DELLEMC

PowerEdge MX7000 Management Module Redundancy

Revisions

Date	Description
Jan 2019	Initial release

Acknowledgements

DELLEMC PowerEdge

Table of contents

DELLEMC PowerEdge

Introduction

The purpose of this whitepaper is to describe the MX7000 Management Module (MM) high availability feature provided by dual MM modules, discuss manual (user initiated) and automatic (system initiated) failovers, physical identification of active/standby MMs for part replacement scenarios and troubleshooting redundancy health.

MM Redundancy

The PowerEdge MX7000 with a recommended configuration has dual MMs, each occupying a slot accessible through the back of the chassis. Redundancy is an optional feature which will automatically establish in a chassis with dual MMs at the same firmware version. PowerEdge MX7000 chassis with redundancy will have the benefit of being more resilient management infrastructure.

MM high availability solution is designed using dual MMs (hardware modules) running a two-node active/standby cluster implemented using

- Linux pacemaker/corosync technologies
- Several cluster aware system services monitored as cluster resources via pacemaker hardware-based fencing agent for STONITH purposes.
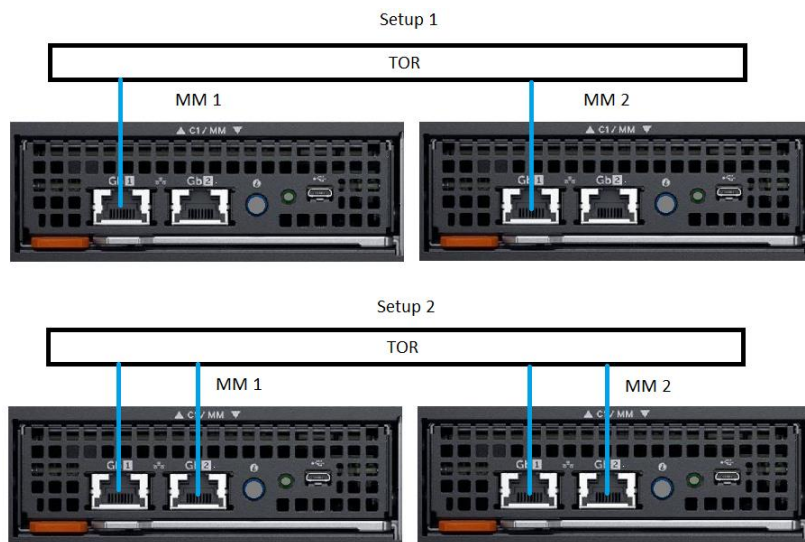

Figure 1   MM network cabling for redundancy

Setup and formation of redundancy

For a single chassis setup, have both MMs inserted into slot 1 and slot 2 on the back of the chassis, connect the MM Gb ports to the top of rack network switch as show in Figure 1. Setup 1 showing the minimum cabling required for redundancy and Setup 2 showing (optional) additional cabling for more cable loss tolerance. Please refer to "*PowerEdge MX7000 Chassis Management Network Cabling*" whitepaper for details on multi chassis setup.

DELLEMC PowerEdge

Establishing redundancy

In a dual MM configuration, on power up, one of the MMs claim and win the active role (more affinity for MM in slot 1) and initiate the boot up. The active MM does the orchestration of initializing the cluster by assuming the active node role and bringing up all monitored resources (services) in active mode, it then onboards the other MM to be the standby node with all its monitored resources (services) in standby mode. Once both the nodes assume their roles, the cluster is fully formed and redundant.

In a single MM configuration, on a power up, the MM will assume the active node role and bring up the cluster and monitored resources in the active role and become fully functional. The active MM is cluster aware with no redundancy at this point, but it is ready to onboard the second (redundant) MM as soon as presence is detected (when the second MM is inserted into the chassis).

Onboarding process of the redundant MM

Onboarding of the redundant MM (either present at the time of chassis power up or inserted into chassis later) goes through the same steps:

- Cluster configuration is pushed to redundant MM from the active MM
- All the data (if any present) on the redundant MM is wiped clean (it is possible that the redundant MM may be re-locating from another chassis)
- New data from the active MM is pushed to the redundant MM to be in sync with active MM
- At this point MM redundancy is established
- Any changes in data on the active MM this point onwards is replicated live to the standby MM

Version compatibility

Both the MMs must be at the same version to be able to form a cluster, if MMs are at different version, then the Active MM will be the only node in the cluster with no redundancy, but fully functional. The MM version mismatch turns the **Chassis Health Critical** and the associated **Chassis MM Subsystem Health Critical**. These critical alerts are displayed in the overview page of the OME-Modular.

To fix the version compatibility and restore MM redundancy:

- Remove the MM (that's not part of the cluster) and insert a compatible MM (same version as the active MM).

OR

- Update the OME Modular firmware version from the active MM. Even though the OME Modular firmware versions are not compatible and are not in cluster, active MM will be able to update the version of the other MM.

Inherent benefits of redundancy

On MM failure

- Continued access to OME-M with approximate downtime of 2.5 minutes. Downtime accounts for detection of failures, promotion of standby to active, reconciliation of inter device communications and network readiness.
- Management network including OME Modular IP addresses continue to function on the new active MM
- All the data (device inventory, configuration, jobs, alerts, logs, etc.) is accessible without any loss (is made possible by live data replication between the MMs)

Failovers

Failover is the process of switching the active MM role to the redundant standby MM to have continued access to OME-Modular after a brief downtime. There are two types of failovers, manual (user initiated) and automatic (system initiated). Figure 2 is depicting the failover process that transitions active node from MM1 to MM2.

- Step 1: While in the redundant state, failover happens (manual or automatic)
- Step 2: MM 1 reboots and at the same time MM2 is preparing to take over active role
- Step 3: MM 2 becomes active node and MM1 after reboot is transitioning to become standby
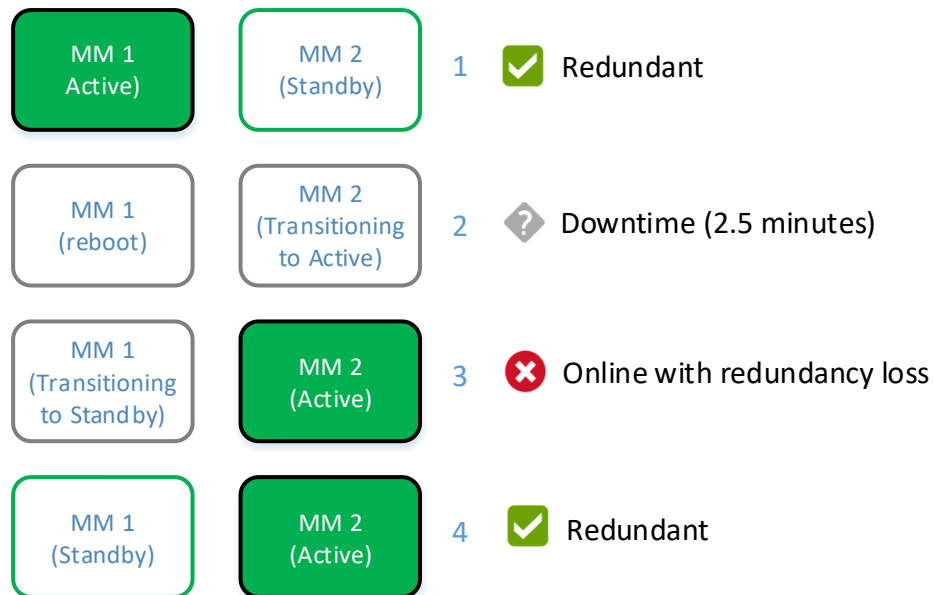- Step 4: Redundancy is completely restored and ready to deal with another failover

MM 1 Active)	MM 2 (Standby)	1	✅	Redundant
MM 1 (reboot)	MM 2 (Transitioning to Active)	2	❓	Downtime (2.5 minutes)
MM 1 (Transitioning to Standby)	MM 2 (Active)	3	❌	Online with redundancy loss
MM 1 (Standby)	MM 2 (Active)	4	✅	Redundant

Figure 2   Process of failover from active MM1 to standby MM2

Manual (user initiated) failover occurs in the following scenarios

- User removes the active MM from the chassis to take it offline
- User initiates the "Failover operation (Figure 3)" or "Reset Management Module operation (Figure 4)" via OME Modular interfaces (GUI, Rest API and Racadm). User may perform these

DELLEMC PowerEdge

operations if there are any observations of persistent issues on the active MM and want to remedy by switching to standby MM.

- – example racadm command: racadm changeover
- – example racadm command: racadm racreset

Automatic (system initiated) failover occurs in the following scenarios

- Long running active MM may eventually develop and manifest failures (software and/or hardware). MM high availability solution is designed to monitor and detect these failures and initiate automatic failover if the solution determines the failure to be hampering normal functioning of the MM.
- OME Modular update on a chassis with dual MM configuration stages the firmware update orchestration in a sequence of steps that requires an automatic failover. Below is a high level sequence of operations with MM1 as active and MM2 as standby:
  - – Update is initiated on Active MM1, MM 1 will push the update to MM2
  - – Standby MM2 is updated
  - – Automatic failover performed to transition the active role from MM1 to MM2
  - – The new active MM2 completes the update process by pushing the update to MM1
  - – Refer to "Management Module Firmware Update" whitepaper for more details
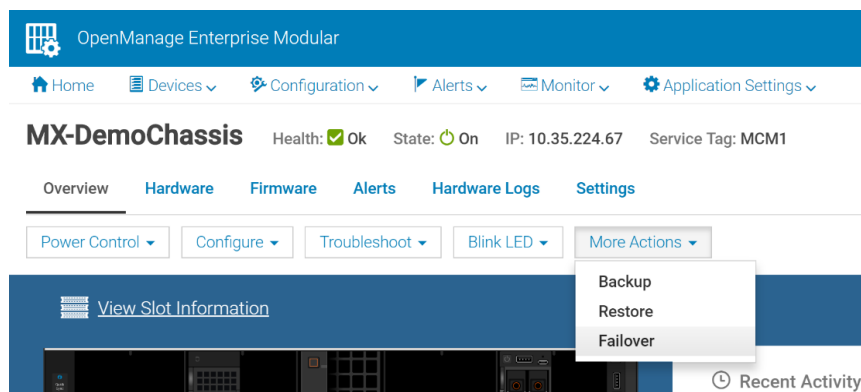

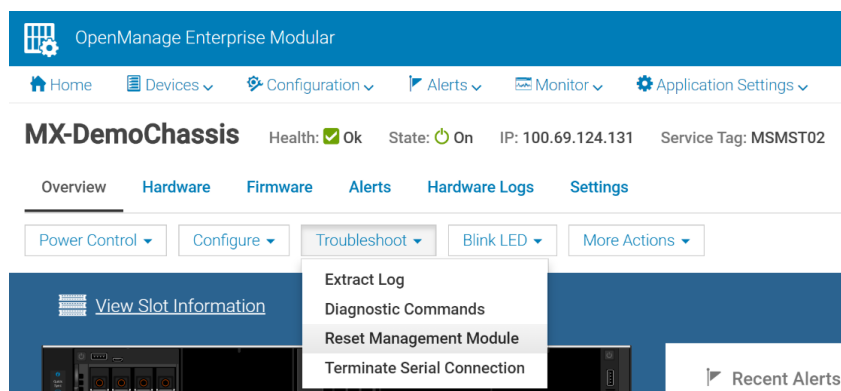Figure 3   Failover using OME Modular GUI

Figure 4   Reset Management Module using OME Modular GUI

Moving/Swapping MMs between chassis

Moving or swapping MMs between chassis could be a typical usecase during maintenance and trouble shooting scenarios. **Use case 1**, please refer to figure 5: A chassis with dual MM configuration and redundancy health OK, fully supports movement of single MM without any configuration or device history data loss. Chassis health will be critical momentarily while the swapped MMs go through the following

- Other chassis data residing on MM is wiped
- Data from the active MM of the new chassis copied
- MM redundancy is restored

Use case 2, please refer to figure 6: There should not be any normal scenarios that require removing or swapping both the MMs of a chassis simultaneously, but should the need arise, please use caution as this action is equivalent to losing both MMs and loss of significant configuration and device history data residing in them. However, there is some minimal configuration data preserved, this data resides on the chassis Right Control Panel (RCP), please refer to table 1 for details on what configuration/data is preserved/lost on both MMs failure or simultaneous replacement.
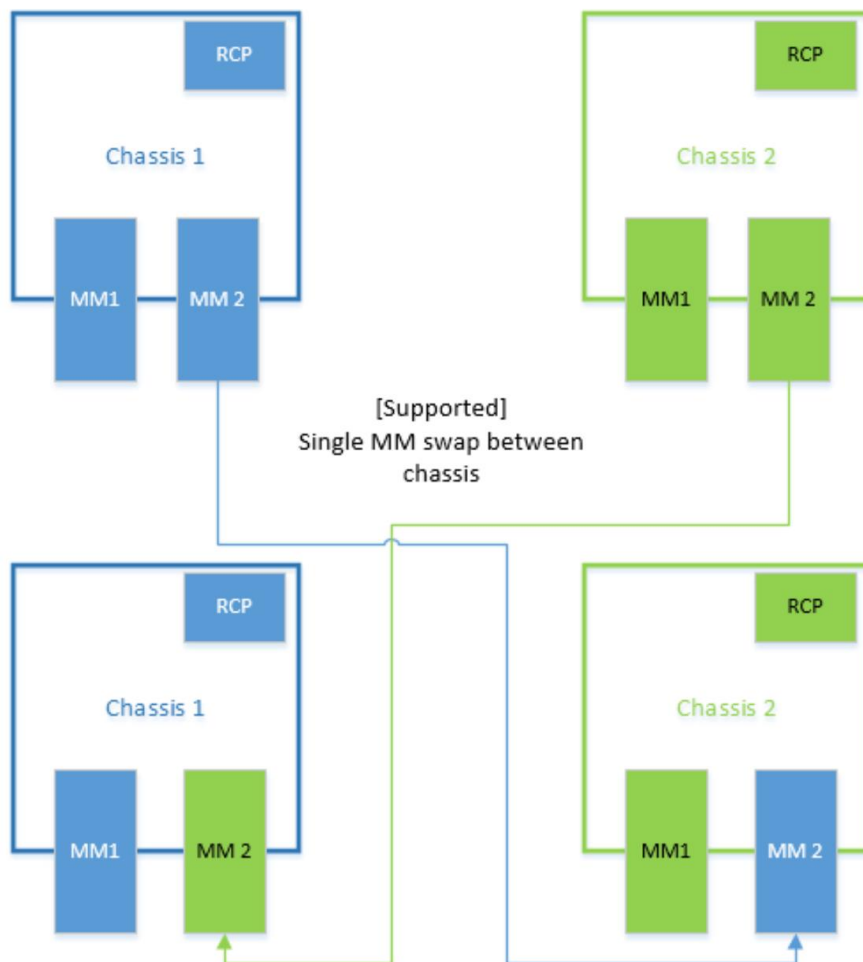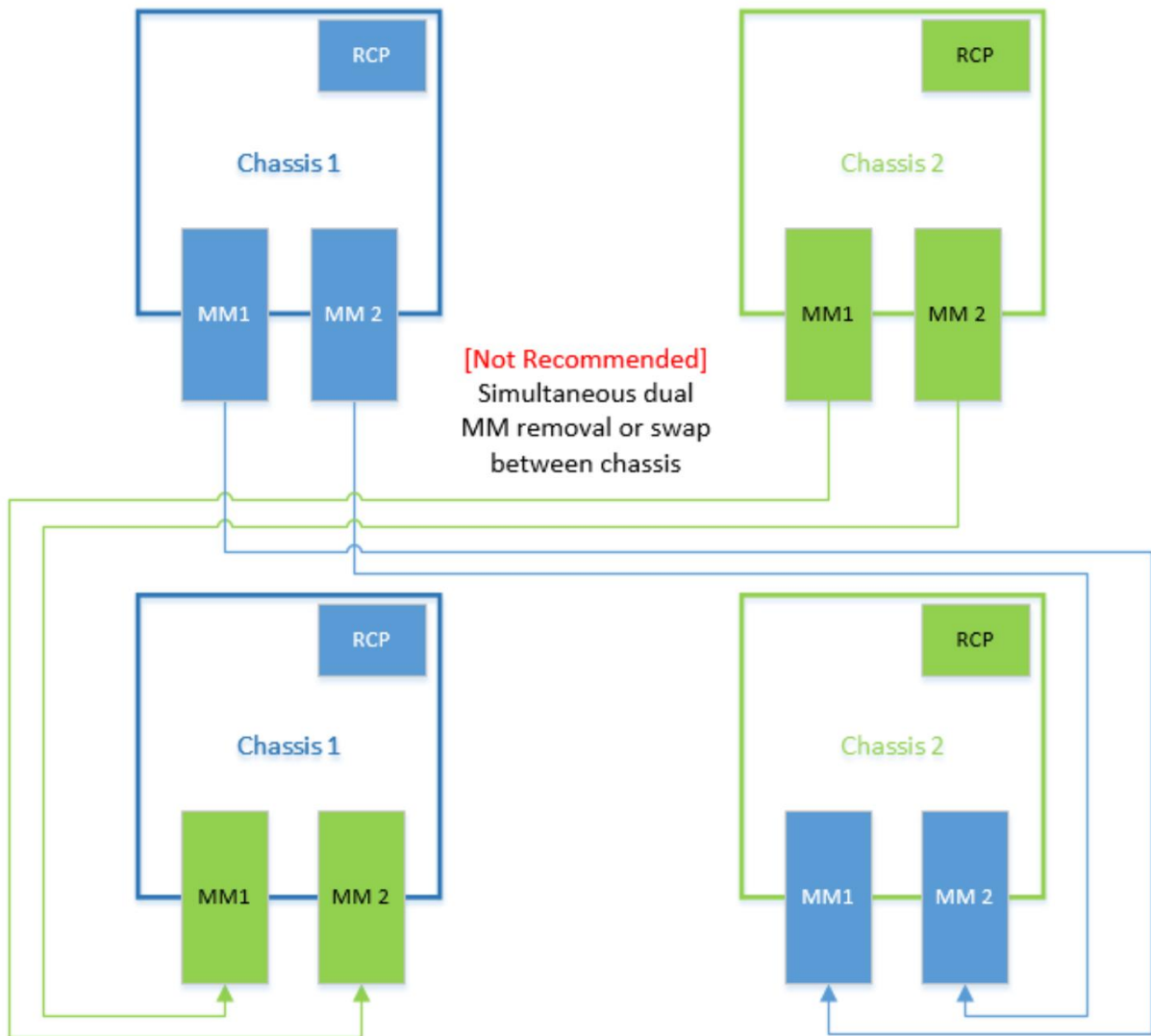

Figure 5   Single MM remove or swap

DELLEMC PowerEdge

Figure 6   Dual MM remove or swap

Identify Active and Standby MMs

Following are two of several ways to identify which MM is active

- Via the OME Modular GUI (Figure 7)
- Physical Identify Combo LED on the back of the chassis (Figure 8). For more details on Identify Combo LED please refer to "PowerEdge MX7000 At-the-box System Identify" whitepaper.
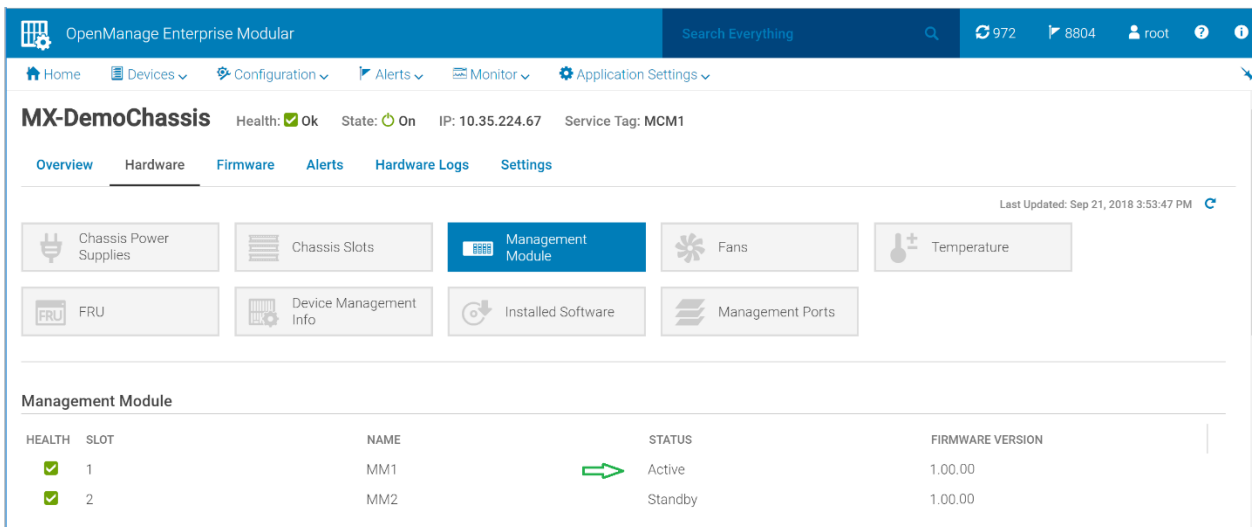
DELLEMC PowerEdge

Figure 7   Identifying Active MM via GUI

Figure 8   Identify Active MM via LED on the MM
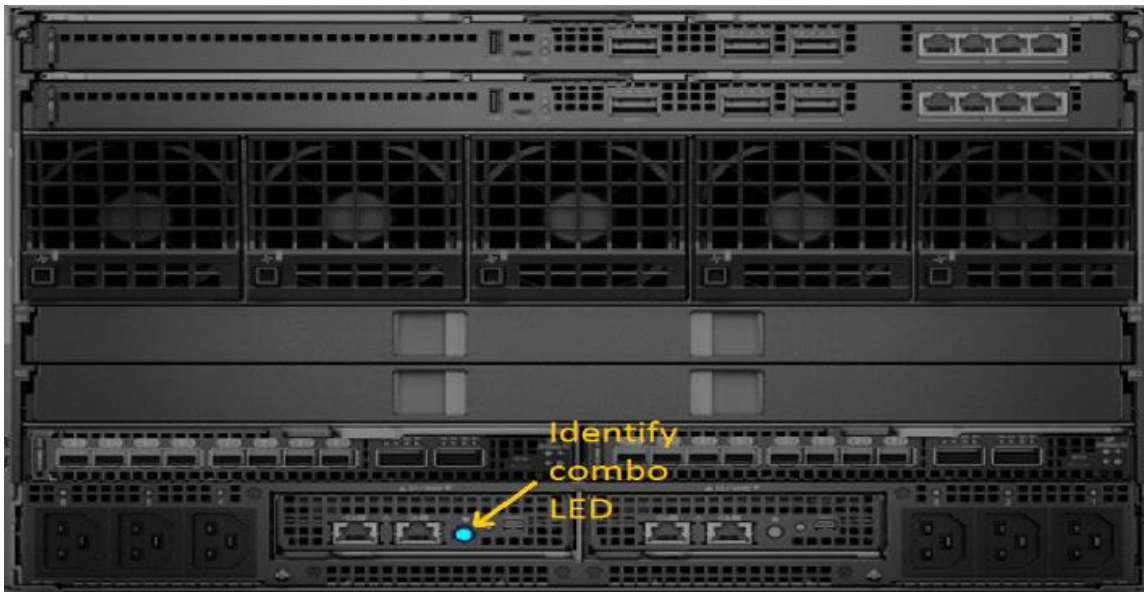
Troubleshooting Redundancy Health Alerts

Following redundancy critical alerts are generated and displayed on OME Modular:

SEL1501: Chassis management controller/Management Module (CMC/MM) redundancy is lost
Reason for the alert: One MM is removed from a dual MM configuration chassis
Recommended action: Insert another MM with same firmware version to restore redundancy

SEL1524: Management Module in Slot [1/2] is offline
Reasons for the alert: One of the MMs is not performing at its optimal level in a dual MM configuration chassis, affected MM will be shown as offline and should self-heal.
Recommended action: If the issue persists for several minutes, perform a reset operation for the Management Module as identified in the message

❌ System Event Log
ManagementModule in slot 1 is offline.
Message ID: SEL1524   Jul 31, 2018 10:08:37 AM

SEL1522: A firmware mismatch detected in Management Modules
Reasons for the alert: In a dual MM configuration, MM versions are not compatible with each other.
Recommended action: Initiate OME Modular firmware update or replace MM with compatible versions.

❌ System Event Log
A firmware mismatch detected in ManagementModules .
Message ID: SEL1522   Jul 31, 2018 10:38:30 AM

SEL1523: Unable to detect network connection of Management Module in slot [1/2]

Reasons for the alert: In a dual MM configuration, one of the MMs has both of its management ports (Gb1/Gb2) in link down state (missing network cable).
Recommended action: Connect the network cables as shown in Figure 1.

❌ System Event Log
Unable to detect network connection of ManagementModule in slot 1 .
Message ID: SEL1523   Jul 31, 2018 10:00:11 AM

SEL1520: Unable to synchronize NVDIMM Settings between Management Modules. Power loss detection redundancy is degraded

Reasons for the alert: NVDIMM Settings could not be synchronized across Management Modules.
Recommended action: Reboot from Management Module GUI, physically reseat or initiate a firmware update on the standby Management Module.

Data	Description	Single MM Failure or Removal	Dual MM Failure or Simultaneous Removal
Historical	Jobs	Preserved	Not preserved
Historical	Alert Log	Preserved	Not preserved
Historical	Audit Logs	Preserved	Not preserved
Historical	Chassis Temperature Statistics	Preserved	Preserved
Historical	Chassis Power Statistics	Preserved	Preserved
Historical	Chassis Hardware Logs	Preserved	Preserved
Configuration	Fabric	Preserved	Preserved
Configuration	Storage Assignments	Preserved	Preserved
Configuration	Group (Multi Chassis)	Preserved	Preserved

DELLEMC PowerEdge

Configuration	Deploy (Templates)	Preserved	Not preserved
Configuration	Identity Pools, Networks	Preserved	Not preserved
Configuration	Firmware Baseline	Preserved	Not preserved
Configuration	Alert Policies	Preserved	Not preserved
Configuration	Chassis Address (ipv4,ipv6,DNS,etc)	Preserved	Preserved
Configuration	Time (NTP, Timezone)	Preserved	Preserved
Configuration	Chassis "root" user password	Preserved	Preserved
Configuration	Chassis Local Access Configuration (Chassis Power Button, Quick Sync, KVM, LCD)	Preserved	Preserved
Configuration	Sled Slot Power Priority	Preserved	Preserved
Configuration	Chassis Power Cap	Preserved	Preserved
Configuration	Chassis Power Redundancy Policy	Preserved	Preserved
Configuration	Sled Slot Name	Preserved	Preserved
Configuration	Sled VLAN ID	Preserved	Preserved
Configuration	Chassis Location	Preserved	Preserved
Configuration	Chassis Name	Preserved	Preserved
Configuration	Chassis Assset Tag	Preserved	Preserved
Configuration	Quick Deploy (iDRAC & IOM)	Preserved	Not preserved
Configuration	Users, Directory Services	Preserved	Not preserved
Configuration	Login IP range, Login Lockout Policy, FIPS	Preserved	Not preserved
Configuration	Web server configuration	Preserved	Not preserved
Configuration	SSL Certificates	Preserved	Not preserved
Configuration	Proxy configuration	Preserved	Not preserved
Configuration	Device Name Preference	Preserved	Not preserved
Configuration	Alerts Configuration (Email, SNMP, Syslog)	Preserved	Not preserved
Configuration	Services Configuration (SNMP, SSH, Remote Racadm)	Preserved	Not preserved

Table 1   MM historical/configuration data state in single and dual MM failure use cases

DELLEMC PowerEdge