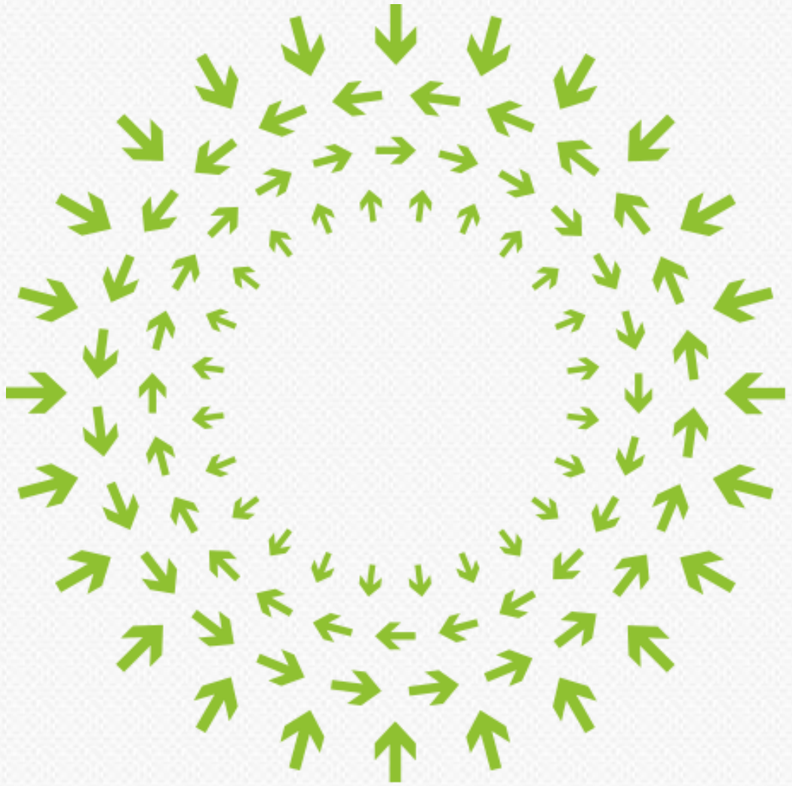


OPEN Compute Summit

March 10–11, 2015

San Jose

OPEN Compute Summit

Open CloudServer v2 specification

Chassis and blade overview

Martin Goldstein

Microsoft

Principal Systems Architect

Microsoft

Chassis

Open CloudServer OCS features

Chassis 12U, EIA 19” Standard Rack Compatibility

- Highly efficient design with shared power, cooling, and managemer
- Cable-free architecture enables simplified installation and repair
- High density: 24 blades / chassis, 96 blades / rack

Flexible Blade Support

- Compute blades – Dual socket, 4 HDD, 4 SSD
- JBOD Blade – scales from 10 to 80 HDDs, 6G or 12G SAS
  - Compatible with v1 JBOD Blade

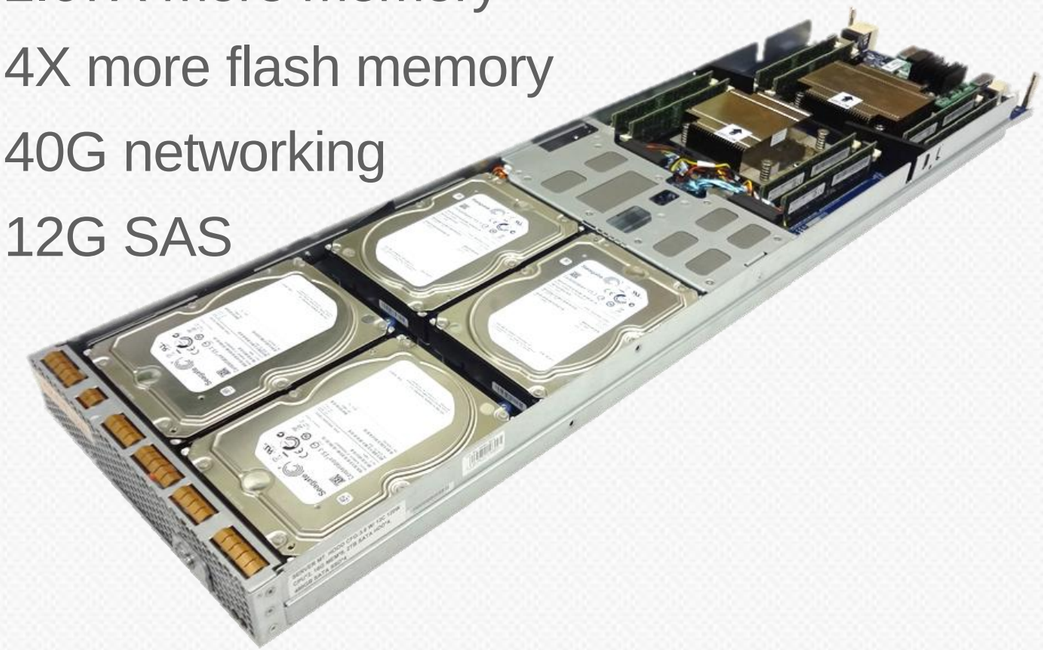
Scale-Optimized Chassis Management

- Secure REST API for out-of-band controls
- Hard-wired interfaces to OOB blade management

Open CloudServer v2 upgrade

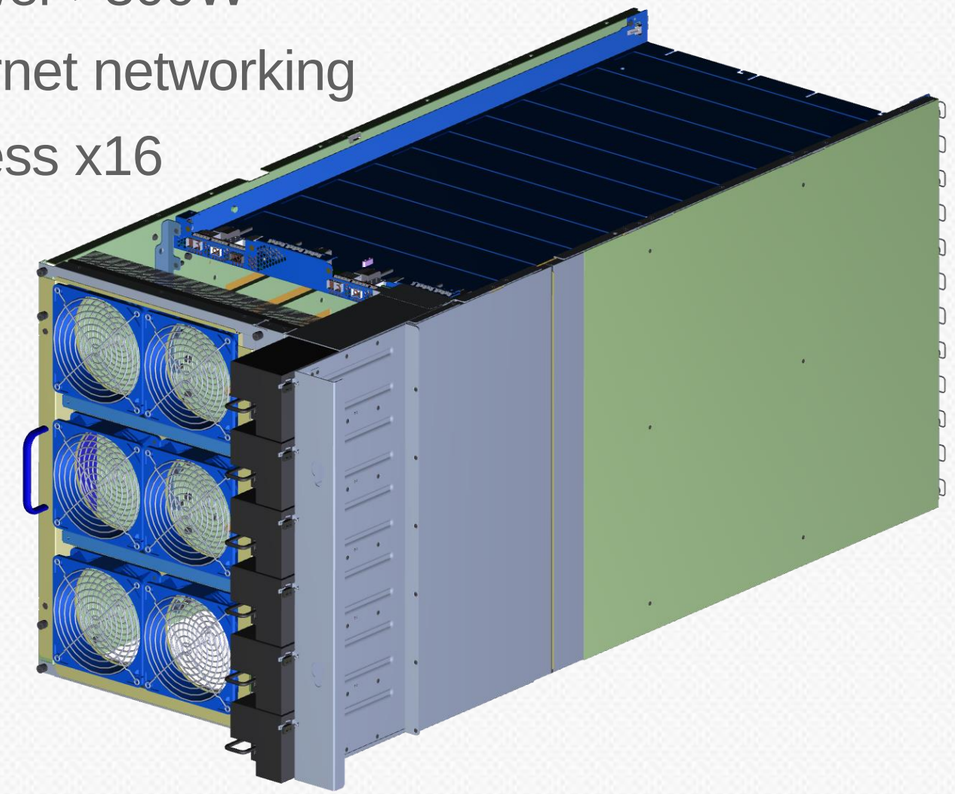
Blade upgrade

- Intel E5-2600 v3
- 36% higher performance
- 2.67X more memory
- 4X more flash memory
- 40G networking
- 12G SAS



High Performance Chassis Upgrade

- New 1600W PSU, 20 millisecond holdup
- Blade power >300W
- 40G Ethernet networking
- PCI-Express x16 expansion



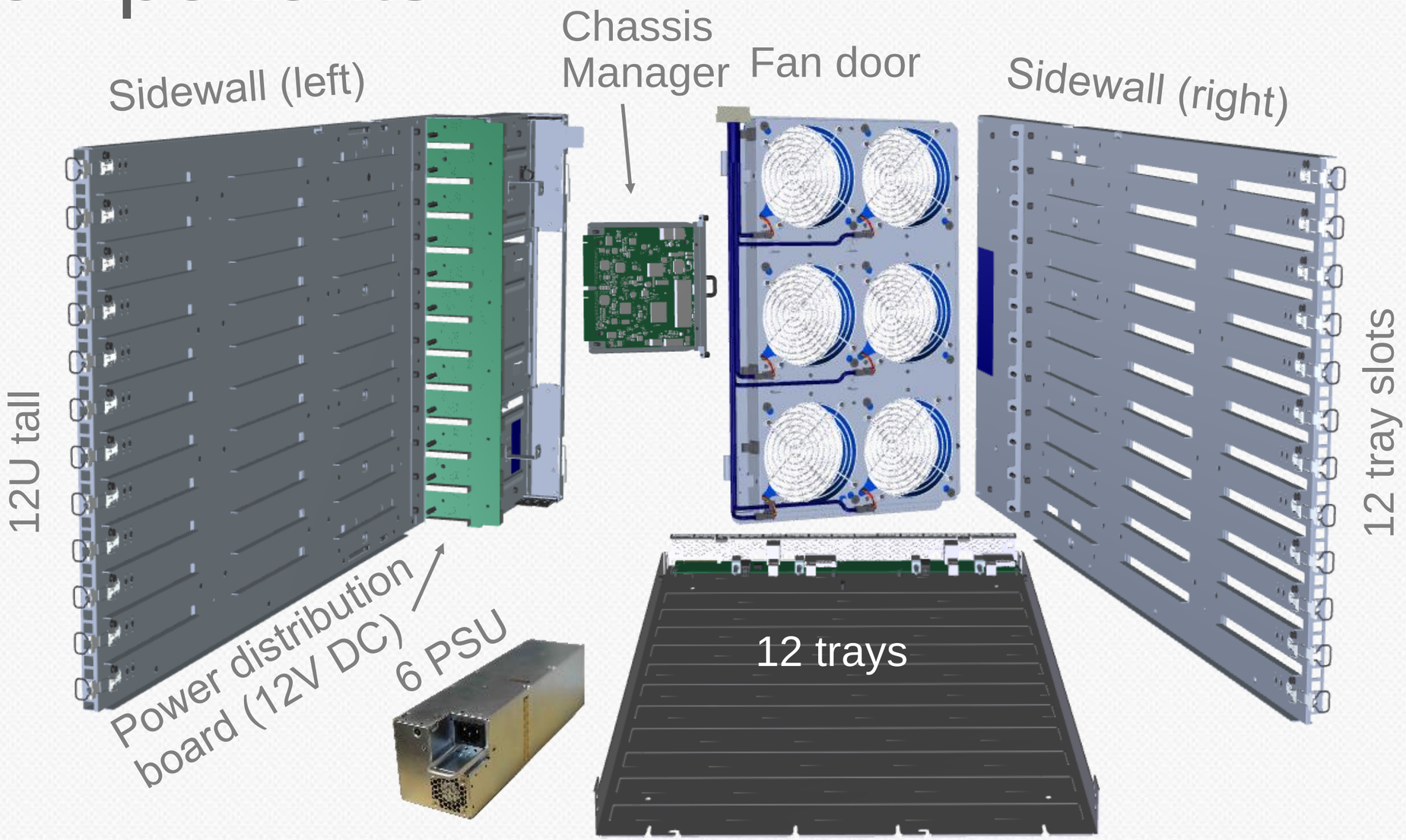
Chassis components

8 kW DC Capacity

- >300W DC blades
- Six 1600W PSU with 20 msec holdup
- Higher CFM fans

Tray upgrades

- 1 x 40Gb + 1 x10Gb
- Mezzanine: x16 Gen3 PCI-Express



Chassis v2 / v1 comparison

	OCS v2	OCS v1
Power Supplies	Six, 1600W PSU, 20 millisecond hold-up time	Six, 1400W PSU, 10 millisecond hold-up time
Blade Power	>= 300W per blade	<= 250W per blade
Fans	New fans match blade power	Match blade power
Tray I/O	PCIe x16 Expansion Mezzanine 10G or 40G Ethernet Dual SAS 12G connectors	N/A Dual 10G Ethernet Dual SAS 6G connectors
Chassis Management	X86 server built into chassis with 4GB memory, 64GB Flash Server 2012R2	X86 server built into chassis with 4GB memory, 64GB Flash Server 2012R2
Blade Support	Up to 24 compute blades JBOD blades, 12G or 6G	Up to 24 compute blades JBOD blades, 6G only

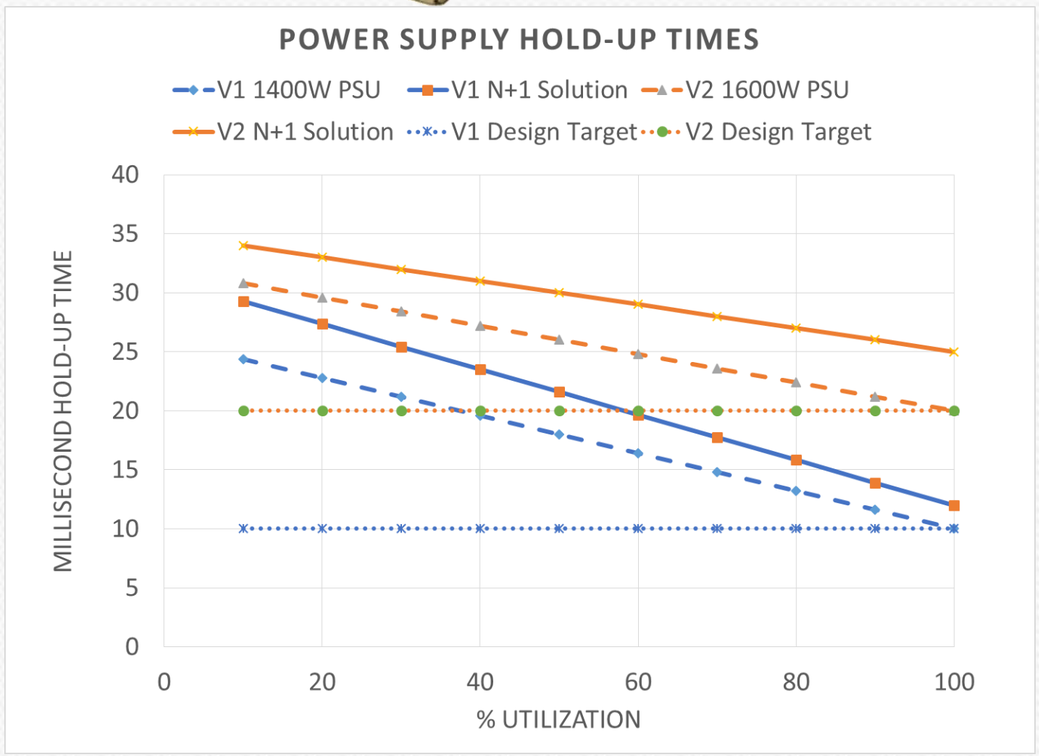
Power Supplies

1600W PSU improves total capacity

- Blade: Increases power from 250W to >300W DC
- Chassis N+1: 8,000 Watts DC
- Chassis N+N: 4,800 Watts DC

Designed for the scale-out Datacenter

- Meets 80 PLUS Platinum 94% efficiency
- Power Factor 0.99+
- Alert added for fast fault notification
- Meets ITIC requirements to enable lower cost datacenter equipment



Power Supply with embedded Battery

Distributed Battery Backup

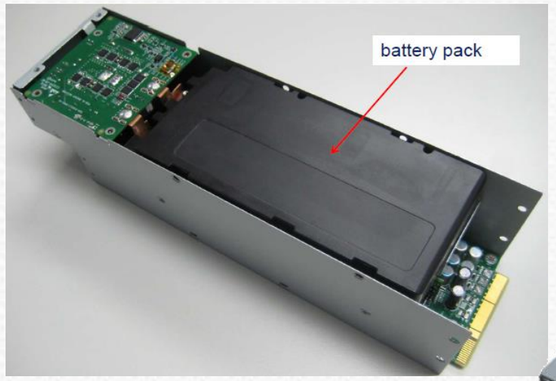
- Eliminates expensive and complexity from centralized UPS solution
- High availability solution

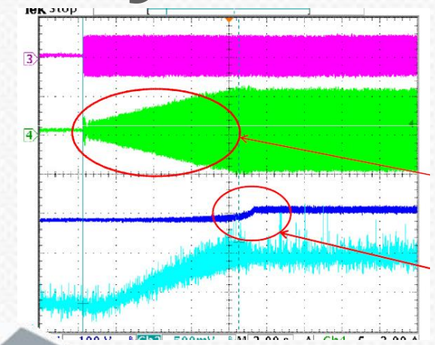
High Efficiency

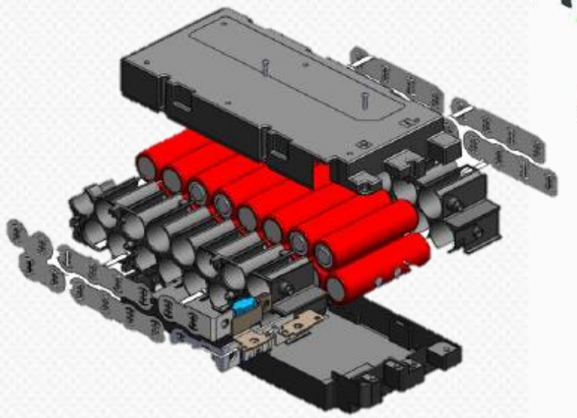
- Total power dedicated to A/C power faults reduced by a factor of three 13% → 4%
- Increases data center efficiency by 9%

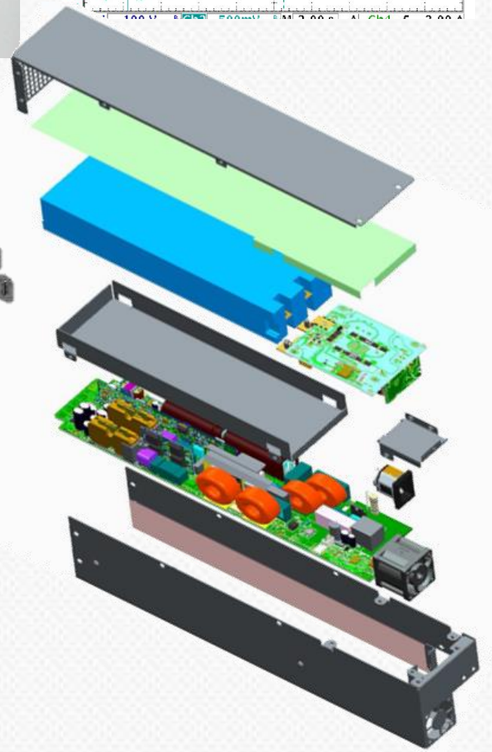
Lower Costs

- Cuts battery room, 25% of DC footprint
- Total costs are half over DC lifetime









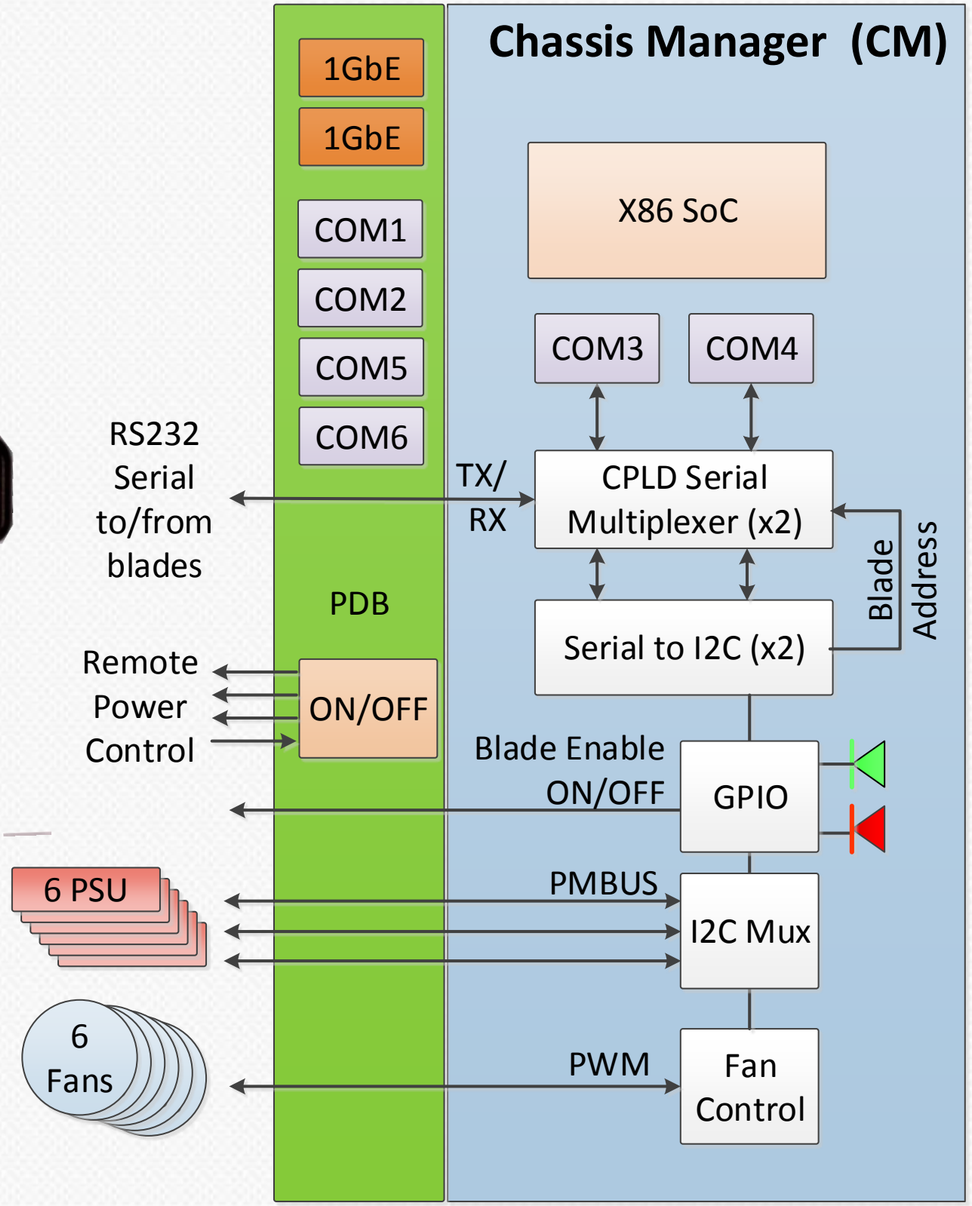
Chassis management

Secure OOB management

- Low-cost embedded x86 SoC
- REST API for machine management
- CLI interface for human operations

Hard-wired management

- On/Off to blade power cut-off circuit
- IPMI-over-serial out of band communication
- Fan and PSU control and monitoring
- Remote switch and CM power control
- Software is being open sourced
- Same hardware as OCS v1



Chassis trays

Blade support

- 12V DC power, management
- Passive PCBA for high reliability

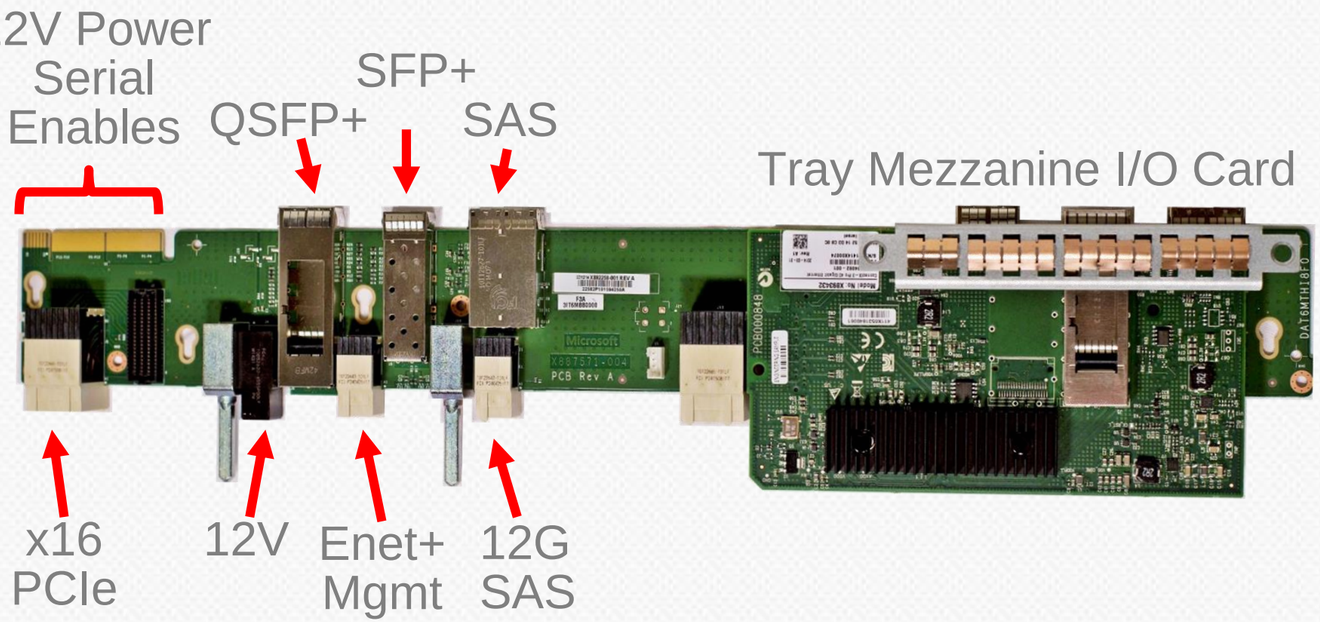
High Speed I/O

- 40G + 10G Ethernet, 12G SAS
- Tray mezzanine: x16 Gen3 PCI-Express

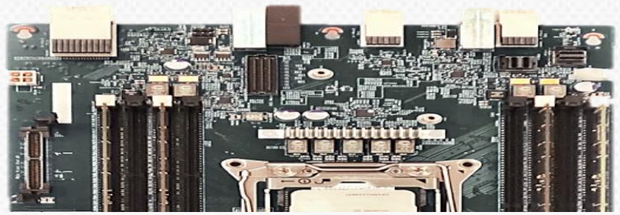
Simplified deployment and operations

- I/O cabling is pre-wired and tested
- Eliminates cabling errors during service
- Reduces need for cabling reseats

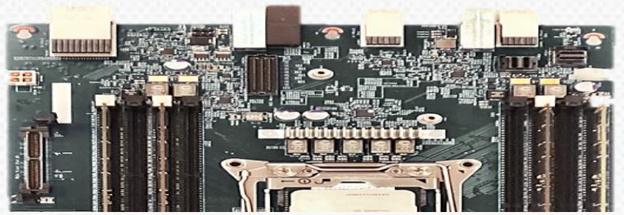
Schematics and gerbers contributed



Left Blade



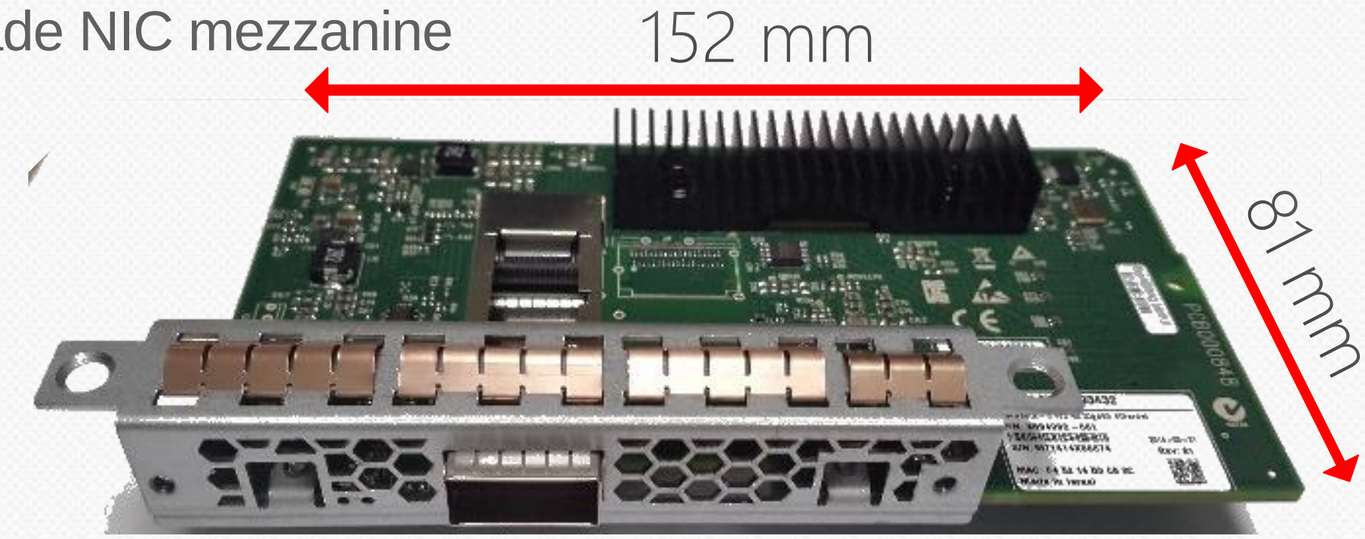
Right Blade



Tray Mezzanine I/O Card

Tray Mezzanine

- Developed for advanced networking – i.e. 25 gbps
- PCI-Express Gen3 x16, bifurcatable to 2 x8 or 4 x4
- Board area is 2.36X area of the v2 blade NIC mezzanine
- Power: 36W maximum



Prototype Single 40G NIC

Compute Blade

Blade design: guiding principles

Simplicity

- No hot plug or redundant components
- Low cost, simplified management

Flexibility

- Three IO card options (LAN, SAS, PCIe)
- LFF SATA HDD and SFF SATA SSD
- M.2 PCI-Express Flash SSD

Serviceability

- Blind-mate connectors simplify server insertion and removal
- Cable-free design minimizes cable-based NTF issues

Total Cost of Ownership

- Density optimized for IT-PAC (container) deployments
- Shared chassis infrastructure amortized across 24 servers

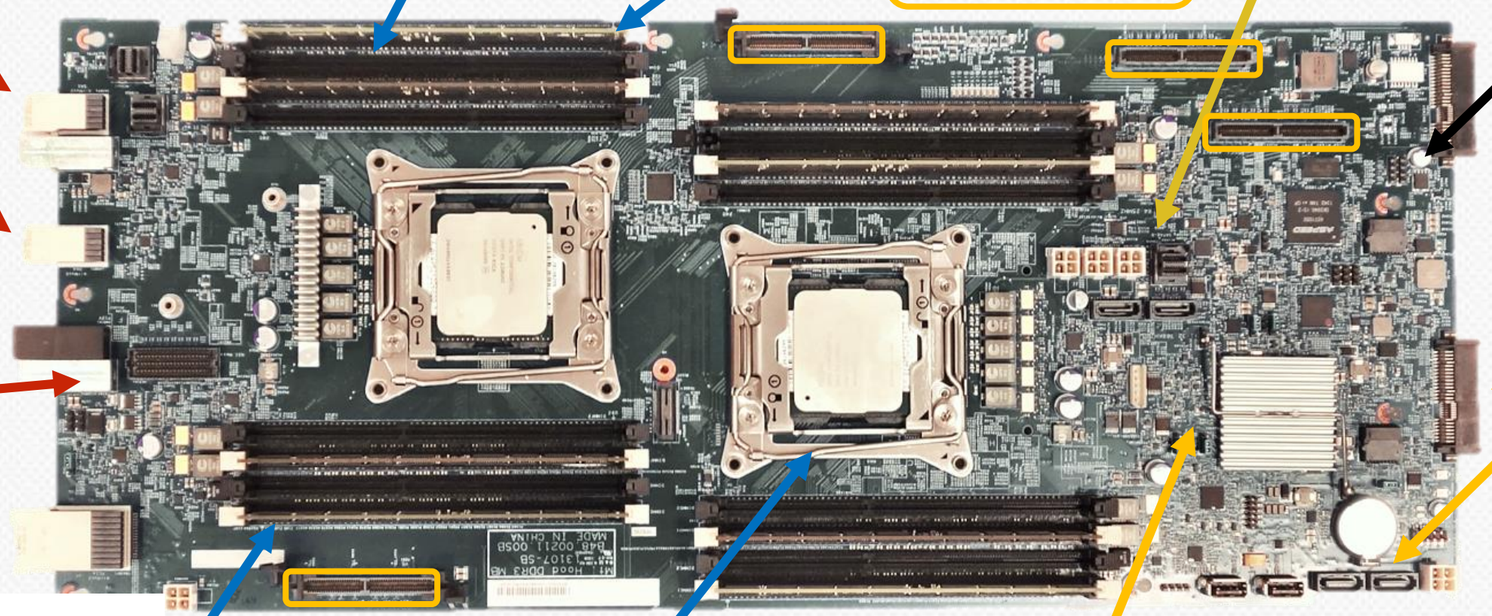
Compute Blade Upgrades (1 of 2)

	OCS v2	OCS v1
CPU	Dual Intel® Xeon® E5-2600 v3 family	Dual Intel® Xeon® E5-2400 v2 family
Core QTY	Up to 14 cores / CPU, 28 / Blade	Up to 10 cores / CPU, 20 / Blade
TDP Wattage	Up to 120W	Up to 95W
Memory Busses and DIMM Slots	8X memory bus per blade 16 DIMM slots per blade	6X memory bus per blade 12 DIMM slots per blade
DIMM Type / Speed	Up to 32GB, 2Rx4, 2133MHz, 1.25V	16GB, 2Rx4, 1333MHz, 1.35V
Capacities Supported	128GB, 192GB, 256GB, 512GB	64GB, 96GB, 128GB, 192GB
Flash devices	Four 2.5” SSD Eight 110mm M.2 PCIe NVME modules	Two 2.5” SSD

Compute Blade Upgrades (2 of 2)

	OCS v2	OCS v1
Chipset	BMC-lite serial thru Chassis Mgr	BMC-lite serial thru Chassis Mgr
Interface	REST API, CLI thru Chassis Manager	REST API, CLI thru Chassis Manager
Version, Vendor	UEFI 2.3.1, AMI	UEFI 2.3.1, AMI
Security	TPM 2.0, Secure Boot	TPM 1.2, Secure Boot
Blade I/O		
SATA	10 ports @ 6.0 Gbps	4 ports @ 3.0 Gb/s 2 ports @ 6.0 Gb/s
PCI-Express Slots	One Gen3 X8 Riser Internal One Gen3 x16 via tray mezzanine	One Gen3 X16 Riser
Networking	Single 10G or 40G Mezzanine	Single or Dual 10G Mezzanine
SAS	Dual 4X SAS @ 12G ports	Dual 4X SAS @ 6G Mezzanine

Compute blade highlights

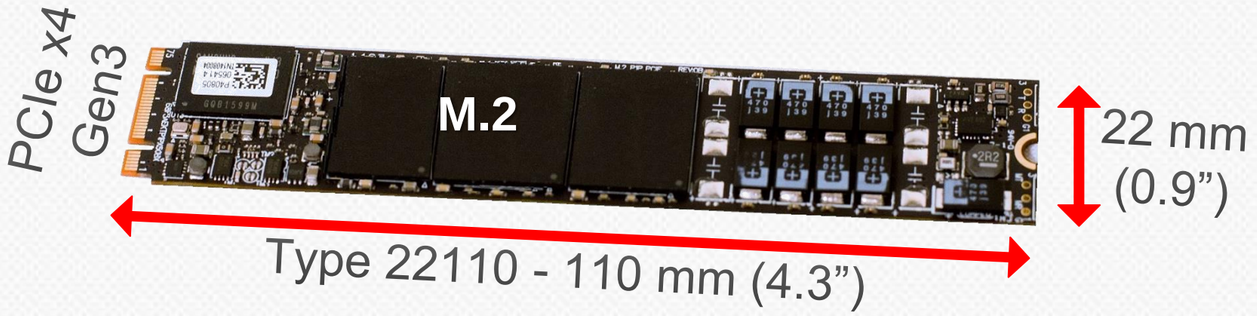


[reference designs shown]

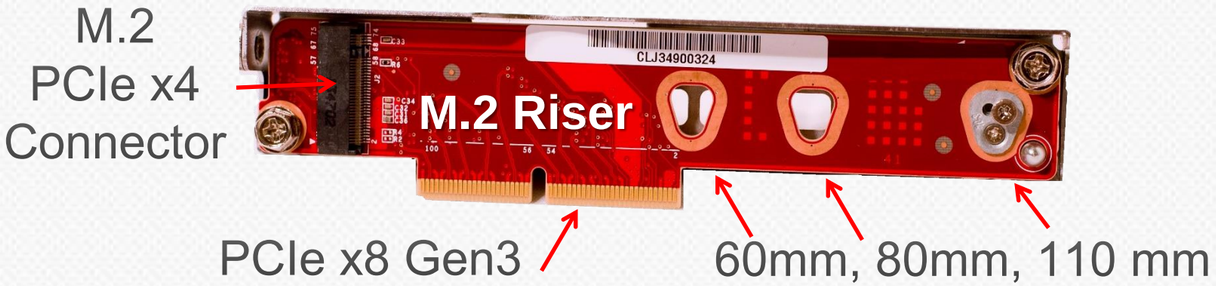
OCS M.2 CloudSSD Optimized Flash

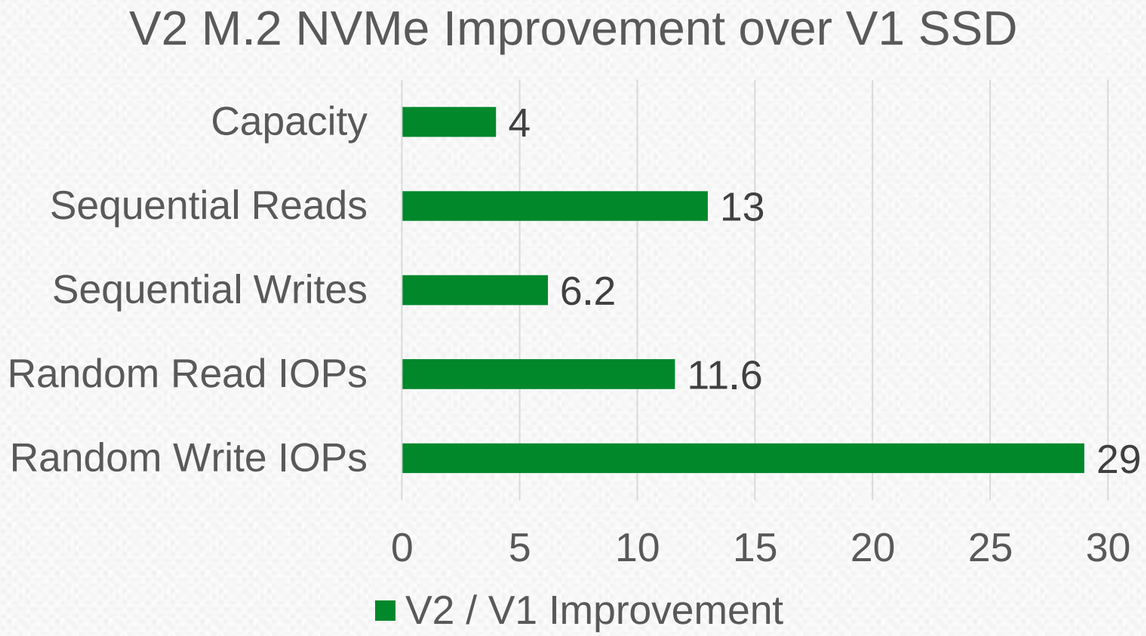
M.2 Flash Drives

- Four risers supporting eight M.2 modules
- PCI-Express Gen3 x4 NVMe & AHCI
- Multiple lengths: 60mm, 80mm, 110mm
- Vertical provides better thermal than SSD
- Low and high endurance capable



M.2 NVMe Emerging Industry Standard





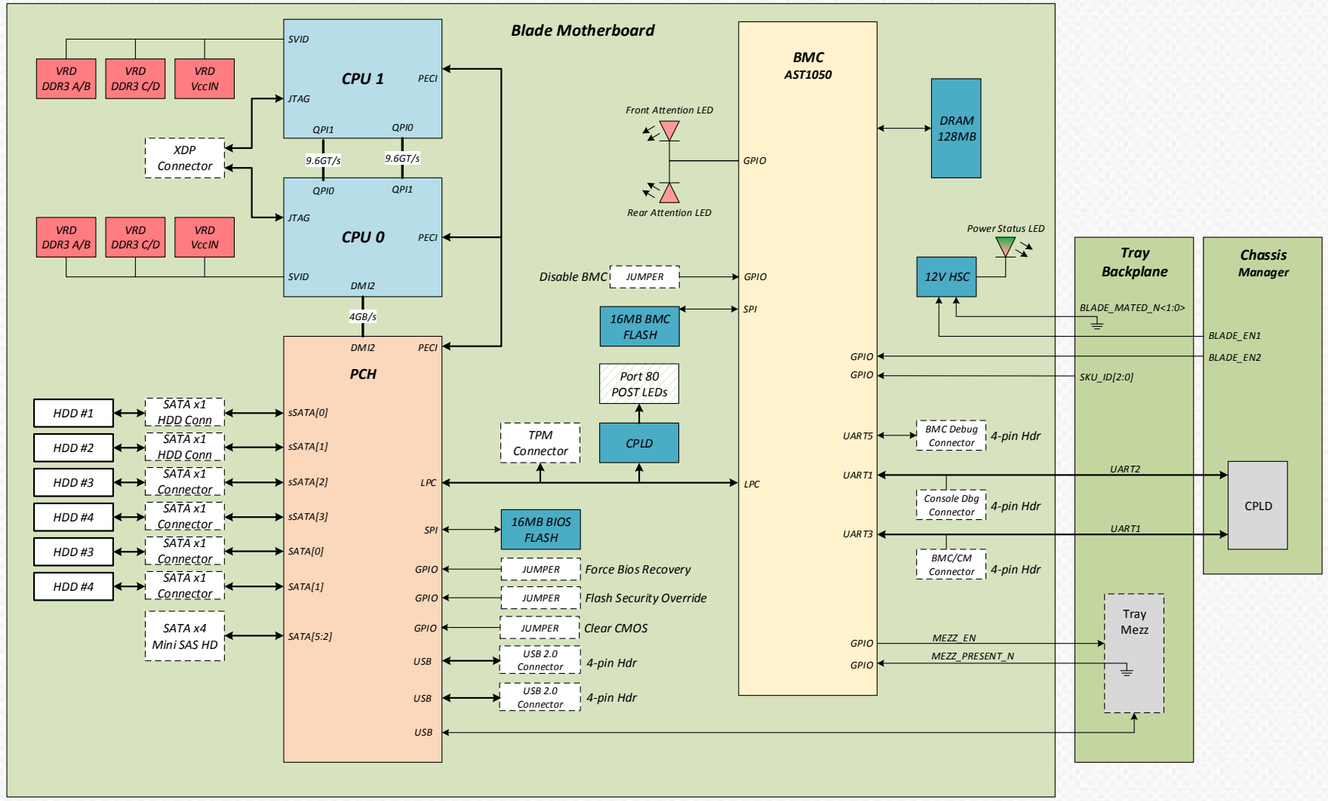
Compute blade Management

Simplicity leverages industry standards

- IPMI basic mode over Serial
- UART I/O
- I2C Master (SDR)
- System Even Log
- Power Control

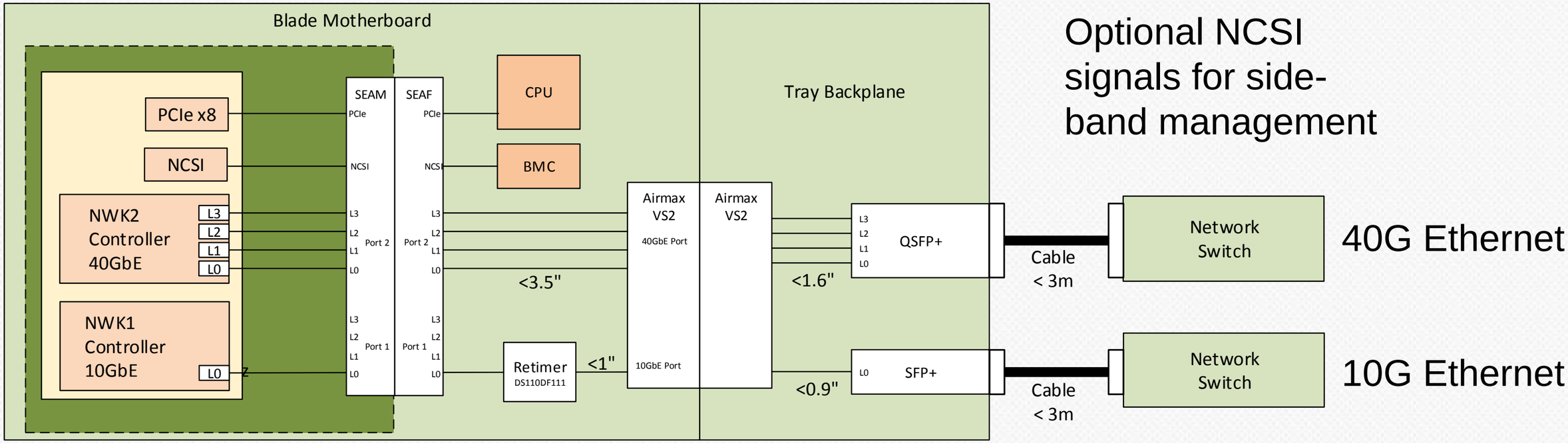
Prescriptive specifications

- Consistency between vendors
- High quality, debugging is cumulative



Compute blade Networking

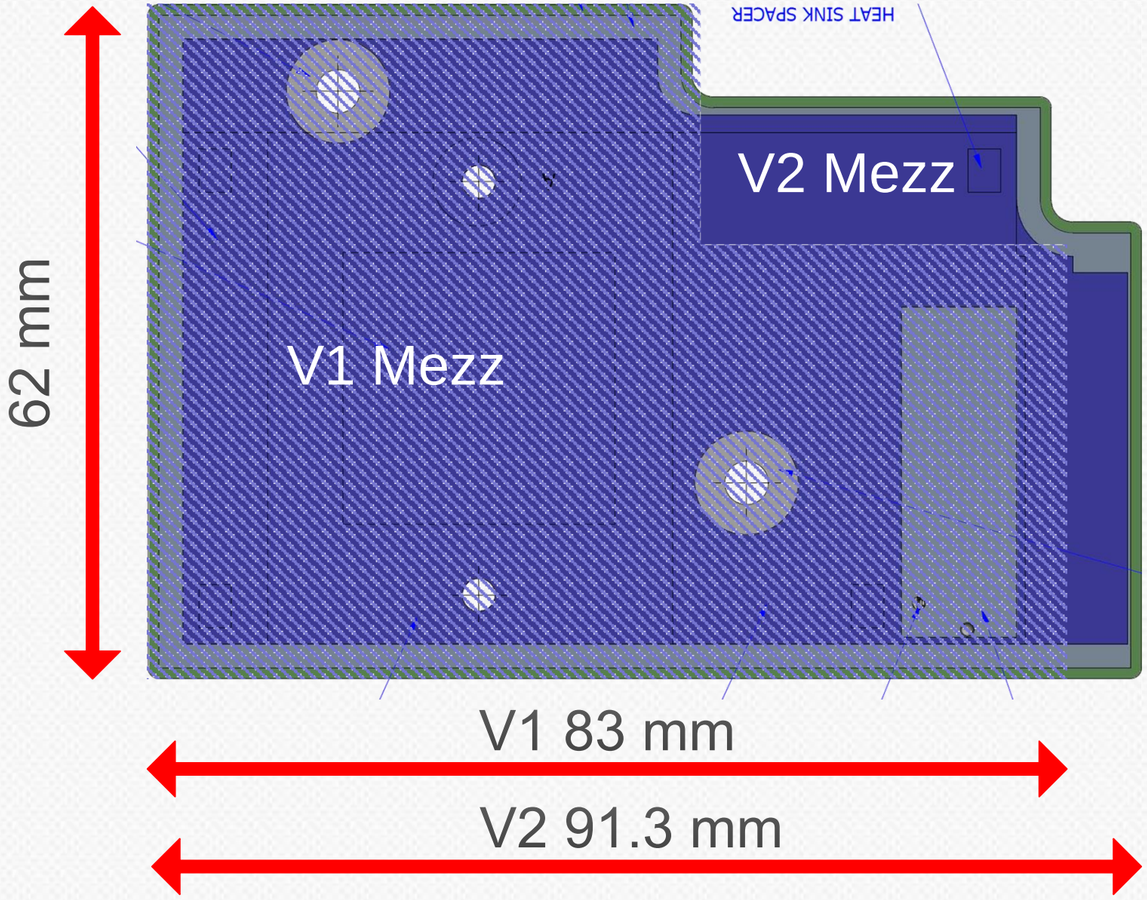
Flexible options to transition from 10G to 40G



Blade NIC Mezzanine

Single 10G or Single 40G Ethernet

- Compatible with v1 Network Mezzanine
- 18% more board area than v1 NIC
- NCSI side-band optional signals added
- Requested by OCP partners
- V1 Network pin-out defined dual 40G
- Dual 10G only cards built for v1
- One of the 40G converted to 10G freeing six diff pairs to feed NCSI

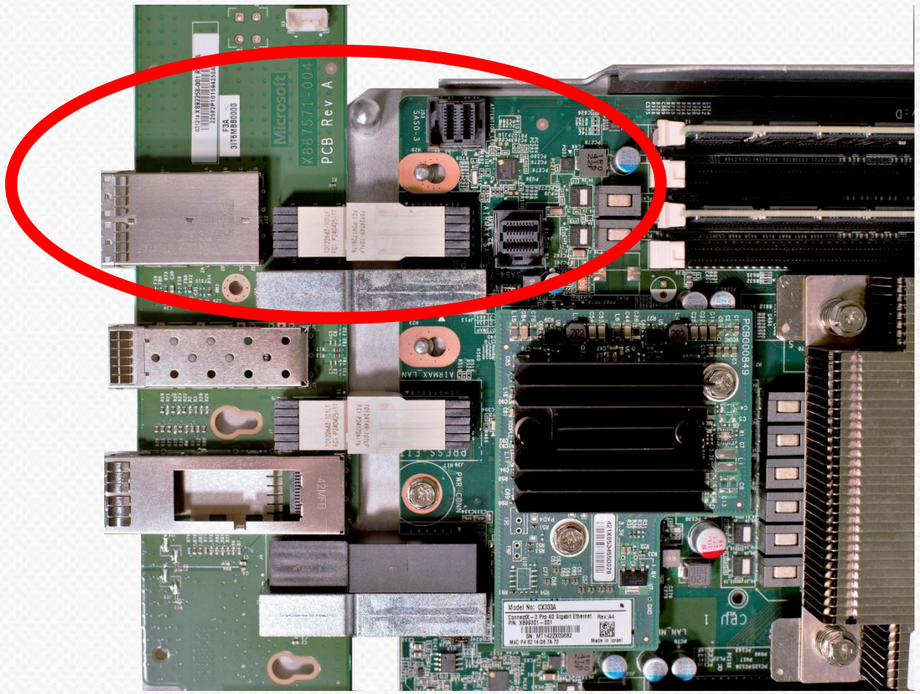


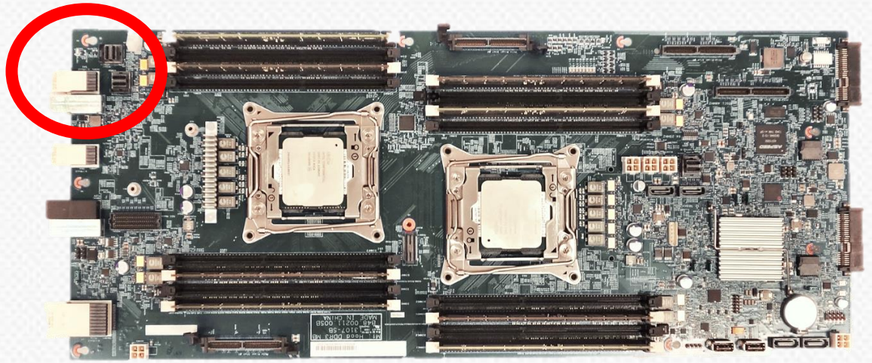
[reference designs shown]

SAS Expansion

Cabled SAS Dual 4X 12G

- Replaces v1 SAS Mezzanine card
- Cables to standard HBA or RAID adapter
- Compatible with v1 JBOD





[reference designs shown]

Safety and compliance

Ready for data centers world-wide

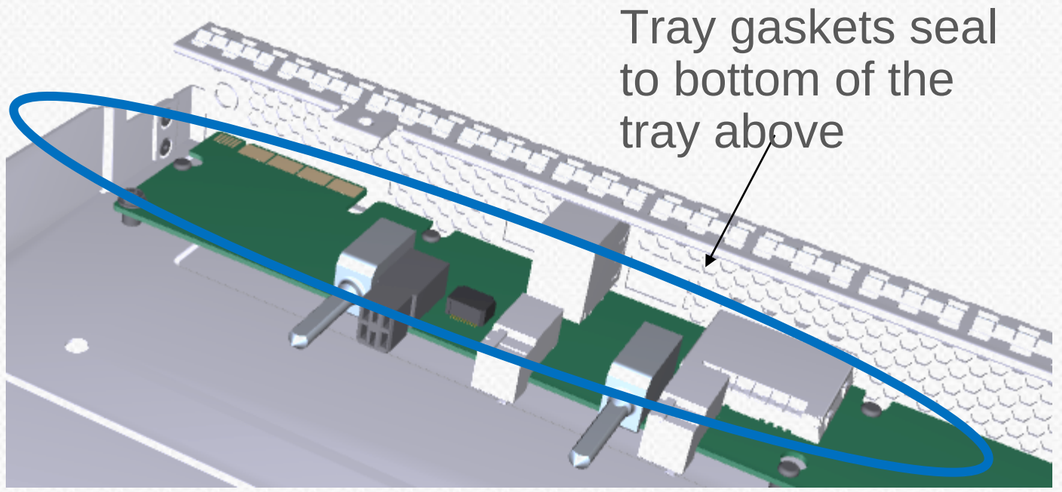
- Microsoft requires full compliance
- Containment at blade and tray
- Chassis is contained for use in EIA racks

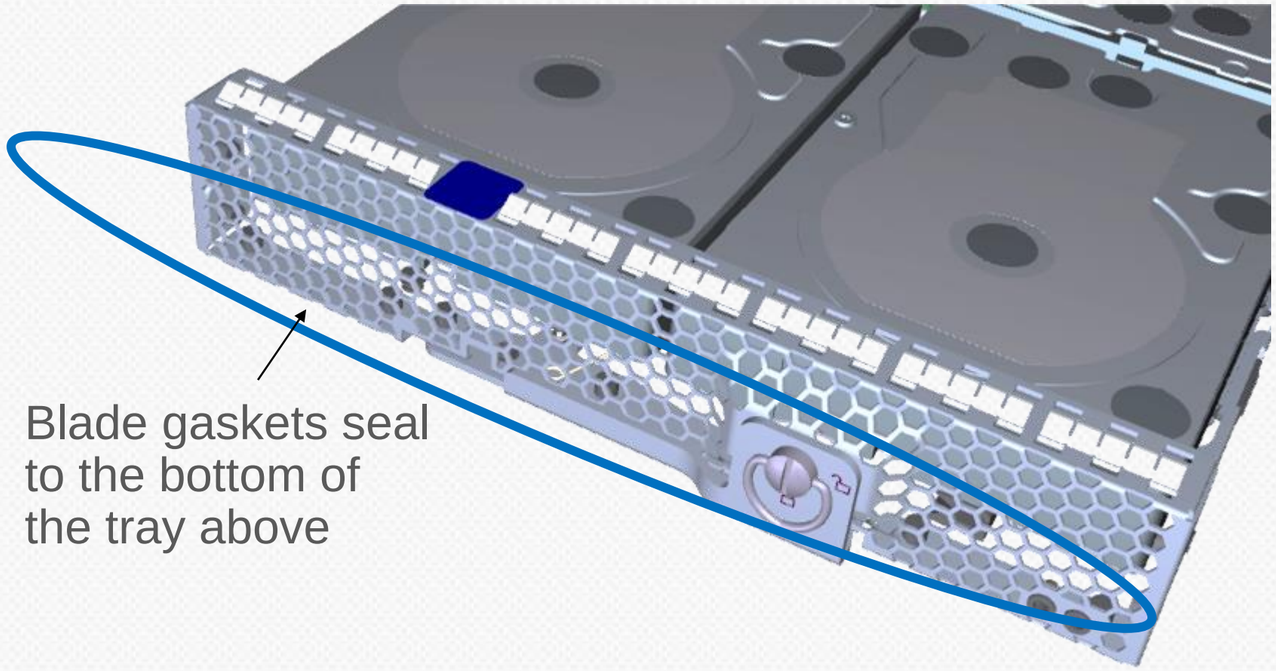
Safety is Microsoft top priority

- UL, IEC, CSA standards among others

EMI Compliance is important

- CISPR, ANSI, IEC standards to start with

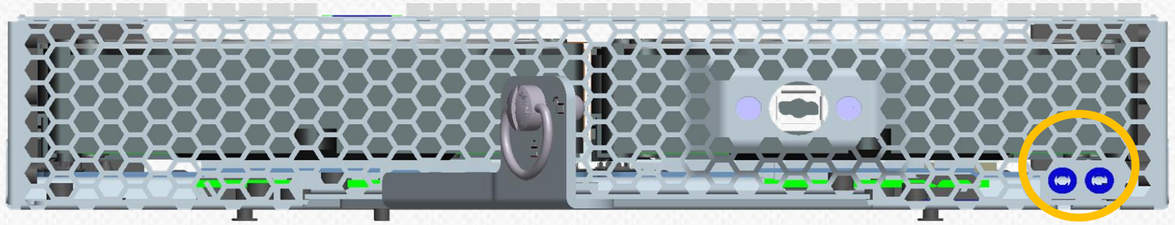


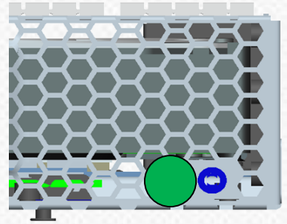


Additional features

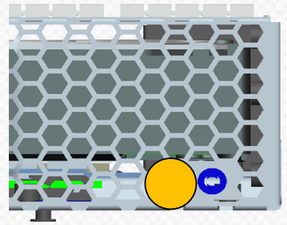
Status LEDs

- Health LED in the front
- Attention LED in the front and back
- Solid colors, no blinking lights!

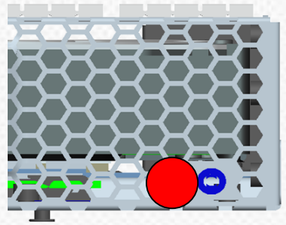




Green Health:
Blade OK



Amber Health:
Blade Fault



Red Attention:
Identify Blade

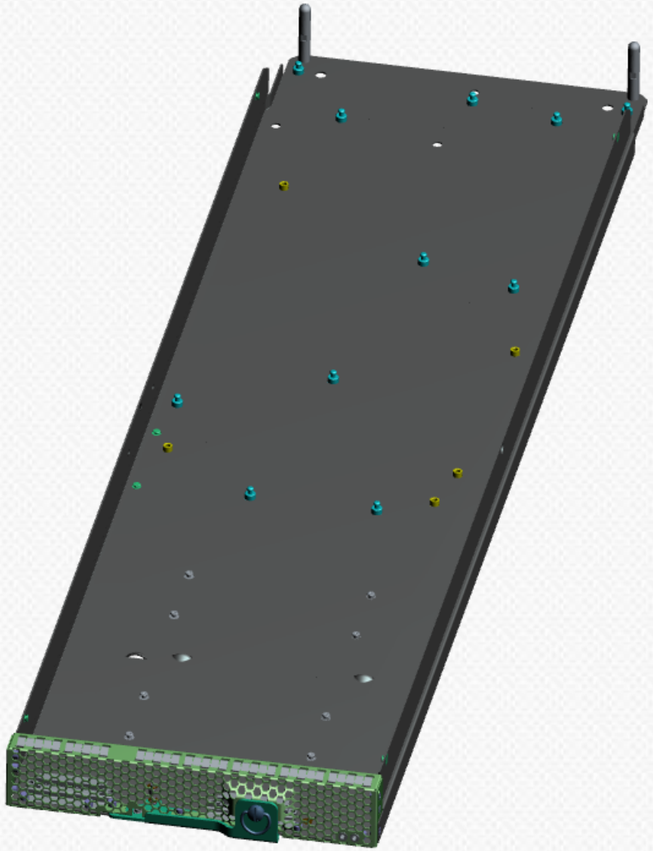
Blade Insertion and Removal

- Front access, tool-less blade extraction
- Rotate latch before engaging release lever
- Two-phase release enables in-rack shipments

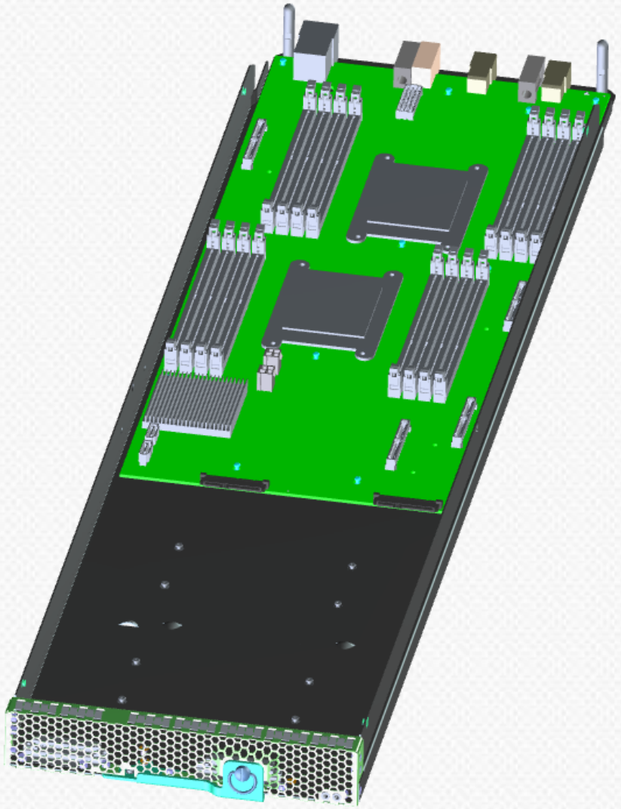
Table 16. Blade Power Status LED Description

LED status	Condition
Off	• Blade is not fully inserted, 12V power is absent, or Blade_EN is de-asserted • Standby and CPU power are off
Solid Amber ON	• Blade is inserted, 12V power is available, and Blade_EN is asserted • Standby power is on, but CPU power is off
Solid Green ON	• Standby and CPU power are turned on

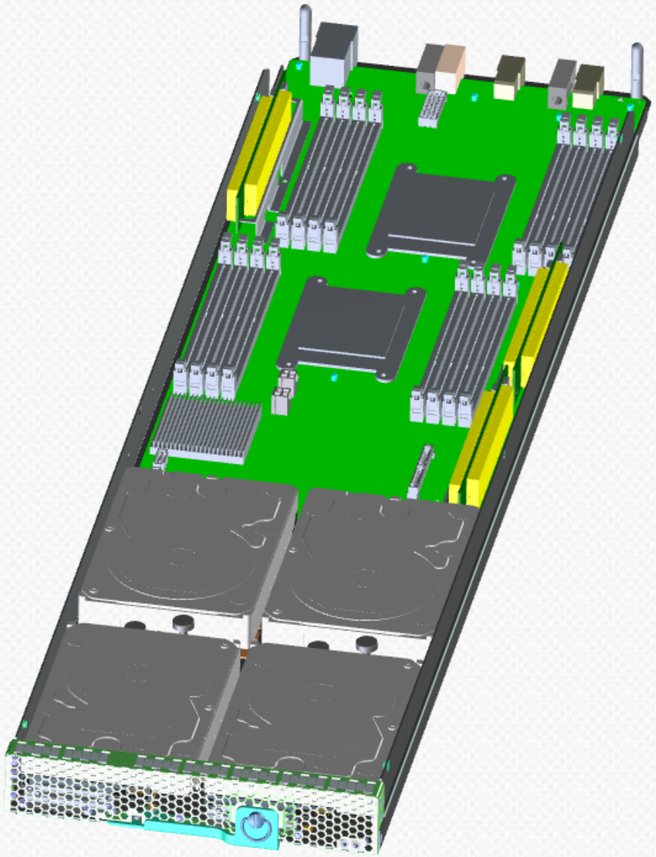
Manufacturing blade build-out



Bare metal



W/ motherboard



W/ 4x3.5” HDDs, 6xM.2

Expansion JBOD

JBOD design: guiding principles

Simplicity

- Scale in 10 HDD blocks
- Direct attach cabling
- v1 JBOD compatibility

Flexibility

- Support 14 to 84 HDD per server
- Eight SAS channels
- Cascaded topologies possible

Serviceability

- Blind-mate connectors simplify JBOD insertion and removal
- Cable-free design minimizes cable-based NTF issues

Total Cost of Ownership

- Density optimized, up to 800 HDDs / rack
- Short cables save cost and weight
- Shared chassis infrastructure is amortized across 24 blade

Expansion v1 JBOD reference design

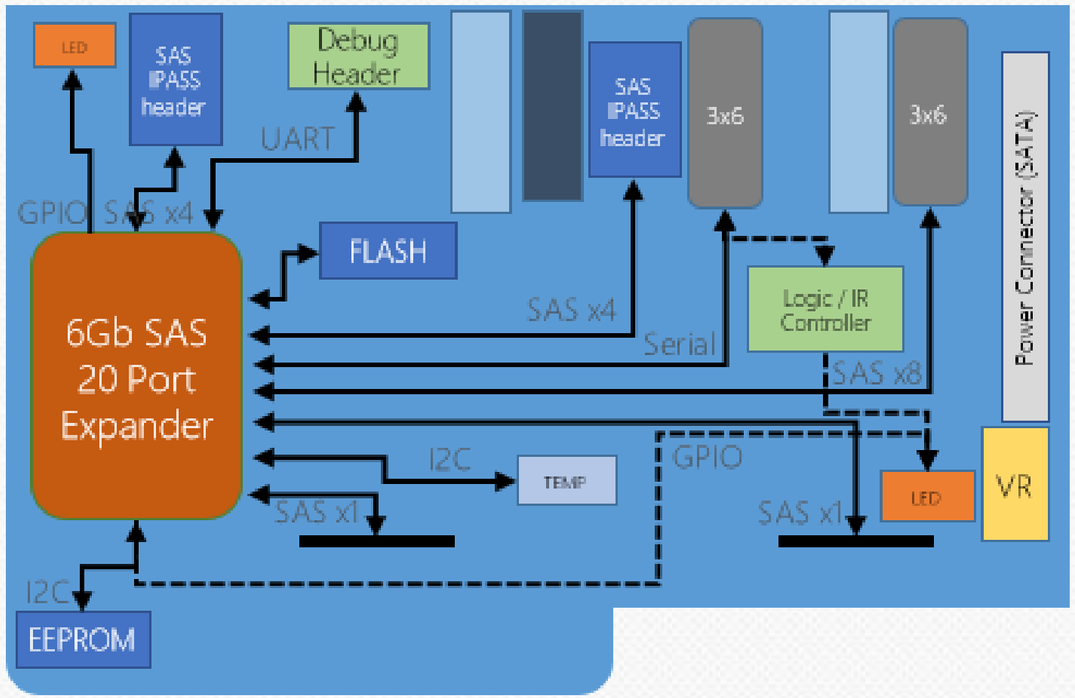
20-lane SAS expander

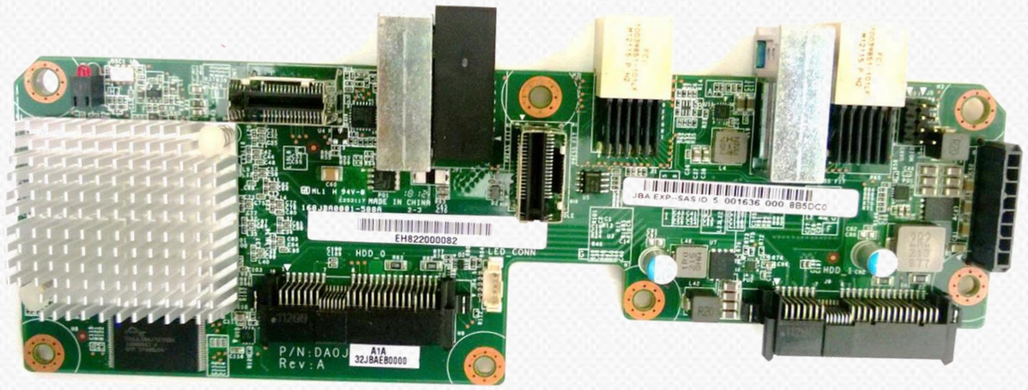
- 10 internal lanes connect to LFF SATA HDDs
- 8 external lanes connect to tray backplane

Expander connects to chassis manager via RS-232 port

- Managed with the same command set as the compute blade

v1 expander board details





Storage expander board

Storage HDD backplane

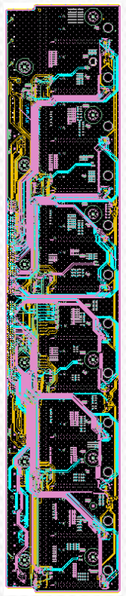
- Blind-mate to tray backplane (SAS, management)
- Direct connect to two 3.5” SATA HDDs
- Cable connect to two storage HDD backplanes

Cable-free attach simplifies drive replacement and eliminates NTFs caused by cable connection issues

[reference designs shown]

Comprehensive Contribution

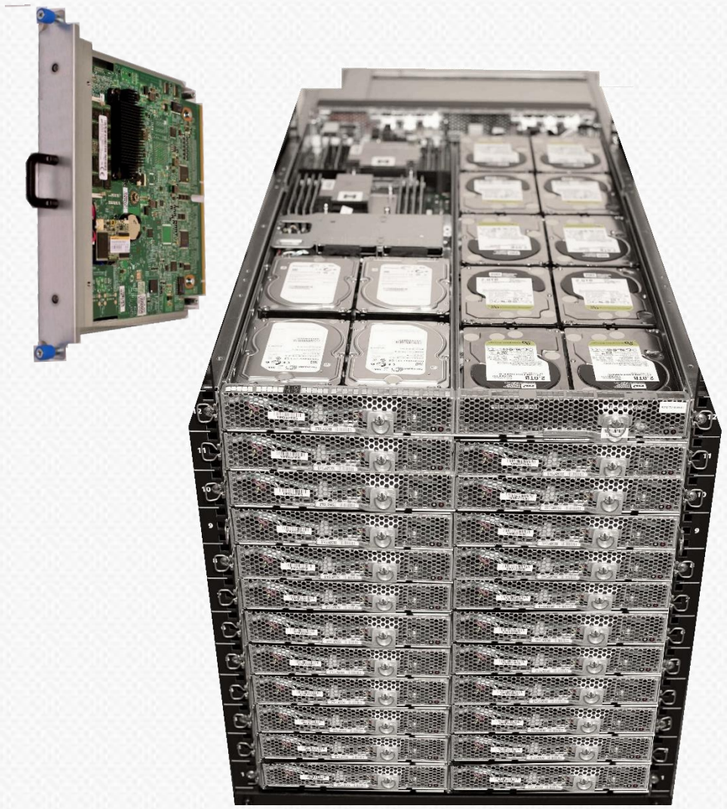
Open Source Code
Chassis management
Operations Toolkit
Interoperability Toolkit



Board Files & Gerbers
Power Distribution Backplane
Tray Backplane

Specifications
Chassis, Blade, Mezzanines
Management APIs
Certification Requirements

Mechanical CAD Models
Chassis, Blade, Mezzanines



Learn more

Visit Microsoft booth

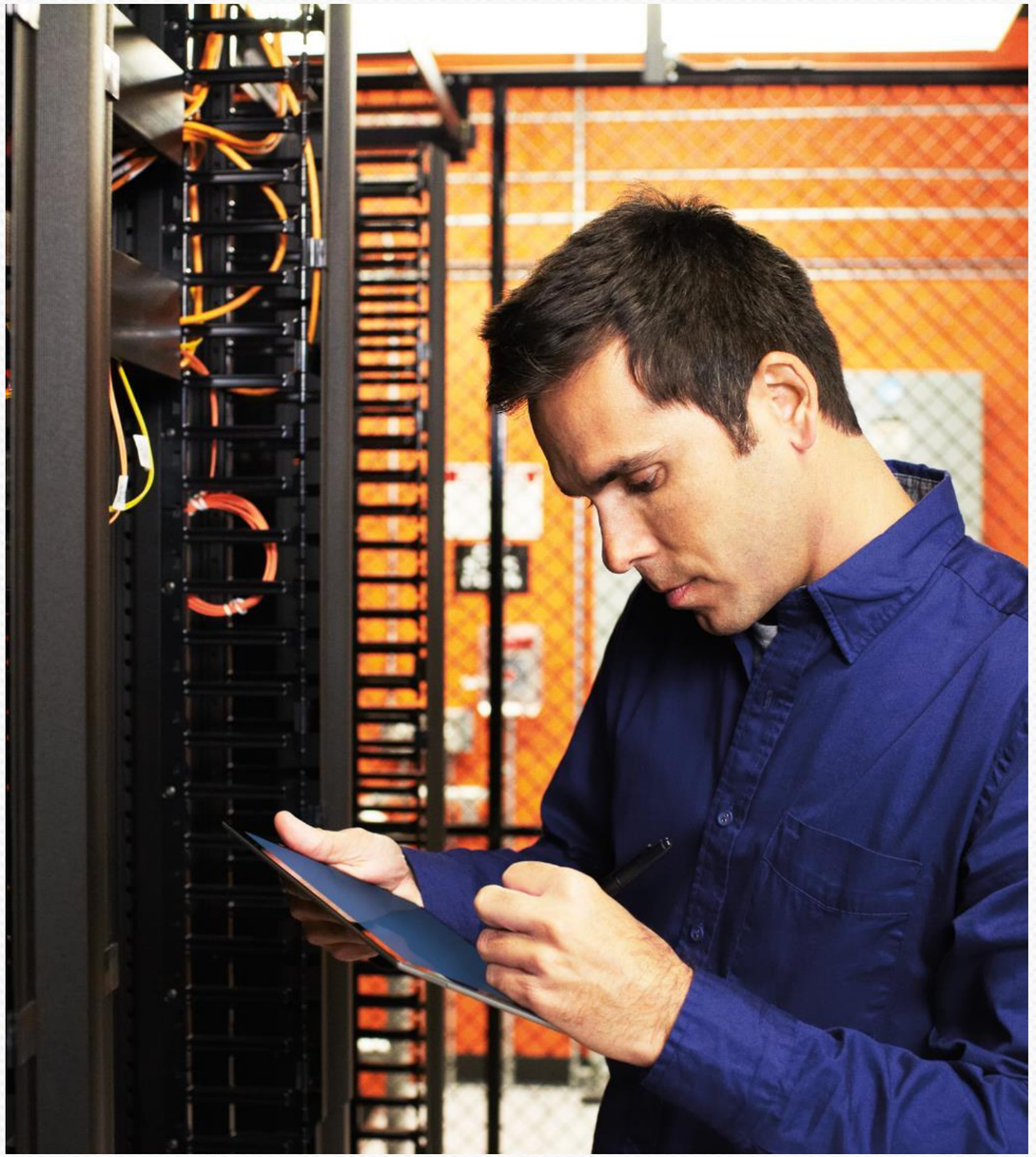
- OCS v2 Systems and Demos on display

Attend executive track session:

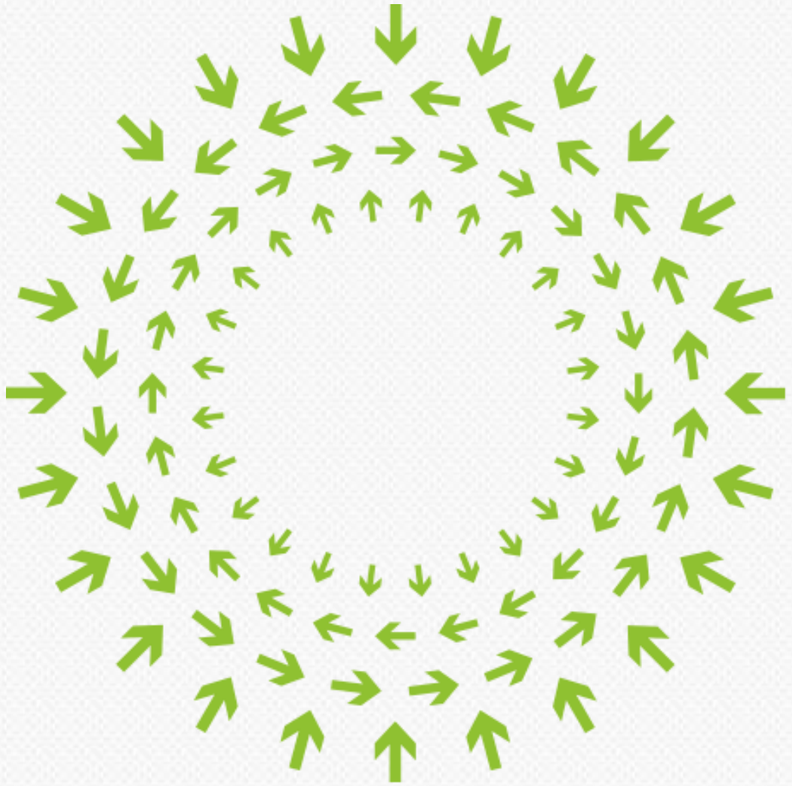
- ?

Attend technical workshops

- OCS v2 Power Supply with Battery, Tues 4:45PM
- OCS v2 CloudSSD, Tues 5:30PM



Q&A



OPEN Compute Summit

March 10–11, 2015

San Jose