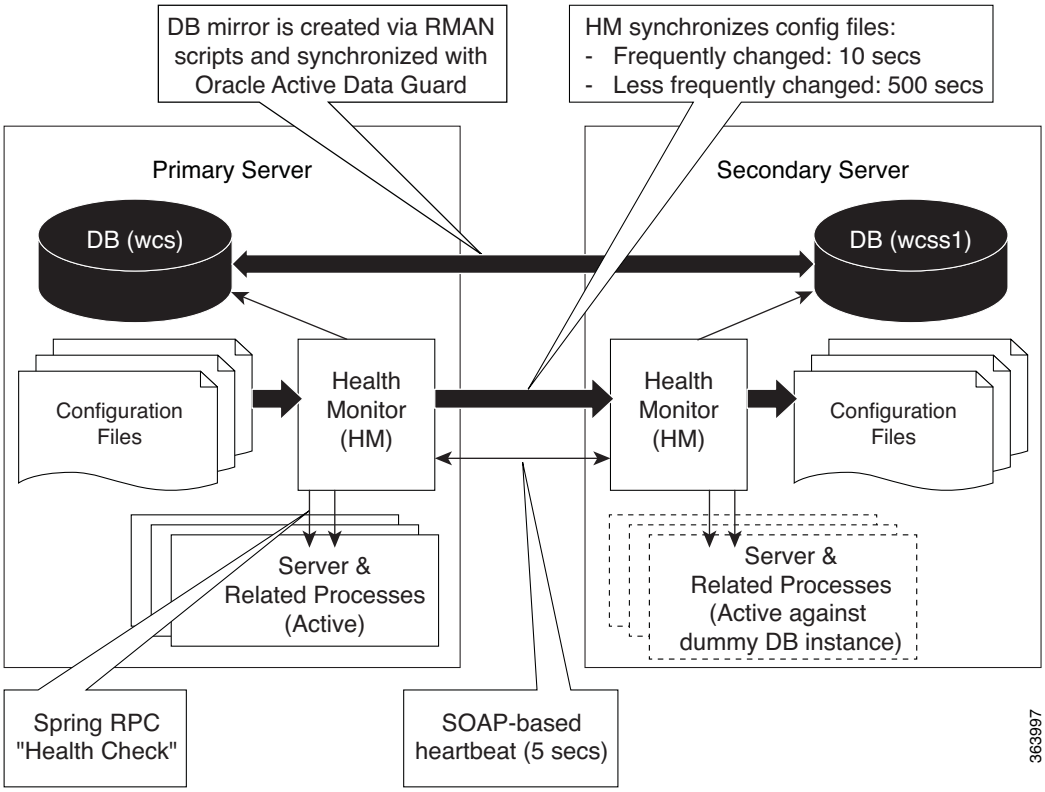
High Availability

How High Availability Works

The Cisco EPN Manager high availability (HA) framework ensures continued system operation in case of failure. HA uses a pair of linked, synchronized Cisco EPN Manager servers to minimize or eliminate the impact of application or hardware failures that may take place on either server.

The following figure shows the main components and process flows for an HA setup.



An HA deployment consists of a primary and a secondary server with Health Monitor (HM) instances (running as application process) on both servers. When the primary server fails (either automatically or because it is manually stopped), the secondary server takes over and manages the network while you restore access to the primary server. When the primary is available again, you can initiate a failback operation. Each primary server must have one dedicated secondary server.

Whenever the HA configuration determines that the primary server has changed, it synchronizes this change with the secondary server. These changes are of two types:

- File changes, which are synchronized using the HTTPS protocol. This includes items such as report configurations, configuration templates, TFTP-root directory, administration settings, licensing files, and the key store. File synchronization is done in:
  - Batches, for files that are not updated frequently, such as license files. These files are synchronized once every 500 seconds.
  - Near real-time, for files that are updated frequently. These files are synchronized once every 10 seconds.
- Database changes, such as updates related to configuration, performance and monitoring data. Oracle Recovery Manager (RMAN) creates the initial standby database and Oracle Active Data Guard synchronizes the databases when there is any change.

The primary and secondary HA servers exchange the following messages to maintain synchronization between the two servers:

- Database Sync: Includes all the information necessary to ensure that the databases on the primary and secondary servers are running and synchronized.
- File Sync: Includes frequently updated configuration files. These are synchronized every 10 seconds, while other infrequently updated configuration files are synchronized every 500 seconds.

- Process Sync: Ensures that application- and database-related processes are running. These messages fall under the Heartbeat category.
- Health Monitor Sync: These messages check for the network, system, and health monitor failure conditions.

Register the Secondary Server on the Primary Server

While the primary server needs no configuration during installation in order to participate in the HA configuration, you must register the secondary server on the primary server.

After the registration is initiated, Cisco EPN Manager begins copying all database and configuration data from the primary to the secondary server. The length of this process depends on the amount of database and configuration data, and the available bandwidth on the network link between the two servers. For a relatively fresh server (in operation for a few days) with 100 devices and network bandwidth of about 100 Obit/sec and a network latency of 10-40 ms, copying will take approximately 25 minutes.

Before You Begin

Make sure you have the following information:

- IP address or host name of the secondary server
- Authentication key from secondary server installation
- An e-mail address for HA state change notifications
- The preferred failover type (manual or automatic)

Note For Failover Type, we recommend that you use manual to avoid failover caused by intermittent network outages. For more information, see the high availability guidelines in the *Cisco Evolved Programmable Network Manager 1.1 Quick Start Guide*.

Step 1 Log in to the Cisco EPN Manager web GUI with a user ID that has administrator privileges.

Step 2 Choose **Administration > System Settings > High Availability** to display the HA status page.

Note You must have access to ROOT-DOMAIN to access the pages under Administration > System Settings.

Step 3 Select **HA Configuration** and then complete the Secondary Server, Authentication Key, Email Address, and Failover Type fields. You can enter multiple comma-separated addresses in the E-mail Address field.

Step 4 If you are using the virtual IP feature, check the **Virtual IP** check box, and then enter the virtual IPv4 or IPv6 address you want both servers to use. Remember that the servers must be on the same subnet to use virtual IP addresses.

Step 5 Click **Save** to save your changes and initiate the HA registration process. The primary and secondary server state will go through the following state transitions:

- Primary server: **Stand Alone** -> **HA Initializing** -> **Primary Active**
- Secondary server: **Stand Alone** -> **HA Initializing** -> **Secondary Syncing**

Step 6 On the HA Configuration page, ensure that the Configuration Mode field displays the value **HA Enabled** to verify that the registration is successful. You can then log in to the Health Monitor.

If HA registration fails, the HA state changes from **HA Initializing** to **Stand Alone** for both primary and secondary servers.

To recover from failed HA registration:

1. Use ping and other tools to check the network connection between the two Cisco EPN Manager servers. Confirm that the secondary server is reachable from the primary, and vice versa.
2. Check that the gateway, subnet mask, virtual IP address (if configured), server hostname, DNS, and NTP settings are all correct.
3. Check that the configured DNS and NTP servers are reachable from the primary and secondary servers, and that both are responding without latency or other network-specific issues.
4. Check that all Cisco EPN Manager licenses are correctly configured.

Once you have remedied any connectivity or setting issues, try the steps in Register the Secondary Server on the Primary Server again.

After registration is initiated, Cisco EPN Manager restarts the database process. While it is restarting, the database will be off line. When the database restarts, the primary and secondary servers and databases are synchronized. Users may observe slow system response until the synchronization is complete. The length of the synchronization is determined by the total database size. Synchronization is handled at the Oracle level by the Oracle Active Data Guard. There is no impact on the execution of user- or system-related activity during the synchronization. During re-registration, Cisco EPN Manager performs a full database replication to the secondary server.

Configure an SSO Server in High Availability Environment

Single Sign-On (SSO) Authentication is used to authenticate and manage users in a multi-user, multi-repository environment. SSO is responsible for storing and retrieving the credentials that are used for logging into different systems. You can set up a Cisco EPN Manager as the SSO server for other instances of Cisco EPN Manager.

To configure an SSO server in the high-availability environment, choose one of the procedures listed in the Table 26-1. See these topics for more information:

- To configure the SSO server, see Add the RADIUS, TACACS+, or SSO Server.
- To configure the high availability servers, see the *Cisco Evolved Programmable Network Manager 1.1 Quick Start Guide*.

Table 26-1 SSO Configuration in a High Availability Deployment

SSO Configuration	Setup SSO Server	Sever Failover Scenario	SSO Server Failure Scenario
SSO as a standalone server	1. Configure the standalone SSO server. 2. Configure the primary and secondary high availability servers.	When the primary server fails, the secondary server is activated. All machines that are connected to the primary server will be redirected to the secondary server.	When the SSO server fails, SSO functionality is disabled. Cisco EPN Manager will use local authentication.
SSO on the secondary Server	1. Configure one server to be the SSO server and the primary server (in other words, the primary server will also be the SSO server). 2. Configure the secondary high availability server.	When the primary server fails, the secondary server is activated. All machines that are connected to primary server will *not* be redirected to the secondary server (because SSO is configured on the primary server).	When the SSO (primary) server fails, the secondary server can be set as the failback option for SSO. This enables all instances to connect to the secondary server. If the secondary is not set to be the SSO server failback option, Cisco EPN Manager will use local authentication.

Launch the Health Monitor

The Health Monitor is one of the main components that manage the HA operations. Health Monitor instances run on both servers as an application process. It performs the following functions:

- Synchronizes database and configuration data related to HA (this excludes databases that synchronize separately using Oracle Data Guard).
- Exchanges heartbeat messages between the primary and secondary servers every 5 seconds, to ensure communications are maintained between the servers.
- Checks the available disk space on both servers at regular intervals, and generates events when storage space runs low.
- Manages, controls and monitors the overall health of the linked HA servers. If there is a failure on the primary server, the Health Monitor activates the secondary server.

Once you have completed HA registration successfully, you can access the Health Monitor web page from the primary or secondary server by entering the following URL on your browser:

`https://ServerIP:8082`

where *ServerIP* is the primary or secondary server's IP address or host name.

Trigger Failover

Failover activates the secondary server in response to a detected failure on the primary server.

The Health Monitor detects failure conditions using the heartbeat messages exchanged between the two servers. If the primary server is not responsive to three consecutive heartbeat messages from the secondary server, the Health Monitor deems the primary server to have failed. During the health check, Health Monitor also checks the application process status and database health. If there is no proper response to these checks, these are also treated as having failed.

LIMITED ORDERABILITY RELEASE

The HA system takes approximately 10 to 15 seconds to detect a process failure on the primary server and initiate a failover. If the secondary server is unable to reach the primary server due to a network issue, it might take more time to initiate a failover. In addition, it may take additional time for the application processes on the secondary server to be fully operational.

As soon as the Health Monitor detects the failure, it sends an email notification. The email includes the failure status along with a link to the secondary server's Health Monitor web page. If HA is configured for automatic failover, the secondary server will activate automatically.

To perform a manual failover:

Step 1 Access the secondary server's Health Monitor web page using the web link given in the email notification, or by entering the following URL on your browser:

`https://ServerIP:8082`

Step 2 Click **Failover**.

Failover Scenarios

- Network is Down
- Process Restart Fails
- Primary Server Restarts During Synchronization
- Secondary Server Restarts During Synchronization
- Both HA Servers Are Down

Network is Down

If there is a loss of network connectivity between the primary and secondary servers, you will get email notifications that each server has lost connectivity to the other server.

If Automatic Failover is Configured:

After the failover is complete, you will receive an email notification that the secondary server is now active.

1. Check on and restore network connectivity between the two servers.
2. Restart the primary server.
3. Trigger a failback from the secondary to the primary server (see Trigger Failback).

If Manual Failover is Configured:

1. Check on and restore network connectivity between the two servers.
2. Use the Health Monitor web page for the secondary server to trigger a failover from the primary to the secondary server (see Launch the Health Monitor).
3. Once you have received email notification that the secondary is now active, trigger a failback from the secondary to the primary server (see Trigger Failback).

LIMITED ORDERABILITY RELEASE

Process Restart Fails

The Cisco EPN Manager Health Monitor process is responsible for attempting to restart any Cisco EPN Manager server processes that have failed. The current state of the primary and secondary servers should be **Primary Active** and **Secondary Syncing** at the time any such failures occur.

If Health Monitor cannot restart a critical process on the primary server, then the primary server is considered to have failed. You will receive an email notification of this failure.

If Automatic Failover is Configured:

After the failover is complete, you will receive an email notification that the secondary server is now active.

1. Restart the primary server and ensure that it is running. When the primary server is up, its HA state will be **Standalone**.
2. Trigger a failback from the secondary to the primary server (see Trigger Failback).

If Manual Failover is Configured:

1. Trigger a failover from the primary to the secondary (see Trigger Failover).
2. Restart the primary server and ensure that it is running. When the primary server is up, its HA state will be **Standalone**.
3. Trigger a failback from the secondary to the primary (see Trigger Failback).

Primary Server Restarts During Synchronization

If Manual Failover is Configured:

If the primary server is restarted while the secondary server is synchronizing, the **Standalone** and the **HA Initializing** states occur immediately after the primary comes back online. No administrator response should be required.

If Automatic Failover is Configured:

If the primary server is restarted while the secondary server is synchronizing, you must trigger a failback from the secondary to the primary (see Trigger Failback).

Secondary Server Restarts During Synchronization

If the secondary server is restarted while syncing with the primary server, you will see the same state transitions regardless of the Failover Type settings. No administrator response should be required.

Both HA Servers Are Down

If both the primary and secondary servers are shut down at the same time, you can recover by bringing them back up in the correct order.

Follow these instructions for both failover types:

1. Restart the secondary server and the instance of Cisco EPN Manager running on it.
2. When Cisco EPN Manager is running on the secondary, access the secondary server's Health Monitor web page You will see the secondary server transition to the state **Secondary Lost Primary**.

LIMITED ORDERABILITY RELEASE

3. Restart the primary server and the instance of Cisco EPN Manager running on it.
4. When Cisco EPN Manager is running on the primary, access the primary server's Health Monitor web page. The states of the primary and secondary servers will be **Primary Active** and **Secondary Syncing**.

Trigger Failback

Failback is the process of re-activating the primary server once it is back online. It also transfers Active status from the secondary server to the primary, and stops active network monitoring processes on the secondary.

When a failback is triggered, the secondary server replicates its current database information and updated files to the primary server. The time it takes to complete the failback from the secondary server to the primary server will depend on the amount of data that needs to be replicated and the available network bandwidth.

Once the data has begun replicating successfully, HA changes the state of the primary server to Primary Active and the state of the secondary server to Secondary Syncing.

During failback, the secondary server is available except during the period when processes are started on the primary and stopped on the secondary. Both servers' Health Monitor web pages are accessible for monitoring the progress of the failback. Additionally, users can also connect to the secondary server to access all normal functionality, except for these caveats:

- Do not initiate configuration or provisioning activity while the failback is in progress.
- Be aware that, after a successful failback, the secondary server will go down and control will switch over to the primary server. During this process, Cisco EPN Manager will be inaccessible to the users for a few moments.

You must always trigger failback manually, as follows:

Step 1 Access the secondary server's Health Monitor web page using the link given in the email notification, or by entering the following URL on your browser:

`https://ServerIP:8082`

Step 2 Click **Failback**.

Respond to a Split Brain Scenario

In a split-brain scenario, both the primary and secondary servers become active at the same time, perhaps due to a network outage or link that temporarily goes down. However, because the primary server constantly checks the secondary server, when the connection is reestablished, the primary server will go down due to the secondary server being active.

Use the primary server and its newly-added data. When the network comes up, the primary server will go down and its HA status will be **Primary Failover**. Do the following:

1. Restart the primary server. Its status changes to **Primary Alone**.
2. Remove the HA configuration, then reregister the secondary server with the primary server.

Use the secondary server and its newly-added data. When the network comes up, the primary server will go down and its HA status will be **Primary Failover**. Do the following:

1. Confirm that a user can log into the Cisco EPN Manager GUI on the secondary server (for example, https://*ip-address*:443). Do not proceed until this access has been verified.
2. Initiate a failback from the secondary server's Health Monitor web page. Users can continue to perform monitoring activities on the secondary server until the switchover to the primary is completed.

Check High Availability States, Transitions, and Events

To view HA states, choose **Administration** > **High Availability** > **HA Status**, from the left sidebar. The current HA status and the HA states for the events are displayed.

The following table lists the HA states, including those that require no response from you. For a list of HA events and instructions for enabling, disabling, and adjusting them, see Control SNMP Traps That Indicate System Problems.

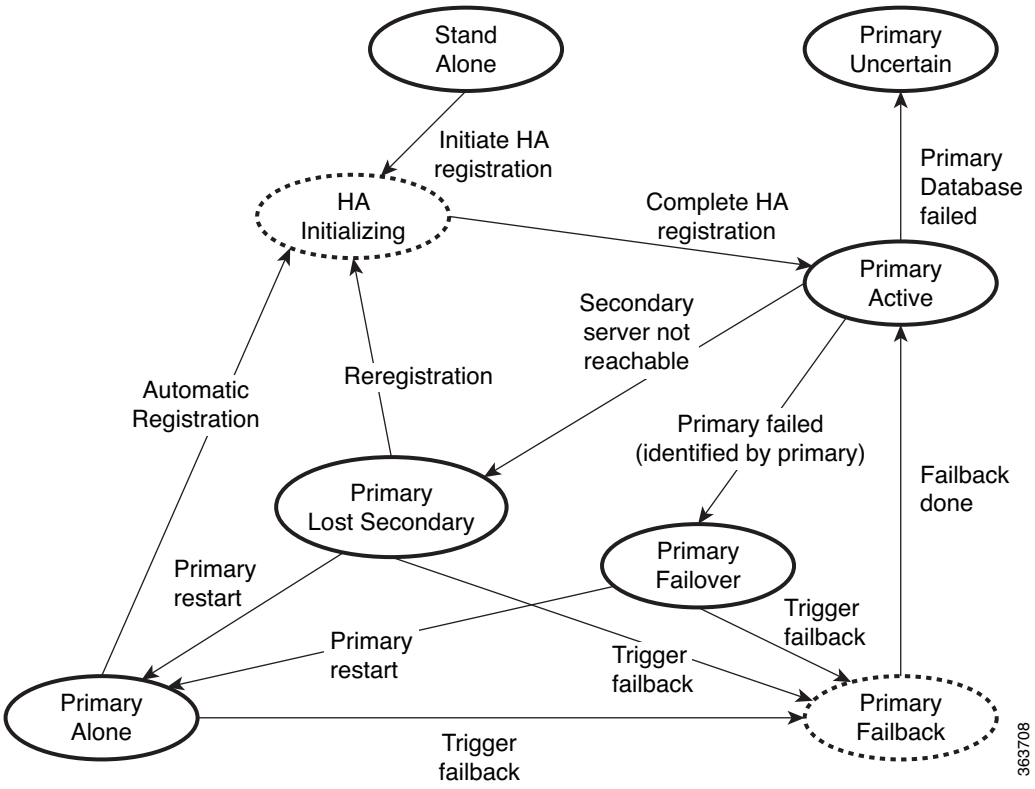
Table 26-2 High Availability States

State	Server	Description
Stand Alone	Both	HA is not configured on this Cisco EPN Manager server
Primary Alone	Primary	Primary restarted after it lost secondary (only Health Monitor is running in this state).
HA Initializing	Both	HA Registration process between the primary and secondary server has started.
Primary Active	Primary	Primary server is now active and is synchronizing with secondary server.
Primary Failover	Primary	Primary server detected a failure.
Primary Failback	Primary	User-triggered failback is currently in progress.
Primary Lost Secondary	Primary	Primary server is unable to communicate with the secondary server.
Primary Uncertain	Primary	Primary server's application processes are not able to connect to its database.
Secondary Alone	Secondary	Primary server is not reachable from secondary server after a primary server restart.
Secondary Syncing	Secondary	Secondary server is synchronizing the database and configuration files from the primary.
Secondary Active	Secondary	Failover from the primary server to the secondary server has completed successfully.
Secondary Lost Primary	Secondary	Secondary server is not able to connect to the primary server (occurs when the primary fails or network connectivity is lost). For automatic failover, the secondary will automatically move to the **Secondary Active** state. For Manual failover, you must trigger the failover to make the secondary active (see Trigger Failover).
Secondary Failover	Secondary	Failover triggered and is in progress.
Secondary Failback	Secondary	Failback triggered and database and file replication is in progress.

Table 26-2 High Availability States (continued)

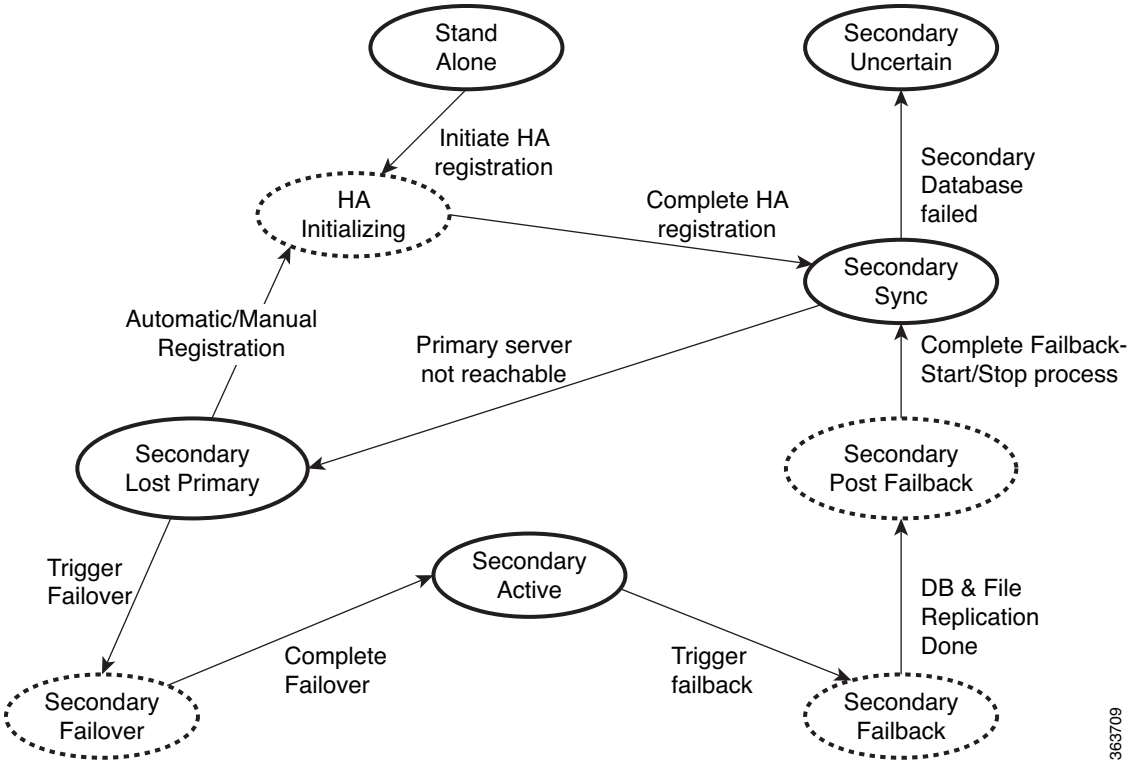
State	Server	Description
Secondary Post Failback	Secondary	Failback triggered; associated process stops and restarts are in progress. Database and configuration files have been replicated from the secondary server to the primary server. The primary server status will change to **Primary Active**, and the secondary server HA status will change to **Secondary Syncing**.
Secondary Uncertain	Secondary	Secondary server's application processes cannot connect to the server's database.

The following figure illustrates the primary server HA state changes.



LIMITED ORDERABILITY RELEASE

This figure illustrates the secondary server HA state changes.



High Availability CLI Command Reference

The following table lists the CLI commands available for HA management.

You must be logged in as the admin CLI user to use these commands. The output reflects the status of the server you are using. In other words, if you run **ha status** from the primary server, Cisco EPN Manager reports the status of the primary server.

Table 26-3 High Availability Commands

Command	Description
ha?	Display the command usage message
ha authkey *newAuthkey*	Update the authentication key to *newAuthKey*
ha remove	Remove the HA configuration
ha status	Get the current status for HA

LIMITED ORDERABILITY RELEASE

Reset the Authentication Key

Users with Administrator privileges can change the HA authentication key using the **ha authkey** command. You will need to ensure that the new authorization key meets the password standards.

Step 1 Connect to the primary server via CLI. Do not enter config mode (see Establish an SSH Session With the Cisco EPN Manager Server).

Step 2 Enter the following at the command line:

```
admin# ha authkey newAuthKey
```

Where *authkeyauthkey* is the new authorization key.

Using HA Error Logging

To save disk space and maximize performance, HA error logging is disabled by default. If you are having trouble with HA, the best place to begin is by enabling error logging and to examine the log files.

Step 1 Launch the Health Monitor on the server that is having trouble (see Launch the Health Monitor).

Step 2 In the Logging area, in the **Message Level** drop-down, select the error-logging level you want.

Step 3 Click **Set Logging**.

Step 4 When you want to download the log files, click **Download** in the Logs area. You can open the downloaded log files using any ASCII text editor.

Reset the Server IP Address or Host Name

Avoid changing the IP address or hostname of the primary or secondary server, if possible. If you must change the IP address or hostname, remove the HA configuration from the primary server before making the change. When finished, re-register HA.