine would be advantageous:

- Server Consolidation: instead of dedicating an entire physical server to VirtualCenter, you can run it in a virtual machine along with others on the same ESX Server host.
- Mobility: by encapsulating the VirtualCenter server in a virtual machine, you can transfer it from one host to another, enabling maintenance and other activities.
- Snapshots: A snapshot of the VirtualCenter virtual machine can be used for backup, archiving, and other similar purposes.
- Availability: using VMware HA, you can provide high availability for the VirtualCenter server

The purpose of this paper is to provide guidelines on how to deploy VirtualCenter in a virtual machine. The rest of this paper is organized as follows
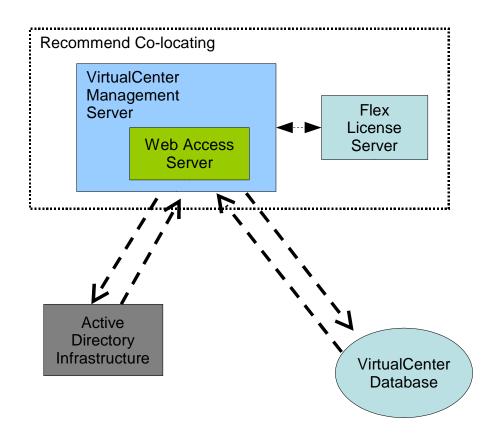
VirtualCenter Architecture

A VirtualCenter deployment can be thought of as consisting of four components, as shown in Figure 1.

- VirtualCenter Management Server — The main component which manages the VMware Infrastructure. It implicitly includes the Web Access server component
- Flex License Server — An independent process which services all licensing requests
- Active Directory infrastructure — Used by VirtualCenter for authentication
- VirtualCenter Database — Stores all state information about the environment, including clusters, resource pools, folders, roles, audit history, and performance & monitoring statistics

Figure 1: VirtualCenter Architecture

The VirtualCenter database communicates with the VirtualCenter Management Server via a standard ODBC connection. The Active Directory infrastructure is a part of the datacenter environment, and also communicates with the VirtualCenter server via a network-based API. Since both of these are only loosely-coupled with the VirtualCenter server, their deployment is independent of the VirtualCenter server and is not considered any further in this paper.

The other two components, the VirtualCenter Management Server and the Flex License Server, can be deployed in a virtual machine. Although they can be installed on separate virtual machines, it is recommended that the Flex License server be treated as an embedded component, similar to the Web Access server, and co-located on the same virtual machine as the VirtualCenter server. This makes maintenance of the deployment easier, ensures that networking issues don't ever contribute to VirtualCenter losing contact with the License server, and reduces the need for an additional dedicated virtual machine.

Sizing and Configuration

The choice of the physical hardware on which to run the VirtualCenter virtual machine, as well as the virtual machine configuration, should be made for VirtualCenter just as it would be for any enterprise application. The resource utilization characteristics of the application largely dictate the hardware and virtual resources that should be assigned. The following guidelines apply to the particular utilization characteristics of VirtualCenter:

- Dedicate the virtual machine to VirtualCenter. In particular, keep the VirtualCenter database on a physical server or separate virtual machine, unless you are only running the MSDE database for demonstration purposes. This will ensure that VirtualCenter will have exclusive access to the resources of the virtual machine's operating system
- Size the VirtualCenter virtual machine based on the physical server requirements
    - Memory: The maximum memory that the VirtualCenter server can use is 4 GB (2 GB for the application, and 2 GB for the operating system), but in environments with fewer than 50 ESX Server hosts, less memory will be required. The minimum requirement is 2GB.
    - CPU: VirtualCenter typically will use only 1 CPU, but for environments greater than 50 ESX Server hosts, it will benefit from 2 CPUs. Configure the number of VCPUs accordingly. Be aware that the virtual machine will use some portion of the CPU to handle virtual networking; the overhead can be as high at 30% for environments above 50 ESX Server hosts.
    - Networking: VirtualCenter will rarely need more than a few hundred KB/sec of network bandwidth
    - Storage: VirtualCenter does not perform any significant I/O operations to its local disk
- Allocate sufficient resources to the VirtualCenter virtual machine on the ESX Server host
- Use resource shares or reservations to ensure that the VirtualCenter virtual machine has the highest priority out of all virtual machines on the ESX Server host, and has guaranteed usage of its configured memory.
- If you choose to run the VirtualCenter virtual machine with 2 VCPUs, make sure that any

ESX Server host on which the virtual machine runs has at least 4 cores or threads, to ensure that the virtual machine never suffers from CPU scheduling starvation (a phenomenon in which a VSMP virtual machine cannot get enough cycles simultaneously on multiple CPUs to run)

- To ensure that you can be notified immediately if there is a problem with the VirtualCenter virtual machine, configure an alarm to send email. Conditions to monitor include:
    - Virtual Machine CPU Usage: trigger if the usage goes above 90%
    - Virtual Machine Memory Usage: trigger if the usage goes above 90%
- Because VirtualCenter manages the entire VMware Infrastructure, make sure to set permissions so that only a very limited set of people can access or manipulate the virtual machine and host on which it runs.

Installation

The most straightforward way to put VirtualCenter on a virtual machine is to install it initially on a standalone ESX Server host, that is, a host that is not being managed by a VirtualCenter instance. When planning for this configuration, note that

- the ESX Server host will need to be enabled with host-based licensing, since server-based licensing is not available without a running instance of VirtualCenter
- once the ESX Server software is deployed onto the host, you can access the host by directly connecting to it with the VI Client, and use the VI Client interface to create the virtual machine for hosting VirtualCenter.

Once the virtual machine for VirtualCenter has been created, follow the standard procedure for installing it as you would in a physical server, as described in Chapter 4 of the Installation and Upgrade Guide. In particular,

- a suitable database for VirtualCenter should already be up and running somewhere in your environment, so that it may be specified during the installation
- make sure to place the license file somewhere on the virtual local disk of the virtual machine before beginning the installation, and during the installation, choose to install the license server on the same system

At the end of the installation process, verify that VirtualCenter is functioning by connecting to it with the VI Client. At that point, you can

- put the original standalone ESX Server host into the VirtualCenter inventory, and start managing it along with the other hosts in the environment.
- convert that host to server-based licensing. Because of the built-in grace period of the licensing, this host and all its virtual machines will continue to operate for 14 days if the VirtualCenter virtual machine goes down.

Limitations

Since certain management tasks for a virtual machine require it to be powered off, you cannot perform most of these tasks on the virtual machine that hosts the VirtualCenter server. In particular, do not attempt the following on the VirtualCenter virtual machine

- Cold-migration
- Cloning
- Any other operation which powers down the virtual machine
- Edit most of the virtual machine settings, such as the hardware properties.

Allowed Functionality

The settings you can modify on a running VirtualCenter virtual machine are the Resource properties, such as shares and reservations. You should change these according to the guidelines given in the Sizing and Configuration section. However, avoid changing the default setting of "Unlimited" for the Limit setting, as setting the limit too low could potentially cause your VirtualCenter to become unresponsive and lock up.

If you wish to modify the hardware properties for the VirtualCenter virtual machine, you will need to schedule downtime for VirtualCenter. Then, you will need to connect to the ESX Server host directly with the VI Client, shut down the VirtualCenter virtual machine, and make the modifications.

You can use VMotion to migrate the running VirtualCenter virtual machine to another host, provided both the hosts meet all the requirements for VMotion. It will continue to operate and will not lose connection with any of the ESX Server hosts or VI Client sessions during the migration.

VirtualCenter in a VMware HA Cluster

If high availability for the VirtualCenter virtual machine is desired, then the ESX Server host on which it is running may be placed in a VMware HA cluster. If that host fails, another will automatically restart the VirtualCenter virtual machine. This will subsequently restart all the VirtualCenter services, including:

- the VirtualCenter server
- the Flex License server
- the VirtualCenter Web Access server

In addition to the standard hardware requirements for a VMware HA cluster (as described in the Resource Management Guide, Chapter 5, Section "Cluster Prerequisites"), the following points should be observed

- Consider using strict admission control, so that additional virtual machines cannot be run on this cluster if they would violate availability constraints in the case of host failure. In all cases, set the restart priority of the VirtualCenter virtual machine to high.
- If VMware DRS is enabled on the cluster, be aware that the host on which the VirtualCenter virtual machine runs might change from time to time as the load is rebalanced. If this virtual machine were to go down unexpectedly, you would need to search manually for the host on which it was last running, since VirtualCenter would be unavailable for you to determine this directly. If the size of the cluster is small, this might not be an issue, since there would be few hosts to search. An alternative is for VMware DRS to be disabled on the VirtualCenter virtual machine, by setting the Automation Level to "Disabled". This way, the location of the VirtualCenter virtual machine is explicitly known at all times, making it easy to access this host via the VI Client directly if there are problems with this virtual machine. This virtual machine can still be migrated manually by an administrator for tasks such as hardware maintenance or manual load rebalancing.
- Along with setting reservations and shares as described earlier, make sure there is enough spare capacity in the cluster to handle a failure and still have enough resources for VirtualCenter to run.
- Make sure that the permissions only allow a limited set of people to access and manage this cluster.