Dell Solution for High Density GPU Infrastructure

李 信 乾 (Clayton Li)

產品技術顧問
HPC@DELL

Collaboration
- Key partnerships & programs
- Customer inputs & new ideas

Innovation
- Core & new technologies
- Critical adoption programs

Acceleration
- Reduced time to implementation
- Increased time to productivity
Solution Component

8-Node HPCC Bundle (Rack Servers)
Meeting Specific Customer Needs with the Industry’s Broadest Server Portfolio

- Customer’s Application
  - Cloud
  - HPC
  - Virtual Farms
  - Ha Clusters
  - Standalone

- Hardware
  - Resiliency and Availability

- Broad Usage Models
  - PowerEdge Core
  - PowerEdge Blades
  - PowerEdge C

- Scale
  - DCS Classic
  - Hyper Scale

- Single Machine Applications

- Design & Features
PowerEdge Server Portfolio

**SMB and Branch Office Optimized**

- **T110 II**
  - Entry

- **T310, T410**
  - 1S & 2S

- **T610, T710**
  - Scalable 2S

**Datacenter and Cloud Optimized**

- **M610, M810, M910, M915**
  - Dense 2S to Scalable 4S

- **M610x**
  - Blade+ Expansion

- **M1000e**
  - Blade Chassis

**Tower Servers**

- **T110 II**
  - Entry

- **T310, T410**
  - 1S & 2S

- **T610, T710**
  - Scalable 2S

**Blade Servers**

- **R610**
  - Dense 2S/1U

- **R510, R515**
  - SMB & Storage Optimized

- **R710, R715, R810, R815**
  - Scalable 2S & 4S /2U

- **C410x**
  - GPU Expansion Chassis

- **R210 II**
  - Entry 1S

- **R310, R410, R415**
  - SMB 1S & 2S (1U & 2U)

- **R910**
  - Highly Scalable 4S

- **R610**
  - Dense 2S/1U

**Rack Servers**

- **R510, R515**
  - SMB & Storage Optimized

- **R710, R715, R810, R815**
  - Scalable 2S & 4S /2U

- **R910**
  - Highly Scalable 4S

- **C410x**
  - GPU Expansion Chassis

- **C1100, C2100, C6100, C6105, C6145**
  - Cloud Optimized

**Intel**  **AMD**
PowerEdge M610x (16 GPU / Chassis)
Intel Based Blade Server with PCIe Expansion

Overview
- 2-socket blade server targeted at high-performance database acceleration, graphics, GPGPU, and other usage models that require PCIe expansion cards to deliver specific capabilities
- The M610x leverages all of the performance and capacities of the M610 half-height blade coupled to a uniquely capable PCIe expansion module in a full-height blade form factor.

Benefits
- Up to 12 DIMM slots delivering up to 192GB of total RAM, a critical component for database performance
- DDR3 memory offers higher bandwidth and low power consumption, with new low-voltage DIMM options for even greater power savings
- High performance Intel Xeon 5600 series processors, power efficient solution for customers that need robust blade for heavy use workloads

Performance
- Up to two Intel Quad- or 6-Core Intel 5600 series processors
- PCI Express I/O Technology
- Full Height Blade
- Buffered DDR-3 SDRAM with LVD options

Availability
- Two 2.5” hot-plug hard drives
- Memory: ECC, SDDC, Spare Bank
- Integrated RAID
- Management: iDRAC
- TPM 1.2
- Hot-plug, redundant power/cooling (chassis)

Expandability, I/O, Storage
- 2 Mezzanine Cards
- 2 x16 PCIe expansion slots
- 2 Gigabit NICs w/ TOE
- Managed Persistent Storage Options

Solution Components
- Dell/Oracle & SQL Database Program Integration
- Dell/VMware Program Integration
- Dell/EMC Integration
PowerEdge M610x
TWO x16 Gen2 PCIe Slots

• PCIe Expansion
  • Maximized Gen2 PCIe expansion is finally realized within a blade
  • PCIe expansion capabilities bring a new dimension of flexibility and performance to the Dell M-Series

• What it is
  • An expansion module containing x16 PCIe Gen2 expansion slots that can support up to two standard full-length/full-height PCIe cards.
  • Supplemental power and modular cooling cradle up to two 250 max wattage cards or one 300 max wattage card.

• Why it’s important
  • The PowerEdge M610x with PCIe expansion module allows you to creatively incorporate a vast array of PCIe-based products into blade deployments
PowerEdge C410x PCIe Expansion Chassis
Maximizing space, weight, energy and cost efficiency, with unprecedented flexibility

PCIe EXPANSION CHASSIS CONNECTING 1-8 HOSTS TO 1-16 PCIe

Great for: HPC including universities, oil & gas, biomed research, design, simulation, mapping, visualization, rendering, and gaming

- **3U chassis**, 19” wide, 143 pounds
- **PCI express modules**: 10 front, 6 rear
- **PCI form factors**: HH/HL and FH/HL
- **Up to 225W per module**
- **PCIe inputs**: 8PCIe x16 IPASS ports
- **PCI fan out options**: x16 to 1 slot, x16 to 2 slot, x16 to 3 slot, x16 to 4 slot
- **GPUs supported**: NVIDIA M2050, M2070, M2070Q
- **Thermals**: high-efficiency 92mm fans; **N + 1 fan redundancy**
- **Management**: On-board BMC; IPMI 2.0; dedicated management port
- **Power supplies**: 4 x 1400W hot-plug, high efficiency PSUs; **N+1 power redundancy**
- **Services vary by region**: IT Consulting, Server and Storage Deployment, Rack Integration (US only), Support Services
PowerEdge C410x PCIe Module

- Serviceable PCIe module capable of supporting half-height, half-length (HH/HL) or full-height/half-length (FH/HL) cards
- Visible power indicator
Highly Serviceable

- N+1 Redundant Hot-plug Power Supplies
- Onboard BMC for system management
- N+1 Redundant Hot-plug fans
- Hot-plug PCI Express Modules

Maximize uptime for business continuity
PowerEdge C410x Configurations

**Summary**

**C6100 “1:1 Sandwich”**
- One Dell C410x (8 GPUs)
- Two C6100 (8 nodes)
- One x16 slot for each node to 1 GPU
- 7U total
- 8 GPUs total
- 8 nodes total (1 GPU per board)

**C6100 “2:1 Sandwich”**
- One Dell C410x (16 GPUs)
- Two C6100 (8 nodes)
- One x16 slot for each node to 2 GPUs
- 7U total
- 16 GPUs total
- 8 nodes total (2 GPUs per board)

**C6100 “4:1 Sandwich”**
- One Dell C410x (16 GPUs)
- One C6100 (4 nodes)
- One x16 slot for each node to 4 GPUs
- 5U total
- 16 GPUs total
- 4 nodes total (4 GPUs per board)

**Summary**

**C6100 “8:1 Sandwich”**
- Two Dell C410x (32 GPUs)
- One C6100 (4 nodes)
- One x16 slot for each node to 8 GPUs
- 8U total
- 32 GPUs total
- 4 nodes total (8 GPUs per board)

**C6145 “16:1 Sandwich”**
- Two Dell C410x (32 GPUs)
- One C6145 (2 nodes)
- Four HIC slots for each node to 16 GPUs
- 8U total
- 32 GPUs total
- 2 nodes total (16 GPUs per board)
PowerEdge C6100
Capacity, performance and flexibility in an eco-friendly package

CLOUD AND CLUSTER OPTIMIZED, SHARED INFRASTRUCTURE SERVER

Great for: Hyperscale-inspired building block for high-performance cluster computing (HPCC), Web 2.0 environments and cloud builders where performance is key

- Up to 4 server nodes in 2U
- 2S Intel 5520 chipset (<100W processors)
- Xeon 5500/5600 series CPUs
- 12 x DDR3 RDIMM
- 24 x 2.5" or 12 x 3.5" HDD
- 2 x GbE Intel 82576 Kawela
- 1 x 16 Gen II (low-profile/half-length)
- 1 x 8 Gen II (custom mezzanine slot)
- IPMI 2.0 management only
- iKVM, DCMI, PXE support
- Hot-swap server nodes & HDD
- Redundant 1100W power supplies
- 470W and 750W options
- Intel Node Manager compliant
PowerEdge C6145 Server
High performance, work-horse with I/O scalability in an efficient 2U

**HIGHEST PERFORMING, SHARED INFRASTRUCTURE SERVER IN THE WORLD**

**Great for:** Power-house building block for HPC applications, video rendering, virtualization, and EDA workloads demanding large core count, high memory density, and massive I/O expansion

- **Up to 2 server nodes in 2U**
- **4S AMD Opteron 6000 series** (Std./HE/SE) series CPUs (65-105W processors)
- **AMD 5670 chipset**
- **32 x DDR3 RDIMMs**
- **24 x 2.5” or 12 x 3.5” HDD**
- **2 x GbE Intel 82576 Kawela**
- **3 x16 GenII (low-profile, half height/half-length)**
- **Micro SD for Embedded Hypervisor**
- **1 x8 Gen II (custom mezzanine slot)**
- **Dedicated x16 HIC Card**
- **IPMI 2.0 management**
- **iKVM, DCMI, PXE support**
- **Hot-swap server nodes & HDD**
- **Redundant 1100W power supplies std.**
- **1400W option**
“We’ve had a very deep relationship with Dell for a number of years, and deployed a lot of their first-of-systems.”

John Towns, director, Persistent Infrastructure, National Center for Supercomputing Applications
THANK YOU

http://www.dellhpcosolutions.com/