Microsoft Office Communications Server 2007 R2

Scale to a Load Balanced Enterprise Edition Pool with WebMux Walkthrough

Published: Sept. 2009

For the most up-to-date version of the Scale to a Load Balanced Enterprise Edition Pool Walkthrough documentation and the complete set of the Microsoft® Office Communications Server 2007 R2 online documentation, see the Office Communications Server TechNet Library at http://go.microsoft.com/fwlink/?LinkID=132106.

Note:

> In order to find topics that are referenced by this document but not contained within it, search for the topic title in the TechNet library at http://go.microsoft.com/fwlink/?LinkID=132106.

Contents

Scale to a Load Balanced Enterprise Edition Pool Walkthrough

Many enterprises begin their Office Communications Server deployment with a single Enterprise Edition Front End Server. As usage of the product becomes increasingly mission critical, it is common to move to a load-balanced configuration to provide high availability and increased capacity. This article describes the procedure for migrating from a single enterprise Front End Server to a load-balanced pool with more than one Front End Server.

Note:

> This procedure assumes that the starting point is an enterprise edition Front End server without a load balancer. If the starting point is a Standard Edition server, similar steps may be followed. However, a new pool would need to be created and users migrated to this new pool as part of the cut over procedure.

In This Document

Walkthrough: Overview of Scaling an Enterprise Pool

Overview

You begin migrating to a load-balanced array of Front End Servers by deploying the Front End Servers that will make up the array. Until you change the FQDN entries in the Active Directory Domain Services, all users will continue to use the existing single Front End Server. Meanwhile, you can test the new Front End Servers individually by using a few host file changes, ensuring minimal downtime.

After you have validated the individual Front End Servers, you insert the load balancer in front of these new servers and configure it. Again, because you have not yet modified the production DNS records, end users are still unaffected. With some additional host file record changes, you can now test the load balanced environment and validate failover behavior on each of the Front End Servers.

After the load balancing behavior has been validated, the final step is to cut over to this new environment by removing the host file entries and changing the correct DNS entries in your production environment.

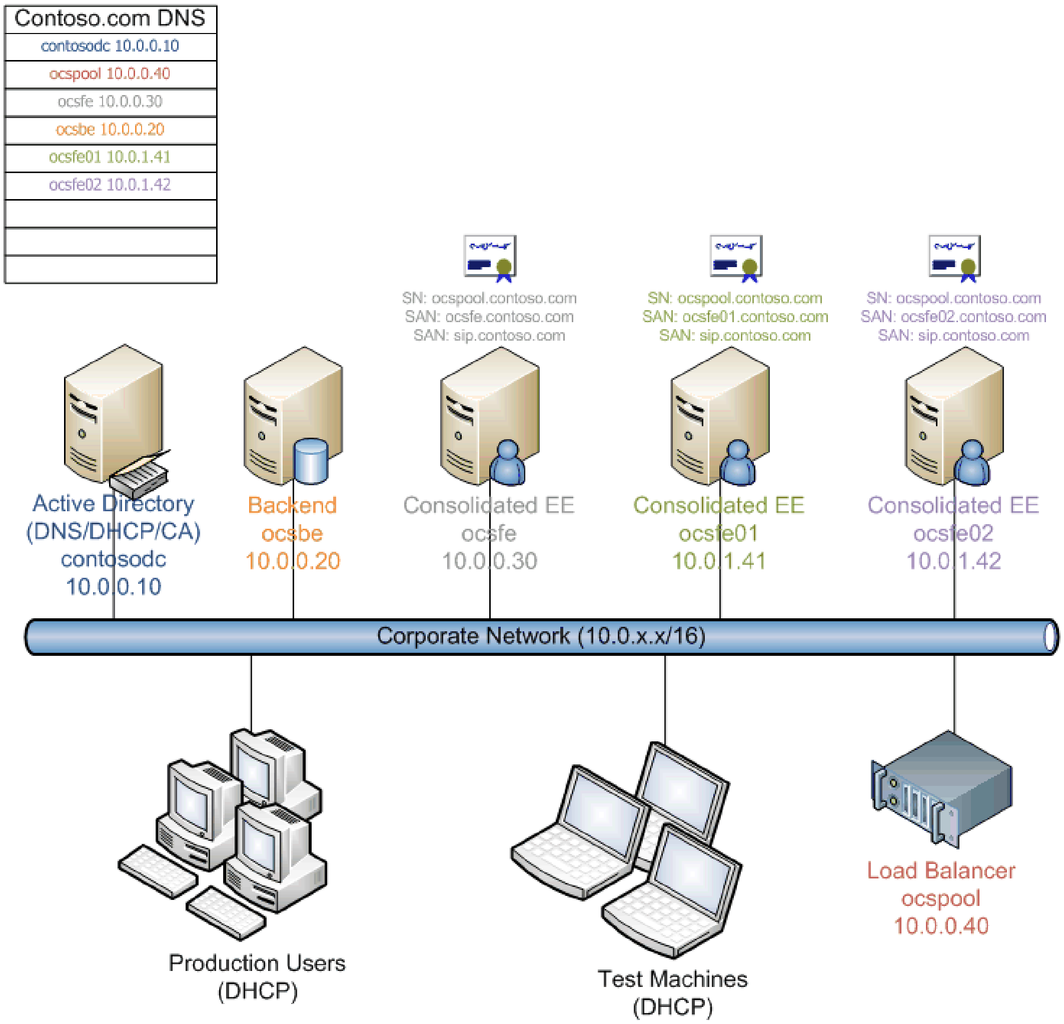
If you encounter any issues during cutover, you can quickly revert to the single Front End Server. When the array is performing correctly, you can decommission the single Front End Server. The rest of this document will describe each phase of this process in detail.

Walkthrough: Supported Topologies for Load Balanced Enterprise Pools

Supported Topologies

For the purposes of this walkthrough, we shall define two sample topologies to illustrate specific steps that would be involved in moving to a load-balanced Enterprise pool. They are shown in the figures below. The first shows a one-armed topology, and the second shows a two-armed topology. These are the only two topologies that are supported in Office Communications Server 2007 R2. Note that the IP addresses of the corresponding servers in both diagrams are the same. The key difference is the networking topology and routing. In particular, notice the difference in subnets between the two diagrams. Please note WebMux has two ways to support two-armed topology: NAT+SNAT and TM+SNAT.

Please note: Microsoft's original document indicates DNAT is not required and not supported. However, if the Destination address is not translated by the load balancer, the Front End server will not accept the clients' request, due to the destination address in the data packet is still the VIP address. To address the server address, the load balancer must have both SNAT and DNAT enabled. DNAT alone will not work with OCS 2007 R2. If you do a TCP capture on the server, you should see both the Source IP address and Destination IP address are changed from the clients' original data packets.

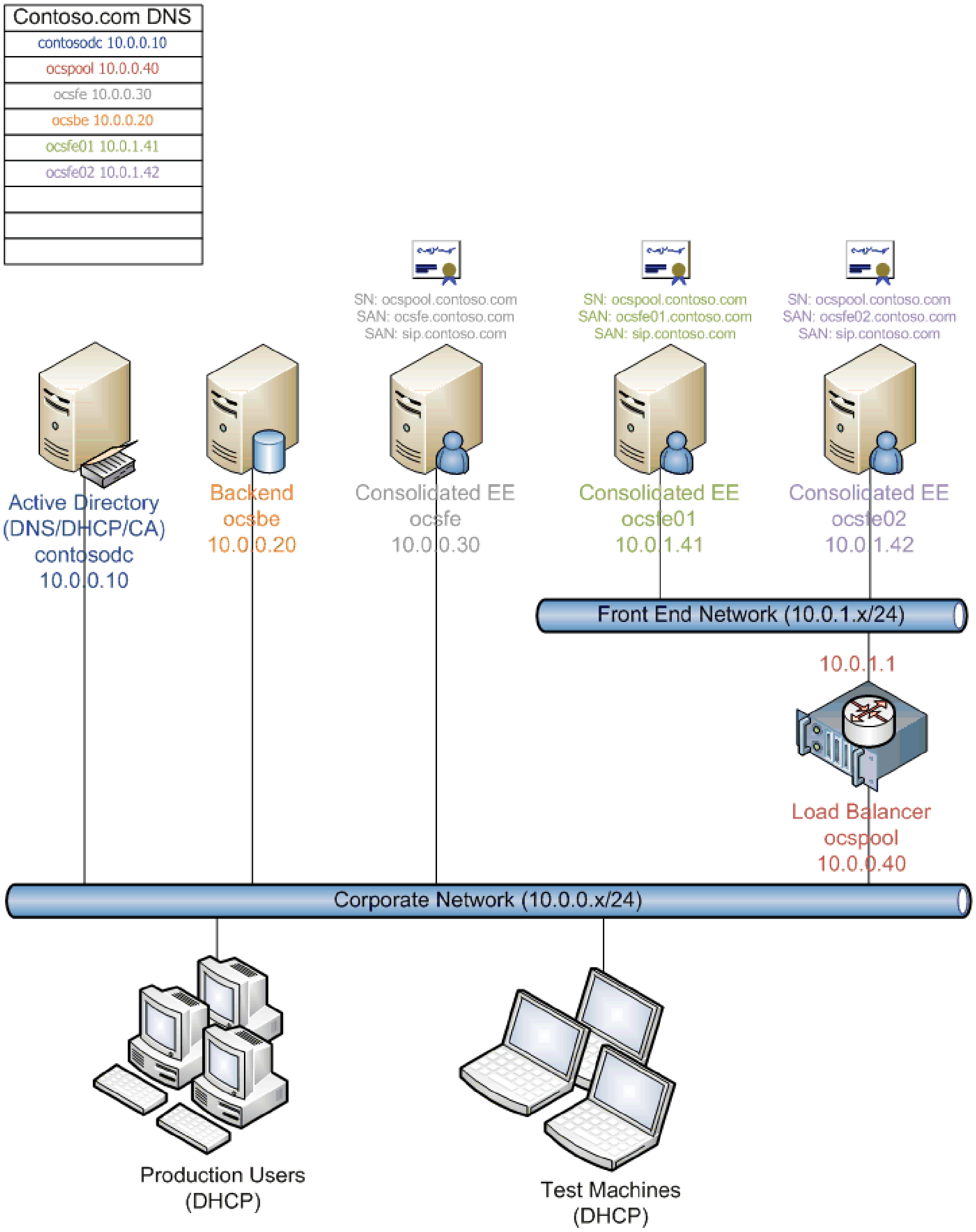


Contoso.com DNS
contosodc 10.0.0.10
ocspool 10.0.0.40
ocsfe 10.0.0.30
ocsbe 10.0.0.20
ocsfe01 10.0.1.41
ocsfe02 10.0.1.42

Figure above: One-Armed topology

Your networking team may have an existing best practice for deploying load balanced services, and that will probably have the biggest impact on which option that you will use. If no precedent exists, here are some factors to consider when deciding between a one-armed or two-armed topology.

<u>One Armed Topology</u>



Contoso.com DNS
contosodc 10.0.0.10
ocspool 10.0.0.40
ocsfe 10.0.0.30
ocsbe 10.0.0.20
ocsfe01 10.0.1.41
ocsfe02 10.0.1.42

Figure above: Two-Armed topology

A one-armed topology is easier to deploy from a networking perspective, because it resides on the same network as the Front End Servers and does not introduce any additional changes in routing; however, not all traffic goes through the virtual IP address (VIP) of the load balancer (media between clients and conferencing servers, for example). If one function of the load balancer is to be a firewall between the corporate network and the load-balanced servers, the one-armed topology will not suffice. One benefit of this topology is that it is easier to the Front End Server functionality independent of the load balancer, because there is no dependency on the routing functionality of the load balancer.

Two-Armed Topology

A two-armed topology means that the Front End Servers reside behind the load balancer on a private network. The intent is to abstract away the Front End Servers from the main networking environment; however, these Front End Servers cannot truly be hidden by the load balancers VIPs alone, because clients need to contact the Front End Servers directly (for example, to establish media with conferencing servers), and the Front End Servers need to interact with other server components in the main network (for example, to look up Active Directory settings). As a result, the networking environment must be altered so that the load balancer is actually routing packets between the main and private networks. Additionally, the private network needs to use an IP address range that is routable within the corporation. This topology does enable the load balancer to be a single point of entry for all packets to and from the Front End Servers, and so performing firewall functionality is possible in the two-armed topology. The networking load will be considerably higher in the two-armed topology, because all traffic destined for the Front End Servers goes through the LOAD BALANCER.

Walkthrough: Enabling Load Balancer Routing

Enabling Load Balancer Routing (Two-Armed only)

If you are using a one-armed load balancer, the load balancer does not need to be in place to install and test the Front End Servers. You can skip this phase and proceed to Walkthrough: Configuring New Front End Servers. If you are using WebMux Transparent Mode with SNAT, you don’t have to enable routing either. Only NAT+SNAT need this step.

If you are using a Two-Armed load balancer topology, you will first need to enable routing between the public network and the private Front End network. To do so, connect the load balancer to both networks and configure the appropriate IP addresses. In our sample topology, the public network IP address is set to 10.0.0.40, and the private Front End network IP address is set to 192.168.0.40. Enable routing between these two networks, but do not configure any virtual IP addresses (VIPs) at this stage by going to “Setup” screen and check “Enable Routing”.

From each Front End Server, verify that you can telnet to the IP address of the Active Directory server on port 135. If this test is successful, you can install and test each Front End Server.

Walkthrough: Configuring New Front End Servers

Configuring New Front End Servers

You are now ready to set up the consolidated Front End Servers that will be used in the load balanced array. These servers will be added to your existing pool, and they will use the same back-end database. Follow the standard setup procedure for an Enterprise Edition server, and add each server to your existing pool. In our example, this procedure would involve the following steps:

1. Obtain a new physical server, assign it an IP address of 10.0.1.41, install the Windows Server 2008 or Windows 2003 operating system with the most current updates, and then install IIS.
2. Name the computer *ocsfe01.contoso.com*, and then add it to your enterprise DNS server.
3. Run the Office Communications Server 2007 R2 Enterprise Edition setup.
4. Select the option to add a consolidated Enterprise Edition server to an existing pool.

   Note:

   > If your existing server is a Standard Edition server, you will need to create an Enterprise Edition pool on the associated back-end database at this stage.

5. Select *ocspool.contoso.com* as the existing pool and add your new consolidated Front End Server to the pool.
6. When you request the certificate, set the Subject Name to *ocspool.contoso.com* and add *ocsfe01.contoso.com* and *sip.contoso.com* as Subject Alternate Names.
7. Run the Activation portion of the setup procedure on *ocsfe01* and during the wizard**, use the existing service account and start all Office Communications Server services.**
8. Configure IIS to allow load balancer FQDN loopback as described in Configuring IIS to Allow Load Balancer FQDN for Loopback in the Deploying Office Communications Server 2007 R2 Enterprise Edition documentation.
9. Repeat the above steps on a second physical server with IP address 10.0.1.42 and named *ocsfe02.contoso.com*.

Ensure the following requirements are met:

1. The two Front End Servers must be capable of routing to each other. There can be no NAT device in this path of communication. Any such device will prevent successful intrapool communication over RPC.
2. Front End Servers must have access to the Active Directory Domain Services environment.
3. Front End Servers must have static IP addresses that can be used to configure them in the load balancer. In addition, these IP addresses must have DNS registrations.

Walkthrough: Validating New Front End Servers

Validating New Front End Servers

At this stage, you have installed and activated two additional consolidated Front End Servers that are fully configured. Although there is no load balancer in place, the Front End Servers themselves should be fully functional and able to handle workloads. In this section, you will validate the basic functionality of your two new Front End Servers so that when you configure the load balancer VIP, you can focus your troubleshooting on the load balancer. In addition, this testing will not affect your production users.

To perform this testing, first add *a temporary host* file entry on each newly installed Front End Server, pointing the FQDN of the existing pool to the IP address of that particular Front End Server. You take this step because Front End Servers use the pool name to contact certain resources. Setting the pool FQDN to resolve to itself ensures that we are using only that server to handle all test scenarios. For instance, in our sample topology, follow these steps:

1. On *ocsfe01.contoso.com*, add a hosts file entry that resolves ocspool.contoso.com to 10.0.0.41.
2. In a command window, run *ipconfig /flushdns* to clear the DNS cache.
3. Restart all services on the Front End Server to ensure that they use the updated IP address.
4. Repeat the above steps on *ocsfe02.contoso.com*, using an IP address of 10.0.0.42.

Next, prepare three test workstations with the Office Communicator, Outlook, and Live Meeting clients and perform the following steps:

1. Add a hosts file entry on each client that resolves *ocspool.contoso.com* to 10.0.0.41.
2. In a command window, run *ipconfig /flushdns* to clear the DNS cache.
3. In Office Communicator, configure connection settings to Manual, specifying *ocspool.contoso.com* as the internal server name.
4. In the Live Meeting Client, configure connection settings to Manual, specifying *ocspool.contoso.com* as the internal server name.
5. Using the test clients, sign in to Office Communicator by using some test accounts that you create, and then verify that all peer-to-peer and conference modalities work, including IM, voice, video, and desktop sharing.
6. Use one of the test clients to schedule a Live Meeting conference by using Outlook. Join the meeting, and then verify that all Web conferencing modalities work.
7. Update the hosts file entry on each client to resolve *ocspool.contoso.com* to 10.0.0.42.
8. Repeat steps 2 through 6 to validate functionality on the second Front End Server. If any issues are discovered, troubleshoot and correct them before you configure a load balancer.

By completing these validation tasks, you can be more confident that when you add a load balancer to the topology, any unexpected behavior is likely to be caused by the Load Balancer and not by the Front End Servers themselves. The final step is to remove the test settings that you configured:

1. On *ocsfe01.contoso.com* and *ocsfe02.contoso.com*, remove the hosts file entry that resolved *ocspool.contoso.com*.
2. In a command window, run *ipconfig /flushdns* to clear the DNS cache.
3. Restart all services on the Front End Server to ensure they use the updated IP address.
4. On the test clients, remove the hosts file entry that resolved *ocspool.contoso.com*.
5. In a command window, run *ipconfig /flushdns* to clear the DNS cache.
6. Reset Office Communicator and Live Meeting clients to use automatic server lookup.
7. Restart the test clients to ensure that they do not use any cached DNS entries.

Walkthrough: Installing and Configuring Load Balancers

Installing and Configuring load balancers

Now that the Front End Servers are functioning correctly, you can install and configure your load balancer. First, confirm that your load balancer meets the requirements for Office Communications Server load balancers, which are described in Load Balancers for Office Communications Server 2007 R2 in the Technical Reference for Office Communications Server 2007 R2. Next, identify the load balancer network topology used in your environment:

1. **One-armed**: In this topology, the Office Communications Server Front End Servers and clients reside on a single network. The load balancer has a single arm, connected to this same network.
2. **Two-armed (routed)**: In this topology, the Office Communications Server Front End Servers and clients reside on two routable networks. The load balancer has two arms, one connected to the client network and the other connected to the network with the Office Communications Server Front End Servers.
3. **Two-armed(TM):** In this topology, OCS servers and clients reside in same network, but OCS servers are isolated from clients and other servers, all traffic must flow through WebMux. WebMux will load balance all related clients traffic, but let other network traffic flow like through a network cable. Wiring is similar to the above Two-Armed (routed) topology.
4. **Out-of-path mode**: This is a form of the one-armed topology in which the initial client request passes through the load balancer, but return traffic travels directly from the Front End Server to the client. This mode is not supported by Office Communications Server.

The mechanism to configure the load balancer varies by manufacturer, but the following core steps are required no matter which model is used:

Connect Load Balancer and Confirm Networking

If your environment uses a one-armed topology, connect the load balancer to the corporate network where your clients and Office Communications Server Front End Servers are located and configure a single virtual IP address (Farm IP/VIP). This static IP address is used by Office Communications Server clients and Front End Servers to provide a single entry point for connecting to resources provided by the Front End Server array. In our sample one-armed topology, the Farm IP is *10.0.0.40*.

If your environment uses a two-armed topology, connect one arm of the load balancer to the corporate network and the other arm to the network where your Office Communications Server Front End Servers reside. Next, configure a Farm IP on the client network and a corresponding IP address on the Front End network. In our sample two-armed topology, *10.0.0.40* is the pool Farm IP and *192.168.0.40* is the corresponding IP address on the Front End network.

Configure load balancer with Front End Pool Servers

Now that the Farm IP has been created, the next step is to configure it to point to the newly created Front End Servers. In our sample one-armed topology, the *10.0.0.40* Farm IP would be configured to use *10.0.1.41* and *10.0.1.42*.

Besides pointing to the IP addresses of the Front End Servers, a number of additional settings need to be enabled in the Farm configuration to ensure proper operation with Office Communications Server:

1. Ensure that the Farm adds the following TCP ports: 5061, 5060, 135, 443, 444, 5065, 5069, 5071, 5072, 5073, 5074, and 8404. It can be done by “add-addr/port” in the farm screen. For details about each port, see Load Balancers for Office Communications Server 2007 R2 in the Technical Reference for Office Communications Server 2007 R2.
2. Configure the Farm to use a TCP idle timeout of 30 minutes in “Setup” screen.
3. Configure the Farm to use a *weighted least connections* algorithm in choosing how to load balance incoming connection requests against the Front End Server array.
4. When add ports to the Farm select service as “generic TCP” for port 5061 and port 444, other ports select service “generic no healthcheck TCP”, each will check Front End Server on port 5061 and 444. This enables the load balancer to detect when one of the Front End Servers goes down and take that server out of the array.
5. Enable Source Network Address Translation (SNAT) in “Setup” screen. This means the load balancer uses one of its IP addresses as the source IP address when it sends the connection to one of the Front End Servers. In Destination Network Address Translation (DNAT) mode, the load balancer uses the source IP address of the endpoint that originated the connection when it sends a connection to one of the Front End Servers. DNAT mode only is not supported.

Note:

> Microsoft's original document indicates DNAT is not required and not supported. However, if the Destination address is not translated by the load balancer, the Front End server will not accept the clients' requests; due to the destination address in the data packet is still the VIP address. To address the server address, the load balancer must have both SNAT and DNAT enabled. DNAT alone will not work with OCS 2007 R2. If you do a TCP capture on the server, you should see both the Source IP address and Destination IP address are changed from the clients' original data packets.

6. If your pool will require more than 65,000 simultaneous connections, configure additional SNAT IP addresses on your VIP. This is because the load balancer would only have one IP address configured on the network containing the Office Communications Server Front End Servers. This would limit the load balancer to roughly 65,000 source ports when making connections to the Front End Server.

It is very important that all of the requirements and configurations listed above be completed correctly. When you are finished, you will validate the load balancer configuration.

Walkthrough: Validating Load Balancers

Validating Load Balancers

Before you use test clients to validate the newly configured load balancers, it is a good practice to use telnet to confirm that the load balancer is at least listening on the virtual IP address (VIP) and directing connections to both Front End Servers. To do so, perform the following steps:

1. On each Front End Server, run Netmon (or a similar protocol analysis tool) and add a filter to only display traffic that is coming in on TCP port 5061.
2. From a test client, telnet to the VIP on port 5061.
3. Close the telnet session and reestablish the telnet session several times. On each Front End Server, the Netmon trace should show TCP connections in an alternating fashion.

To validate the load balancer configuration by using test clients, we will employ similar configuration changes to the hosts file similar to those that were used to validate the Front End Server installation. The only difference is that the load balancer is now configured, and so we can point the pool FQDN to the load balancer Farm. This test simulates the behavior that will occur when the pool FQDN is updated in the production DNS server.

First, add *a temporary host* file entry on each newly installed Front End Server, pointing the FQDN of the existing pool to the IP address of the corresponding load balancer Farm. For instance, in our sample topology you would do the following:

1. On *ocsfe01.contoso.com*, add a hosts file entry that resolves ocspool.contoso.com to 10.0.0.40.
2. In a command window, run *ipconfig /flushdns* to clear the DNS cache.
3. Restart all services on the Front End Server to ensure that they use the updated IP address.
4. Repeat the above steps on *ocsfe02.contoso.com*, setting it to the same VIP address.

Next, prepare three test workstations with the Office Communicator, Outlook, and Live Meeting clients, and then perform the following steps:

1. Add a hosts file entry on each client that resolves *ocspool.contoso.com* to the corresponding load balancer VIP. In the sample topology, this would be 10.0.0.40.
2. In a command window, run *ipconfig /flushdns* to clear the DNS cache.
3. In Office Communicator, configure connection settings to Manual, specifying *ocspool.contoso.com* as the internal server name.
4. In the Live Meeting Client, configure connection settings to Manual, specifying *ocspool.contoso.com* as the internal server name.
5. Using the test clients, sign in to Office Communicator by using some test accounts and verify that all peer-to-peer and conference modalities work, including IM, voice, video, and desktop sharing.
6. Use one of the test clients to schedule a Live Meeting conference by using Outlook. Join the meeting and verify that all Live Meeting conference modalities work.
7. Update the hosts file entry on each client to resolve *ocspool.contoso.com* to 10.0.0.42, and then rerun *ipconfig /flushdns* to clear the DNS cache.
8. Repeat steps 2 through 6 to validate functionality on the second Front End Server.
9. If you discover any issues, troubleshoot and correct them before you proceed.

At this stage, you have validated that the functionality of the newly added load balancers works with the Front End Servers. By completing these validation tasks, you can be more confident when you perform the actual cutover to the production DNS. Before you do that, however, take some time to simulate failure of a Front End Server. For example, stop the Front End service, or pull the network cable. Identify the client behavior, and document the procedure for responding to such a situation. Clients typically use their inbuilt keep-alive mechanisms, client retry logic, and server reconnection randomizations to detect when their Front End Server goes down, and they will connect seamlessly to another Front End Server that is available behind the load balancer.

The final step is to remove the test settings that you configured:

1. On *ocsfe01.contoso.com* and *ocsfe02.contoso.com*, remove the hosts file entry that resolved *ocspool.contoso.com*.
2. In a command window, run *ipconfig /flushdns* to clear the DNS cache.
3. Restart all services on the Front End Server to ensure that they use the updated IP address.
4. On the test clients, remove the hosts file entry that resolved *ocspool.contoso.com*.
5. In a command window, run *ipconfig /flushdns* to clear the DNS cache.
6. Reset Office Communicator and Live Meeting clients to use automatic server lookup.

7. Restart the test clients to ensure that they do not use any cached DNS entries.

Walkthrough: Decommissioning Non-Load Balanced Servers

Cutting over and decommissioning old servers

You are now ready to cut over production users to the new load-balanced environment. To do so, update the pool FQDN in the production DNS to point to the virtual IP address (VIP) of the new load balancer. We recommend that you cut over at the start of a maintenance window in order to provide ample time for any client DNS caches to expire. There may be a period of time when some users may sign in to the old server and some to the new server. If any issues arise as a result of moving to the load-balanced pool, simply revert to the production DNS pool entry. After the load-balanced system has been functioning under load in a stable manner, you can decommission and repurpose the old single Front End Server.