CHAPTER 3

About a Cisco Unity Connection Cluster

This chapter contains the following sections:

How a Cluster Works in Cisco Unity Connection

Revised May 2009

The Cisco Unity Connection cluster feature provides high availability voice messaging through two Connection servers that are configured in a cluster. Under normal conditions, the Connection servers are both active so that:

- The cluster can be assigned a DNS name that is shared by the Connection servers.
- Clients such as email applications and the web tools available through the Cisco Personal Communications Assistant (PCA) can connect to either Connection server.
- Phone systems can send calls to either Connection server.
- Incoming phone traffic load is balanced between the Connection servers by the phone system, PIMG/TIMG units, or other gateways that are required for the phone system integration.

Each server in the cluster is responsible for handling a share of the incoming calls for the cluster (answering phone calls and taking messages). The server with Primary status is responsible for the following functions:

- Homing and publishing the database and message store, which are both replicated to the other server.

- Sending message notifications and MWI requests (the Connection Notifier service is activated).
- Sending SMTP notifications and VPIM messages (the Connection Message Transfer Agent service is activated).

When one of the servers stops functioning (for example, when it is shut down for maintenance), the remaining server assumes responsibility for handling all incoming calls for the cluster. The remaining server also assumes responsibility for the database and message store, which are both replicated to the other server when the connection and its functionality are restored.

When the server that stopped functioning is able to resume its normal functions and is activated, it resumes responsibility for handling its share of incoming calls for the cluster.

To monitor the status of the servers, the Connection Server Role Manager service runs in Cisco Unity Connection Serviceability on both servers. This service performs the following functions:

- Starts the applicable services on each server, depending on server status.
- Determines whether critical processes (such as voice message processing, database replication, and message store replication) are functioning normally.
- Initiates changes to server status when the server with Primary status is not functioning or when critical services are not running.

Note the following limitations when the publisher server is not functioning:

- If the Connection cluster is integrated with an LDAP directory, directory synchronization does not occur, although authentication continues to work when only the subscriber server is functioning. When the publisher server is functioning again, directory synchronization resumes.
- If a Digital Network includes the Connection cluster, directory updates do not occur, although messages continue to be sent to and from the cluster when only the subscriber server is functioning. When the publisher server is functioning again, directory updates resume.

Licenses for a Cluster

Added May 2009

A Cisco Unity Connection cluster requires a license for each Connection server. The license that has the MAC address of the publisher server must be installed on the publisher server. The license that has the MAC address of the subscriber server must be installed on the subscriber server.

For information on managing licenses, see the "Managing Licenses" chapter of the *System Administration Guide for Cisco Unity Connection Release 7.x* at http://www.cisco.com/en/US/docs/voice_ip_comm/connection/7x/administration/guide/7xcucsagx.html.

About the Publisher Server

The first Cisco Unity Connection server that is configured in the cluster is the publisher server. The Cluster Management page in Cisco Unity Connection Serviceability identifies the publisher server.

The publisher server assumes responsibility for publishing the database and message store when the cluster is functioning normally.

When the publisher server does not have Primary status (for example, when the administrator manually changes the status of the other server to Primary, which automatically changes the status of the publisher server to Secondary), the other server assumes responsibility for publishing the database and message store.

The publisher server cannot be removed from the cluster.

Server Status Functions in the Cluster

Revised May 2009

Each server in the cluster has a status that appears on the Cluster Management page of Cisco Unity Connection Serviceability. The status indicates the functions that the server is currently performing in the cluster, as described in Table 3-1.

Table 3-1 Server Status Functions in the Cisco Unity Connection Cluster

Server Status	Functions in Connection Cluster
Primary	• Publishes the database and message store, which are both replicated to the other server in the cluster. • Receives replicated data from the other server when that server is able to share its data. • Displays and accepts changes to the administrative interfaces (such as Connection Administration). This data is replicated to the other server in the cluster. • Answers phone calls and takes messages. • Sends message notifications and MWI requests (the Connection Notifier service is activated). • Sends SMTP notifications and VPIM messages (the Connection Message Transfer Agent service is activated). • Connects with clients such as email applications and the web tools available through the Cisco PCA. • When the cluster is functioning normally, the publisher server has Primary status. **Note** A server with Primary status cannot be deactivated.
Secondary	• Receives replicated data from the server with Primary status. Data includes the database and message store. • Replicates data to the server with Primary status. • Displays and accepts changes to the administrative interfaces (such as Connection Administration). The data is replicated to the server with Primary status. • Answers phone calls and takes messages. • Connects with clients such as email applications and the web tools available through the Cisco PCA. **Note** Only a server with Secondary status can be deactivated.

Table 3-1 Server Status Functions in the Cisco Unity Connection Cluster (continued)

Server Status	Functions in Connection Cluster
Deactivated	• Receives replicated data from the server with Primary status. Data includes the database and message store. • Displays and accepts changes to the administrative interfaces (such as Connection Administration). The data is replicated to the server with Primary status. • Does not answer phone calls or take messages. • Does not connect with clients such as email applications and the web tools available through the Cisco PCA.
Not Functioning	• Does not receive replicated data from the server with Primary status. • Does not replicate data to the server with Primary status. • Does not display the administrative interfaces (such as Connection Administration). • Does not answer phone calls or take messages. **Note** A server with Not Functioning status usually is shut down.
Starting	• Receives replicated data from the server with Primary status. Data includes the database and message store. • Replicates data to the server with Primary status. • Does not answer phone calls or take messages. **Note** This status lasts only a few minutes, after which the server takes the applicable status.
Replicating Data	• Sends and receives data from the cluster. • Temporarily does not answer phone calls or take messages. • Temporarily does not connect with clients such as email applications and the web tools available through the Cisco PCA. **Note** This status lasts only a few minutes, after which the previous status resumes for the server.
Split Brain Recovery	• *After detecting two servers with Primary status:* Assigns Primary status to the publisher server. • Updates the database and message store on the server that is determined to have Primary status. • Replicates data to the other server. • Temporarily does not answer phone calls or take messages. • Temporarily does not connect with clients such as email applications and the web tools available through the Cisco PCA. **Note** This status lasts only a few minutes, after which the previous status resumes for the server.

Server Assignments and Usage of Voice Messaging Ports

Revised May 2009

In a Cisco Unity Connection cluster, the servers share the same phone system integrations. Each server is responsible for handling a share of the incoming calls for the cluster (answering phone calls and taking messages).

Depending on the phone system integration, each voice messaging port is either assigned to a specific server or used by both servers. Table 3-2 describes the port assignments.

Table 3-2 Server Assignments and Usage of Voice Messaging Ports in the Cisco Unity Connection Cluster

Integration Type	Server Assignments and Usage of Voice Messaging Ports
Integration by Skinny Client Control Protocol (SCCP) with Cisco Unified Communications Manager or Cisco Unified Communications Manager Express	• The phone system is set up with twice the number of SCCP voice mail port devices that are needed to handle the voice messaging traffic. (For example, if 16 voice mail port devices are needed to handle all voice messaging traffic, 32 voice mail port devices must be set up on the phone system.) • In Cisco Unity Connection Administration, the voice messaging ports are configured so that half the number of the ports set up on the phone system are assigned to each server in the cluster. (For example, each server in the cluster has 16 voice messaging ports.) • On the phone system, a line group, hunt list, and hunt group are configured so that the subscriber server will answer most incoming calls for the cluster. • If one of the servers stops functioning (for example, when it is shut down for maintenance), the remaining server assumes responsibility for handling all incoming calls for the cluster. • When the server that stopped functioning is able to resume its normal functions and is activated, it resumes responsibility for handling its share of incoming calls for the cluster.
Integration through a SIP Trunk with Cisco Unified Communications Manager or Cisco Unified Communications Manager Express	• In Cisco Unity Connection Administration, half the number of voice messaging ports that are needed to handle voice messaging traffic are assigned to each server in the cluster. (For example, if 16 voice messaging ports are needed to handle all voice messaging traffic for the cluster, each server in the cluster is assigned 8 voice messaging ports.) • On the phone system, a route group, route list, and route pattern are configured to distribute calls equally to both servers in the cluster. • If one of the servers stops functioning (for example, when it is shut down for maintenance), the remaining server assumes responsibility for handling all incoming calls for the cluster. • When the server that stopped functioning is able to resume its normal functions and is activated, it resumes responsibility for handling its share of incoming calls for the cluster.

Table 3-2 Server Assignments and Usage of Voice Messaging Ports in the Cisco Unity Connection Cluster

Integration Type	Server Assignments and Usage of Voice Messaging Ports
Integration through PIMG/TIMG units	• The number of ports set up on the phone system is the same as the number of voice messaging ports on each server in the cluster so that the servers share all the voice messaging ports. (For example, if the phone system is set up with 16 voice messaging ports, each server in the cluster must have the same 16 voice messaging ports.) • On the phone system, a hunt group is configured to distribute calls equally to both servers in the cluster. • The PIMG/TIMG units are configured to balance the voice messaging traffic between the servers. • If one of the servers stops functioning (for example, when it is shut down for maintenance), the remaining server assumes responsibility for handling all incoming calls for the cluster. • When the server that stopped functioning is able to resume its normal functions and is activated, it resumes responsibility for handling its share of incoming calls for the cluster.
Other integrations that use SIP	• In Cisco Unity Connection Administration, half the number of voice messaging ports that are needed to handle voice messaging traffic are assigned to each server in the cluster. (For example, if 16 voice messaging ports are needed to handle all voice messaging traffic for the cluster, each server in the cluster has 8 voice messaging ports.) • On the phone system, a hunt group is configured to distribute calls equally to both servers in the cluster. • If one of the servers stops functioning (for example, when it is shut down for maintenance), the remaining server assumes responsibility for handling all incoming calls for the cluster. • When the server that stopped functioning is able to resume its normal functions, it resumes responsibility for handling its share of incoming calls for the cluster.

Requirements for a Cisco Unity Connection Cluster

For current Cisco Unity Connection cluster requirements, refer to *System Requirements for Cisco Unity Connection Release 7.x* at http://www.cisco.com/en/US/docs/voice_ip_comm/connection/7x/requirements/7xcucsysreqs.html.

Effects on Calls in Progress When Server Status Changes

When the status of a Cisco Unity Connection server changes, the effects on calls in progress depend on the final status of the server that is handling a call and on the condition of the network. Table 3-3 describes the effects.

Table 3-3 Effects on Calls in Progress When Server Status Changes

Status Change	Effects
Primary to Secondary	When the status change is initiated manually, calls in progress are not affected. When the status change is automatic, effects on calls in progress depend on the critical service that stopped.
Secondary to Primary	When the status change is initiated manually, calls in progress are not affected. When the status change is automatic, effects on calls in progress depend on the critical service that stopped.
Secondary to Deactivated	Calls in progress are dropped. To prevent dropped calls, on the Cluster Management page in Cisco Unity Connection Serviceability, click Stop Taking Calls for the server, wait until all calls have ended, and deactivate the server.
Primary or Secondary to Replicating Data	Calls in progress are not affected.
Primary or Secondary to Split Brain Recovery	Calls in progress are not affected.

If network connections are lost, then calls in progress may be dropped, depending on the nature of the network problem.

Effects on Cisco Unity Connection Web Applications When Server Status Changes

Normal functioning of the following web applications are not affected when the server status changes:

- Cisco Unity Connection Administration
- Cisco Unity Connection Serviceability
- Cisco Unity Connection web tools accessed through the Cisco PCA (the Cisco Unity Assistant, the Cisco Unity Inbox, and the Cisco Unity Personal Call Transfer Rules web tools)

Effects of Stopping a Critical Service on a Server

Revised May 2009

Critical services are necessary for the normal functioning of the Cisco Unity Connection system. The effects of stopping a critical service depend on the server and its status. Table 3-4 describes the effects.

Table 3-4 Effects of Stopping a Critical Service on a Server

Server	Effects
Publisher	• When the server has Primary status, stopping a critical service in Cisco Unity Connection Serviceability causes the server status to change to Secondary and degrades the ability of the server to function normally. The status of the subscriber server changes to Primary if it does not have the Disabled or Not Functioning status. • When the server has Secondary status, stopping a critical service in Cisco Unity Connection Serviceability degrades the ability of the server to function normally. The status of the servers does not change.
Subscriber	When the server has Primary status, stopping a critical service in Cisco Unity Connection Serviceability degrades the ability of the server to function normally. The status of the servers does not change.

Effects of a Split-Brain Condition

Revised May 2009

When the servers in a Cisco Unity Connection cluster have Primary status at the same time (for example, when the servers have lost their connection with each other), both servers handle incoming calls (answer phone calls and take messages), send message notifications, send MWI requests, and accept changes to the administrative interfaces (such as Connection Administration). However, the servers do not replicate the database and message store to each other and do not receive replicated data from each other.

When the connection between the servers is restored, the status of the servers temporarily changes to Split Brain Recovery while the data is replicated between the servers and MWI settings are coordinated. When the recovery process is complete, the publisher server has Primary status and the other server has Secondary status.

Events When Server Status Changes

This section describes the events that take place when server status changes in the following situations:

- Automatic Change of Server Status Initiated by Server with Primary Status, page 3-8
- Automatic Change of Server Status Initiated by Server with Secondary Status, page 3-9
- Manual Change of Server Status Initiated by Administrator, page 3-9

Automatic Change of Server Status Initiated by Server with Primary Status

Revised May 2009

1. The Connection Server Role Manager service on the server with Primary status detects an unrecoverable failure (for example, the database fails or a critical service is stopped).
2. The Connection Server Role Manager service on the server with Primary status notifies the Connection Server Role Manager service on the other server to change its status.
3. The Connection Server Role Manager service on both servers posts alarms that it is initiating a change of status.

4. The Connection Server Role Manager service on the server with Primary status sets its status in the database to Secondary.
5. The Connection Server Role Manager service on the other server (the server that originally had Secondary status) sets its status in the database to Primary.
6. The Connection Server Role Manager service on the server that now has Primary status starts the critical services on that server.
7. The data connector detects the changed server status and sets the connections to use the database on the server that now has Primary status.
8. If possible, database and message store replication continues between the servers.
9. The Connection Server Role Manager service on the server that now has Primary status posts an alarm that the change of status is complete.

Automatic Change of Server Status Initiated by Server with Secondary Status

1. The Connection Server Role Manager service on the server with Secondary status does not receive contact from the Connection Server Role Manager service on the server with Primary status.
2. The Connection Server Role Manager service on the server with Secondary status confirms its network connection by pinging the local host and other known remote servers.
3. If the network connection is confirmed, the Connection Server Role Manager service on the server with Secondary status posts an alarm that it is initiating a change of status.

   If the network connection is not available, the status does not change and the remaining events do not occur.
4. The Connection Server Role Manager service on the server with Secondary status sets its status in the database to Primary.
5. The Connection Server Role Manager service on the server that now has Primary status starts the critical services on that server.
6. The data connector detects the changed status and sets the connections to use the database on the server that now has Primary status.
7. If possible, database and message store replication continues between the servers.
8. The Connection Server Role Manager service on the server that now has Primary status posts an alarm that the change of status is complete.

Manual Change of Server Status Initiated by Administrator

Revised May 2009

1. In Cisco Unity Connection Serviceability, the administrator manually initiates a change of server status.
2. The Connection Server Role Manager service on the server with Secondary status notifies the Connection Server Role Manager service on the server with Primary status to initiate change of status.
3. The Connection Server Role Manager service on the both servers posts alarms that the change of status is being initiated.
4. The Connection Server Role Manager service on the server with Primary status sets its status in the database to Secondary.

5. The Connection Server Role Manager service on the other server (the server that originally had Secondary status) sets its status in the database to Primary.
6. The Connection Server Role Manager service on the server that now has Primary status starts the critical services on that server.
7. The data connector detects the changed status and sets the connections to use the database on the server that now has Primary status.
8. Database and file replication continues between the servers.
9. The Connection Server Role Manager service on the server that now has Primary status posts an alarm that the change of status is complete.