Understanding Redundancy in Dell PowerEdge Blade Servers

When administrators plan blade server deployment, they should carefully review redundant system configuration options. Depending on the data center’s design and IT policies, blade server systems may require redundancy at different levels. Dell™ PowerEdge™ blade servers offer redundancy options for various functions—ranging from power redundancy to chassis-management redundancy.

By Narayan Devireddy and Sanjeev S. Singh

By consolidating servers, infrastructure components, and management within a single chassis, Dell PowerEdge blade servers can help achieve high efficiency in the data center and provide an optimized rack environment. Blade server deployment requires careful planning of data center resources such as power, networking, infrastructure fabric, cooling, and management access—and the blade server architecture differs from monolithic servers in its possible redundant configurations.

The system chassis used in today’s monolithic servers hosts a single compute node with one or more CPUs, local storage, network, and other infrastructure components such as management access, power, and cooling. Redundancy options for monolithic servers range from cooling to network redundancy, thereby eliminating single points of failure. Several Dell PowerEdge servers offer redundant hot-pluggable cooling fans and power supplies, redundant memory with failover memory banks for memory mirroring, RAID controllers for redundant storage configuration, and dual-port integrated network interface cards (NICs) with network teaming software for redundant network configuration. When ordered with power, cooling, or memory redundancy options, Dell PowerEdge servers arrive fully configured and these options do not require any additional configuration at the deployment site. Network and storage redundancy options, however, require additional configuration using management applications.

The system chassis of Dell PowerEdge blade servers, known as the Dell Modular Server Enclosure, can host up to 10 compute nodes (server blades), plus shared infrastructure components. The common infrastructure components include cooling fans, power supplies, network switches, I/O modules, a KVM (keyboard, video, mouse) module, and chassis management in the form of the Dell Remote Access Controller/Modular Chassis (DRAC/MC). These components are shared through the midplane, which is passive to help ensure high reliability—that is, the midplane contains no active logic, just connectors and traces. Each server blade in a modular server chassis typically offers the same redundancy options as a monolithic server.

The primary reason for implementing redundant configurations within servers is to avoid single points of failure. Depending on the type of device and the design of the redundancy algorithm, setting up a redundant...
configuration can simply require installing two modules and powering up the system—or it may require using a management interface to configure the redundancy options. In a redundant setup, when one device fails, the second device assumes control and service continues uninterrupted. When a failover occurs, most devices typically transmit informational events that alert IT management about the redundancy failover. The Dell Modular Server Enclosure offers redundant configurations for power supplies, fans, I/O modules, network connections, and management modules.

**Power supply redundancy**

In the Dell Modular Server Enclosure, the DRAC/MC is responsible for system power budgeting and management. Power supply redundancy allows uninterrupted system operation in the event of failure of one or two power supplies in the chassis. This means that the DRAC/MC allows the blade server system to be powered up only if the required power is less than the available power based on the redundancy policy selection. The DRAC/MC lets the administrators select the redundancy configuration that best meets their requirements.

The following redundancy policies are available with DRAC/MC firmware version 1.3 (see Figure 1):

- **No redundancy**: In this mode, the total available power from the installed power supplies is used to power up the server blades and chassis components. Failure of one power supply may cause the chassis to power down based on the power consumption and available power at the time of failure.
- **3+1 redundancy**: In this mode, the power capacity of one power supply is kept in reserve while powering up the server blades; failure of any one power supply does not cause the chassis to power down.
- **2+2 redundancy**: In this mode, the capacity of two power supplies is kept in reserve while powering up the server blades; failure of any two power supplies does not cause the chassis to power down.

![Figure 2. DRAC/MC Power Budget screen](image)

Note: If administrators have already deployed a Dell Modular Server Enclosure with the DRAC/MC installed, they can upgrade the firmware to version 1.3. For more information about obtaining DRAC/MC firmware version 1.3, visit www.support.dell.com.

The desired power supply redundancy policy can be selected from the Power Budget page of the DRAC/MC Web interface (see Figure 2). The same selections can be made through the DRAC/MC command-line interface using the following command:

```
racadm config -g cfgChassisPower -o cfgChassisRedundancyPolicy value
```

Valid values for the `cfgChassisRedundancyPolicy` option are 0 (no redundancy), 1 (3+1 redundancy), and 2 (2+2 redundancy).

The redundancy policy selection is available only if four 2,100-watt power supplies are installed.¹ If a power supply is not installed, fails, or is removed, the redundancy policy selection is grayed out on the Power Budget page of the DRAC/MC Web interface.

<table>
<thead>
<tr>
<th>Redundancy policy</th>
<th>Power supply wattage</th>
<th>Support for power supply redundancy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>If all four power supplies are installed and working</td>
</tr>
<tr>
<td>3+1</td>
<td>1,200</td>
<td>Yes if only single-core Dell PowerEdge 1855 server blades are installed (If dual-core PowerEdge 1855 or PowerEdge 1955 server blades are installed, redundancy is not supported)¹</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2,100</td>
</tr>
<tr>
<td>2+2</td>
<td>1,200</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>2,100</td>
<td>Yes</td>
</tr>
</tbody>
</table>

¹Dual-core Dell PowerEdge 1855 and Dell PowerEdge 1955 blade server systems do not support 1,200-watt power supplies.
AC versus DC redundancy

For AC redundancy, power supplies must be connected to different AC grids so that failure of one AC grid does not cause a total loss of power to the chassis. Thus, in a configuration with four 2,100-watt power supplies, two power supplies should be connected to one AC circuit, while the other two power supplies should be connected to a different AC circuit. If all four power supplies are connected to the same AC circuit, then the system has DC redundancy because it is protected against power supply failures but not against AC failure.

For maximum benefit, Dell best practices recommend that administrators use redundant AC power connections when selecting the 2 + 2 redundancy policy. This means that administrators should configure a separate AC connection for each pair of power supplies. In this configuration, the system can continue to function if one AC panel fails.

Fan module redundancy

The Dell Modular Server Enclosure provides two cooling fan modules. These are located in the middle of the rear of the chassis and provide redundant cooling to the chassis. Each fan module contains two fans that provide redundancy within the individual modules. In addition to these fan modules, the chassis’s power supplies contain fans to cool the chassis. Administrators should install dummy power supplies in any empty power supply bay and not leave any bay unoccupied. Leaving a bay unoccupied affects cooling and can cause the blades to be throttled.

When a fan failure occurs in one of the modules, the status is communicated in two ways:

- The fan’s fault indicator light turns amber.
- An entry is made in the DRAC/MC system event log, and if configured to do so, the system sends alerts to the appropriate management consoles and e-mail accounts.

The fans require no special configuration to be set up for redundancy. Administrators can monitor the fan status through the DRAC/MC interface. For more information, see the Dell Remote Access Controller/Modular Chassis User’s Guide at support.dell.com/support/edocs/software/smdrac3/dracmc.

I/O module redundancy

The Dell Modular Server Enclosure supports a maximum of four I/O modules. I/O bays 1 and 2 can support only Ethernet-based I/O modules, whereas bays 3 and 4 can support any type of I/O module provided that I/O configuration rules are followed.²

Installing the I/O modules in pairs allows the modules to be configured redundantly. The redundancy is achieved through proper software configuration of the daughtercards. Installing a second Fibre Channel module in the Dell Modular Server Enclosure can create data-path redundancy and double performance by providing two paths for data to travel to and from the system. This redundant data-path configuration helps ensure server-to-storage connectivity if one of the data path fails.

Network connection redundancy

Each server blade has two embedded network LAN on Motherboards (LOMs) that have a dedicated circuit to the internal ports of the integrated switches or pass-through modules. The LOMs provide dedicated 1,000 Mbps full-duplex connections. LOM 1 on each server blade connects to an internal port of switch 1 or pass-through module 1, and LOM 2 on each server blade connects to the counterpart port of switch 2 or pass-through module 2 (see Figure 3).

Installing a second switch or pass-through module is optional. However, two installed switches or pass-through modules can enable additional connectivity or network redundancy and fault tolerance.

An important distinction between a blade server and other types of servers is that the connection between the LOM and the internal ports of the integrated I/O module (switch or pass-through) is hardwired through the midplane. This design enables the link between the LOM and the integrated I/O module to remain in a connected state—unless either a LOM or the I/O port fails. The link remains active even in the absence of a network connection between the external uplink ports on the integrated switch and the external network. Because of this, failures of cables or switch ports outside the enclosure do not trigger failover events in a network-teaming scenario in which an integrated switch is used. To overcome this limitation, administrators can use the nic-redundancy command available in Dell PowerConnect™ 5316M switch firmware version

1.0.0.35. This global configuration command enables the switch to support NIC teaming.

By default, NIC redundancy is disabled. Enabling the nic-redundancy option triggers a failover in teaming software if an external link on the integrated switch fails. The switch brings down the internal port links if there is no external port link and brings them back up once an external link is reestablished. The following command enables the nic-redundancy feature:

```
console(config)# nic-redundancy
```

The following command disables the nic-redundancy feature:

```
console(config)# no nic-redundancy
```

The scenario is different if a pass-through module is used. With a pass-through module, the link is in a connected state only if a network connection exists between the external ports on the pass-through module and the switch port outside the enclosure—as is true for a stand-alone server. Thus, for the pass-through module, the teaming software triggers a failover event if the LOM, pass-through port, cable, or external switch port fails.³

Management module redundancy

IT administrators can build chassis management redundancy into Dell PowerEdge blade servers by installing a redundant DRAC/MC in the chassis to serve as a standby for the primary DRAC/MC. The primary DRAC/MC actively monitors the chassis, while the standby DRAC/MC monitors the active signal from the primary DRAC/MC module. The standby DRAC/MC becomes the active primary DRAC/MC if a failure endures for more than five seconds.

Failures can be caused by any of the following conditions:

- The primary DRAC/MC network connection is broken. For example, the network cable has been disconnected or is broken.
- An administrator removes the primary DRAC/MC from the chassis.
- The primary DRAC/MC is rebooting or an administrator initiates a DRAC/MC reset.
- The primary DRAC/MC is nonresponsive and fails to exchange a heartbeat signal with the standby DRAC/MC.
- The firmware is being updated, causing a temporary failover. In this case, because the primary and standby DRAC/MCs use the same IP address, the console, Telnet, and user interface are rendered inactive. During a firmware update, the standby DRAC/MC monitors the chassis while the primary DRAC/MC updates its firmware. When the primary DRAC/MC completes its firmware update, the standby DRAC/MC continues its Trivial FTP (TFTP) update. The DRAC/MC network interface is unavailable until the firmware update is complete.

When a failover occurs, the primary and standby DRAC/MC share the same IP address and the same Media Access Control (MAC) addresses. At any point in time, only the current primary DRAC/MC provides chassis management and responds to an administrator’s request from the network; the standby DRAC/MC does not perform any chassis management and does not respond to any administrator request from the network.

Designating the primary DRAC/MC

When the chassis powers up for the first time, the DRAC/MC that is located above power supply 1 becomes the primary module. The chassis orientation assumes that the administrator is viewing the chassis from the back. Viewed from the back, the primary DRAC/MC is the one located on the right side of the chassis during initial power-up.

High availability for blade server environments

The Dell Modular Server Enclosure provides multiple redundancy options for Dell PowerEdge blade servers. Configuring these options properly can help reduce downtime by increasing the availability and performance of Dell PowerEdge blade servers. Each component within the enclosure should be individually considered and configured to achieve optimal results.

Narayan Devireddy is a development manager in the Dell Enterprise Systems Management Software organization. He has 14 years of systems management product development experience. Before joining Dell, Narayan worked for Novell, Compaq, Cheyenne Software, and Computer Associates in different capacities. He has an M.S. in Computer Sciences from Alabama A&M University.

Sanjeev S. Singh is a software engineer in the Dell Enterprise Systems Management Software organization. Previously, he was a software engineer at Hewlett-Packard and NCR. He has a B.S. in Electrical Engineering, and he has an M.S. in Computer Engineering from North Carolina State University.