

OPEN

Compute Engineering Workshop

March 9, 2015

San Jose

Monday	
1:00 PM	Microsoft Open CloudServer OCS V2 Overview
2:00 PM	Microsoft OCS 1U Quad Server System
2:45 PM	OCP Mezz 2.0 Update
3:15 PM	Mellanox ConnectX-4
3:45 PM	Break
4:00 PM	Barreleye: OCP / OpenPOWER server
4:30 PM	Characterizing Server Performance
5:00 PM	DaaP: An experiment with ExpressFabric
Tuesday	
4:00 PM	Microsoft Open CloudServer OCS V2 Chassis and Blade Hardware
4:45 PM	Microsoft OCS Local Energy Storage Power Supply
5:30 PM	Microsoft OCS Cloud M.2 SSD - Optimizing flash storage for Hyperscale
Wednesday	
1:00 PM	Facebook's 1S server Overview
1:30 PM	Facebook's Multi-Node Server Platform Overview



Open CloudServer v2 specification

Systems Overview

Mark Shaw

Microsoft

Director of Hardware Engineering



Open CloudServer OCS features

Chassis 12U, EIA 19" Standard Rack Compatibility

- Highly efficient design with shared power, cooling, and management
- 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, 8 M.2 CloudSSD
- JBOD Blade – scales from 10 to 80 HDDs, 6G or 12G SAS
 - Compatible with v1 JBOD Blade

Scale-Optimized Chassis Management

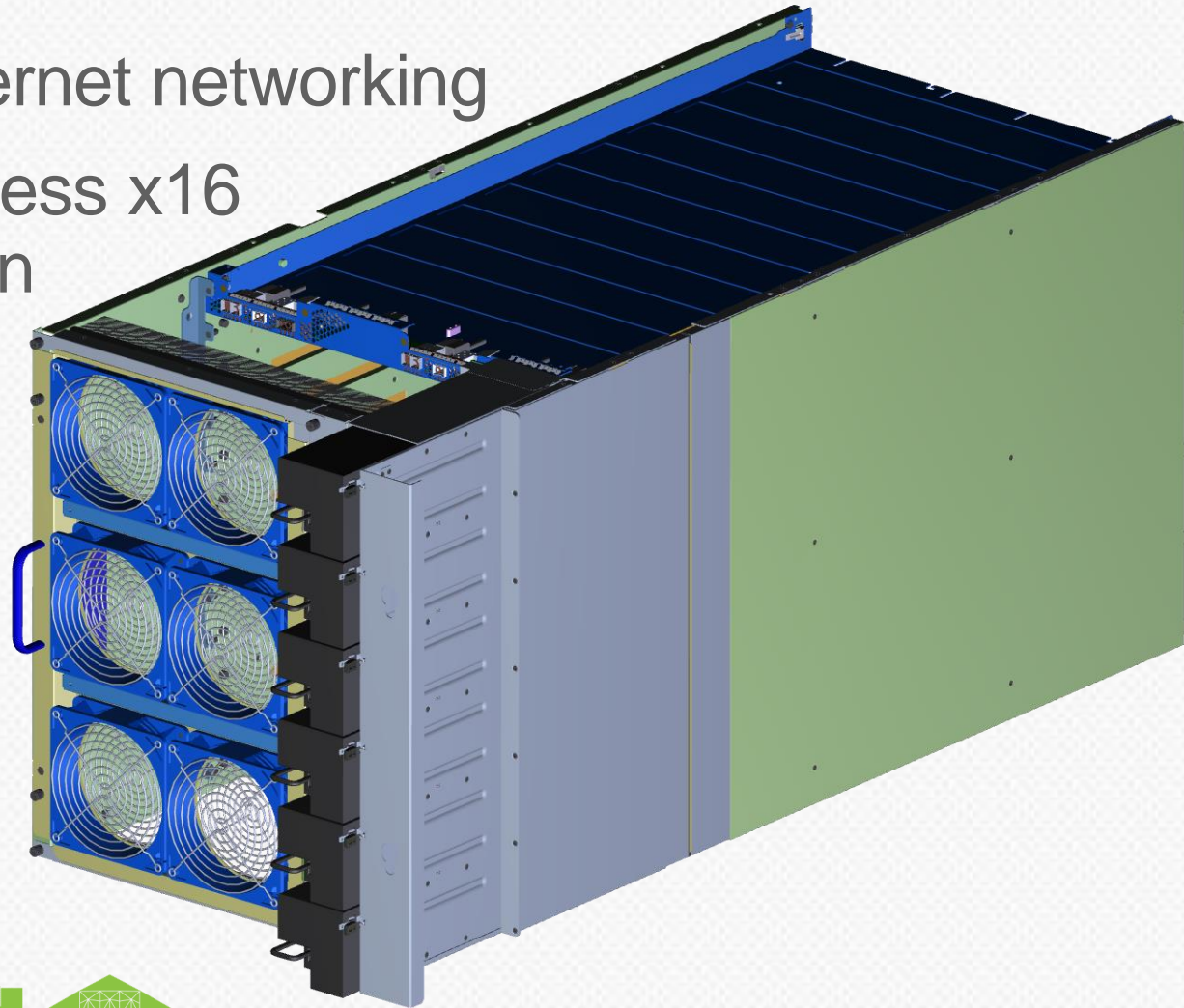
- Secure REST API for out-of-band controls
- Hard-wired interfaces to OOB blade management



Open CloudServer v2 upgrade

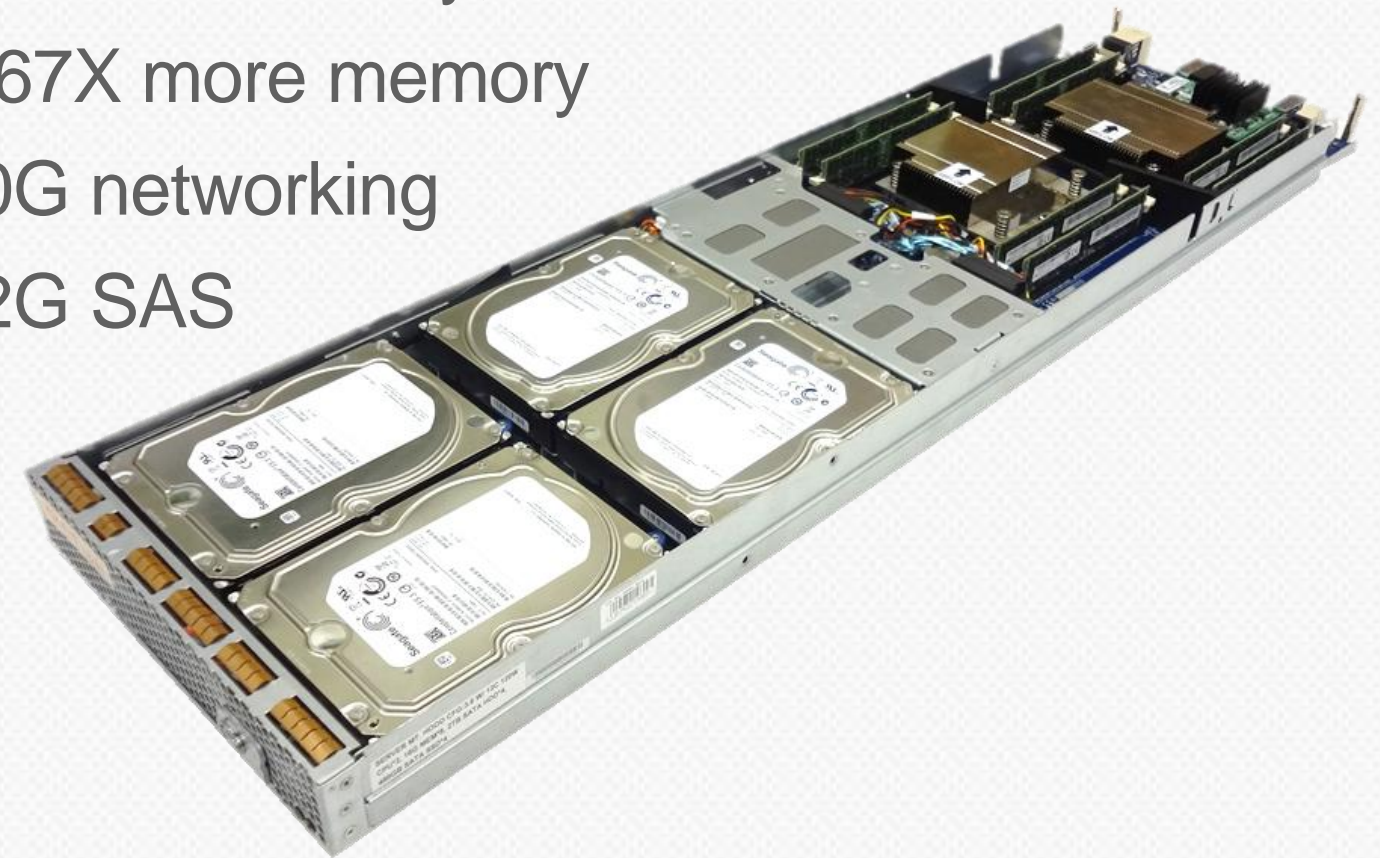
High Performance Chassis Upgrade

- Local Energy Storage 1600W PSU
- Blade power >300W
- 40G Ethernet networking
- PCI-Express x16 expansion



Blade upgrade

- Intel E5-2600 v3
- 36% higher performance
- 4X flash memory, 8 slots M.2 CloudSSD
- 2.67X more memory
- 40G networking
- 12G SAS



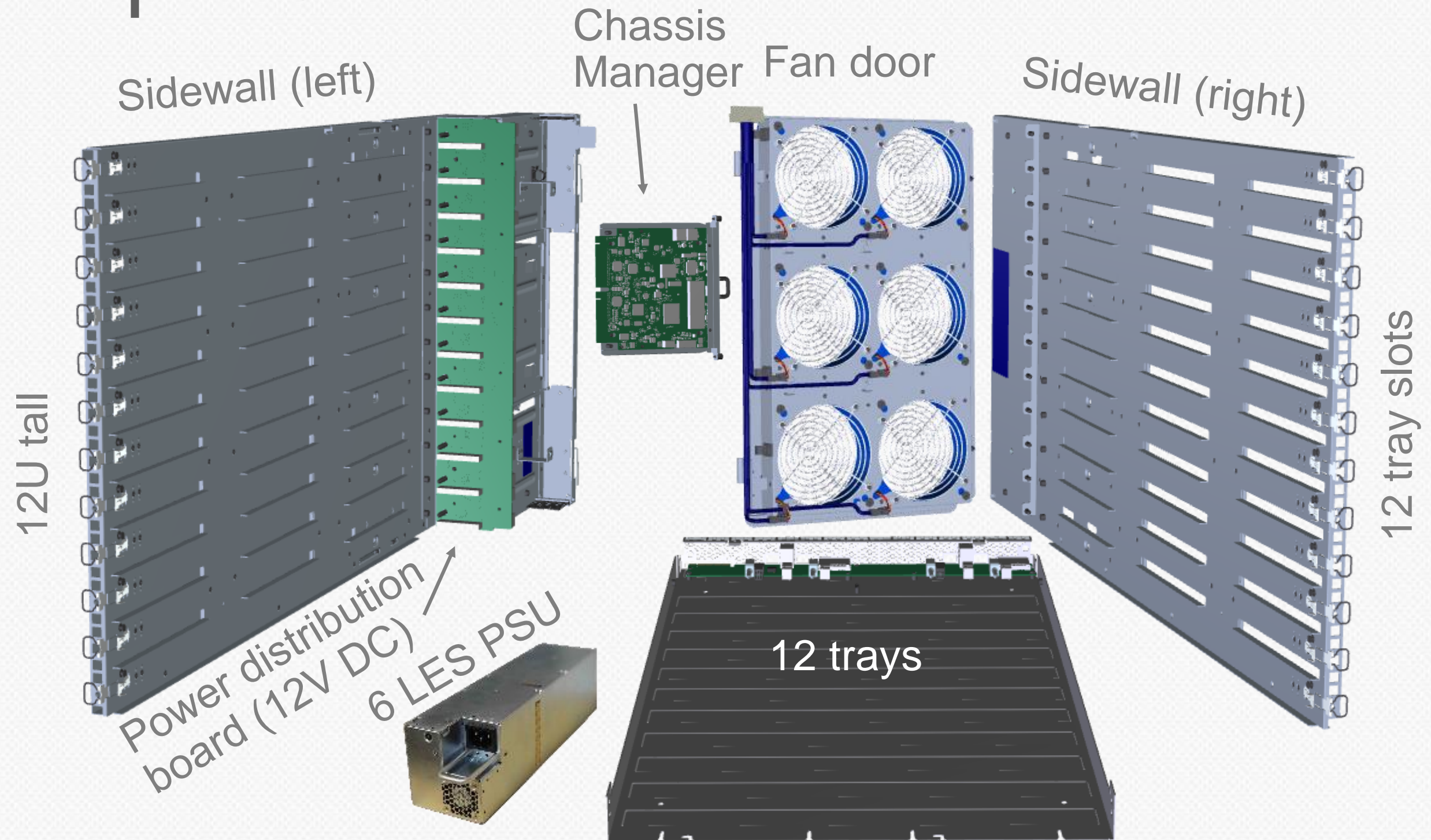
Chassis components

8 kW DC Capacity

- >300W DC blades
- Six LES 1600W PSU with embedded battery
- Higher CFM fans

Tray upgrades

- 1 x 40Gb + 1 x 10Gb
- Mezzanine: x16 Gen3 PCI-Express



Local Energy Storage Power Supply

Local Battery Backup

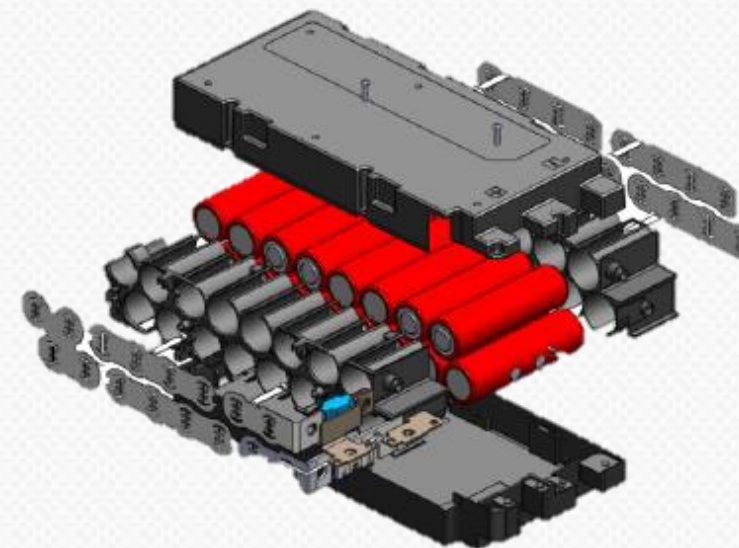
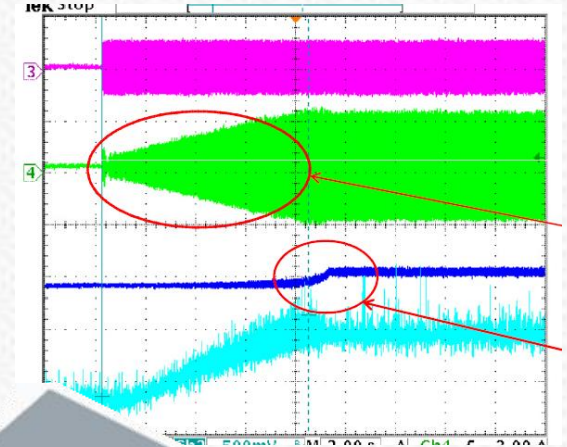
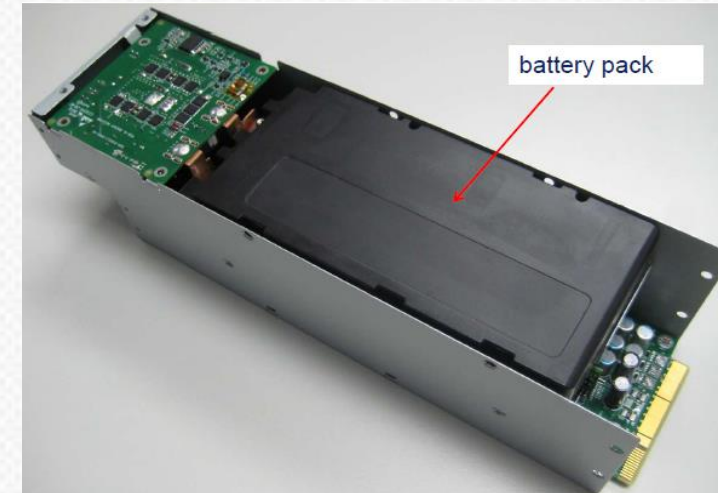
- Eliminates expensive and complexity from centralized UPS solution
- High availability solution

High Efficiency

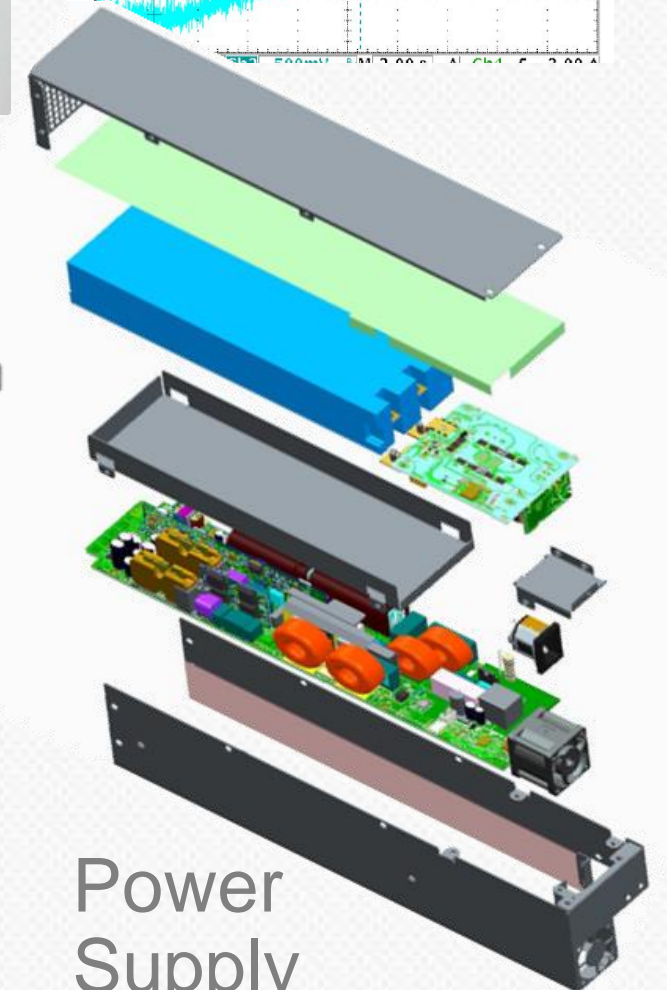
- Total power dedicated to A/C power faults reduced by a factor of three 13% → 4%
- Increases data center efficiency by 9%

Lower Costs

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



Battery Pack



Power Supply



Open CloudServer OCS v2.0 blade

Motherboard

- Intel © E5-2600 v3, 2S, 120W
- 16 x DDR4 DIMMs, 512GB memory

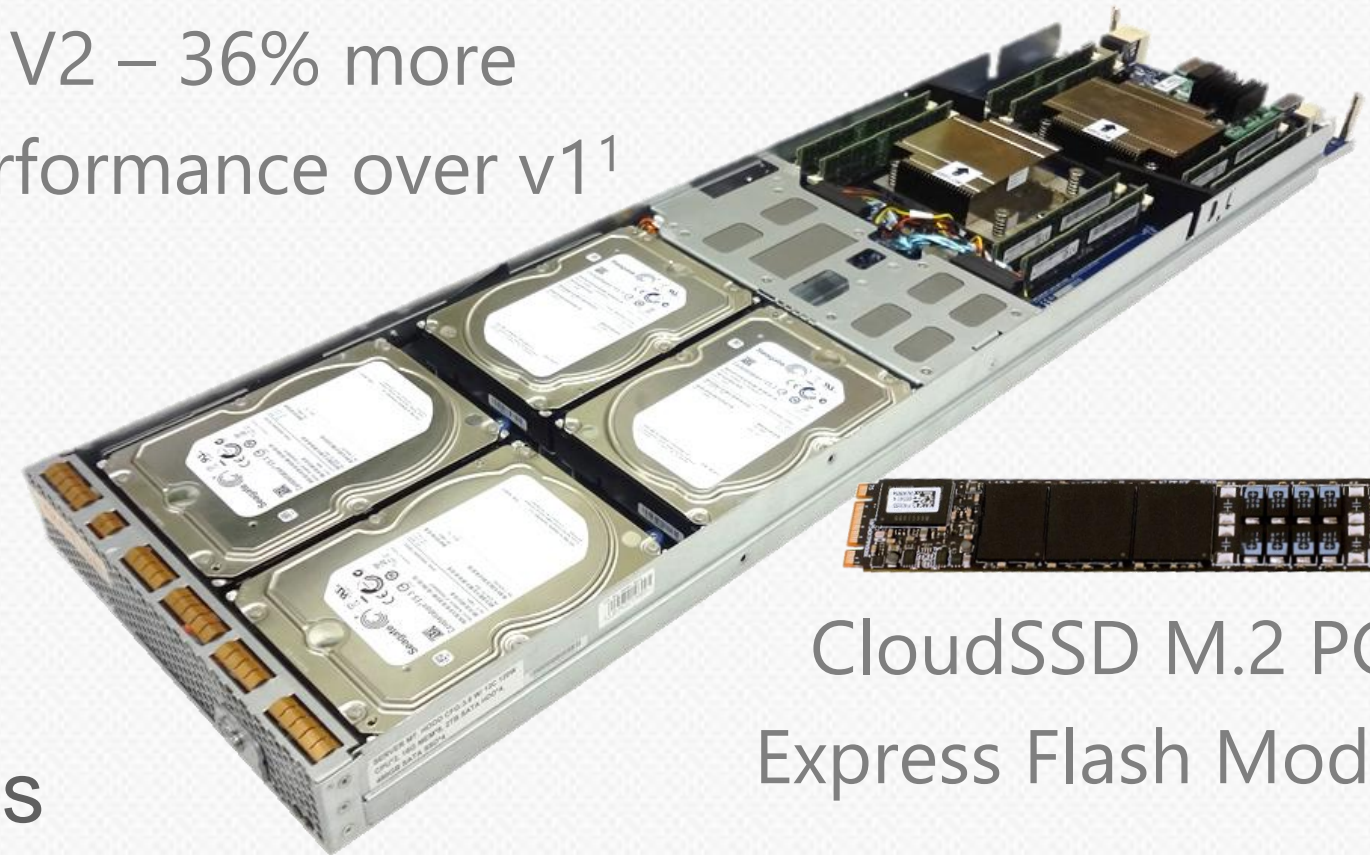
Storage

- 4 x 3.5" HDD
- 4 x 2.5" SSD –or–
- 8 x CloudSSD PCIe x4 M.2 Flash Modules

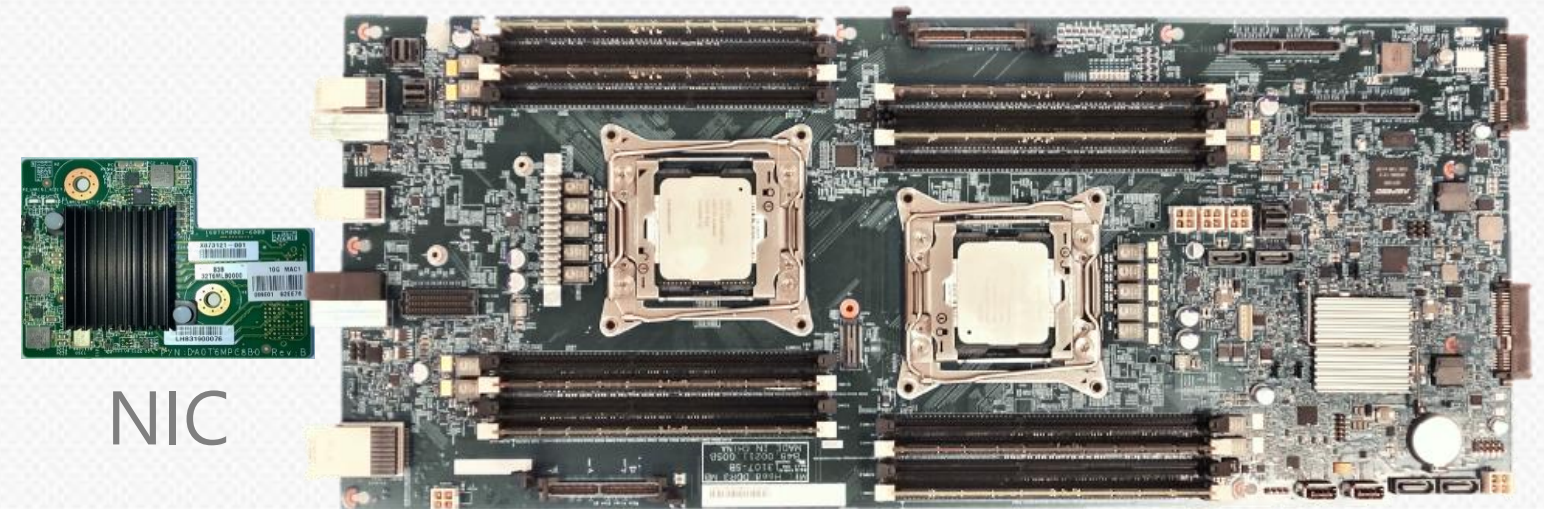
I/O

- Blade single 40G or 10G NIC mezzanine
- Tray PCI-Express x16 Gen3 Mezzanine
- Dual 4X SAS 12G

V2 – 36% more
performance over v1¹



CloudSSD M.2 PCI-
Express Flash Modules



NIC

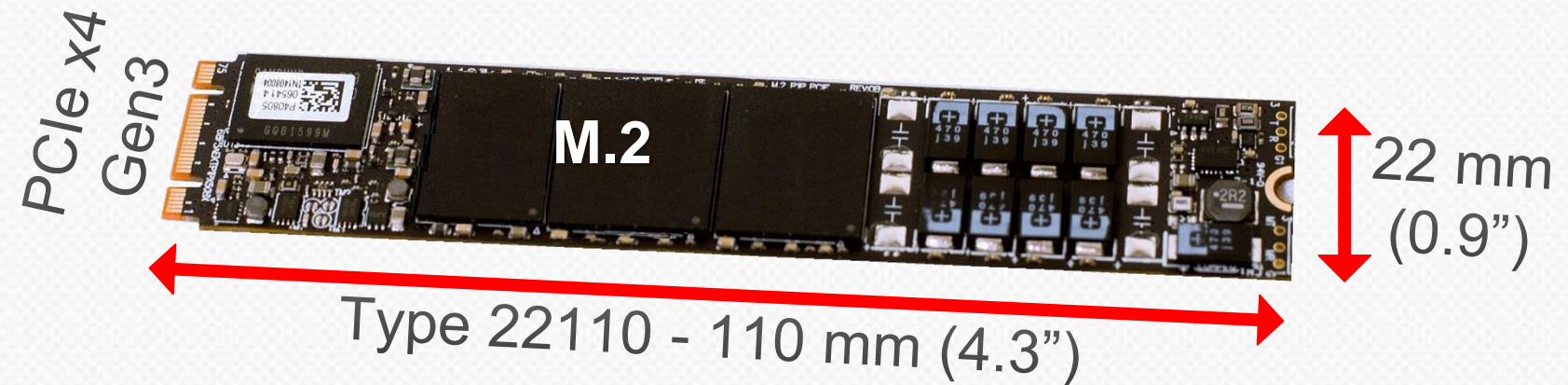
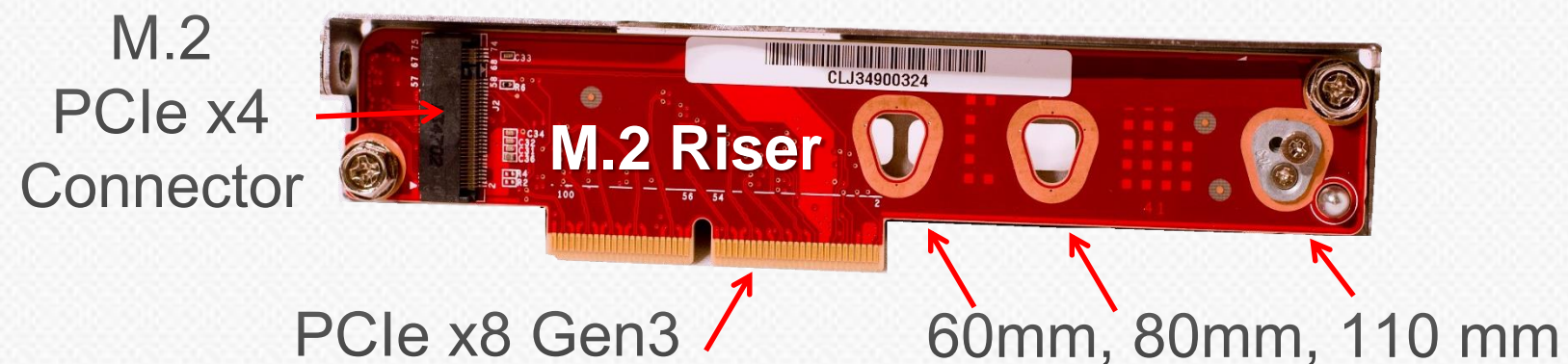
¹SpecInt Rate E5-2670 v3 120W vs. E5-2470 v2 95W

OCS M.2 CloudSSD Optimized Flash

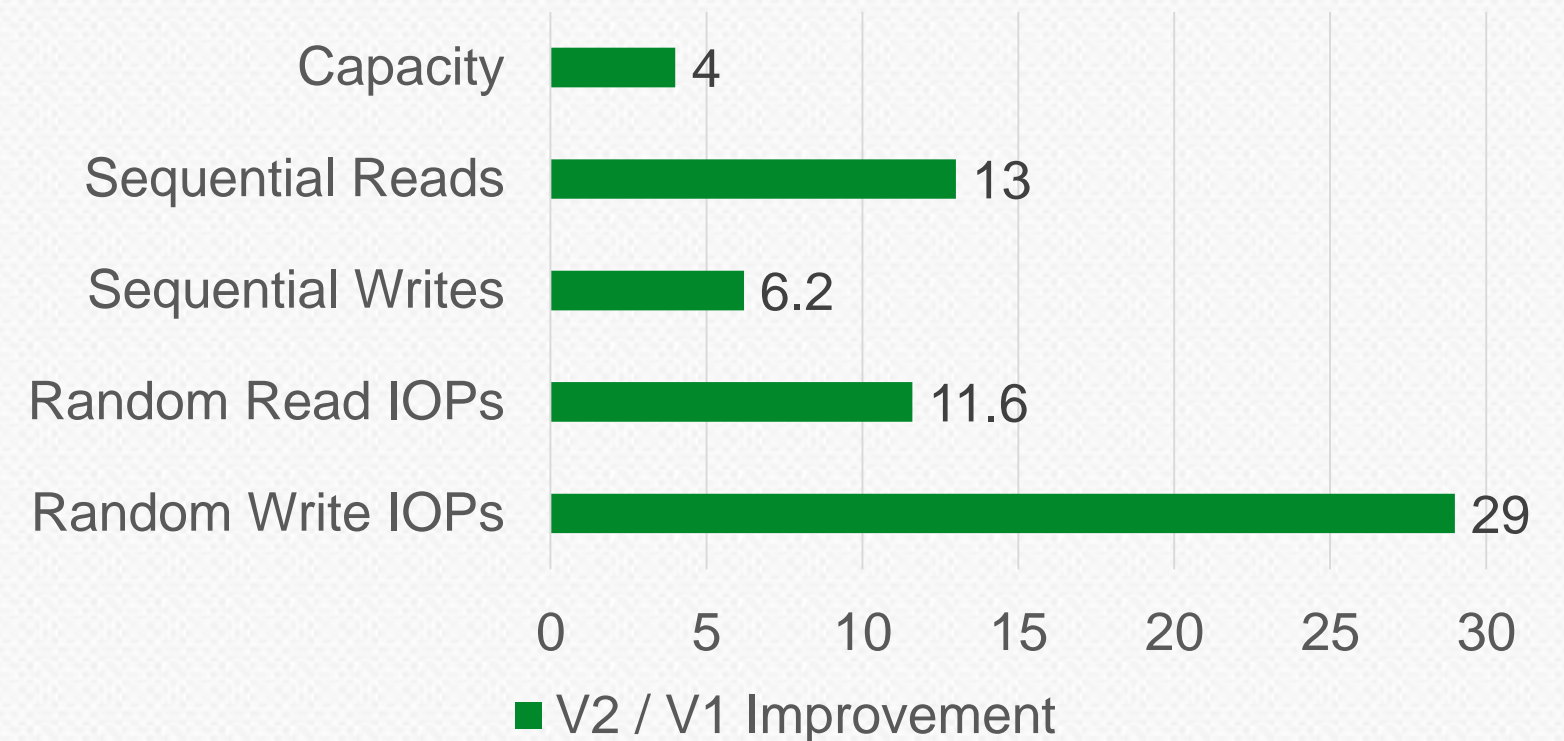
M.2 CloudSSD Flash Drives

- 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



V2 M.2 NVMe Improvement over V1 SSD



Chassis trays

Blade support

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

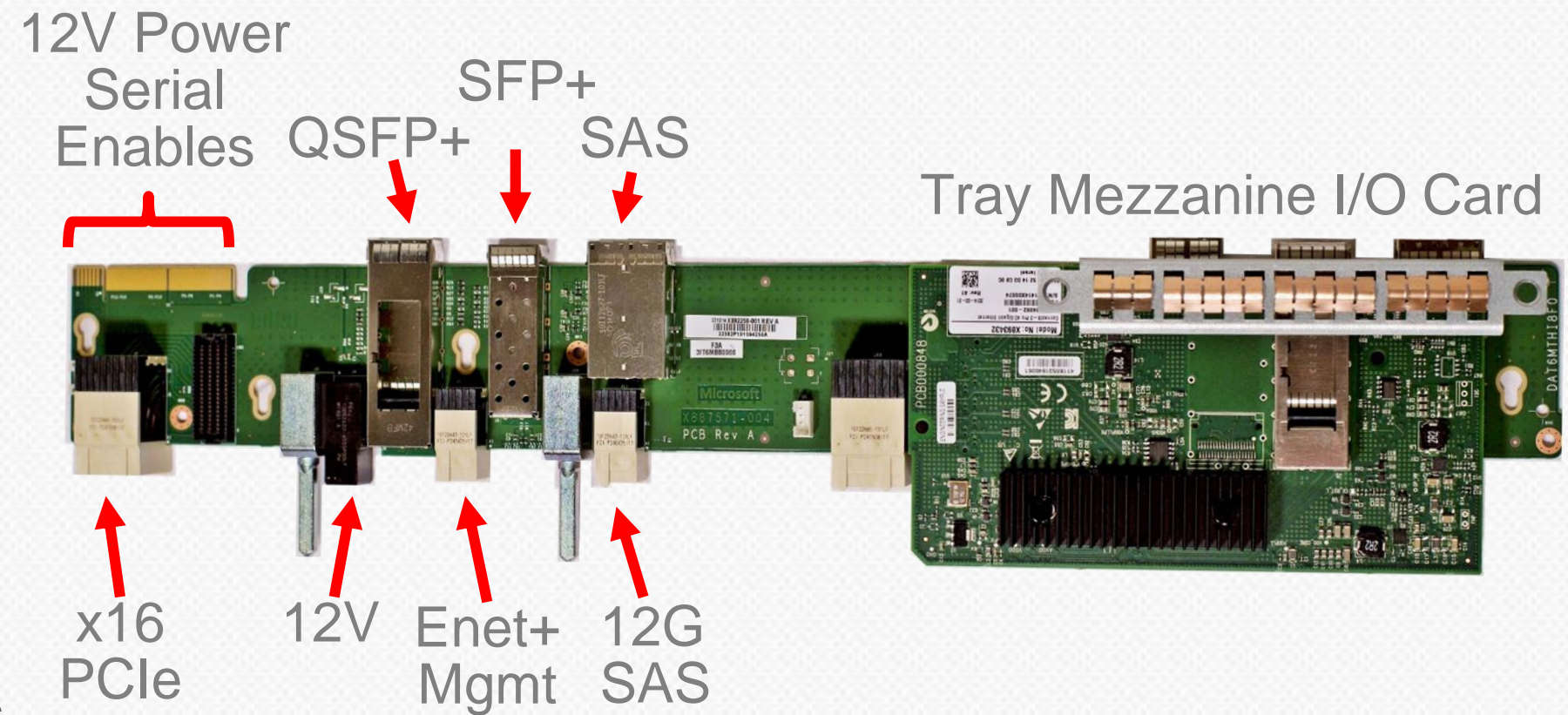
High Speed I/O

- 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



Tray mezzanine: designed for advanced networking such as 25 Gbps



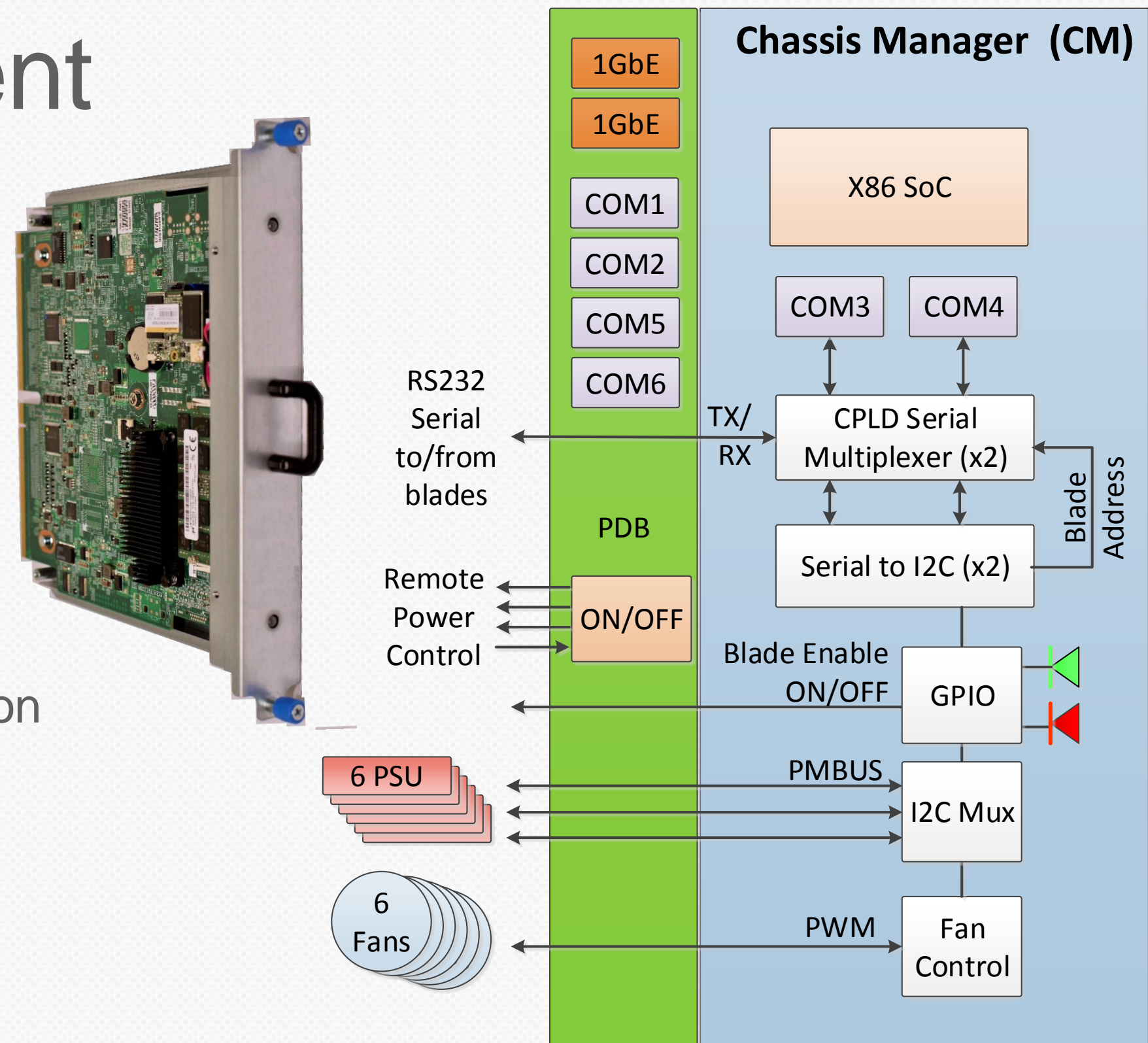
Chassis management

Secure OOB management

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

Hard-wired management

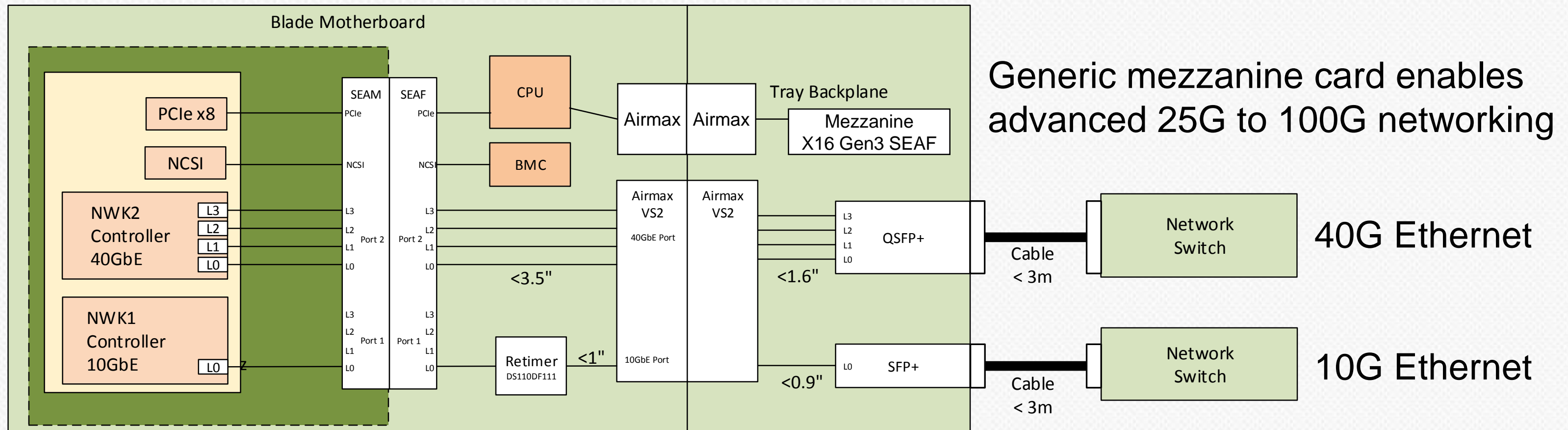
- On/Off to blade power cut-off circuit
- IPMI-over-serial out of band communication
- Fan and PSU control and monitoring
- Remote switch and CM power control
- Software is being open sourced
- Same hardware as OCS v1



Compute blade Networking

Flexible options to transition from 10G to 40G

Optional NCSI signals for side-band management



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



Safety and compliance

Ready for data centers world-wide

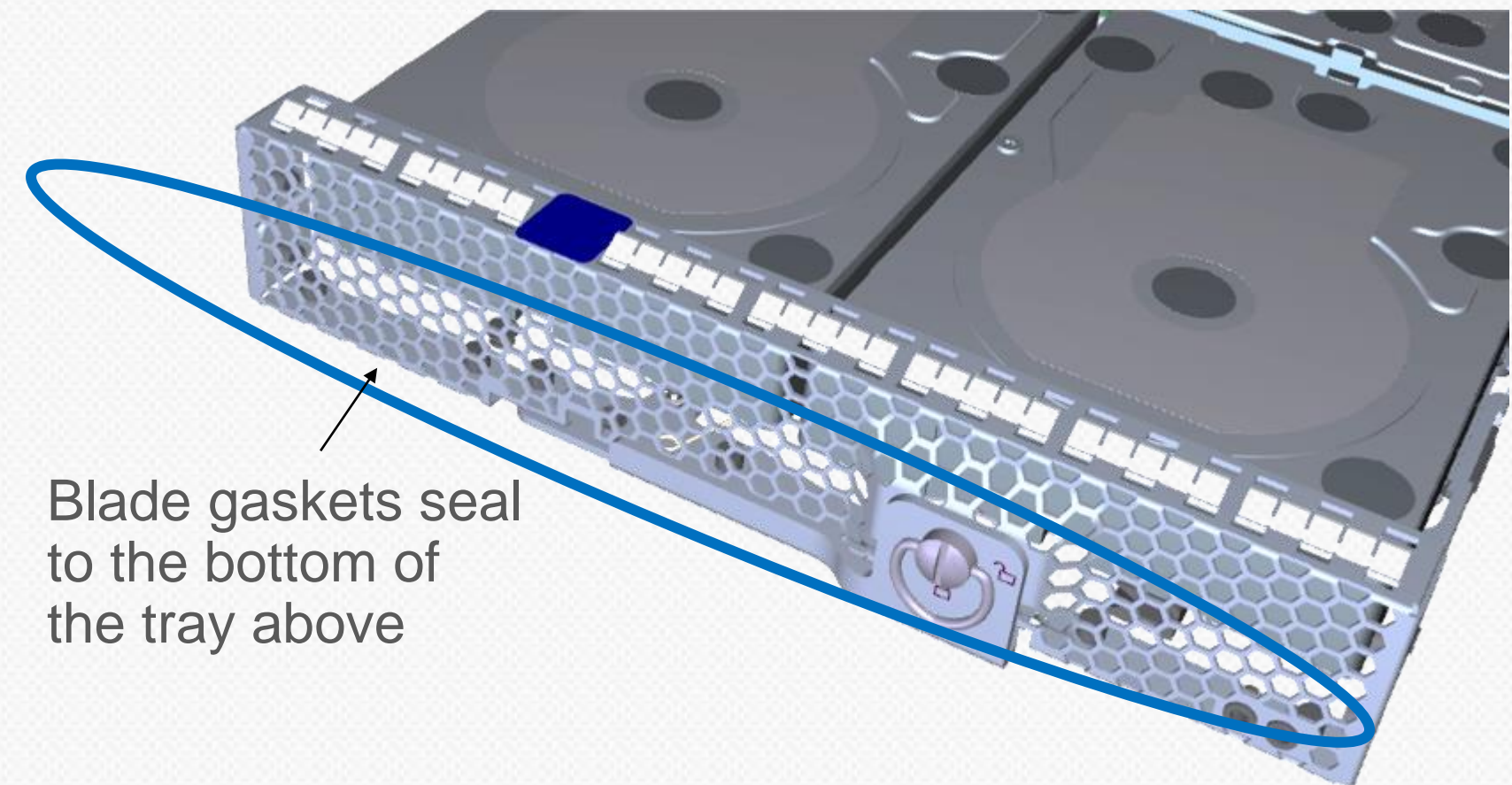
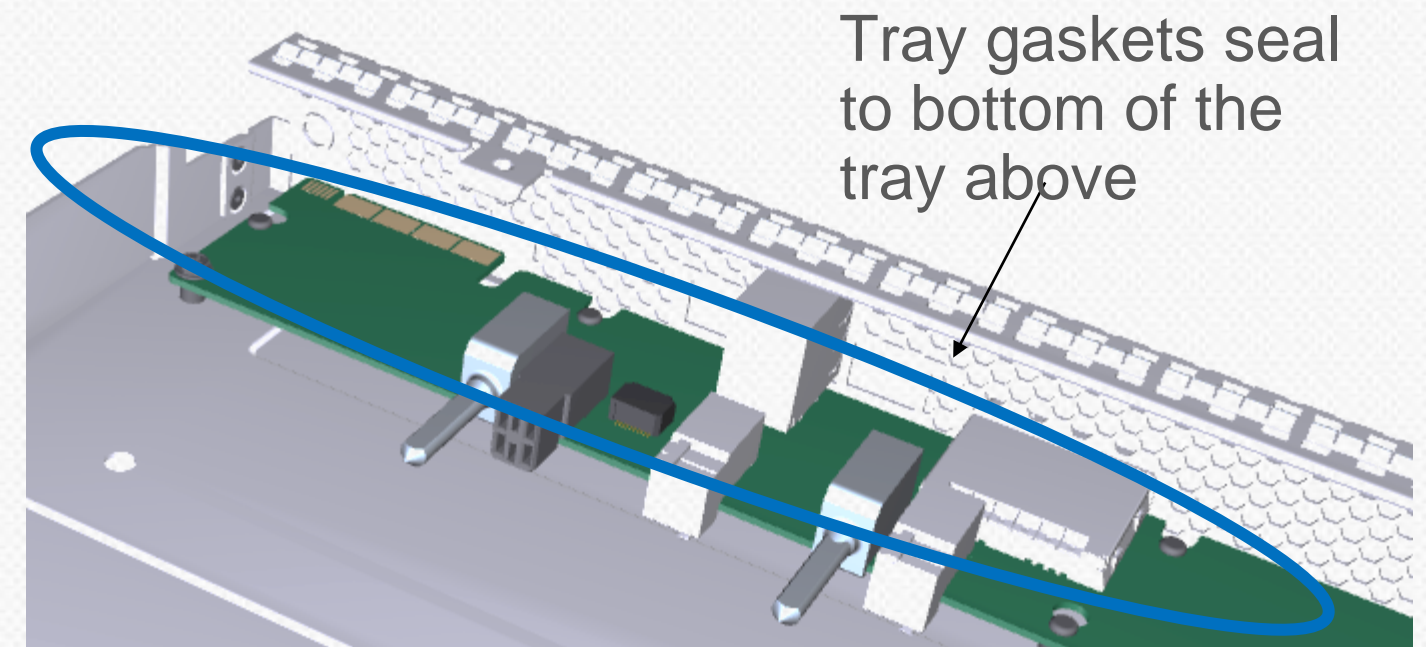
- Microsoft requires full compliance
- Containment at blade and tray
- Chassis is contained for use in EIA racks

Safety is Microsoft top priority

- UL, IEC, CSA standards among others

EMI Compliance is important

- CISPR, ANSI, IEC standards to start with



SW submissions

Operations Toolkit

Chassis Manager Service v2



Operations Toolkit

Component Updates

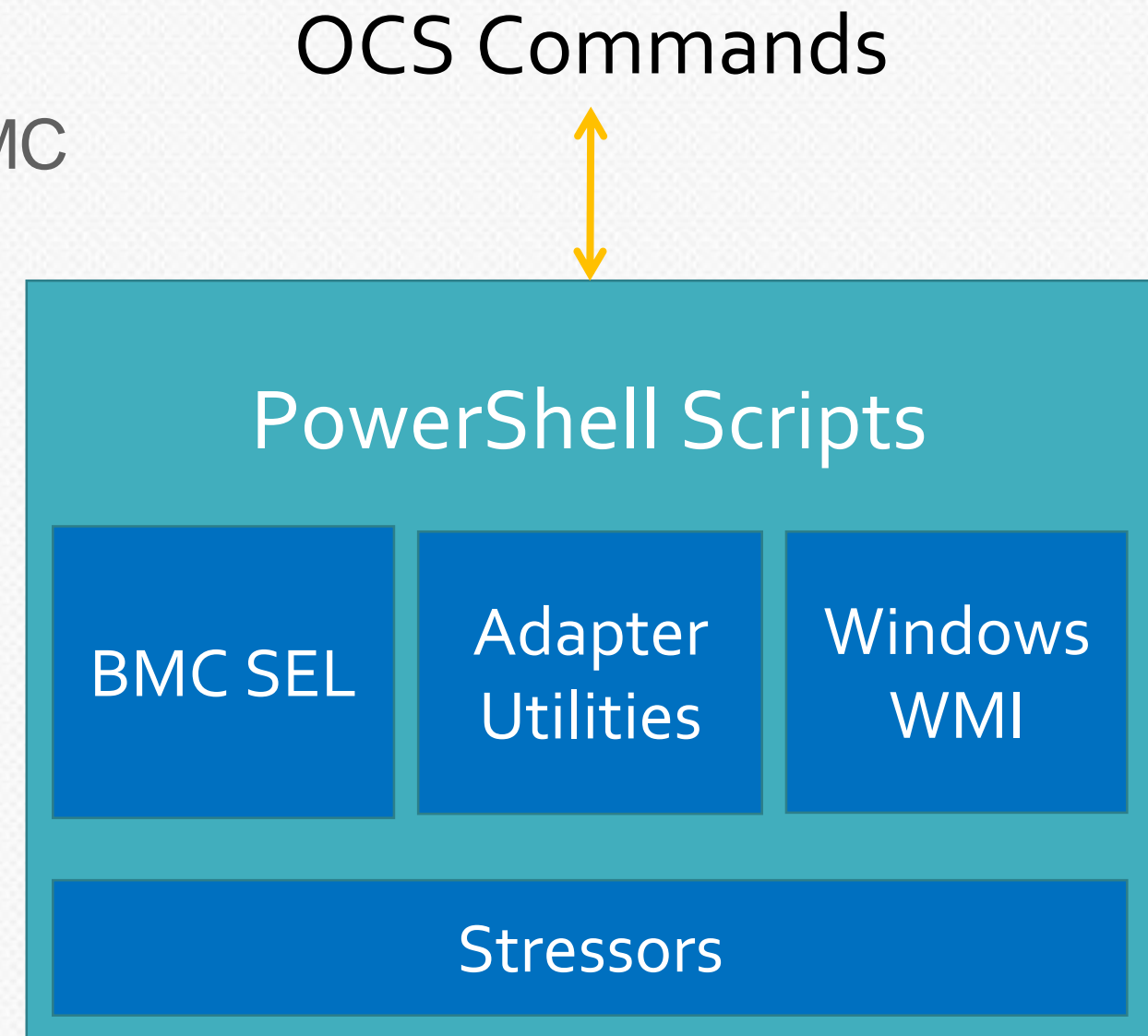
- Update programmable components such as BIOS and BMC
- Batch update of all programmable components

Diagnostics

- Identify defective components (HDD, DIMM, CPU)
- View, Log, and Compare Configurations
- Read, clear and log errors
- Execute IPMI and REST commands

Functional Tests

- System stress tests to identify intermittent problems
- Component specific stress tests (HDD, DIMM, and CPU)



Chassis Manager Service v2

Better User Interface for CLI

- CLI auto-completion/scroll up-down/Function keys enabling
- VT100 support for serial sessions/DIGI sessions

Adding DIGI support

- CLI over serial
- Serviceability over Serial (crash cart)

QA enhancement

- Serial session hardening and state management
- Bug fixes



Comprehensive Contribution

Open Source Code

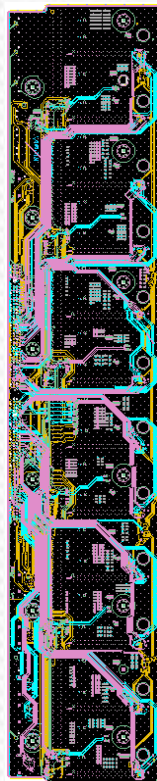
Chassis management
Operations Toolkit
Interoperability Toolkit

Specifications

Chassis, Blade, Mezzanines
Management APIs
Certification Requirements

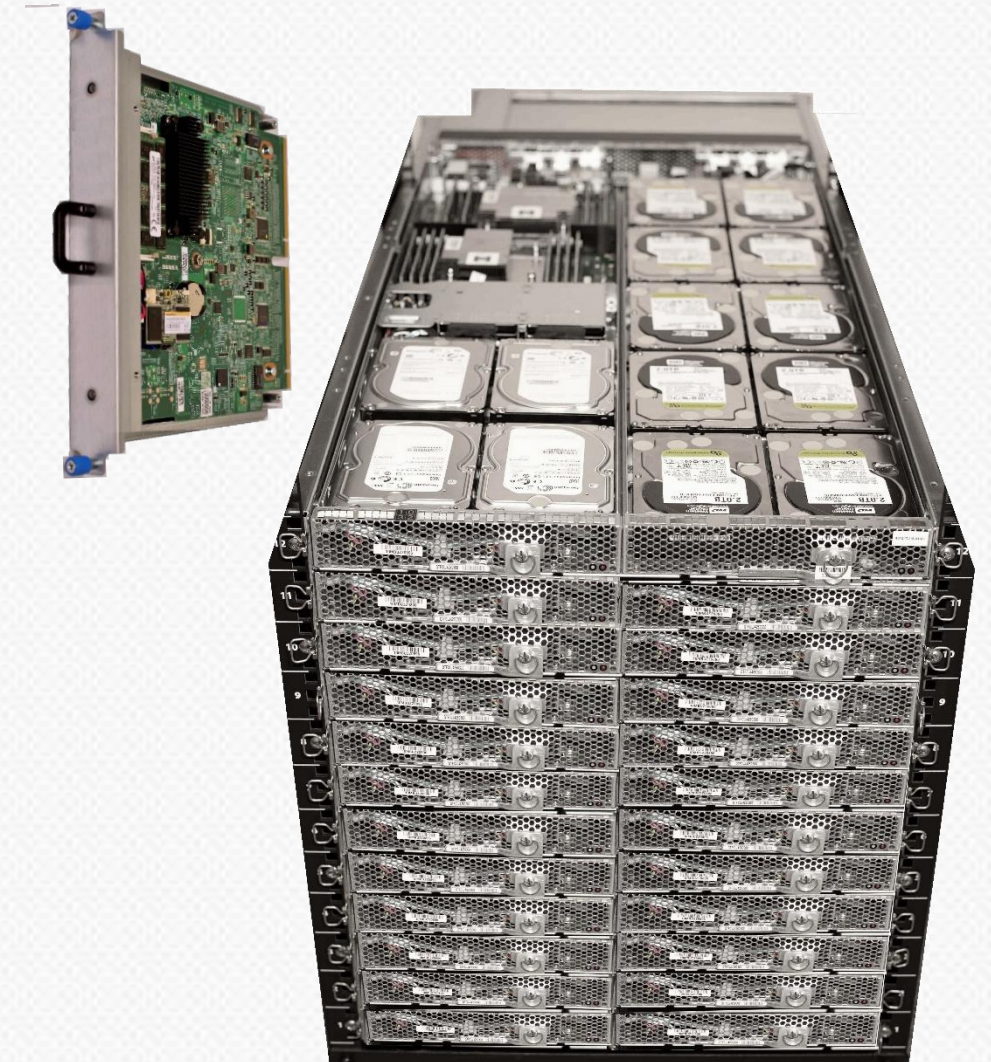
Mechanical CAD Models

Chassis, Blade, Mezzanines



Board Files & Gerbers

Power Distribution Backplane
Tray Backplane



Learn more

Visit Microsoft booth

- OCS v2 Systems and Demos on display

Attend executive track session:

- Microsoft, Kushagra Vaid, Wed 9:30

Attend technical workshops

- OCS 1U Quad Server System, Mon 2:00PM
- HW Management workshop (multi-node management), Mon 4:00PM
- OCS v2 Hardware Overview, Tues 4:00PM
- OCS v2 Power Supply with Battery, Tues 4:45PM
- OCS v2 CloudSSD, Tues 5:30PM
- OCS v2 Operations Toolkit, 2:00PM?



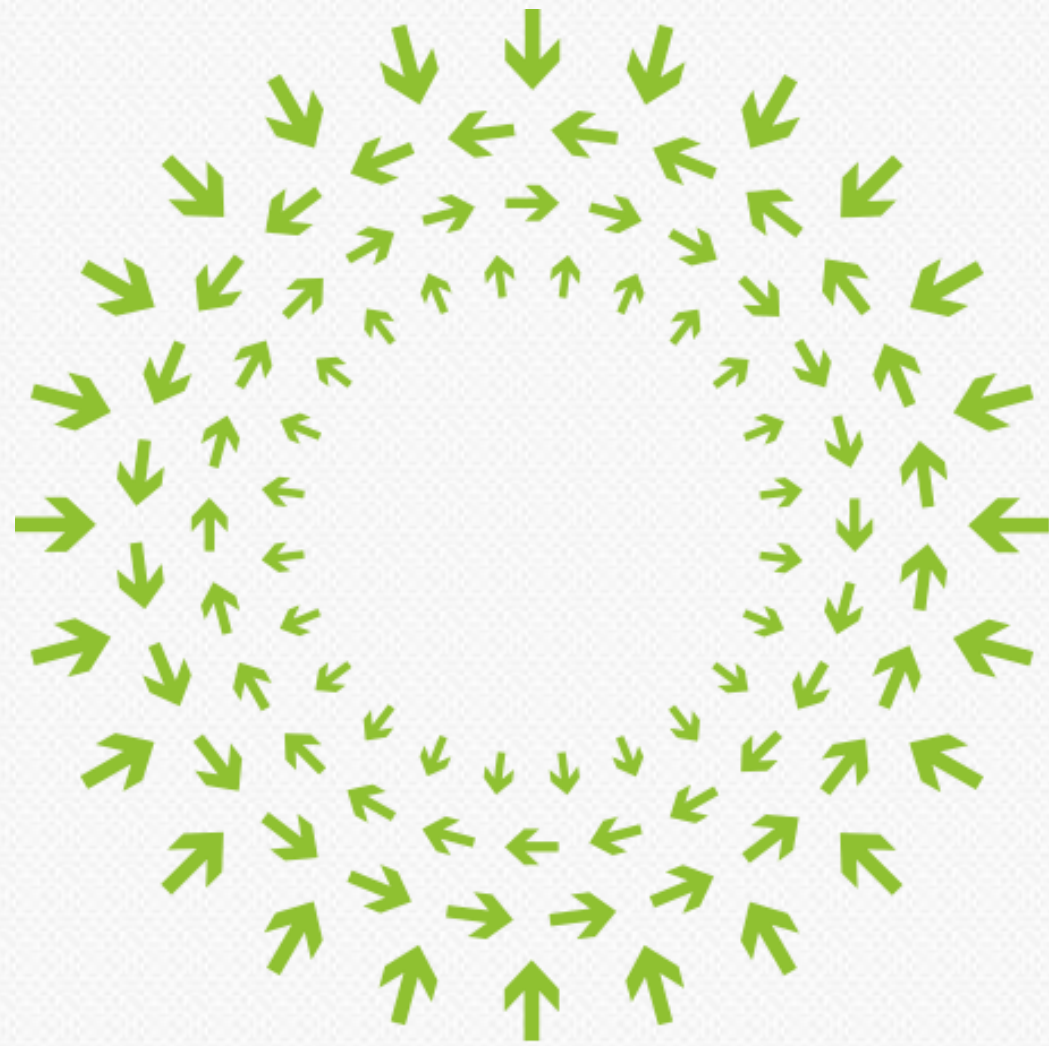
Q&A





© 2014 Microsoft Corporation. All rights reserved. The information herein is for informational purposes only and represents the current view of Microsoft Corporation as of the date of this presentation. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information provided after the date of this presentation. MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS PRESENTATION.





OPEN

Compute Engineering Workshop

March 9, 2015

San Jose