CHAPTER 1

Virtualizing Cisco Unity Messaging on VMware ESX

This chapter contains the following sections:

Introduction

Revised April 21, 2010

Virtualization has emerged as a leading trend in the data center. Organizations are embracing the virtualized data center for a variety of reasons, including total cost of ownership (TCO), consolidation, operational efficiency and flexibility, disaster recovery and business continuity, security, and reduced carbon footprint.

Cisco is committed to leading the transition toward the virtualized data center through innovative products and solutions, including support for Cisco Unity messaging on the VMware ESX platform, part of the VMware Infrastructure suite.

Deploying Cisco Unity messaging on VMware ESX delivers the following advantages:

- Allows customers to extend the benefits of their virtualization and storage data center investments to Cisco Unity messaging.
- Maintains the same predictable scalability provided by running the Cisco Unity application on a physical server.

Cisco Unity messaging is a real-time application, which makes it more difficult to virtualize than traditional data-centric applications, such as database and email servers. (For example, to support 144 concurrent voice sessions, Cisco Unity messaging must place 7,200 packets on the wire at a precise 20 ms interval.) Delivering this level of performance in a reliable, predictable, and serviceable manner requires some concessions, primarily surrounding CPU Affinity.

You can virtualize the Cisco Unity system in a Unified Messaging or Voice Messaging configuration:

- In a Unified Messaging configuration, you can also virtualize Microsoft Exchange, Active Directory, or IBM Lotus Domino servers. However, this guide provides information about virtualizing the Cisco Unity application and voice-recognition servers and not the others. In addition, Cisco does not provide technical support for message-store servers or for domain controllers/global catalog servers for Unified Messaging deployments.

Note Virtualizing a Domino server is supported only for use with Cisco Unity 7.0(2). With the release of Cisco Unity 8.x, Domino is no longer supported.

- In a Voice Messaging configuration, you must use the Voice Mail Run-Time Edition of Exchange Server 2003 (with or without Active Directory installed). This guide provides information about virtualizing the Cisco Unity application, voice-recognition, and Voice Mail Run-Time servers.

You can also run a mix of virtual and nonvirtual machines, including the servers in a Cisco Unity failover pair.

Also note that to virtualize Cisco Unity messaging, there is an additional support burden. Customers provide the hardware, VMware ESX software, and Microsoft Windows Server software, and are responsible for coordinating support for these components, including provisioning and performance troubleshooting.

Customers who are unwilling to take on such support may be better candidates for deploying Cisco Unity messaging on physical servers.

Requirements

Revised August 12, 2011

Cisco supports virtualizing the following servers:

- Cisco Unity application servers—Version 7.0(2), or version 8.0(3) or later.
- Cisco Unity voice-recognition servers—Version 7.0(2), or version 8.0(3) or later.
- Voice Mail Run-Time Edition of Microsoft Exchange Server 2003 servers (with or without Active Directory installed).

Note This guide collectively refers to the three types of virtualized servers mentioned above as "Cisco Unity virtual machines."

The following requirements apply to all virtual machines running the Cisco-provided software, unless otherwise noted:

- VMware ESX 3.5 or later, or VMware ESXi 3.5 or later.

Note This guide refers to both VMware ESX and VMware ESXi as VMware ESX.

- VMware ESX must be deployed in a configuration supported by VMware.

  VMware provides comprehensive compatibility guides on its website for systems (servers), I/O devices, and storage and storage-area networks.

- The following VMware ESX host server CPU models are supported:

Intel Xeon E5540	Intel Xeon W5580
Intel Xeon E5472	Intel Xeon X7350
Intel Xeon E5462	Intel Xeon X5570
Intel Xeon E5450	Intel Xeon X5560
Intel Xeon E5440	Intel Xeon X5550
Intel Xeon E5430	Intel Xeon X5492
Intel Xeon E5420	Intel Xeon X5482
Intel Xeon E5640	Intel Xeon X5472
Intel Xeon E6540	Intel Xeon X5470
Intel Xeon L5430	Intel Xeon X5460
Intel Xeon L5420	Intel Xeon X5450

  If a CPU model is not listed, it is not supported.

- The Cisco Unity virtual machine hard disks must reside on shared Fibre Channel storage. Traditional Fibre Channel host bus adapters and Fibre Channel over Ethernet converged network adapters are supported.

  NAS/NFS, iSCSI, and direct-attached storage are not supported.

- All Cisco Unity virtual machine configuration properties must match a virtual machine template. (See the "Cisco Unity Virtual Machine Templates" section on page 1-5).

- The CPU Affinity feature of VMware ESX is required to dedicate CPU cores to the Cisco Unity virtual machines. (See "Processor" in the "Provisioning of Host Resources for Virtual Machines" section on page 1-8.)

- The Memory Reservation feature is required for each Cisco Unity virtual machine. (See "Memory" in the "Provisioning of Host Resources for Virtual Machines" section on page 1-8.)

- Disk I/O performance must meet or exceed the required level. (See "Disk" in the "Provisioning of Host Resources for Virtual Machines" section on page 1-8.)

- The IP network design must ensure that Cisco Unity virtual machine traffic never encounters contention or delay when leaving the VMware ESX host. (See "Network" in the "Provisioning of Host Resources for Virtual Machines" section on page 1-8. In addition, the appendix "Performance Test Information" includes network I/O examples to aid design.)

  Cisco Technical Assistance Center (TAC) asks for the results before troubleshooting performance issues.

- Cisco Unity application virtual machines must run the Cisco Unity Performance Information and Diagnostics utility at all times with the appropriate configuration file. Cisco Technical Assistance Center (TAC) asks for the results before troubleshooting performance issues.

  On Cisco Unity 7.0 voice-recognition virtual machines and on Exchange 2003 Voice Mail Run-Time Edition virtual machines, the utility is optional but recommended.

(The Windows-based utility is not supported with Cisco Unity 8.x voice-recognition software, which runs on the Linux operating system.)

- Read-only VMware Infrastructure Client access to the VMware ESX host.

  Cisco TAC requires access to confirm that configuration requirements are met.

- The NIC on a Cisco Unity application virtual machine must have a static MAC address. (Cisco Unity voice-recognition and Exchange 2003 Voice Mail Run-Time Edition virtual machines do not need to have static addresses.)
- The Cisco Unity virtual machines must have VMware Tools installed.
- The Cisco Unity virtual machines' time (clock) must be reliably kept in sync with true time.

  VMware provides documentation on timekeeping in the "Timekeeping in VMware Virtual Machines" document available on its website.

- Cisco Unity Voice Connector for Microsoft Exchange can be installed on an Exchange server in both Unified Messaging and Voice Messaging configurations.

VMware Infrastructure Feature Support

VMware Infrastructure includes a number of innovative features such as VMotion that are not available with traditional physical hardware installations. Not all of these features lend themselves to real-time streaming media applications such as Cisco Unity messaging. Table 1-1 lists VMware Infrastructure features and their current support status for use with Cisco Unity virtual machines.

Note that "supported" means that no problems are expected when the feature is used with Cisco Unity virtual machines. Although some supported features were tested with Cisco Unity virtual machines, not all were tested.

Table 1-1 Support Status of VMware Infrastructure Features for Use with Cisco Unity Virtual Machines

Feature	Support Status
VMware Consolidated Backup	Supported.
VMware Dynamic Resource Scheduler (DRS)	Not supported. (Incompatible with CPU Affinity, and it requires VMotion, which also is not supported.)
VMware High Availability (HA)	Not supported. (Incompatible with CPU Affinity.)
VMware Site Recovery Manager	Supported.
VMware Snapshots	Supported.
VMware Snapshots with Memory	Not supported. (Interrupts Cisco Unity performance.)
VMware Storage VMotion	Not supported. (Incompatible with CPU Affinity, and it leverages VMotion, which also is not supported.)
VMware VMotion	Not supported. (Incompatible with CPU Affinity, and it is service interrupting.) See also "VMware VMotion" in the following "Feature Limitations and Restrictions" section.
VMware vCenter Update Manager	Supported with limitations. See "VMware vCenter Update Manager" in the following "Feature Limitations and Restrictions" section.

Feature Limitations and Restrictions

VMware vCenter Update Manager

The standalone Cisco Security Agent for Cisco Unity is not supported for use with vCenter Update Manager.

Customers who do not install Cisco Security Agent for Cisco Unity can use vCenter Update Manager.

Cisco Unity upgrades and engineering specials are not delivered through vCenter Update Manager.

VMware VMotion

VMotion and features that leverage it are not supported because VMotion presents the following significant challenges for the Cisco Unity application:

- A virtual machine with CPU Affinity enabled cannot be migrated with VMotion. Therefore CPU Affinity must be disabled prior to migration, a condition that is known to dramatically affect voice quality.
- During the VMotion cutover, the system is paused. For a real-time streaming media application such as Cisco Unity messaging, this creates service interruption. While testing shows that calls are not dropped during a VMotion migration, voice calls that were in progress experience degraded voice quality after the migration.

We recommend that customers deploying Cisco Unity application virtual machines do so with the Cisco Unity failover feature. Manual Cisco Unity failovers do not interrupt service, thus mitigating the need to migrate a live virtual machine.

VMware ESX Software Updates

Software updates for VMware ESX are supported at the time of release. This includes major releases, minor releases, and patches. In the event that an update negatively impacts Cisco Unity messaging, Cisco will work to correct the issue in a reasonable time frame.

To mitigate risk associated with VMware ESX software updates and Cisco Unity virtual machines, we highly recommend that customers maintain sufficient VMware ESX host resources running the known compatible version of VMware ESX. If there is a compatibility issue, Cisco Unity virtual machines can be migrated back to the known compatible VMware ESX version until Cisco provides a resolution.

Cisco Unity Virtual Machine Templates

Cisco Unity messaging on VMware ESX provides predictable scalability through minimum hardware requirements, minimum performance requirements, and predefined virtual machine templates.

A virtual machine template defines the configuration of the virtual machine hardware. The configuration of a Cisco Unity virtual machine must match a supported virtual machine template defined in this section:

- Cisco Unity Template 1, page 1-6
- Cisco Unity Template 2, page 1-6
- Exchange 2003 Voice Mail Run-Time Edition Template, page 1-7

- Cisco Unity Voice-Recognition Template, page 1-7

Caution Cisco conducted all testing on VMware ESX hosts with the default set of Advanced Settings. Settings other than the defaults could potentially impact performance of Cisco Unity virtual machines, resulting in undesirable performance.

Cisco Unity Template 1

Configuration

- 4 vCPU
- 4 GB RAM
- 1 vNIC with static MAC address
- 4 vDisks:
  - vDisk 1 = 24 GB—Operating system
  - vDisk 2 = 24 GB—Cisco Unity binaries
  - vDisk 3 = 24 GB—Cisco Unity logs, SQL Server transaction logs
  - vDisk 4 = 24 GB—Cisco Unity Message Repository (UMR), SQL Server database

Limits

- 144 ports
- 15,000 users
- 36 text-to-speech ports

When assigning a static MAC address, we recommend choosing a complex address that another customer is unlikely to use. This can prevent accidental licensing overlap between different customers.

An example of a complex MAC address is 00:50:56:01:3B:9F.

A less complex MAC address is 00:50:56:11:11:11, because of the repeating 1s.

Cisco Unity Template 2

Added April 2, 2009

Configuration

- 2 vCPU
- 4 GB RAM
- 1 vNIC with static MAC address
- 4 vDisks:
  - vDisk 1 = 24 GB—Operating system
  - vDisk 2 = 24 GB—Cisco Unity binaries
  - vDisk 3 = 24 GB—Cisco Unity logs, SQL Server transaction logs
  - vDisk 4 = 24 GB—Cisco Unity Message Repository (UMR), SQL Server database

Limits

- 48 ports
- 5,000 users
- 36 text-to-speech ports

When assigning a static MAC address, we recommend choosing a complex address that another customer is unlikely to use. This can prevent accidental licensing overlap between different customers.

An example of a complex MAC address is 00:50:56:01:3B:9F.

A less complex MAC address is 00:50:56:11:11:11, because of the repeating 1s.

Exchange 2003 Voice Mail Run-Time Edition Template

Revised April 21, 2010

Configuration

- 1 vCPU
- 4 GB RAM
- 1 vNIC
- 3 vDisks
  - vDisk 1 = 24 GB—Operating system, Exchange binaries
  - vDisk 2 = 24 GB—Active Directory transaction logs, Exchange transaction logs
  - vDisk 3 = <User defined size>—Active Directory database, Exchange mailbox store

Limits

- 7,500 Exchange mailboxes (with or without IMAP)

For vDisk 3, you can set the disk size based on your message storage and directory needs.

Cisco provides sizing numbers for the various supported codecs in the white paper *Audio Codecs and Cisco Unity* at http://www.cisco.com/en/US/docs/voice_ip_comm/unity/white/paper/cuaudiocodecs.html. Sizing guidance for Active Directory is available in the white paper *Active Directory Capacity Planning for Cisco Unity* at http://www.cisco.com/en/US/docs/voice_ip_comm/unity/white/paper/5xcuadsizing.html.

Cisco Unity Voice-Recognition Template

Configuration

- 1 vCPU
- 2 GB RAM
- 1 vNIC
- 1 vDisk = 24 GB—Operating system, Cisco Unity voice-recognition binaries, Cisco Unity voice-recognition logs

Software Installation and System Setup

Revised April 21, 2010

Installation of Cisco Unity software and the associated applications, and system setup follow the same processes described in the applicable Cisco Unity installation documentation, available at http://www.cisco.com/en/US/products/sw/voicesw/ps2237/prod_installation_guides_list.html.

Note the following virtualization exceptions to content in the *Installation Guide for Cisco Unity*:

- Create a single partition for each virtual disk.
- Follow the layout of disks defined in the "Cisco Unity Virtual Machine Templates" section on page 1-5 of this guide.
- For Windows Server, follow the *Installation Guide* instructions for installing Windows Server by using a retail Windows Server disc. (The other options in the *Installation Guide* do not apply in the virtualization context.)

Note For Cisco Unity 7.0, you use the Release 5.x versions of the guides. Content and instructions in the 5.x guides apply to version 7.0 as well. See also the "New Functionality" and "Changed Functionality" sections of *Release Notes for Cisco Unity Release 7.0(2)* at http://www.cisco.com/en/US/docs/voice_ip_comm/unity/7x/release/notes/702curelnotes.html.

Provisioning of Host Resources for Virtual Machines

This section addresses the following resources:

Processor

Revised May 7, 2010

Note The information in this section applies to virtualized Cisco Unity application servers and to virtualized Cisco Unity voice-recognition servers. For virtualized servers running the Voice Mail Run-Time Edition of Microsoft Exchange Server 2003, the settings are recommended.

Each Cisco Unity virtual machine must have physical CPU cores dedicated to the virtual machine. Without dedicated CPU cores, call quality degrades noticeably.

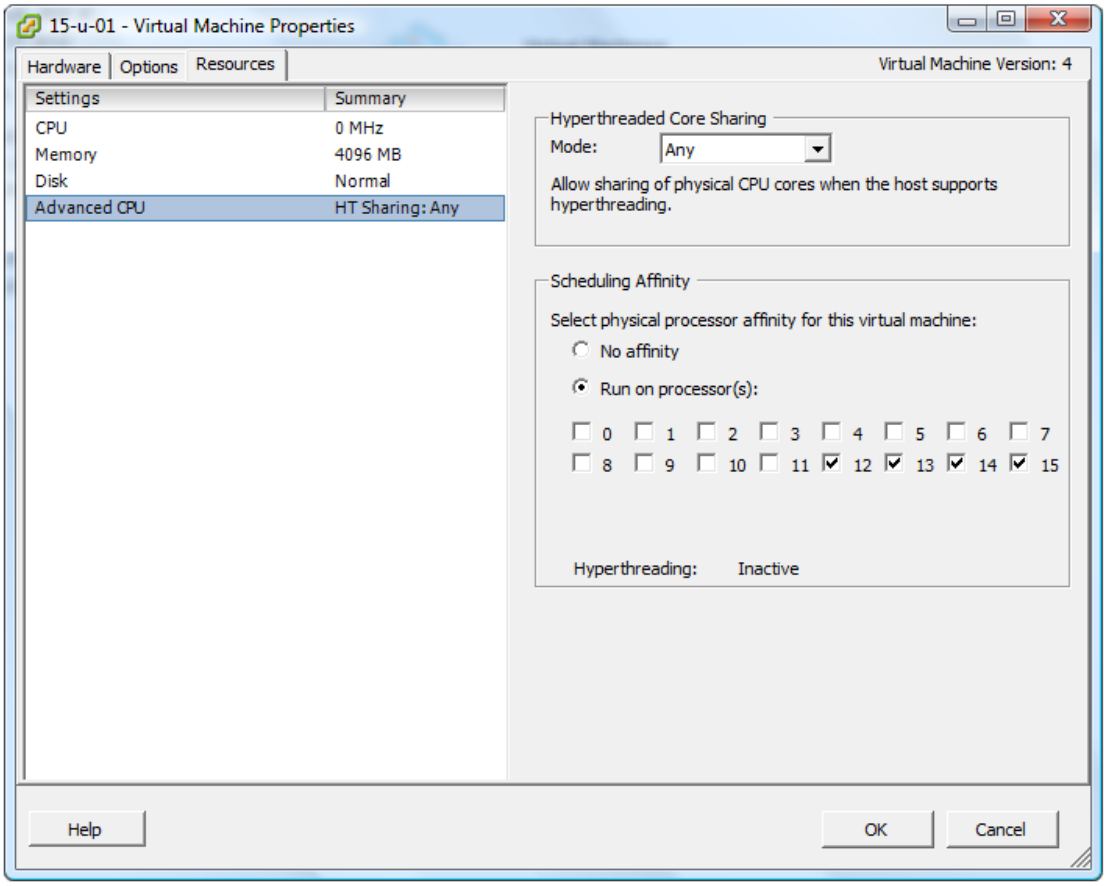
The CPU Affinity feature of VMware ESX is the mechanism used to dedicate CPU cores. CPU Affinity is applied at the virtual-machine level and restricts the CPU cores on which a given virtual machine is allowed to execute.

For CPU Affinity to effectively partition CPU resources so that the Cisco Unity virtual machines have a dedicated one-to-one mapping of physical CPU core to vCPU, the following conditions must be met:

- All virtual machines running on the VMware ESX host must have CPU Affinity configured.
- Each Cisco Unity virtual machine must be configured to execute on as many cores as vCPUs configured.
- No other virtual machine running on the ESX host can be assigned a core that is assigned to a Cisco Unity virtual machine.

Figure 1-1 shows CPU Affinity settings for a Cisco Unity virtual machine with four vCPUs.

Figure 1-1 CPU Affinity Settings



If unallocated CPU cores exist after provisioning the Cisco Unity virtual machines, then third-party virtual machines are permitted to run on the same VMware ESX host. Cisco does not require the third-party virtual machines to have dedicated one-to-one mappings of physical CPU core to vCPU core, as is required for the Cisco Unity virtual machines. The unallocated CPU cores can be pooled together and shared between the third-party virtual machines without undermining the performance of the Cisco Unity virtual machines running on the same VMware ESX host.

When configuring CPU Affinity with a multiprocessor virtual machine, the assigned CPU cores must be on the same CPU die. Doing so greatly improves performance by keeping interprocessor communication on die and is required to achieve the scalability numbers described in the "Cisco Unity Virtual Machine Templates" section on page 1-5.

With VMware ESX, the console operating system runs on CPU 0. We recommend that you do not configure Cisco Unity virtual machines to use CPU 0, particularly if you run scripts or agents in the console operating system, as it may impact Cisco Unity voice quality and capacity.

Memory

Revised May 7, 2010

Note The information in this section applies to virtualized Cisco Unity application servers and to virtualized Cisco Unity voice-recognition servers. For virtualized servers running the Voice Mail Run-Time Edition of Microsoft Exchange Server 2003, the settings are recommended.

Each Cisco Unity virtual machine must have the Memory Reservation for the full amount of memory allocated to the virtual machine. For example, if the virtual machine has 4 GB of RAM allocated, the Memory Reservation must be set to 4096 MB.

Disk

Revised May 7, 2010

Note The information in this section applies to virtualized Cisco Unity application servers and to virtualized Cisco Unity 7.0 voice-recognition servers. (The Windows-based utilities referred to are not supported with Cisco Unity 8.x voice-recognition software, which runs on the Linux operating system.) For virtualized servers running the Voice Mail Run-Time Edition of Microsoft Exchange Server 2003, the settings are recommended.

Table 1-2 lists the minimum performance requirements that each Cisco Unity virtual machine hard disk must meet, as reported by Microsoft Windows Performance Monitor or by the Cisco Unity Performance Information and Diagnostics utility.

Table 1-2 Minimum Performance Requirements

Performance Monitor Object	Requirement
PhysicalDisk\% Idle Time	Average greater than 25 (during a busy hour)
PhysicalDisk\Avg. Disk sec/Read	Less than 0.25
PhysicalDisk\Avg. Disk sec/Write	Less than 0.25
PhysicalDisk\Current Disk Queue Length	Less than 4

While troubleshooting a Cisco Unity virtual machine, Cisco TAC asks for log output from the Cisco Unity Performance Information and Diagnostics utility. If any disk is not performing at or better than the minimum requirements, Cisco TAC will decline support until the system is performing at or above the minimum requirements.

The virtual disk layouts for Cisco Unity virtual machines defined in the "Cisco Unity Virtual Machine Templates" section on page 1-5 were designed to provide maximum flexibility with regard to LUN placement and I/O segmentation. In addition, the appendix "Performance Test Information" provides disk I/O characterization data to aid the storage architect with successful design.

Storage design is left to your discretion, with the exception of requiring shared Fibre Channel storage.

Network

Revised May 7, 2010

Note The information in this section applies to virtualized Cisco Unity application servers and to virtualized Cisco Unity voice-recognition servers. For virtualized servers running the Voice Mail Run-Time Edition of Microsoft Exchange Server 2003, the settings are recommended.

Network design is left to your discretion, with the exception of requiring gigabit or faster NICs.

When multiple virtual machines share the same VMware ESX host physical network interface, there is the possibility of contention for inbound and outbound bandwidth. Inbound bandwidth contention can be managed with Quality of Service (QoS) in a hardware switch. The VMware ESX soft switch does not support the same QoS capabilities for outbound traffic, making contention more difficult to manage dynamically.

One way to avoid outbound contention completely is to dedicate a NIC on the VMware ESX hosts to the Cisco Unity virtual machines. This, however, is not a requirement. Deploying 10 Gigabit Ethernet and/or link aggregation with Link Aggregation Control Protocol or Cisco EtherChannel can help reduce or eliminate the possibility of outbound contention as well. In testing, the VMware ESX software switch rate-limiting feature also proved capable, although using the feature might be impractical for most customers.

The appendix "Performance Test Information" provides network I/O characterization data to aid the network architect with design. Additional factors the architect should consider are any third-party virtual machines that will share the outbound interface with the Cisco Unity virtual machines. A file server might not be the best candidate to share bandwidth with Cisco Unity virtual machines, whereas a low-traffic print server might be. Examining the historical traffic patterns of coresident virtual machines through the VMware Infrastructure client could prove helpful.

Unfortunately, there is no simple method to detect outbound contention from within a virtual machine. In situations where inbound or outbound contention is suspected, troubleshooting will need to occur from the switch(es) to which the VMware ESX host links.

Redundancy

Revised May 7, 2010

Note The information in this section applies to virtualized Cisco Unity application servers.

With storage fabric redundancy, failover is not instantaneous. During a failover event, the Cisco Unity system cannot write to disk, and it pauses—typically for 1 to 2 minutes—until the event completes.

For this reason, we recommend that customers deploying Cisco Unity application virtual machines do so with the Cisco Unity failover feature. To get the full benefits of failover, there must be careful consideration given to redundancy of the underlying server, network, and storage infrastructure.

Note If a failover event happens because of a storage redundancy event, there is no information within the Cisco Unity virtual machine to determine that the failover event was caused by the storage event. You need to coordinate with the teams that manage the components of the storage infrastructure to conduct cause analysis.

Disaster Recovery

Disaster recovery for Cisco Unity virtual machines supports the same in-host techniques as Cisco Unity messaging on physical servers: the Cisco-provided Cisco Unity Disaster Recovery tools (DiRT) and optional third-party backup applications.

Virtualization and storage networks allow for new out-of-host disaster-recovery techniques. Examples include snapshots, entire virtual machine backups, and multisite recovery scripts with products such as VMware Site Recovery Manager. You can use out-of-host disaster-recovery techniques, provided disk-performance minimums are maintained.

Caution Disaster-recovery techniques that rely on snapshot or synchronous interdata-center I/O writes can have a negative impact on disk performance.

Licensing

Licensing of Cisco Unity messaging on VMware ESX is identical to Cisco Unity licensing on physical hardware. (See also the "Technical Support" section on page 1-12.)

Platform Migration

Revised April 21, 2010

Migrating a Cisco Unity physical server to a virtual machine is supported only by using the process for replacing a server, as described in the Cisco Unity documentation:

- For Cisco Unity 8.x—See the "Replacing or Converting a Cisco Unity 8.x Server, or Upgrading to Windows 2003" chapter of the *Reconfiguration and Upgrade Guide for Cisco Unity Release 8.x* at http://www.cisco.com/en/US/docs/voice_ip_comm/unity/8x/upgrade/guide/8xcurugx.html.
- For Cisco Unity 7.0—See the "Replacing or Converting a Cisco Unity 5.x Server, or Upgrading to Windows 2003" chapter of the applicable *Reconfiguration and Upgrade Guide for Cisco Unity Release 5.x* at http://www.cisco.com/en/US/products/sw/voicesw/ps2237/prod_installation_guides_list.html.

Note Content and instructions in the 5.x *Reconfiguration and Upgrade Guides* apply to reconfiguring version 7.0 as well. See also the "New Functionality," "Changed Functionality," and "Documentation Updates" sections of *Release Notes for Cisco Unity Release 7.0(2)* at http://www.cisco.com/en/US/docs/voice_ip_comm/unity/7x/release/notes/702curelnotes.html.

Converting or migrating a Cisco Unity physical server to a virtual machine in any other way—including by using physical-to-virtual migration tools from VMware or other third parties—is not supported.

If the Cisco Unity physical server has a Cisco-provided embedded Windows Server license, it is not transferable to the virtual machine.

Technical Support

Revised April 21, 2010

To virtualize Cisco Unity messaging, customers provide the hardware, VMware ESX software, and Microsoft Windows Server software, and are responsible for coordinating technical support for these components, including provisioning and performance troubleshooting.

Cisco TAC does not provide technical support for provisioning the VMware ESX environment or the Cisco Unity virtual machines. The expectation is that the customer has sufficient experience to complete these processes.

When Cisco Unity is installed in a Unified Messaging configuration, IBM Lotus Domino or Microsoft Exchange can be installed on any hypervisor supported by IBM or Microsoft, respectively. In addition, Active Directory domain controllers/global catalog servers (DC/GCs) can be installed on any hypervisor supported by Microsoft. (Cisco does not provide technical support for message-store servers or for DC/GCs.)

Note For virtualization, Domino is supported only for use with Cisco Unity 7.0(2). With the release of Cisco Unity 8.x, Domino is no longer supported.

Cisco TAC provides technical support only for Cisco Unity software and the related Cisco-provided components.