One Dell Way
Round Rock, Texas 78682
www.dell.com

Dell Enterprise White Paper

Dell M1000e Modular Enclosure Architecture

By John Loffink, Engineer/Strategist

Contents

Key Benefits

- 16 server modules
- 3x2 (Redundant) I/O fabrics
- Redundant Power
- Redundant Cooling
- Redundant Chassis Management Controllers
- Avocent iKVM interface
- LCD Control Panel

Summary

The Dell PowerEdge M1000e Modular Server Enclosure is a breakthrough in enterprise server architecture. The enclosure and its components spring from a revolutionary, ground up design incorporating the latest advances in power, cooling, I/O, and management technologies. These technologies are packed into a highly available rack dense package that integrates into standard Dell and 3rd party 2000mm depth racks.

The enclosure is 10U high and supports:

- Up to 16 server modules.
- Up to 6 network & storage I/O interconnect modules.
- A high speed passive midplane connects the server modules in the front and power, I/O, and management infrastructure in the rear of the enclosure.
- Comprehensive I/O options support dual links of 20 Gigabits per second today (with 4x DDR InfiniBand) with future support of even higher bandwidth I/O devices when those technologies become available. This provides high-speed server module connectivity to the network and storage now and well into the future.
- Thorough power management capabilities including delivering shared power to ensure full capacity of the power supplies available to all server modules.
- Broad management ability including private Ethernet, serial, USB, and low level management connectivity between the Chassis Management Controller (CMC), Keyboard/Video/ Mouse switch, and server modules.
- Up to two Chassis Management Controllers (CMC-1 is standard, 2nd provides optional redundancy) and 1 optional integrated Keyboard/Video/Mouse (iKVM) switch.
- Up to 6 hot pluggable, redundant Power Supplies and 9 hot pluggable, N+1 redundant Fan Modules.
- System Front Control panel w/ LCD panel and two USB Keyboard/Mouse and one Video "crash cart" connections.

Index of Figures

Index of Tables

Acronyms and Definitions

1000BASE-KX – IEEE standard for transmission of 1 Gbps Ethernet over backplanes
1000M – 1000 Megabit per second Ethernet
100BaseT – 100 Megabit per second Ethernet over Twisted Pair
100M – 100 Megabit per second Ethernet
10GBASE-KR – IEEE standard for transmission of 10 Gbps Ethernet over backplanes using 1 differential lane pair
10GBASE-KX4 – IEEE standard for transmission of 10 Gbps Ethernet over backplanes using 4 differential lane pairs
10GE – 10 Gigabit Ethernet
10M – 10 Megabit per second Ethernet
3G – 3 Gbps
4G – 4 Gbps
6G – 6 Gbps
8G – 8 Gbps
AC – Alternating Current
ACI – Analog Console Interface
ASIC – Application Specific Integrated Circuit
Bel – 10 Decibels
BER – Bit Error Rate
BMC – Baseboard Management Controller
CD – Compact Disk
CERC6 – Cost Effective RAID Controller version 6
CFM – Cubic Feet per Minute
CIM – Common Information Model
CMC – Chassis Management Controller
CLI – Command Line Interface
CPU – Central Processing Unit
DDR – Double Data Rate
DDR2 – Double Data Rate 2 Dynamic Random Access Memory
DHCP – Dynamic Host Configuration Protocol
DIMM – Dual Inline Memory Module
DVD – Digital Video Disk
ECC – Error Correcting Code
EIA – Electronic Industries Alliance
FBDIMM – Fully Buffered DIMM
FC – Fibre Channel
FP – Full Power

GB - Gigabyte
Gbps – Gigabits per second
GE – Gigabit Ethernet
Gen1 – Generation 1
Gen2 – Generation 2
GPIO – General Purpose I/O
GUI – Graphical User Interface
HA – High Availability
HBA – Host Bus Adapter
HCA – Host Controller Adapter
HDD – Hard Disk Drive
HPCC – High Performance Computing Cluster
IB - InfiniBand
ICH – Input/Output Controller Hub
I/O – Input/Output
iDRAC – Integrated Dell Remote Access Controller
iKVM – Integrated KVM
I/O – Input/Output
IOH – Input/Output Hub
IOM – Input/Output Module
IPC – Interprocessor Communications
IPMI – Intelligent Platform Management Interface
IPv6 – Internet Protocol version 6
IR – Integrated RAID
iSCSI – Internet SCSI
ISO – International Organization for Standardization
KVM – Keyboard Video Mouse
LAN – Local Area Network
LCD – Liquid Crystal Display
LED – Light Emitting Diode
LOM – LAN on Motherboard
LR – Long Reach
LV – Low Voltage
Mbps – Megabits per second
NIC – Network Interface Card
MB - Megabyte
MCH – Memory Controller Hub
MHz – MegaHertz
MV – Mid Voltage
N+0 – N (1, 2, 3…) plus zero, no fault tolerance
N+1 – N (1, 2, 3…) plus one, fault tolerance with one backup unit
N+N – N (1, 2, 3…) plus N (1, 2, 3…) with N backup units
NPIV – N_Port ID Virtualization
OPSF – Open Shortest Path First
OS – Operating System

PCIe – PCI Express, or Peripheral Component Interconnect Express
PDU – Power Distribution Unit
PMBus – Power Management Bus
QDR – Quad Data Rate
QOS – Quality of Service
R2 – Release 2
RACADM – Remote Access Controller and Administrator
RAID – Redundant Array of Inexpensive Disks
RIP – Routing Information Protocol
SAN – Storage Area Network
SAS – Serial Attached Small Computer System Interface, or Serial SCSI
SAS6/IR – SAS I/O Card 6 with integrated RAID
SATA – Serial Advanced Technology Attachment, or Serial ATA
SDDC – Single Device Data Correction
SDR – Single Data Rate
SFP – Small Form-factor Pluggable Transceiver
SMASH-CLP – Systems Management Architecture for Server Hardware Command Line Protocol
SNMP – Simple Network Management Protocol
SOL – Serial over LAN
SR – Short Reach
SSH – Secure Shell
SSL – Secure Sockets Layer
TCP/IP – Transmission Control Protocol/Internet Protocol
TFTP – Trivial File Transfer Protocol
TOE – TCP/IP Offload Engine
USB – Universal Serial Bus
VGA – Video Graphics Array
vKVM – Virtual KVM
VLAN – Virtual LAN
vMedia – Virtual Media
VRRP – Virtual Router Redundancy Protocol
W2K3 – Windows Server 2003 Operating System
WOL – Wake On LAN
WSMAN – Web Services for Management
XML – Extensible Markup Language
XFP – 10 Gigabit Small Form Factor Pluggable Transceiver

Background

IT IS ALL ABOUT EFFICIENCY. The new PowerEdge M1000e is designed to help you be more efficient with your time, your power and cooling, your investment, and your system's performance. It is a breakthrough Dell engineered and patent pending design that maximizes flexibility, power and thermal efficiency, system wide availability, performance, and manageability. The chassis integrates the latest in management, I/O, power and cooling technologies in a modular, easy to use package. Designed from the ground up to support current and future generations of server, storage, networking, and management technologies, the PowerEdge M1000e includes the headroom necessary to scale your environment for the future.

The PowerEdge M1000e Modular Server Enclosure solution supports server modules, network, storage, and cluster interconnect modules (switches and passthrough modules), a high performance and highly available passive midplane that connects server modules to the infrastructure components, power supplies, fans, integrated KVM and Chassis Management Controllers (CMC). The PowerEdge M1000e uses redundant and hot-pluggable components throughout to provide maximum uptime.

Introducing new levels of modularity to server based computing, the M1000e provides identical and symmetric fabric options B and C for each modular server. Ethernet I/O switches support I/O submodules that provide external I/O flexibility of stacking ports, 10GE copper ports, or 10GE optical ports. True modularity at the system and subsystem level provides simplicity of extension and enhancement, now and in the future.

Server Modules

Virtually unlimited in scalability, the PowerEdge M1000e Enclosure provides ultimate flexibility in server processor and chipset architectures. Both Intel and AMD server architectures are planned for introduction into the M1000e infrastructure, while cutting edge mechanical, electrical and software interface definitions enable multi-generational server support and expansion.

The M1000e enclosure supports up to 16 Half-Height server modules, each occupying a slot accessible in the front of the enclosure. Server Modules other than standard Half-Height form factor are also supported in the future, occupying full slot heights, dual slot widths, or both. The mechanical slots in the enclosure can support servers that are twice the height and/or width of the Half-Height module, as shown in Figure 1.

Server Modules can be freely located within each 2 x 2 Half Height quadrant. The mechanical design of the M1000e has support structures for Half Height Server Modules above or below Double Width Server Modules, and for Half Height Server Modules side by side with Full Height Server Modules, as shown in Figure 2.

The first two modular servers available for the PowerEdge M1000e Modular Enclosure are Half-Height 2 CPU socket, 8 DIMM capable systems, with two 2.5 inch hot pluggable Hard Disk slots, and two flexible fabric mezzanine cards in addition to the two on-board 1 Gigabit Ethernet LAN on Motherboard (LOMs). A simplified block diagram of the PowerEdge M600 modular server's architecture, based on the Intel processor/chipset architecture, is shown in Figure 3.

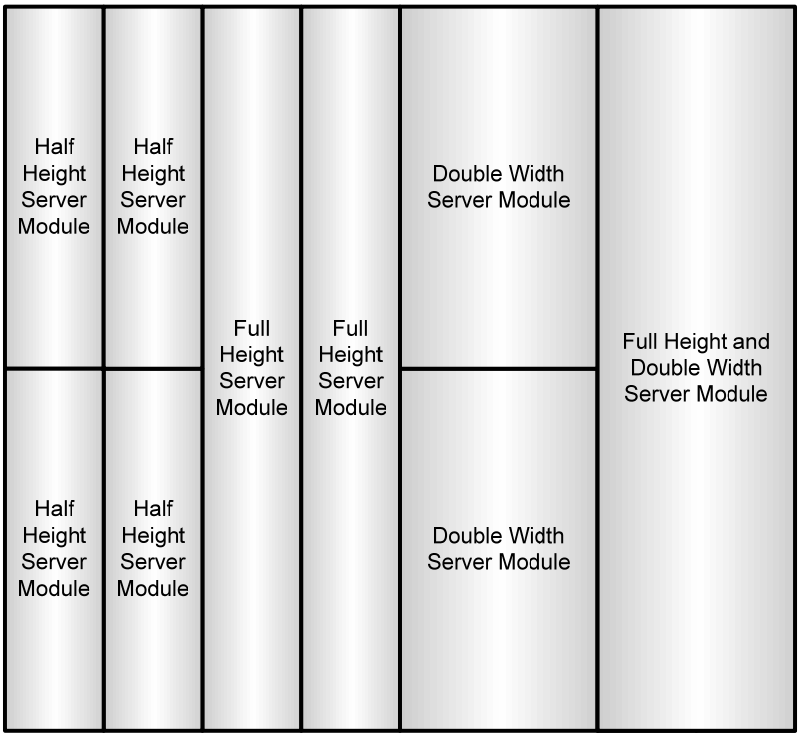


Figure 1 Possible Server Module Sizes, Front Panel View

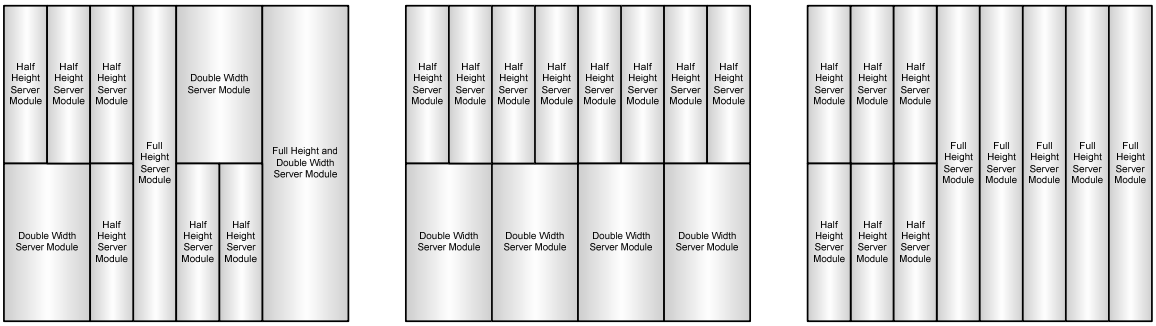


Figure 2 Example Server Module Configurations

Server Processors

The first generation of Half-Height server modules support a full range of dual and quad core processors in either Low Voltage (40-68W), Medium Voltage (80-95W) or Full Power (120W) profiles. Some restrictions on memory support may apply to the 120W quad core full power processors, based upon the system power envelope and cooling capabilities. Both AMD and Intel processor architectures are supported.

Server Memory

Memory support in the first generation Half-Height server modules is comprised of 8 DIMM slots. Intel based architectures use Fully Buffered DIMM technology, while AMD based architectures use DDR-2 DIMM technology. Interleaving (2 way lockstep), ECC, memory sparing

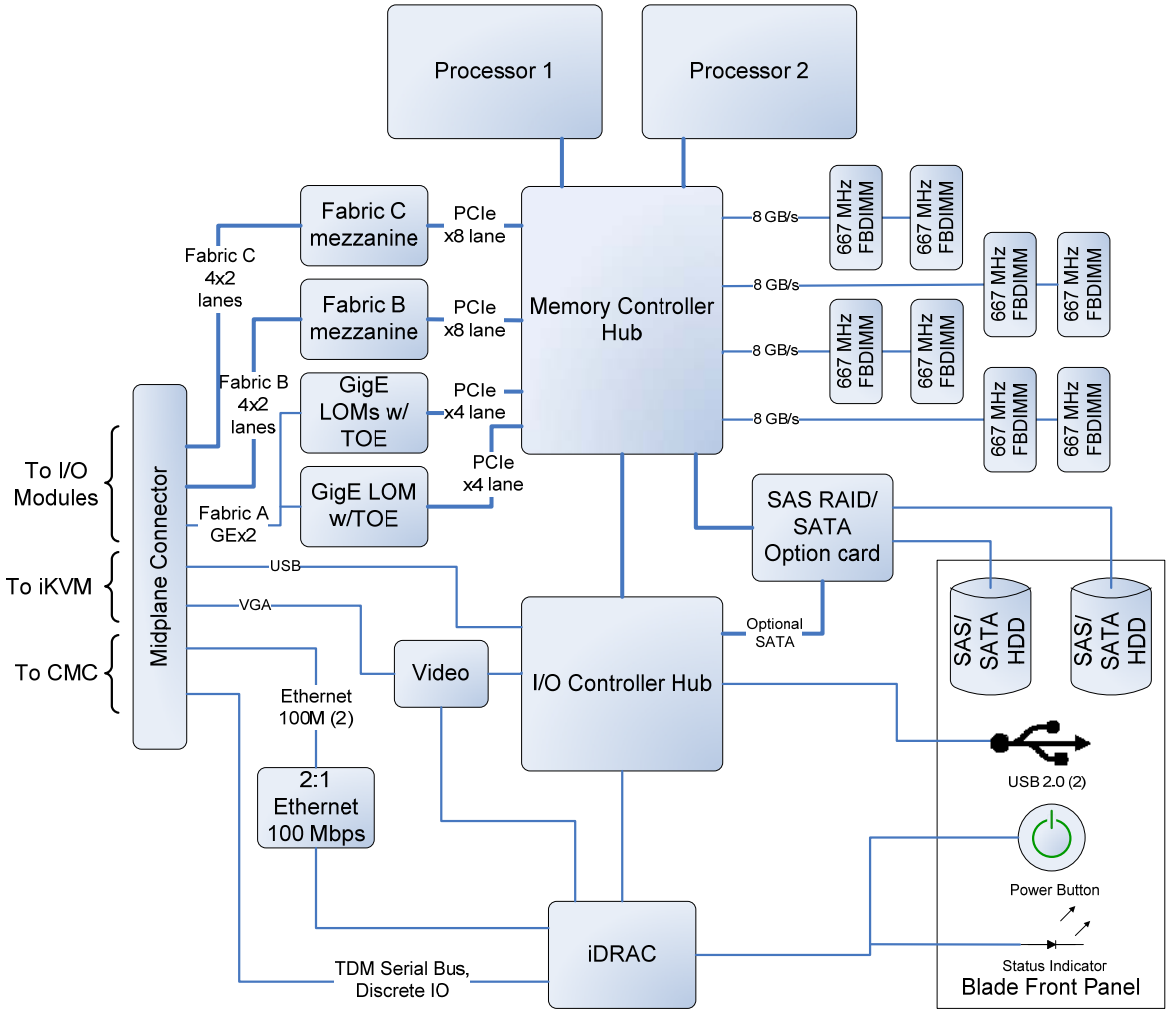


Figure 3 Dual Socket Modular Server Architecture

and ChipKill/SDDC are supported with both AMD and Intel based designs. Memory mirroring is supported on the PowerEdge M600 Intel based design. Supporting a range of DIMM capacities from 512 MB to 8 GB, each server module may contain a maximum of 64 GB of memory.

Server I/O

Each server module connects to traditional network topologies, while providing sufficient bandwidth for multi-generational product lifecycle upgrades. I/O fabric integration encompasses networking, storage, and interprocessor communications (IPC). Before starting, let's establish some terminology used throughout this white paper.

A ***fabric*** is defined as a method of encoding, transporting, and synchronizing data between devices. Examples of fabrics are Gigabit Ethernet (GE), Fibre Channel (FC) or InfiniBand (IB). Fabrics are carried inside the M1000e system, between server module and I/O Modules through the midplane. They are also carried to the outside world through the physical copper or optical interfaces on the I/O modules.

A ***lane*** is defined as a single fabric data transport path between I/O end devices. In modern high speed serial interfaces each lane is comprised of one transmit and one receive differential pair. In reality, a single lane is four wires in a cable or traces of copper on a printed circuit board, a transmit positive signal, a transmit negative signal, a receive positive signal, and a receive negative signal. Differential pair signaling provides improved noise immunity for these high speed lanes. Various terminology is used by fabric standards when referring to lanes. PCIe calls this a lane, InfiniBand calls it a physical lane, and Fibre Channel and Ethernet call it a link.

A ***link*** is defined here as a collection of multiple fabric lanes used to form a single communication transport path between I/O end devices. Examples are two, four and eight lane PCIe, or four lane 10GBASE-KX4. PCIe, InfiniBand and Ethernet call this a link. The differentiation has been made here between lane and link to prevent confusion over Ethernet's use of the term link for both single and multiple lane fabric transports. Some fabrics such as Fibre Channel do not define links as they simply run multiple lanes as individual transports for increased bandwidth. A link as defined here provides synchronization across the multiple lanes, so they effectively act together as a single transport.

A ***port*** is defined as the physical I/O end interface of a device to a link. A port can have single or multiple lanes of fabric I/O connected to it.

There are three supported high speed fabrics per M1000e Half-Height server module, with two flexible fabrics using optional plug in mezzanine cards on the server. The ports on the server module connect via the midplane to the associated I/O Modules (IOM) in the rear of the enclosure, which then connect to the customer's LAN/SAN/IPC networks.

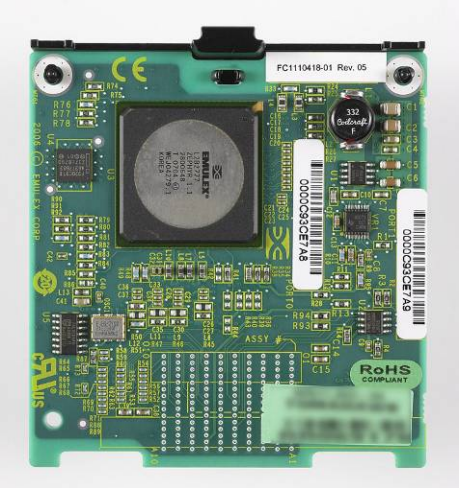


Figure 4 Modular Server Fabric I/O Mezzanine Card, Fibre Channel 4 Gbps

The first embedded high speed fabric (Fabric A) is comprised of dual Gigabit Ethernet LOMs and their associated IOMs in the chassis. The LOMs are based on the Broadcom 5708 NetXtreme II Ethernet controller, supporting TCP/IP Offload Engine (TOE) and iSCSI boot capability. In addition, full iSCSI HBA capability with full protocol offload and a broader array of OS boot support is planned to be available in early 2008. While initially servers released in this series

have a dual GE LOM configuration, the system midplane is designed to allow future support for up to quad GE LOMs in the Half-Height form factor.

Optional customer configured fabrics are supported by adding up to two dual port I/O Mezzanine cards. Each Half-Height server module supports two identical I/O Mezzanine card connectors which enable connectivity to I/O fabrics B and C. These optional Mezzanine cards offer a wide array of Ethernet (including iSCSI), Fibre Channel, and InfiniBand technologies, with others possible in the future. Figure 4 shows an example of a dual port Fibre Channel 4 Gigabits per second (Gbps) mezzanine card. Like the connector, I/O Mezzanine card form factors are common to both fabric B and C, for the highest level of flexibility in fabric configuration.

The optional mezzanine cards are designed to connect via 8 lane PCIe to the server module's chipset. For PCIe Gen1 this provides up to 16 Gbps of data bandwidth per mezzanine card. Mezzanine cards may have either one dual port ASIC with 4 or 8 lane PCIe interfaces or dual ASICs, each 4 lane PCIe interfaces. Both PCIe fabrics and external going fabrics are routed through high speed, 10 Gigabit per second capable air dielectric connector pins through the planar and midplane. For best signal integrity the signals isolate transmit and receive signals for minimum crosstalk. Differential pairs are isolated with ground pins and signal connector columns are staggered to minimize signal coupling.

Wake on LAN (WOL) is supported by the server module's LOMs. Ethernet WOL for Fabrics B and C is also supported. Boot from SAN is supported by Fibre Channel and iSCSI enabled cards.

The M1000e system management hardware and software includes Fabric Consistency Checking, preventing the accidental activation of any misconfigured fabric device on a server module. Since mezzanine to I/O Module connectivity is hardwired yet fully flexible, a user could inadvertently hot plug a server module with the wrong mezzanine into the system. For instance, if Fibre Channel I/O Modules are located in Fabric C I/O Slots, then all server modules must have either no mezzanine in fabric C or only Fibre Channel cards in fabric C. If a GE mezzanine card is in a Mezzanine C slot, the system will automatically detect this misconfiguration and alert the user of the error. No damage occurs to the system, and the user will have the ability to reconfigure the faulted module.

Table 1 shows options for the initial release of fabric mezzanines in the M1000e product line. The same mezzanine card can be used in Fabric B or C. Table 2 shows some typical server module

Fabric Type	Ports	Speed per port	Features
Gigabit Ethernet	2	1.25 Gbps	TOE, iSCSI boot, Jumbo Frames, WOL
Fibre Channel 4G	2	4.25 Gbps	Boot from SAN, NPIV
Fibre Channel 8G*	2	8.50 Gbps	Boot from SAN, NPIV
4x Double Data Rate (DDR) InfiniBand*	2	20.00 Gbps	
10 Gigabit Ethernet*	2	10.3125 Gbps	TOE, iSCSI offload, Jumbo Frames

***planned**

Table 1 Modular Server Fabric I/O Mezzanine Fabric B and C Options

Typical Fabric Configurations	Fabric A	Fabric B	Fabric C
Low Cost Compute Cluster Node	2 x 1GE LOM	Empty	Empty
Ethernet/SAN server or HA Cluster Node	2 x 1GE LOM	2 x 1GE Mezzanine	2 x FC4 Mezzanine
High Performance Cluster Node	2 x 1GE LOM	2 x 4 DDR InfiniBand	Empty
Unified Fabric	2 x 1GE LOM	2 x 1GE Mezzanine	2 x 10GE Mezzanine
High Performance Unified Fabric	2 x 1GE LOM	2 x 10GE Mezzanine	2 x 10GE Mezzanine

Table 2 Typical Modular Server Fabric Configuraitons

fabric configurations, based upon current and future architectural models. Future mezzanine releases will support speed and lane count increases.

Local Storage

Server module local storage is comprised of up to two 2.5 inch Hard Drives. Both hot pluggable SAS and non-hot pluggable SATA hard drives are supported. Local storage controller options include low cost chipset SATA I/O, SAS Daughtercard with integrated RAID (SAS6/IR) or SAS RAID Daughtercard with cache and high performance RAID processor (CERC6/i).

Integrated Server Management

The M1000e supports the introduction of Dell's new Integrated Dell Remote Access Controller (iDRAC) which is integrated on each M600/605 server module. iDRAC contains features that are typically add-on options to standard monolithic rack based servers. The iDRAC adds virtual Media (vMedia) and virtual KVM (vKVM) functions as well as out of band GUI status/inventory. Traditional IPMI based Baseboard Management Controller (BMC) features like hardware monitoring and power control are supported. Additionally, onboard graphics and keyboard/mouse USB connects to an optional system level Integrated KVM (iKVM) module for local KVM access. Full USB access is available through the server module front panel.

The new iDRAC is connected to the CMC via dedicated, fully redundant 100 Mbps Ethernet connections wired through the midplane to a dedicated 24-port ethernet switch on the CMC, and exposed to the outside world through the CMC's external Management Ethernet interface (10/100/1000M). This connection is distinct from the three redundant data Fabrics A, B and C. Unlike previous generations of Dell server modules, the iDRAC's connectivity is independent of, and in addition to, the onboard GE LOMs on the server module. Each server module's iDRAC has its own IP address and can be accessed, if security settings allow, directly through a supported browser, telnet, SSH, or IPMI client on the management station.

Server Module and Packaging

The server module is a sheet metal module that houses the server planar and hot plug hard drives and is designed to be tool-less for any assembly or disassembly operation. Any part removal from the server is intuitive and requires no tools. This includes the top cover, planar

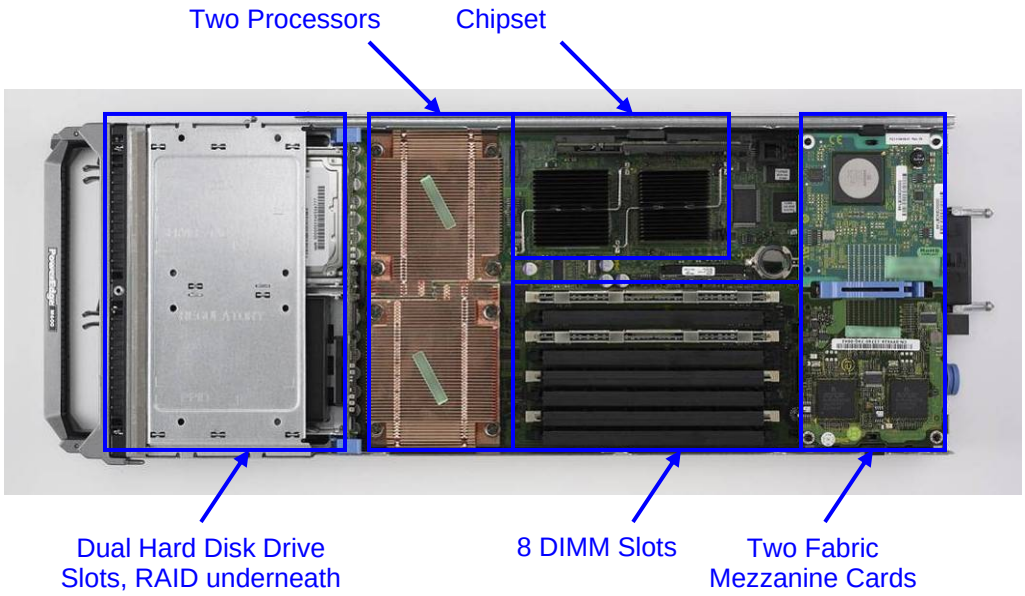


Figure 5 M600 Dual Socket Modular Server with Intel Processors

installation, hard drives, fabric cards and removable Hard Drive cage. Photos of the M600 and M605 blades are shown in Figure 5 and Figure 6 respectively, while overall blade dimensions are shown in Table 3.

Blade Measurement	Dimension
Length	520 mm
Height	193 mm
Width	50 mm

Table 3 Half Height Modular Server Dimensions

Server Module Front Panel

The Half Height server module's front panel provides access to two hot pluggable hard drives, two bootable USB connectors, one server power button and one status indicator. All server modules support hot pluggable 2.5 inch hard disk drives. The front panel is visible in the photo in Figure 7.

Server Module Removal and Translating Handle

The server module is manually installed by aligning it with one of the open front slot locations and sliding it until the connectors begin to engage with the midplane. A translating handle, which travels in the same direction as the server module, is then used to fully engage and lock it in place. To remove the server module, a button releases the translating handle latch. The handle is then used for pulling the server out of the system, disengaging the internal latch as the handle is pulled outwards and unmating the connectors.

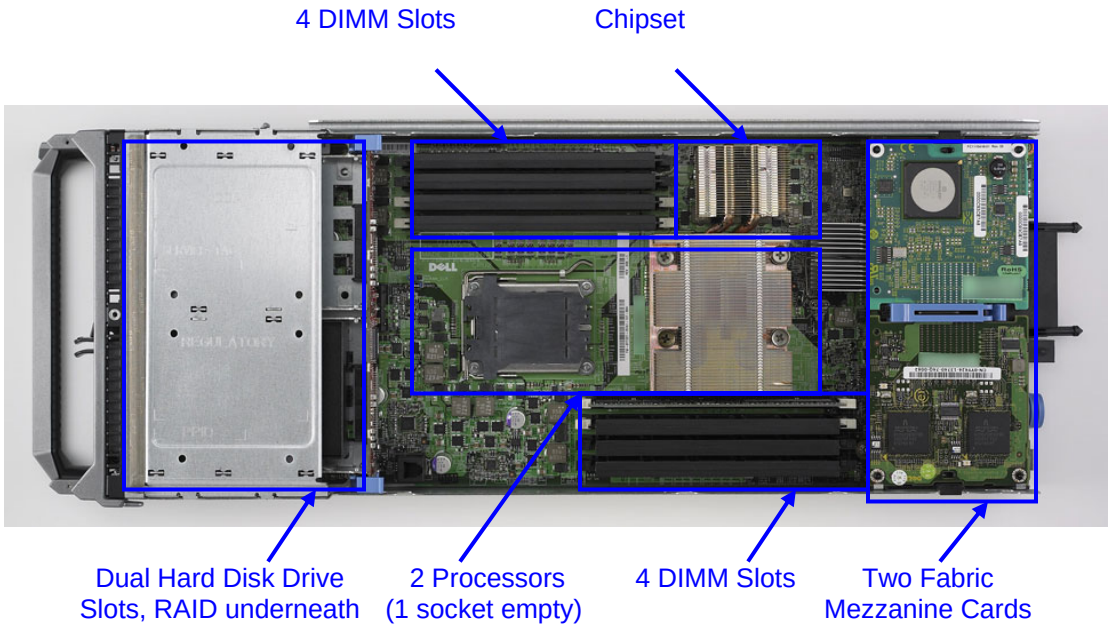


Figure 6 M605 Dual Socket Modular Server with AMD Processors

Figure 7 Half Height Modular Server Front Panel View

There are guide pins on both sides of the server module signal connector for precise alignment to the chassis. Keying features prevent the server module from installing upside down. A defined maximum sliding force insures ergonomic friendliness, while precision point of contact positioning and camming features ensure proper connector wipe for maximum signal quality. Modular server power is received through a dedicated 2x3 power block, with wide power pins designated in such a way to prevent power rail shorting in case a pin bends for any reason.

Figure 8 Modular Server Removal

Server Module Summary

Table 4 summarizes the features for the first two server modules shipping for the M1000e. Additional server modules are expected to ship following the launch of the M1000e, and throughout the lifetime of the modular system.

Server Module	M600	M605
Processor	Intel Xeon Dual and Quad Core 40W, 65W, 80W, and 120W options	AMD Opteron 2000 series Dual and Quad Core 68W and 95W options
Chipset	Intel 5000P (Blackford)	NVIDIA MCP55
Memory Slots	8 Fully Buffered DIMMs (667 MHz)	8 DDR2 (667/800 MHz)
Memory Capacity	32GB (4GB x 8) at launch 64 GB (8GB x 8) planned Q108	32GB (4GB x 8) at launch 64 GB (8GB x 8) planned Q108
LOM	2 x GE with hardware TCP/IP Offload Engine and iSCSI Firmware Boot Upgrade to full iSCSI offload via license key	2 x GE with hardware TCP/IP Offload Engine and iSCSI Firmware Boot Upgrade to full iSCSI offload via license key

Server Module	M600	M605
Fabric Expansion	2 x 8 lane PCIe mezzanine daughtercards 1. Dual port GE w/ TOE 2. Dual Port FC4 (Emulex & Qlogic) 3. Dual Port 4x DDR InfiniBand	2 x 8 lane PCIe mezzanine daughtercards 1. Dual port GE w/ TOE (*due to PCIe controller limitations the dual 5708S TOE based card cannot be used in Fabric C. Dell plans to release an integrated dual TOE solution that will resolve this limitation.) 2. Dual Port FC4 (Emulex & Qlogic) 3. Dual Port 4x DDR InfiniBand
Baseboard Management	iDRAC w/ IPMI 2.0 + vMedia + vKVM	iDRAC w/ IPMI 2.0 + vMedia + vKVM
Local Storage Controller Options	SATA (chipset based- no RAID or hotplug) SAS6/IR (R0/1) CERC6/i (R0/1 w/ Cache)	SATA (chipset based- no RAID or hotplug) SAS6/IR (R0/1) CERC6/i (R0/1 w/ Cache)
Local Storage HDD	2x 2.5 inch hot pluggable SAS or SATA	2x 2.5 inch hot pluggable SAS or SATA
Video	ATI RN50	ATI RN50
USB	2x USB 2.0 bootable ports on front panel for floppy/CD/DVD/Memory	2x USB 2.0 bootable ports on front panel for USB floppy/CD/DVD/Memory
Console	• Virtual KVM via iDRAC • IPMI Serial over LAN (SoL) via iDRAC • Rear mounted iKVM switch ports (tierable) • Front KVM ports on modular enclosure control panel	• Virtual KVM via iDRAC • IPMI Serial over LAN (SoL) via iDRAC • Rear mounted iKVM switch ports (tierable) • Front KVM ports on modular enclosure control panel
HA Clustering	Fibre Channel and iSCSI based clustering options	Fibre Channel and iSCSI based clustering options
Operating Systems	Microsoft W2K3 and W2K3 R2, Red Hat Enterprise Linux 4/5, SuSE Linux Enterprise Server 9/10	Microsoft W2K3 and W2K3 R2, Red Hat Enterprise Linux 4/5, SuSE Linux Enterprise Server 9/10

Table 4 Half Height Server Module Options

M1000e Midplane and I/O

M1000e Front View

Server modules are accessible from the front of the M1000e enclosure. At the bottom of the enclosure is a flip out multiple angle LCD screen for local systems management configuration, system information, and status. The front of the enclosure also contains two USB connections for USB keyboard and mouse, a video connection and the system power button. The front control panel's USB and video ports work only when the iKVM module is installed, as the iKVM provides the capability to switch the KVM between the blades.

Figure 9 M1000e Front View

The Dell M1000e supports up to sixteen Half Height server modules. The chassis guide and retention features are designed such that alternative module form factors are possible. See the Server Module section for more details. The chassis architecture is flexible enough that server, storage or other types of front loading modules are possible.

Not visibly obvious, but important nonetheless, are fresh air plenums at both top and bottom of the chassis. The bottom fresh air plenum provides non-preheated air to the M1000e power supplies, enabling ground breaking levels of power density. The top fresh air plenum provides non-preheated air to the CMC, iKVM and I/O Modules. See the Cooling section for more details.

Feature	Parameter
Chassis Size	10U high rack mount
Blades per Chassis	16 Half Height, 8 Full Height
Total Blades in a 42U Rack	64 Half Height, 32 Full Height
Total I/O Module Bays	6 (3 redundant or dual fabrics)
Total Power Supplies	6 (3+3 redundant)
Total Fan Modules	9 (8+1 redundant)
Management Modules and Interfaces	2 CMCs (1+1 redundant), 1 iKVM, Front Control Panel, Graphical LCD Control Panel
Width, not including rack ears	447.5 mm
Height	440.5 mm
Depth, Rear of EIA Flange to Rear of Chassis	753.6 mm
Total System Depth (Front Bezel to PS Latch):	835.99 mm

Table 5 M1000e Modular System Features and Parameters

M1000e Midplane

Though hidden from view in an actively running system, the midplane is the focal point for all connectivity within the M1000e Modular System. The midplane is a large printed circuit board providing power distribution, fabric connectivity and system management infrastructure. Additionally it allows airflow paths for the front to back cooling system through ventilation holes.

As is requisite for fault tolerant systems, the M1000e midplane is completely passive, with no hidden stacking midplanes or interposers with active components. I/O Fabrics and system management are fully redundant from each hot pluggable item. The system management Ethernet fabric is fully redundant when two CMCs are installed, with two point to point connections from each server module.

I/O fabrics are routed through 10 Gbps capable high speed connectors and dielectric material. All traces are differential 100 ohm characteristic impedance. Placement of connectors and pluggable components was determined based upon minimum routing distance for high speed critical nets. High speed I/O connector holes are backdrilled in the midplane to minimize electrical stub length and improve high speed signal integrity. Midplane material is an improved FR408 dielectric material, providing better high frequency transmission characteristics and lower attenuation. The I/O Channels have been simulated to 10GBASE-KR channel models. Per industry standard requirements, fabrics internally support a Bit Error Rate of 10^{-12} or better. At the system level this means that a high level of investment has been made to insure scalable bandwidth, for current and future generations of server and infrastructure.

The midplane serves as transport for a patent pending time division multiplexed serial bus for General Purpose I/O reduction. This serial bus contributes greatly to the midplane's I/O lane count reduction, which is typically burdened with a significant I/O pin and routing channel count of largely static or low speed functions. For instance, all Fibre Channel I/O Passthrough module LED and SFP status information is carried over this bus, which alone eliminates over one hundred point to point connections that would otherwise be required. The time division

multiplexed serial bus is fully redundant, with health monitoring, separate links per CMC and error checking across all data.

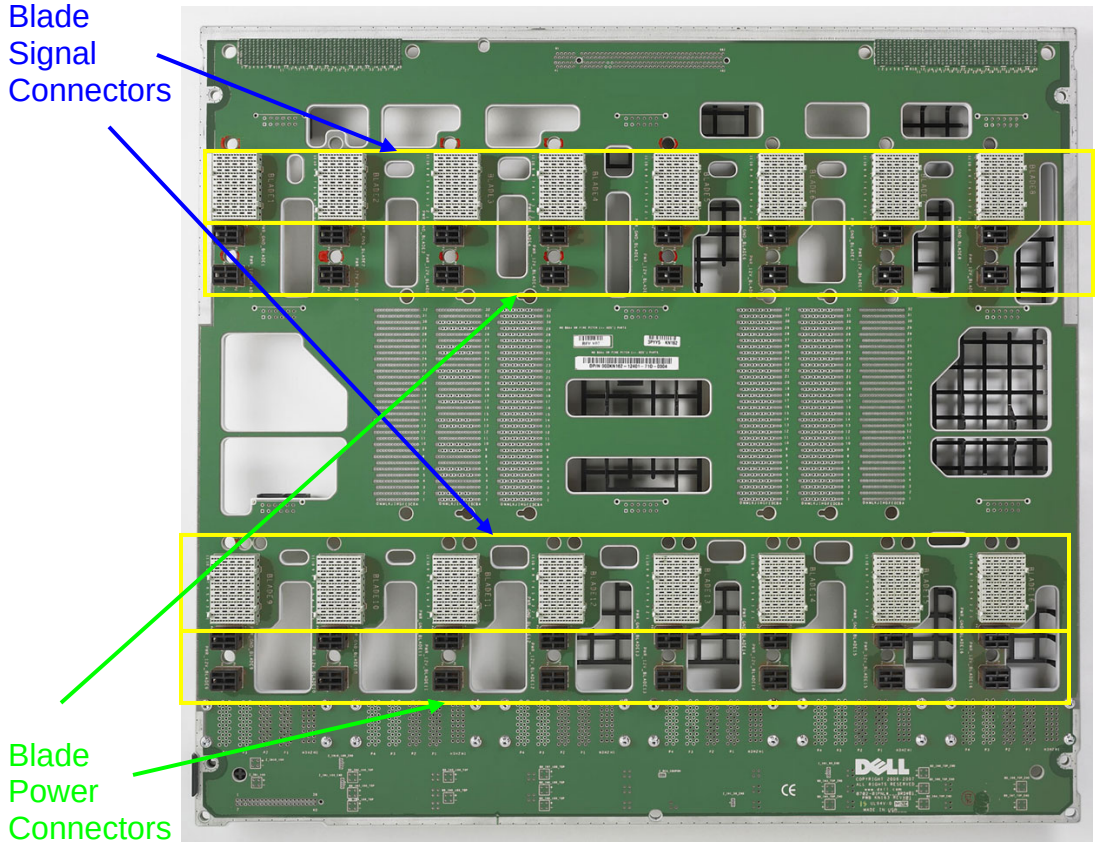


Figure 10 M1000e Midplane Front View

The system is designed for receptacles on all midplane connectors and pins on all pluggable components, so any potential for bent pins is limited to the pluggable field replaceable unit, not to the system. This contributes to the high reliability and uptime of the M1000e modular system.

The midplane is physically attached to the enclosure front structural element. It is aligned by guide-pins and edges in all 3 axes. This provides close tolerance alignment between the server modules and their midplane connections.

The midplane has been carefully designed to minimize the impact to the overall system airflow. The midplane design was evaluated at various points throughout the design cycle to make tradeoffs between thermal performance and midplane costs. Early analysis of the midplane showed large increases in airflow due to a move from 15% to 20% midplane opening. The ventilation holes resulting from this expansion are seen in Figure 10 and Figure 11. These changes added to the demanding requirements of the midplane design, but provided the best overall system solution.

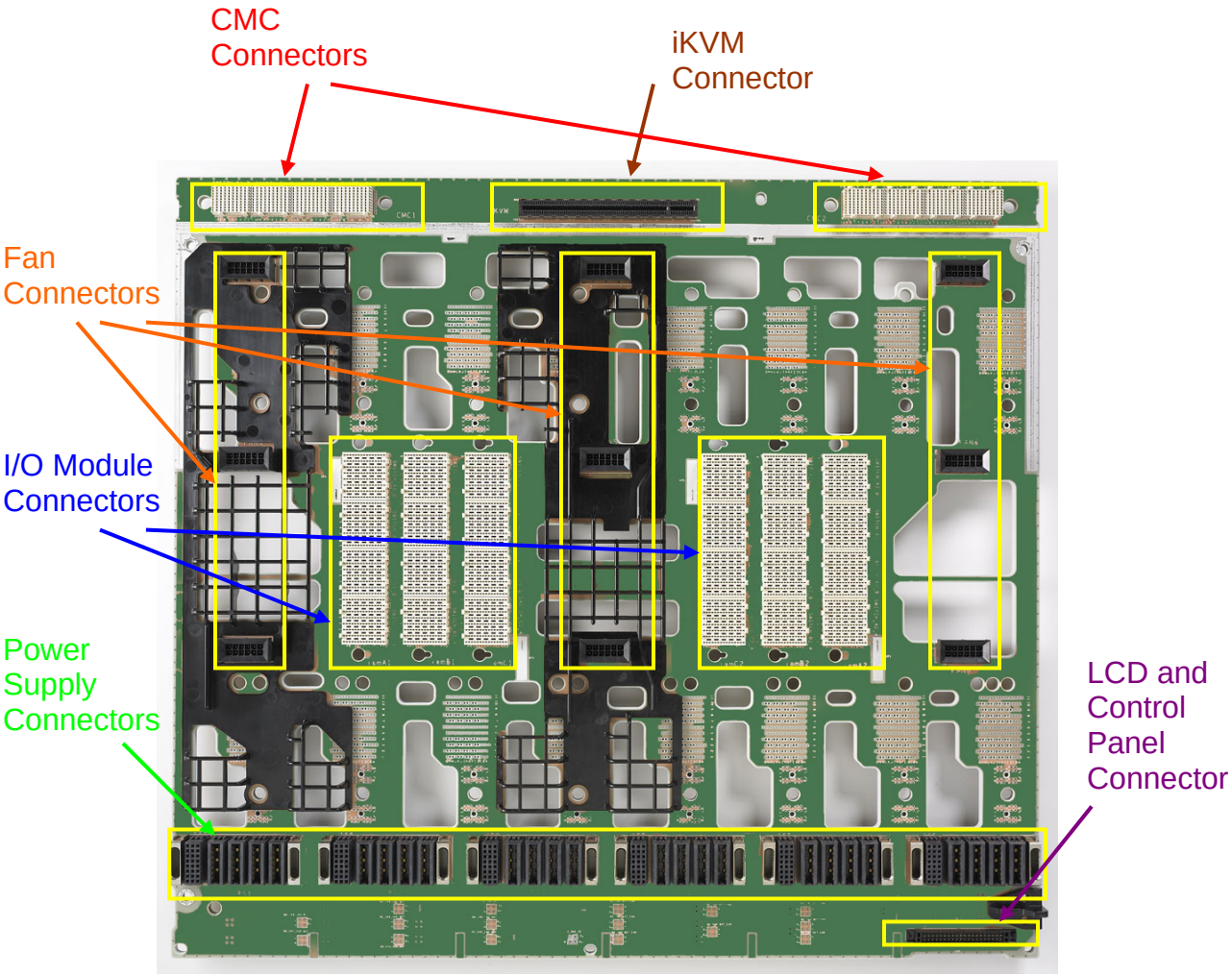


Figure 11 M1000e Midplane Rear View

While you might think that multiple midplanes are better for the product, there really is no system advantage if the midplane is properly designed. All M1000e midplane routing is fully isolated, supporting all chassis power, fabric, system management and fault tolerance requirements. The single integrated midplane also simplifies manufacturing assembly, reducing net acquisition cost to the customer.

M1000e I/O

M1000e I/O is fully scalable to current and future generations of server modules and I/O Modules. There are three redundant multi-lane fabrics in the system, as illustrated in Figure 12.

Fabric A is dedicated to Gigabit Ethernet. Though initial server module releases are designed as dual Gigabit Ethernet LOM controllers on the server module planar, the midplane is enabled to support up to four Gigabit Ethernet links per server module on Fabric A. Potential data bandwidth for Fabric A is 4 Gbps per Half Height server module.

One important new capability in the M1000e is full 10/100/1000M Ethernet support when using Ethernet passthrough modules. In the past, customers were limited to connecting only to external switches with 1000M ports. Now you will be able to connect to any legacy infrastructure

whether using Ethernet passthrough or switch technology. This technical advance uses in band signaling on 1000BASE-KX transport, and requires no user interaction for enablement.

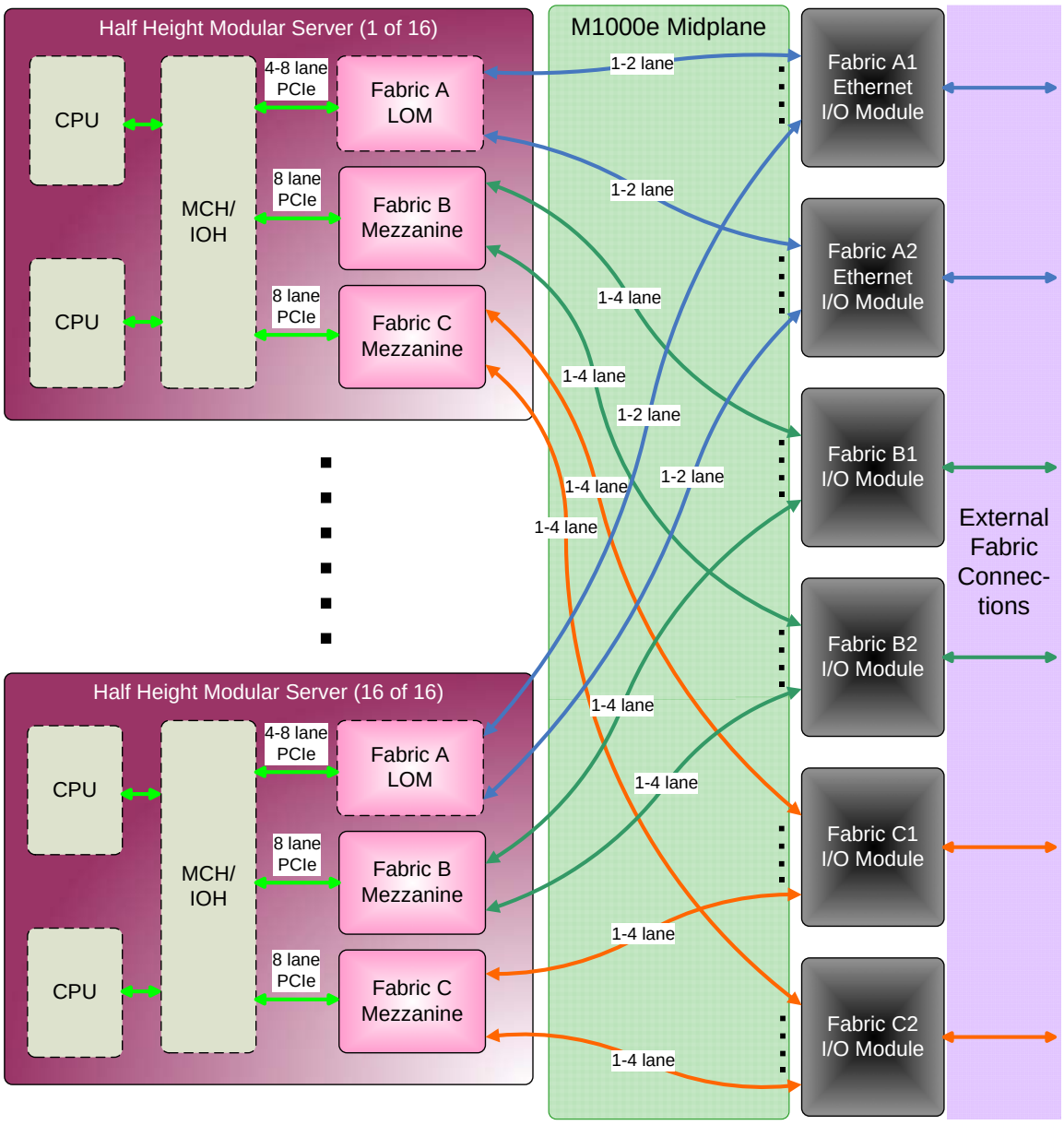


Figure 12 High Speed I/O Architecture

Fabric B and C are identical, fully customizable fabrics, routed as two sets of four lanes from mezzanine cards on the server modules to the I/O Modules in the rear of the chassis. Supported bandwidth ranges from 1 to 10 Gbps per lane depending on the fabric type used. Table 6 shows comparative bandwidth between current and forthcoming fabrics. Symbol Rate is defined as the maximum fabric speed with all data encoded as per its encoding mechanism. Data Rate is the

maximum amount of actual data bits that can be transported on the fabric. Some fabrics are defined as single lane only, while others are defined in various lane increments for link to lane aggregation, in order to provide increased total bandwidth.

As each mezzanine card is connected by an 8 lane PCIe link to the server chipset, there are no throttle points in the system where I/O bandwidth is constricted. When multi-lane 10GBASE-KR becomes a reality, server modules will have moved to PCIe Gen2 or better, providing full end to end I/O bandwidth from the server modules to the I/O Modules.

Fabric	Encoding	Symbol Rate Per Lane (Gbps)	Data Rate Per Lane (Gbps)	Data Rate Per Link (Gbps)	Lanes Per Link Per Industry Specification
PCIe Gen1	8B/10B	2.5	2	8 (4 lane)	1, 2, 4, 8, 12, 16, 32
PCIe Gen2	8B/10B	5	4	16 (4 lane)	1, 2, 4, 8, 12, 16, 32
PCIe Gen3	scrambling	8	8	32 (4 lane)	1, 2, 4, 8, 12, 16, 32
SATA 3G	8B/10B	3	2.4	2.4	1
SATA 6G	8B/10B	6	4.8	4.8	1
SAS 3G	8B/10B	3	2.4	2.4	1-Any
SAS 6G	8B/10B	6	4.8	4.8	1-Any
FC 4G	8B/10B	4.25	3.4	3.4	1
FC 8G	8B/10B	8.5	6.8	6.8	1
IB SDR	8B/10B	2.5	2	8 (4 lane)	4, 12
IB DDR	8B/10B	5	4	16 (4 lane)	4, 12
IB QDR	8B/10B	10	8	32 (4 lane)	4, 12
GE: 1000BASE-KX	8B/10B	1.25	1	1	1
10GE: 10GBASE-KX4	8B/10B	3.125	2.5	10 (4 lane)	4
10GE: 10GBASE-KR	64B/66B	10.3125	10	10	1

Table 6 Relative Comparison of Industry Standard High Speed I/O

Ethernet Bandwidth Migration Path

The bandwidth growth path for Ethernet is from GE to 10GE, supported through three backplane standards defined by the IEEE 802.3ap Committee. 1000BASE-KX supports GE through 1.25 Gbps differential backplane signaling. 10GBASE-KX4 supports 10GE through four lanes of 3.125 Gbps differential backplane signaling. The four lanes are aggregated to provide the sum total 10 Gbps bandwidth requirement. 10GBASE-KR is an emerging IEEE 802.3ap standard that supports 10GE through a single lane of 10.3125 Gbps differential backplane signaling. In terms of customer impact, this means that today's modular systems using 10GBASE-KX4 are limited to one 10GE channel per 4 lanes of internal differential signaling, while in the future those same lanes can carry multiple lanes of 10GBASE-KR and provide a resultant increase in total system throughput. See Figure 13 for an illustration of the 10GE growth path. This growth path applies to both Fabrics B and C in the M1000e.

GE fabrics use single lanes per link over 1000BASE-KX. Initially 10GE will be supported through 10GBASE-KX4, which like InifiniBand uses four lanes per link. As technology matures and enables high level integration, 10GBASE-KR will replace 10GBASE-KX4 as the 10GE backplane transport of choice, enabling 10 Gbps over single lanes within the system midplane. Finally,

multiple lanes of 10GBASE-R per mezzanine will allow expanded Ethernet bandwidth up to 80 Gbps per fabric mezzanine.

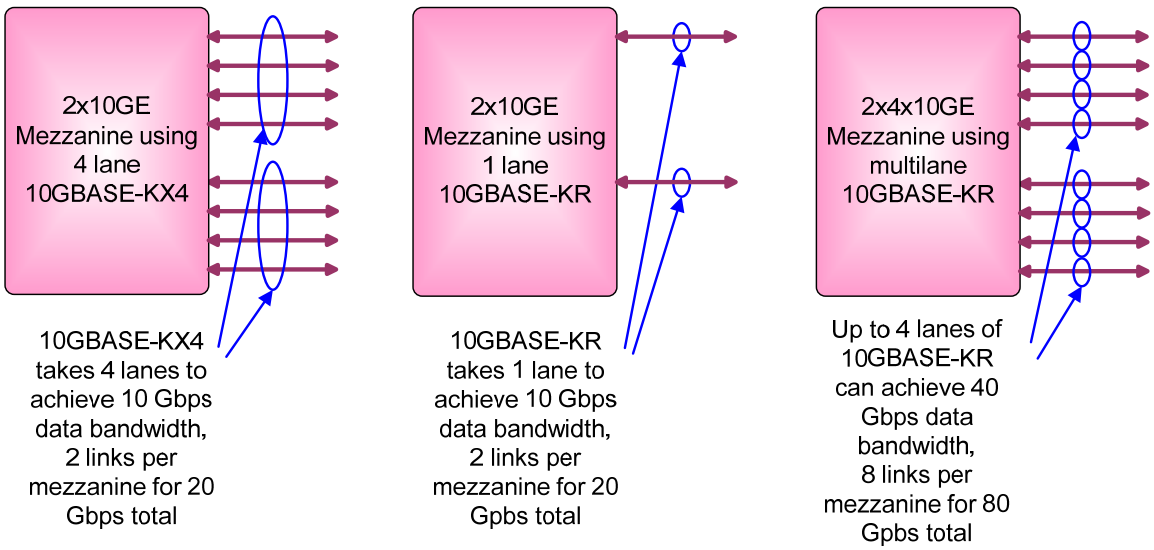


Figure 13 10GE Potential Growth Path

Total System Bandwidth

In most cases, only a small percentage of this bandwidth is used in a system, as illustrated in Figure 14. A typical configuration is defined as 4 lanes of GE and 2 lanes of Fibre Channel per server module, and uses less than 10% of the M1000e potential bandwidth. An example of a unified Ethernet configuration, aggregating all network, storage and interprocessor communications on Ethernet links, is shown as 2 lanes of GE and 4 lanes of 10GE per server module, using 25.7% of M1000e potential bandwidth. An example of a High Performance

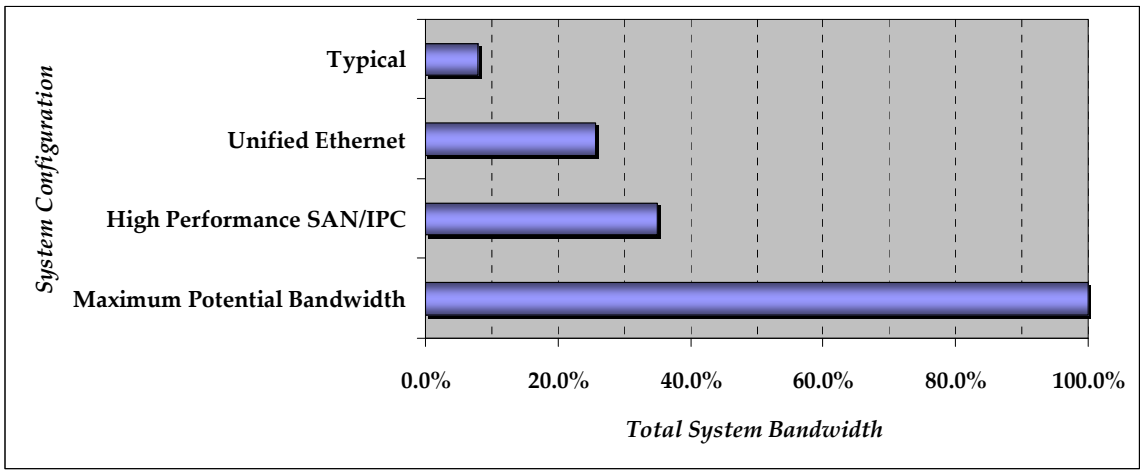


Figure 14 M1000e Modular Server System Bandwidth

SAN/IPC configuration is shown as 2 lanes of GE, 2 lanes of FC8, and 2 lanes of QDR InfiniBand, and still uses only 35% of M1000e potential bandwidth. The results of the graph show that the M1000e is fully scalable for I/O infrastructure needs for many generations of servers and switches.

Assuming a full population of GE and 10GbE lanes in Fabrics A, B and C, the M1000e can deliver theoretical total bandwidth of 5.44 Terabits per second. However, this extreme bandwidth comparison may not ever be fully realized in real world application usage. It also does not address factors such as flexibility of fabric migration paths.

The M1000e is designed for full support of all near, medium and long term I/O infrastructure needs. While the M1000e system's bandwidth capabilities lead the industry, the M1000e is also intelligently designed for maximum cost, flexibility and performance benefit. An example of this is the routing of dual 4 lane paths for Fabrics B and C, and dual 2 lane paths in Fabric A. In the near term, 10GBASE-KX4 routing supports 10GE connectivity. 10GBASE-KX4 routing uses all 4 lanes, each running at a 2.5Gbps data rate to achieve total link data bandwidth of 10 Gbps. Today and in the near term future, 10GBASE-KX4 is the most cost effective, ubiquitous solution for 10GE fabrics over midplanes. A fully configured M1000e system supports 2x2 GE and 2x2 10GE implementations with 10GBASE-KX4 routing, supporting the leading edge of unified network topologies. Such a configuration could, for instance, provide redundant 10GE links per blade for traditional network traffic, and another set of redundant 10GE links for iSCSI or Fibre Channel Over Ethernet (FCoE) networked storage.

I/O Modules are fully compatible across slots. While Fabric A is dedicated to the server module LOMs, requiring Ethernet switch or passthrough modules for I/O slots A1 and A2, Fabrics B and C can be freely populated with Ethernet, Fibre Channel or InfiniBand solutions. As part of Dell's effort to simplify IT, there is no confusing support matrix to follow for mezzanines or I/O modules and no multiplication of mezzanine and I/O Module form factors and design standards, with resultant compatibility or configuration concerns. Dell supports one mezzanine design standard and one I/O Module design standard for true modular computing.

M1000e Rear View

The rear of the M1000e Enclosure contains system management, cooling, power and I/O components. At the top of the enclosure are slots for two Chassis Management Cards and one integrated Keyboard/Video/Mouse switch. The enclosure ships by default with a single CMC, with the customer having the option of adding a second CMC to provide a fully redundant, active-standby fault tolerant solution for management access and control.

Interleaved in the center of the chassis are fans and I/O Modules. This arrangement optimizes the balance of airflow through the system, allowing lower pressure buildup in the system and resulting in lower airflow requirements for the fans.

I/O Modules are used as pairs, with two modules servicing each server module fabric and fully redundant. I/O Modules may be passthroughs or switches. Passthrough modules provide direct 1:1 connectivity from each LOM/mezzanine card port on each server module to the external network. Switches provide an efficient way to consolidate links from the LOM or Mezzanine cards on the server modules to uplinks into the customer's network. The Enclosure above is shown with 10/100/1000Mb Ethernet and FC4 Passthrough Modules. Figure 16 shows the

difference between these two types of IOMs. Figure 17 shows the Dell Gigabit Ethernet Passthrough Module.

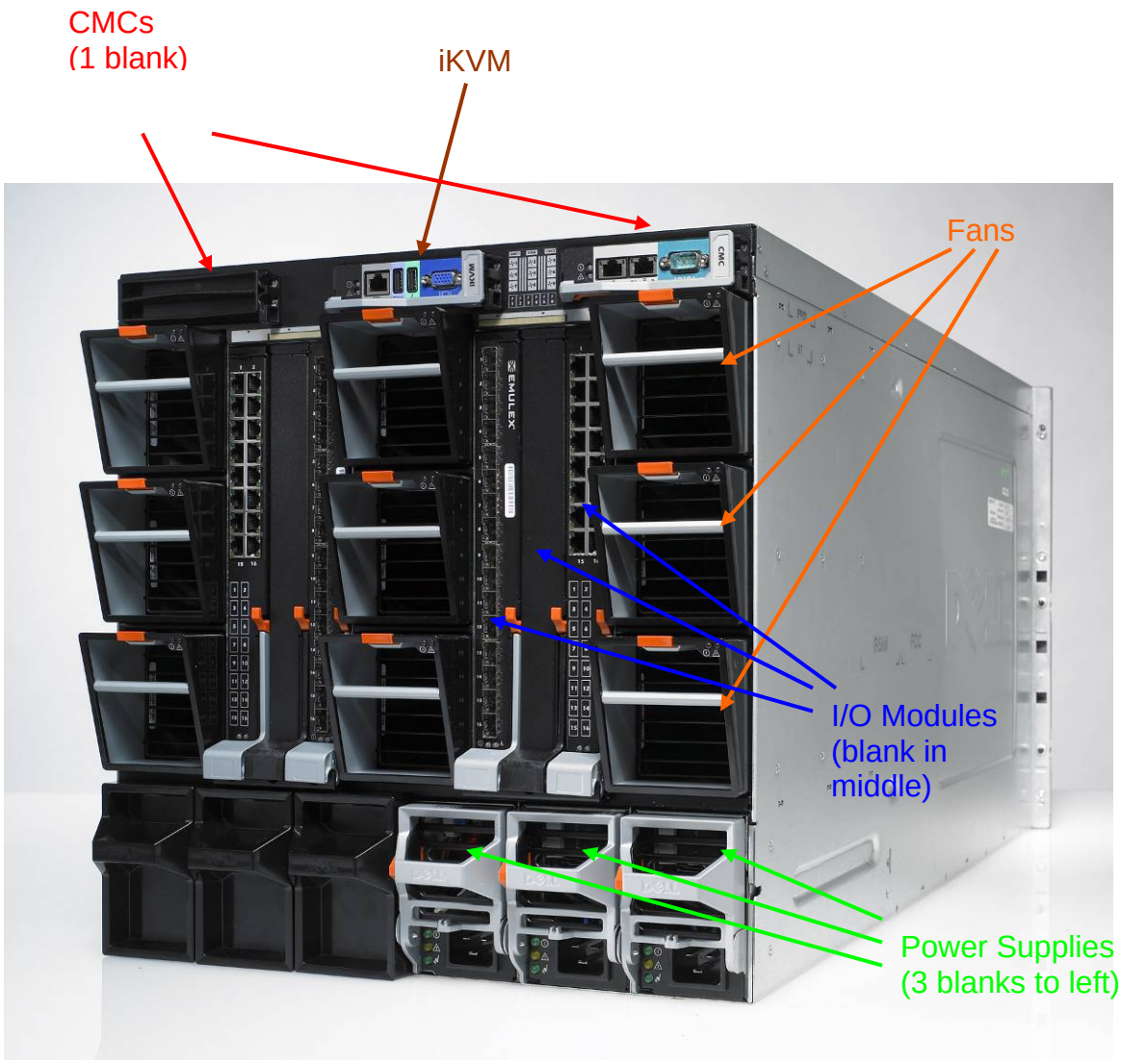


Figure 15 M1000e Rear View

The Dell and Cisco Ethernet switches for the M1000e enclosure uniquely demonstrate the level of forethought that went into the design of the I/O modules. These Ethernet switch modules introduce submodule I/O expansion for maximum flexibility and extension, a capability never before offered in modular systems. They provide a single hardware design that allows scaling from a low cost all GE solution, to a solution that utilizes switch stacking technologies to interconnect multiple switches within or between chassis, and/or to a solution that supports all of these plus 10GE uplinks to the core network with flexible interface types, both copper and optical. Figure 18 shows the Dell PowerConnect Gigabit Ethernet Switch Module with its two submodule bays available for switch I/O expansion.

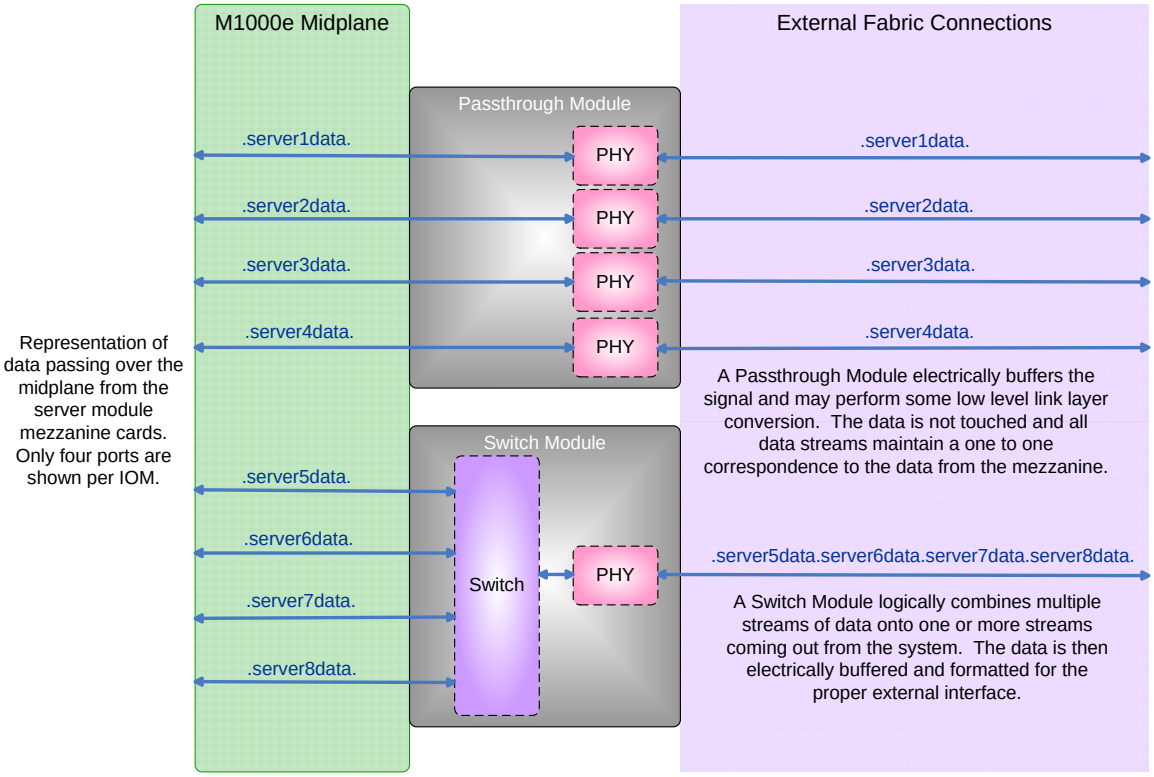


Figure 16 Difference between Passthrough and Switch Modules

Figure 17 I/O Module, Dell PowerEdge Gigabit Ethernet Passthrough

Figure 18 I/O Module, Dell PowerEdge Gigabit Ethernet Switch

The initial options for I/O Modules are outlined in Table 7, Table 8 and Table 9. Dell plans future I/O Module releases supporting all technology upgrade paths, including 8G FC, QDR IB and 10GE.

I/O Module / Features	10/100/ 1000M Ethernet Passthrough	Dell Power Connect M6220 Ethernet Switch	Cisco Catalyst Blade Switch 3032*, 3130S-G*, 3130S-X*
Internal ports	16	16	16
External ports	16 x 10/100/1000M	4 x fixed copper 10/100/1000M plus 2 of the following option modules: 1) 2 x 24G stacking ports 2) 2 x 10G Optical (XFP-SR/LR) uplinks 3) 2 x 10G copper CX4 uplinks 4) 2 x 10GBASE-T copper uplinks*	**3032**: 4 fixed copper 10/100/1000M + 2 optional bays each supporting 2 x 1G copper or optical SFPs **3130S-G**: Above + 2 x 32G Stackwise Plus Stacking ports **3130S-X**: Above + up to 2 x X2 modules for 10G CX4 or SR/LR uplinks
Speed	10/100/1000M	Internal ports- 1G External ports Fixed RJ-45- 10/100/1000M CX4 uplink- 10GE Optical XFP- 10GE 10GBASE-T – 10GE	Internal ports- 1G External ports Fixed RJ-45- 10/100/1000M X2 CX4 uplink- 10GE X2 SR/LRM Optical - 10GE
Manage-ability	N/A	CLI/Web	CLI/Web, Cisco mgmt tools
Features	Support for 10/100/1000M speeds (hard set and auto-negotiation)	Layer 3 routing (OSPF, RIP, VRRP), 10GE connections, Stacking, IPv6, Layer 2/3 QOS, Access Control Lists	L2 switching, Base L3 routing (static routes, RIP), Access Control lists, L2/3 QoS Optional upgrades to IP Services (Adv. L3 protocol support) and Advanced IP Services (IPv6)

***planned**

Table 7 Ethernet I/O Module Options

I/O Module / Features	Emulex FC4 Passthrough	Brocade 4424 Access Gateway and FC4 Switch
Internal ports	16	8/16
External ports	16 x 4G SFP	4/8 x 4G SFP
Speed	1/2/4G	1/2/4G
Manageability	N/A	CLI/Web, Brocade and EMC Fabric Management tools
Features	N/A	Access Gateway Mode enables NPIV functionality on external ports (enhanced interoperability, simplified setup, doesn't consume an FC domain) 12 or 24 port enablement options

Table 8 Fibre Channel I/O Module Options

I/O Module / Features	Cisco SFS M7000 InfiniBand Switch*
Internal ports	16
External ports	8 Copper or Optical ports
Speed	4x DDR
Manageability	Fabric based management

***planned**

Table 9 InfiniBand I/O Module Option

At the bottom of the chassis are the system power supplies. See the power section for more details on power supply operation and the power system infrastructure.

Cable Management

One of the advantages of a modular server system is the reduction in cable management needs within a rack system. The inclusion of fabric switches, integrated KVM and system management aggregation at the CMCs provides six-fold or better cable reduction. Table 10 shows a comparison of a typical reduction available when using the M1000e Modular system with integrated switches, compared to traditional "rack and stack" components. The configuration in the table assumes a server with four Ethernet ports and two Fibre Channel ports. In support of the M1000e Dell is releasing an improved modular system cable management system to ease system installation in Dell or other industry standard racks.

Component	Rack Height	AC power cables	Ethernet Cables	FC Cables	KVM Cables
2 socket server	1Ux16	2x16	4x16	2x16	USBx16 + VGAx16
KVM	1U	1	-	-	USBx1 + VGAx1
Ethernet Switches	1Ux4	1x4	4x4	-	-
FC Switches	1Ux2	1x2	-	2x2	-
Total Rack	**23U height**	**39 AC cables**	**72 Ethernet Cables**	**36 FC Cables**	**USBx17 + VGAx17**
M1000e Equivalent	**10U height**	**6 AC cables**	**16 Ethernet Cables**	**4 FC Cables**	**USBx1 + VGAx1**

Table 10 Typical Modular Server System Rack Height and Cable Reduction

Power

Dell leads the industry in power/performance efficiency with its EnergySmart products, and the M1000e continues this tradition. A modular system has many advantages over standard rack mount servers in terms of power optimization, and this aspect was a focal point throughout the M1000e's conceptualization and development. The key areas of interest are Power Delivery and Power Management.

Power Delivery

In support of the M1000e, Dell introduces a Single Phase 60 Amp PDU and a Three Phase 30 Amp PDU. Because three phase 30 Amp power is the most common data center high power feed, the M1000e has been sized to fit perfectly within that envelope, taking one three phase 30 Amp feed per set of three power supplies under any loading condition. The three phase power grid is complemented by the 3+3 redundant power system of the M1000e.

Features for the PDUs are shown in Table 11. PDUs support worldwide standard high power connections for ease of installation, and standard zero U vertical mounting within Dell racks. Typical configurations for the M1000e with single phase and three phase PDUs are shown in Figure 19 and Figure 20.

Feature	Single Phase PDU	Three Phase PDU
Mounting	1U Horizontal or zero U vertical	
Outlets	3 x C19	
Input Voltage	Single phase 200 to 240 VAC nominal	3-phase 200 to 240 VAC nominal
Line Frequency	47 to 63 Hertz	
Recommended AC service	60 Amps	30 Amps (NA/Japan), 32 Amps (International)
Fixed Input Plug/Cord Rating	IEC-309 60 A Pin & Sleeve Plug	NEMA L15-30P (NA/Japan), IEC 309 4 pole, 4 wire, 380-415 VAC, 32A (International)
Output Rating Voltage	200-240 VAC 60/50 Hz, 1-phase	
Output Rating Current (IEC320 C19)	16 Amps	
Circuit Breaker, Over Current Protection	20 A, per outlet receptacle	

Table 11 M1000e Single Phase and Three Phase PDU Options

The power distribution inside the M1000e Modular Server System consists of a 3+3 redundant power supply system, located in the rear bottom of the chassis. Each power supply is rated at 2360W. With current sharing between power supplies, total system redundant power is approximately 6700W in a 3+3 power supply configuration.

The M1000e provides industry leading power efficiency and density, accomplished through highly efficient components, improved design techniques and a fresh air plenum that reduces the air temperature to the power supply components. Lower operating temperature equates to higher power density for the power supply, exceeding 21 Watts per cubic inch, and higher power efficiency, better than 86% at 20% load, and higher at heavier loads, approaching 91% efficiency under normal operating conditions.

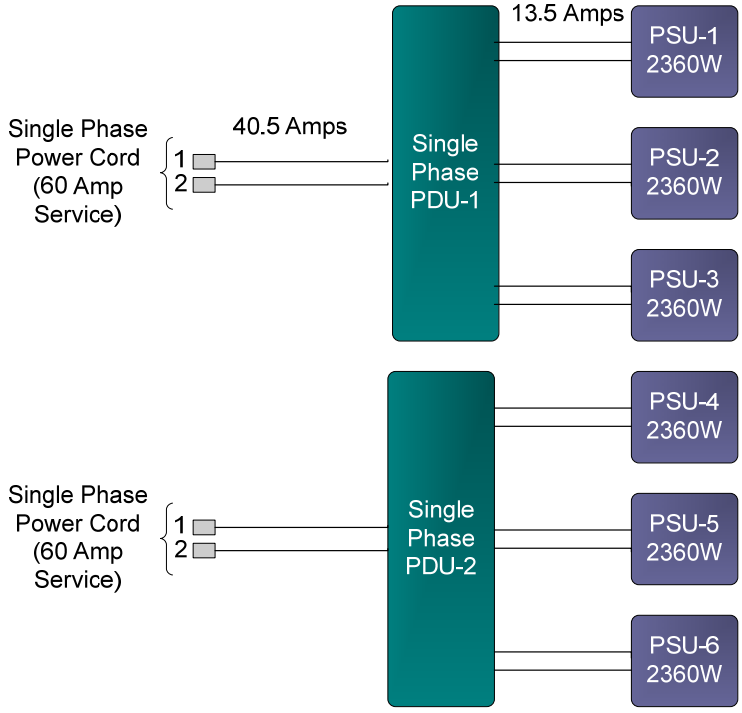


Figure 19 M1000e Single Phase PDU Power System

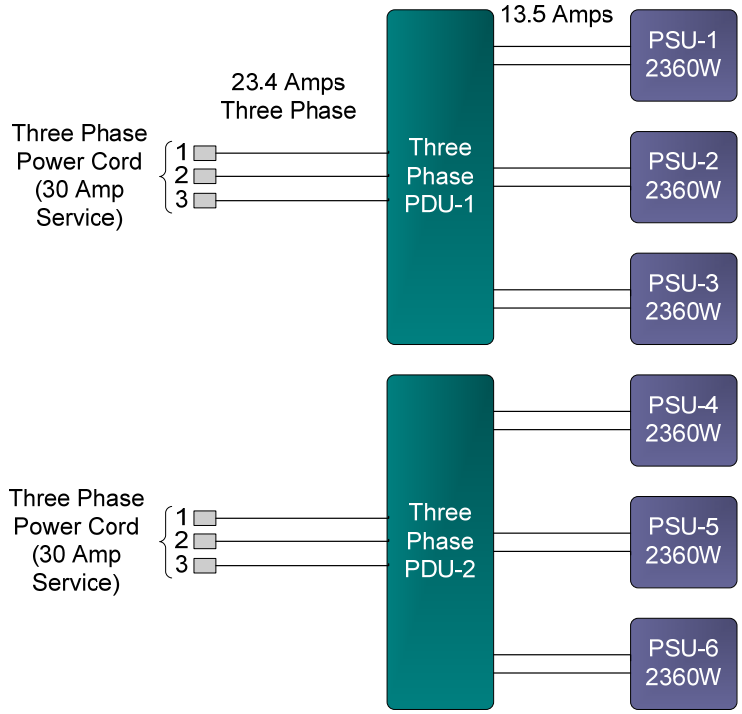


Figure 20 M1000e Three Phase PDU Power System

Figure 21 M1000e 2360 Watt Power Supply

Power efficiency in the M1000e does not stop with the power supply. Every aspect of efficiency has been tweaked and improved from previous designs, adding more copper to PC board power planes to reduce I^2R losses, improved inductors and other components, increasing efficiencies of DC-DC converters, and replacing some linear voltage regulators with more efficient switching regulators.

Power redundancy in the M1000e supports any necessary usage model. The M1000e requires three 2360 Watt power supplies to power a fully populated system or six power supplies in a fully redundant system. Figure 22 shows typical power redundancy models. In the N+N power supply configuration, the system will provide protection against AC grid loss or power supply failures. If one power grid fails, three power supplies lose their AC source, and the three power supplies on the other grid remain powered, providing sufficient power for the system to continue running. In the N+1 configuration only power supply failures are protected, not grid failures. The likelihood of multiple power supplies failing at the same time is remote. In the N+0 configuration there is no power protection and any protection must be provided at the node or chassis level. Typically this case is an HPCC or other clustered environment where redundant power is not a concern, since the parallelism of the processing nodes across multiple system chassis provides all the redundancy that is necessary.

The midplane carries all 12 Volt DC power for the system, both main power and standby power. The CMCs, LCD and Control Panel are powered solely by 12 Volt Standby power, insuring that chassis level management is operational in the chassis standby state, whenever AC power is present. The server modules, I/O Modules, Fans, and iKVM are powered solely by 12 Volt Main power.

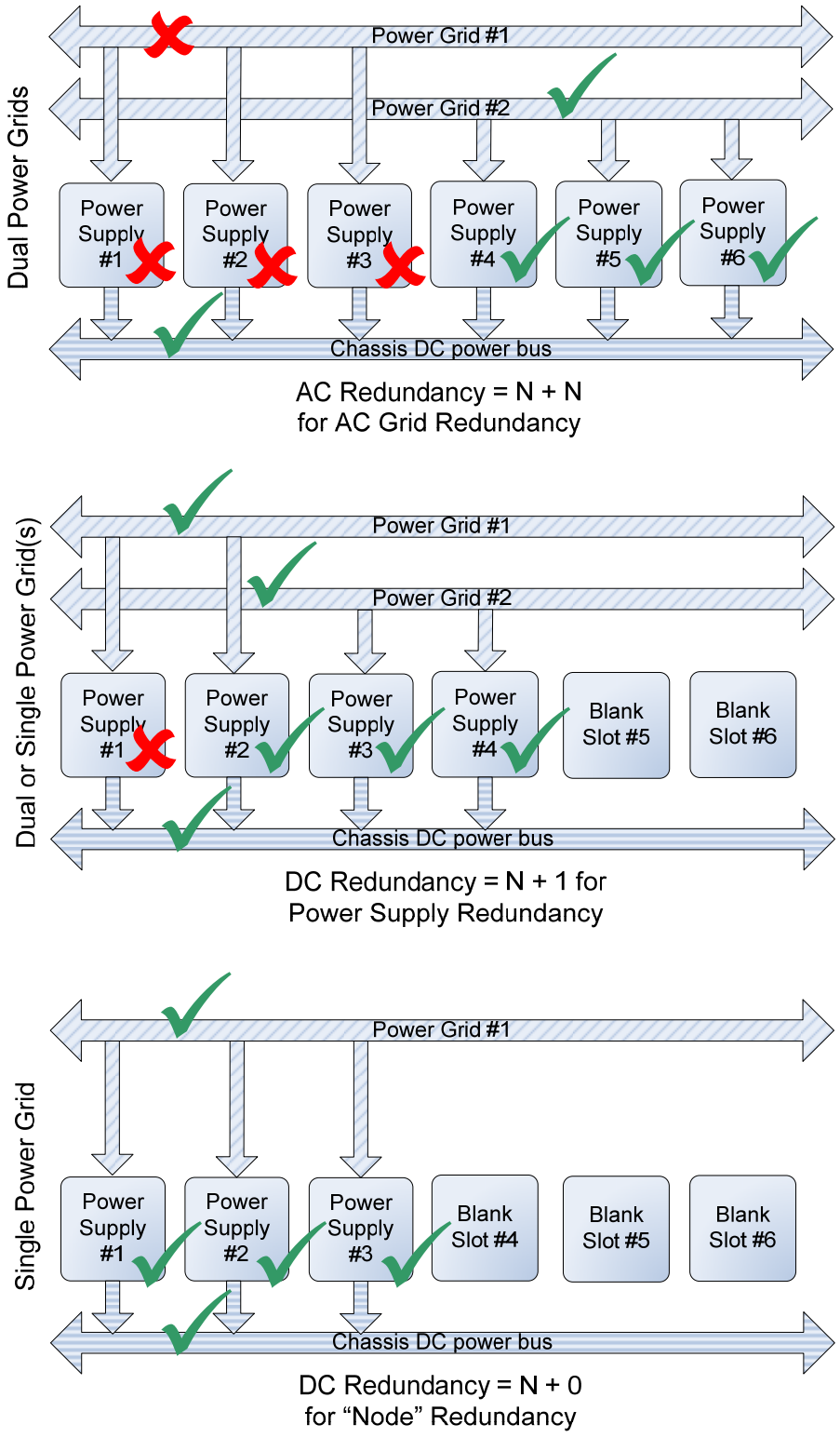


Figure 22 Power Redundancy Modes

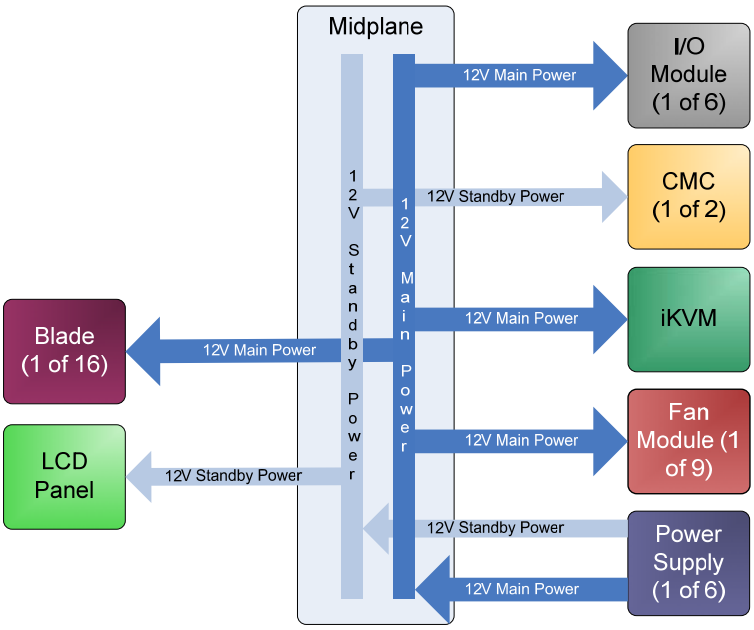


Figure 23 Power Architecture

Power Management

Power is no longer just about power delivery, it is also about power management. The M1000e System adds a number of advanced power management features. Most of these features operate transparently to the user, while others require only a one time selection of desired operating modes.

Shared power takes advantage of the large number of resources in the modular server, distributing power across the system without the excess margin required in dedicated rack mount servers and switches. The M1000e introduces an advanced power budgeting feature, controlled by the CMC and negotiated in conjunction with the iDRAC on every server module. Prior to any server module powering up, through any of its power up mechanisms such as AC recovery, WOL or a simple power button press, the server module iDRAC performs a sophisticated power budget inventory for the server module, based upon its configuration of CPUs, memory, I/O and local storage. Once this number is generated, the iDRAC communicates the power budget inventory to the CMC, which confirms the availability of power from the system level, based upon a total chassis power inventory, including power supplies, iKVM, I/O Modules, fans and server modules. Since the CMC controls when every modular system element powers on, it can now set power policies on a system level.

In coordination with the CMC, iDRAC hardware constantly monitors actual power consumption at each server module. This power measurement is used locally by the server module to insure that its instantaneous power consumption never exceeds the budgeted amount. While the system administrator may never notice these features in action, what they enable is a more aggressive utilization of the shared system power resources. No longer is the system "flying blind" in regards to power consumption, and there is no danger of exceeding power capacity

availability, which could result in a spontaneous activation of power supply over current protection without these features.

The system administrator can also set priorities for each server module. The priority works in conjunction with the CMC power budgeting and iDRAC power monitoring to insure that the lowest priority blades are the first to enter any power optimization mode, should conditions warrant the activation of this feature.

The M1000e introduces compliance to PMBus Specification 1.1, using this new power management standard for status, measurement and control. M1000e power supplies continuously monitor AC input current, voltage and power, enabling exposure of data to Dell OpenManage IT Assistant or to other enterprise level management tools. Real time power consumption is now viewable per system.

Cooling

The cooling strategy for the M1000e supports a low-impedance, high-efficiency design philosophy. Driving lower airflow impedance allows the M1000e to draw air through the system at a lower operating pressure than competitive systems. In some cases, this can be up to 40% less backpressure than similarly configured competitive products. The lower backpressure reduces the system fan power consumed to meet the airflow requirements of the system.

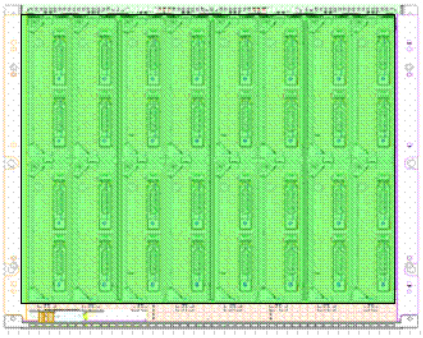
The low impedance design is coupled with a high-efficiency air moving device designed explicitly for the M1000e. The efficiency of an air moving device is defined as the work output of the fan as compared to the electrical power required to run the fan. The M1000e fan operates at efficiencies up to 40% greater than typical axial fans on the market, which correlates directly into savings in the customer's required power-to-cool.

Figure 24 M1000e Fan

The high-efficiency design philosophy also extends into the layout of the subsystems within the M1000e. The Server Modules, I/O Modules, and Power Supplies are incorporated into the system

with independent airflow paths. This isolates these components from pre-heated air, reducing the required airflow consumptions of each module.

Server Module Inlet
(M1000e front view)



Server Module Cooling Air Profile
(M1000e side view)

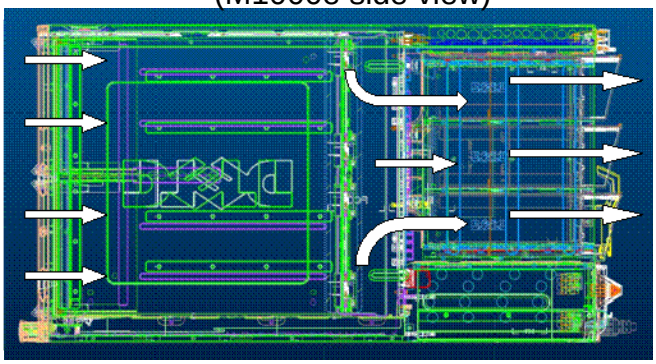
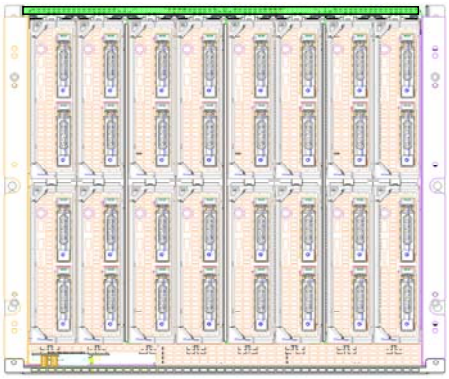


Figure 25 Server Module Cooling Air Profile

I/O Module Inlet
(M1000e front view)



IOM Locations and Airflow Direction
(M1000e back view)

I/O Cooling Air Profile
(M1000e side view)

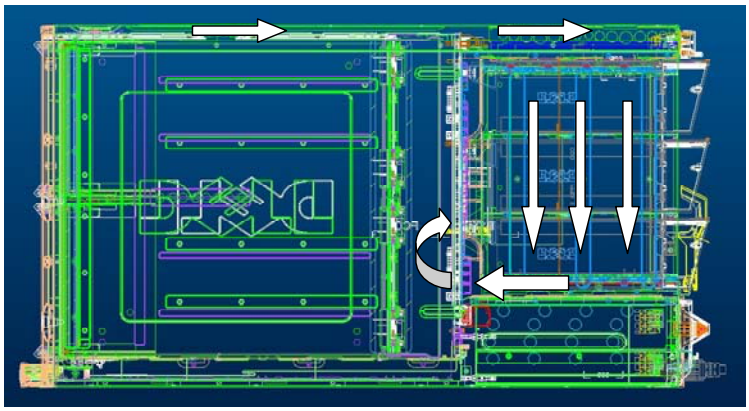


Figure 26 I/O Module Cooling Air Profile

The Server Modules are cooled with traditional front-to-back cooling. As shown in Figure 25, the front of the system is dominated by inlet area for the individual server modules. The air passes through the server modules, through venting holes in the midplane, and is then drawn into the fans which exhaust the air from the chassis. There are plenums both upstream of the midplane, between the midplane and the blades, and downstream of the midplane, between the midplane and the fans, to more evenly distribute the cooling potential from the three columns of fans across the server modules.

The I/O Modules use a bypass duct to draw ambient air from the front of the system to the I/O Module inlet, as seen in Figure 26. This duct is located above the server modules. This cool air is then drawn down through the I/O Modules in a top to bottom flow path and into the plenum between the midplane and fans, from where it is exhausted from the system.

The Power Supplies, located in the rear of the system, use basic front-to-back cooling, but draw their inlet air from a duct located beneath the server modules, as seen in Figure 27. This insures that the power supplies receive ambient temperature air.

Power Supply Inlet
(M1000e front view)

Power Supply Cooling Air Profile
(M1000e side view)

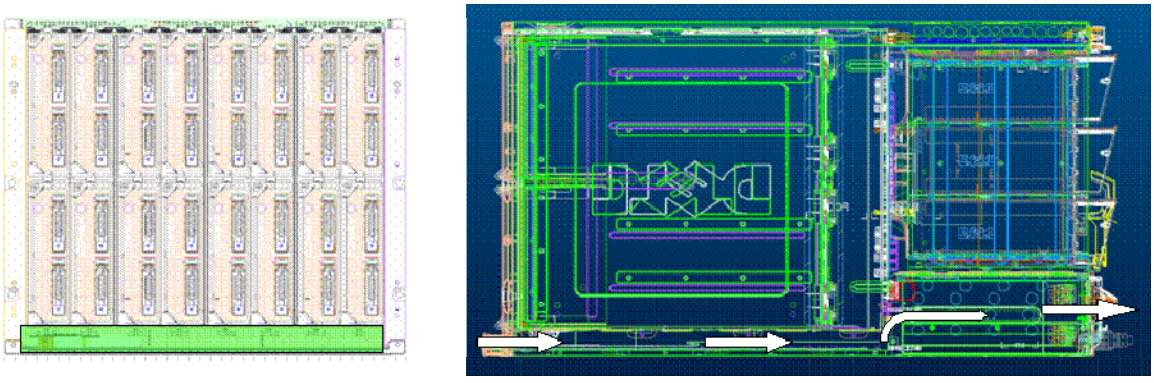


Figure 27 Power Supply Cooling Air Profile

This hardware design is coupled with a thermal cooling algorithm that incorporates the following:

- Server Module level thermal monitoring by the iDRAC
- I/O Module thermal health monitors
- Fan control and monitoring by the CMC

The iDRAC on each server modules calculates the amount of airflow required on an individual server module level and sends a request to the CMC. This request is based on temperature conditions on the server module, as well as passive requirements due to hardware configuration. Concurrently, each IOM can send a request to the CMC to increase or decrease cooling to the I/O subsystem. The CMC interprets these requests, and can control the fans as required to maintain Server and I/O Module airflow at optimal levels.

Fans are N+1 redundant, meaning that any single fan can fail without impacting system uptime or reliability. In the event of a fan failure, system behavior is dependent on the resultant temperatures of the system, as monitored by the Server Module iDRAC and I/O Modules. The CMC continues to interpret the airflow needs of each server and I/O module to control the fan speeds appropriately. The system will not ramp the fans to full speed in the event of a fan failure unless deemed necessary by on-board monitoring.

The M1000e is engineered for good sound quality in accordance with the Dell Enterprise acoustical specification. Compared to previous generations of products the fans have more levels of control, allowing much finer tuning of the fan behavior. Firmware is optimized to choose the lowest fan speeds and therefore the lowest acoustical output for any condition and configuration.

System Management

Dell's M1000e modular server delivers major enhancements in management features. These features are based on direct customer feedback and assist customers deploying modular servers in their environment. Each subsystem has been reviewed and adjusted to optimize efficiencies, minimizing the impacts to existing management tools and processes, and providing future growth opportunities to standards based management.

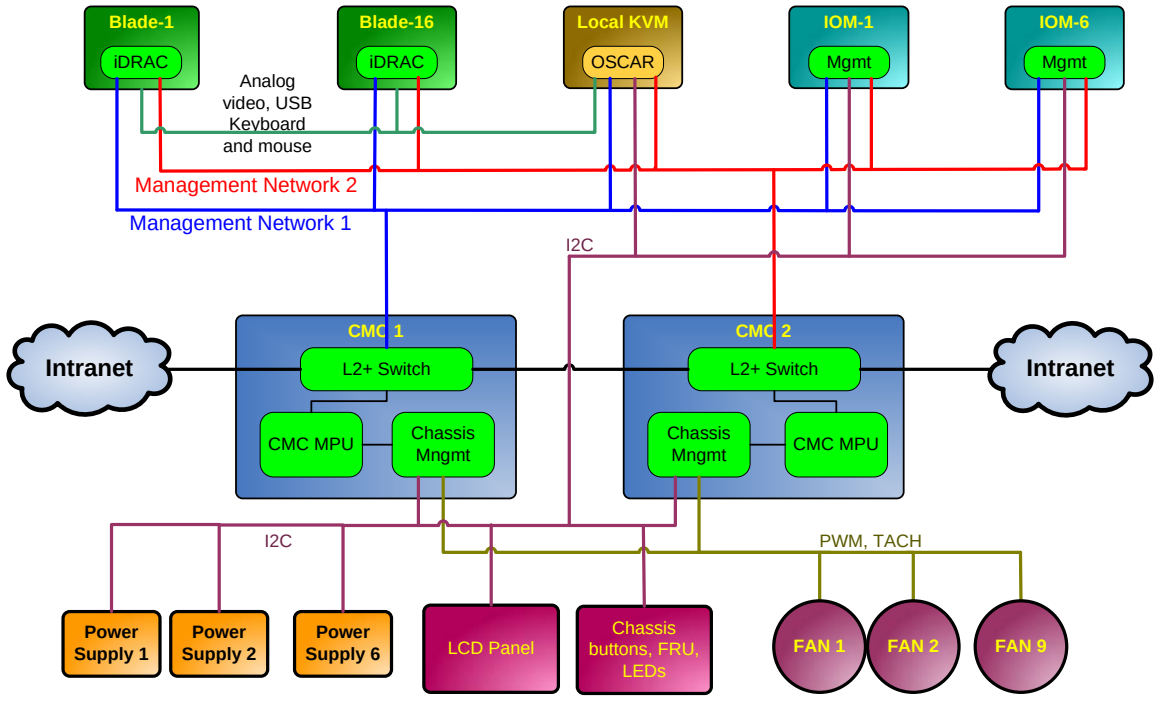


Figure 28 System Management Architecture Simplified Block Diagram

Figure 28 illustrates the management connections which transfer health and control traffic throughout chassis. The system management fabric is architected for 100BaseT Ethernet over differential pairs routed to each module. There are two 100BaseT interfaces between CMCs, one switched and one unswitched. All system management Ethernet is routed for 100 Mbps

signaling. Every module has a management network link to each CMC, with redundancy provided at the module level. Failure of any individual link will cause fail over to the redundant CMC.

Chassis Management Controller (CMC)

Chassis management is accessed and managed by the Chassis Management Controller or CMC. The M1000e possesses at least one CMC and supports an optional redundant module, each occupying a slot accessible through the rear of the chassis. Redundancy is provided in an Active – Standby pairing of the modules and fail over occurs when the active module has failed or degraded. The CMC interfaces through dual stacking 10/100/1000 Ethernet ports and one serial port. The CMC serial port interface provides common management of up to six I/O modules through a single connection.

Figure 29 M1000e Chassis Management Controller

Figure 30 Chassis Management Controller Front Panel

The CMC provides secure remote management access to the chassis and installed modules. It manages or facilitates the management of the following:

- Status, Inventory and Alerting for server modules, chassis infrastructure and I/O Modules
- Centralized configuration for iDRAC, I/O Modules and CMC
- SSL/SSH
- Power sequencing of modules in conjunction with the defined chassis power states
- Power budget management and allocation
- Configuration of the embedded management switch, which facilitates external access to manageable modules.
- Remote user management
- User interface entry point (web, telnet, SSH, serial)
- Monitoring and alerting for chassis environmental conditions or component health thresholds. This includes but is not limited to the following:
    - Real time power consumption
    - Power supplies
    - Fans
    - Power allocation
    - Temperature
    - CMC redundancy
    - I/O fabric consistency
- SMASH-CLP features providing chassis server module power control functions
- WSMAN
- CIM XML
- SNMP
- RACADM
- Ethernet traffic management (firewall)
- Ethernet switch management
- Private VLAN
- Private DHCP service
- Private TFTP service
- Virtual Media
- Virtual KVM interface between the IP network and the server module iDRAC vKVM
- Firmware management of CMC, IOM, iDRAC, iKVM

Integrated Dell Remote Access Controller (iDRAC)

The server module base management solution includes additional features for efficient deployment and management of servers in a modular server form factor. The base circuit, which integrates the BMC function with hardware support for Virtual KVM (vKVM) and Virtual Media (vMedia), is called the integrated Dell Remote Access Controller or iDRAC. iDRAC has two Ethernet connections, one for each CMC, providing system management interface redundancy.

Dell modular servers now include vKVM as a standard feature, routing the operator's keyboard output, mouse output and video between the target server module and a console located on the system management IP network. With up to two simultaneous vKVM sessions per blade, remote management now satisfies virtually any usage model. vMedia is also now standard, providing emulation of USB DVD-R/W, USB CD-R/W, USB Flash Drive, USB ISO image and USB Floppy

over an IP interface. Connection to vKVM and vMedia is through the CMC, with encryption available on a per stream basis.

Highlights of the iDRAC solution include the following:

- Dedicated management interface for high-performance management functions
- Virtual Media
- Virtual KVM
- IPMI 2.0 Out Of Band management
- Serial over LAN redirection
- SMASH CLP
- Blade status and inventory
- Active power management
- Integration with Active Directory
- Security, Local and Active Directory

Integrated Keyboard and Mouse Controller

The modular enclosure supports one optional Integrated KVM (iKVM) module. This module occupies a single slot accessible through the rear of the chassis. The iKVM redirects local server module video, keyboard, and mouse electrical interfaces to either the iKVM local ports or the M1000e front panel ports. The iKVM allows a single dongleless VGA monitor, USB keyboard, and USB mouse to be attached and then moved between server modules as needed through an onscreen display selection menu. The iKVM also has an Analog Console Interface (ACI) compatible RJ45 port that allows the iKVM to tie the interface to a KVM appliance upstream of the iKVM via CAT5 cabling. Designed with Avocent technology, the ACI port reduces cost and complexity by giving access for sixteen servers using only one port on an external KVM Switch.

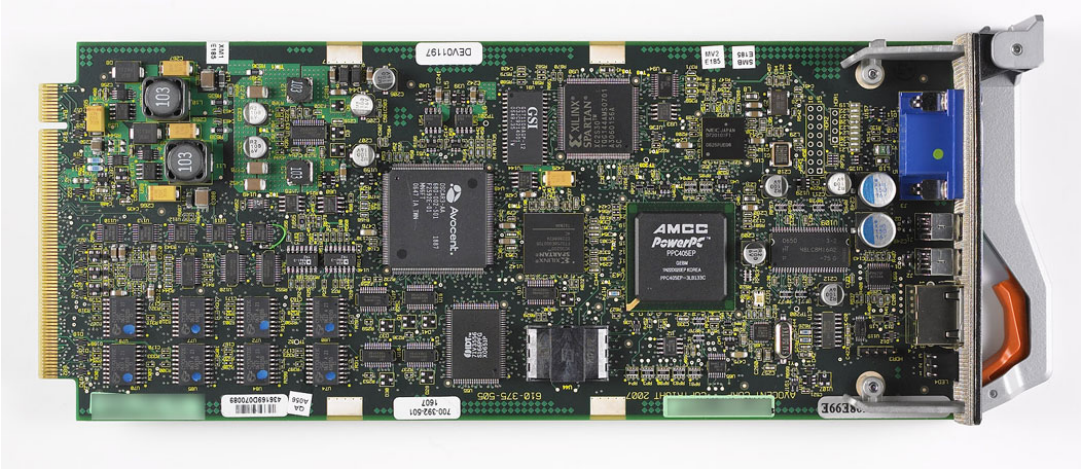


Figure 31 M1000e iKVM Module

The iKVM contains a new "seventeenth blade" feature, connecting the CMC Command Line Interface via the KVM switch and allowing text based deployment wizards on VGA monitors. iKVM firmware is updated through the CMC.

Figure 32 iKVM Front Panel

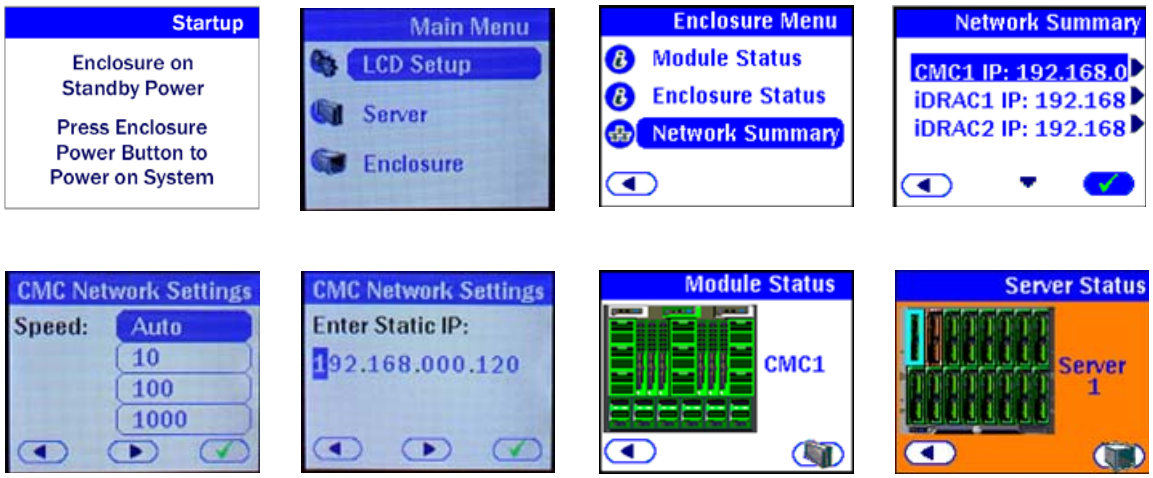
Control Panel and LCD

The control panel contains the local user interface. Functions include chassis level diagnostic LEDs, LCD display and power button. This device is not hot pluggable and is always powered, even in chassis standby mode.

Figure 33 M1000e LCD Panel Recessed Position

Figure 34 M1000e LCD Panel During Usage

The Chassis LCD shows chassis information, chassis status and faults, major server module faults and provides an interface for server module and CMC network configuration. Included is an error message when no CMC is present in the chassis. Service Tag, Asset Tag and IP Address are now easily visible on every M1000e modular system through the LCD display. The LCD panel can be retracted into the chassis body, or extended and angled once deployed for full visibility no matter where the M1000e is mounted in the rack.



*** Graphic images may differ from final product**

Figure 35 LCD Graphic Examples

Conclusion

The PowerEdge M1000e Modular Server Enclosure is a breakthrough in enterprise class server design. Dell optimized the new PowerEdge M1000e Modular Server Enclosure and M600/605 Server Modules to:

- Maximize flexibility- modular I/O, power, cooling, and management architecture.
- Maximize longevity- optimized power and cooling design supports current and future generations of server modules and I/O. I/O bandwidth to support not only today's generation of 10Gb Ethernet, 20Gbps InfiniBand and 4Gbps Fibre Channel, but up to 40Gbps QDR InfiniBand, 10Gbps Serial Ethernet, and 8Gbps Fibre Channel.
- Lower TCO- lower cost than monolithic servers with equivalent features. Best in class power and cooling efficiency.

Acknowledgments:

The author would like to thank Richard Crisp, Mike Roberts and K.C. Coxe for their contributions to this white paper.