Round-Robin Load Balancing

VMware ESX Server 3.5, VMware ESX Server 3i version 3.5, VMware VirtualCenter 2.5

VMware® ESX Server 3.5 and ESX Server 3i version 3.5 enhance ESX Server native multipathing by providing experimental support for round-robin load balancing. This technical note explains how round-robin load balancing works and how to set it.

NOTE   Round-robin load balancing is experimental and not supported for production use.

Round-robin load balancing is available from the VI Client but is marked as experimental.

Understanding Round-Robin Load Balancing

ESX Server hosts can use multipathing for failover. When one path from the ESX Server host to the SAN becomes unavailable, the host switches to another path. ESX Server hosts can also use multipathing for load balancing. To achieve better load balancing across paths, administrators can specify that the ESX Server host should switch paths under certain circumstances. Different settable options determine when the ESX Server host switches paths and what paths are chosen.

- **When to switch** – Specify that the ESX Server host should attempt a path switch after a specified number of I/O blocks have been issued on a path or after a specified number of read or write commands have been issued on a path. If another path exists that meets the specified path policy for the target, the active path to the target is switched to the new path. The `--custom-max-commands` and `--custom-max-blocks` options specify when to switch.

- **Which target to use** – Specify that the next path should be on the preferred target, the most recently used target, or any target. The `--custom-target-policy` option specifies which target to use.

- **Which HBA to use** – Specify that the next path should be on the preferred HBA, the most recently used HBA, the HBA with the minimum outstanding I/O requests, or any HBA. The `--custom-HBA-policy` option specifies which HBA to use.

Setting the Path Switching Policy

You can set the path-switching policy for failover and for load balancing by using the `esxcfg-mpath` command.

NOTE   You can set path switching for failover in the VI Client. See the *SAN Configuration Guide*. You can set load balancing only from the command line.

You can set the path switching policy on a per-LUN basis by using the `esxcfg-mpath` command's `--policy custom` option. If you specify `--policy custom`, you must also specify one of the custom policy options. Because the path switching policy is set on a per-LUN basis, you must always specify the LUN using the `--lun` option.

The following options are supported and discussed in more detail in Table 1.

```
esxcfg-mpath --lun <lun> [--policy [mru|rr|fixed|custom]]
                         [--custom-hba-policy|-H [mru|preferred|any|minq]] |
                         [--custom-target-policy|-T [mru|preferred|any]] |
                         [--custom-max-commands|-C <max_commands>] |
                         [--custom-max-blocks|-B <max-blocks>]
```

Table 1. Policy Options for esxcfg-mpath

Option	Description
--policy -p [mru|rr|fixed|custom]	Set the policy for a given LUN to mru, rr, fixed, or custom. This option requires that the --lun option is also passed to indicate the LUN to set the policy for. ■ Most recently used (mru) selects the path most recently used to send I/O to a device. ■ Round robin (rr) uses the mru target selection policy and the any HBA selection policy to select paths. ■ Fixed (fixed) uses only the active path. ■ Custom (custom) sets the LUN to expect a custom policy. **NOTE** After you've set the policy to custom for a LUN, you can specify one or more of the other options. For any LUN, you set the policy to custom only once.
--custom-hba-policy -H	Selection policy for the HBA. Use one of the following: ■ preferred — Use the HBA that's part of the preferred path. ■ mru – Use the most recently used HBA. ■ minq – HBA that has the smallest number of outstanding I/O commands when the ESX Server system does the path selection. ■ any – Use any available HBA.
--custom-target-policy -T	Selection policy for the target. Use one of the following: ■ preferred – Use the target that is part of the preferred path. ■ mru – Use the most recently used target. ■ any – Use any available target.
--custom-max-commands -C	Maximum number of I/O requests to be issued on a given path before the system tries to select a different path. A setting of zero (0) disables switching based on commands. Specify the number of commands as an integer. Default is 50.
--custom-max-blocks -B	Maximum number of I/O blocks to be issued on a given path before the system tries to select a different path. A setting of zero (0) disables switching based on blocks. Specify the number of blocks as an integer. Default is 2048.

Notes

If you set the custom-max-blocks and custom-max-commands, options, the system attempts to switch paths as soon as one of the limits is reached.

If you set the target or the HBA policy to preferred, the system chooses the target or the HBA of the preferred path when possible. If a preferred policy is set on an active/passive SAN array, and the preferred target is not on the active SP (Storage Processor), the system does not select the preferred target but a target on the active SP.

Path switching is not performed if an outstanding SCSI reservation is on the target, or if a path failover is underway. Path switching is delayed until an I/O request is performed when no reservations or path failovers are pending.

Examples

`esxcfg-mpath --lun=vmhba0:0:0 -p custom`	Sets this LUN to use a custom policy. After the policy has been set to `custom` for a LUN, you do not have to include the `-p custom` option when you further specify the custom policy.
`esxcfg-mpath --lun=vmhba0:0:0 -H any -B 4000`	Changes to a different HBA after 4000 blocks have been issued on the current path. Uses any available HBA.
`esxcfg-mpath -q --lun=vmhba0:0:0`	Displays the following.

```
Disk vmhba0:0:0 /dev/sda (34732MB) has 1 paths and policy of
Custom: maxCmds=50 maxBlks=2048 hbaPolicy=any
targetPolicy=mru Local 1:4.0 vmhba0:0:0 On active preferred
```

`# esxcfg-mpath --lun=vmhba0:0:0 --custom-max-commands=100`	Sets to 100 the maximum number of commands that can be issued before the system switches to a different path.
`# esxcfg-mpath -q --lun=vmhba0:0:0`	Displays the following (as a result of the preceding command).

```
Disk vmhba0:0:0 /dev/sda (34732MB) has 1 paths and policy of
Custom: maxCmds=100 maxBlks=2048 hbaPolicy=any
targetPolicy=mru Local 1:4.0 vmhba0:0:0 On active preferred
```

Global Disk Configuration Options

Two global disk configuration options determine the default settings for `--custom-max-blocks` and `--custom-max-commands` when the round-robin policy is set:

- `SPBlksToSwitch` — Number of blocks sent over a given path before a path switch.

- `SPCmdsToSwitch` — Number of I/O commands sent over a given path before a path switch.

You can use `esxcfg-advcfg` to view and set these options, for example:

```
# esxcfg-advcfg -g /Disk/SPCmdsToSwitch
Value of SPCmdsToSwitch is 50

# esxcfg-advcfg -s 200 /Disk/SPCmdsToSwitch
Value of SPCmdsToSwitch is 200

# esxcfg-advcfg -g /Disk/SPCmdsToSwitch
Value of SPCmdsToSwitch is 200
```