OPEN
Compute Project

Microsoft Project Olympus V1.0 Overview

- Microsoft Azure Cloud Hardware Infrastructure
- Presentation to OCP Server Project
- 11/29/2017

Talk Outline

- Open-compute and Azure at Cloud-scale
- OCP Server Project Approach
- Microsoft Implementation
- Project Olympus Modular Architecture
  - Rack Structure, Power, and Management
  - Servers Blades
    - Base Specification
    - Various Servers
  - Server Firmware
  - Expansion Chassis
    - JBOD, JBOF, PCIe Expansion Chassis
  - List of Contributed Material
  - Project Olympus Sub-group within OCP Server Project

Microsoft

Open Compute Enables Cloud-Scale

AT MICROSOFT AZURE

- *Cloud-scale, Cloud-speed!*
- Azure servers diverse set of customers
- Efficiency is key at all levels
  - Space, Power, Cooling, Management, Material, Cost, Time
  - Performance
- Fail-Fast Model
  - Many opportunities, everchanging requirements
  - Explore several concepts concurrently
- At Cloud-scale, we need everybody’s help!
  - No one partner can execute on all angles at the same time
  - Need to work with competing partners

Microsoft

OCP Approach

AT OCP SERVER PROJECT

- Detailed Specification Contribution
  - Base Specification
    - Architectural Tenets guide several Designs
  - Design Specification
    - Clear Module Boundaries for a specific product based on the Base Specification
  - Product Contribution
    - Design Package, IP, and Collaterals
    - Orderable SKU
- Encourage complementary contributions
  - At System level
  - At Module level

Microsoft

Microsoft Implementation

MICROSOFT APPROACH

- Modular Interoperability for Innovation
- Copy-exact for Controlled Production
- Assurance of Supply through multi-sourcing, etc.

- Early Contribution to OCP, encouraging feedback
  - Detailed specification
  - Promote Independent Collaboration

Microsoft

PROJECT OLYMPUS – OPEN SOURCE HARDWARE

Next-gen Cloud Hardware

Open-sourcing

leading edge

Hyperscale

Cloud Hardware

Development Model

New collaboration model with OCP community:

co-develop open-hardware at *Cloud Speed*

Industry Ecosystem

Bootstrap a vibrant ecosystem in OCP for the next generation of Datacenter Hardware

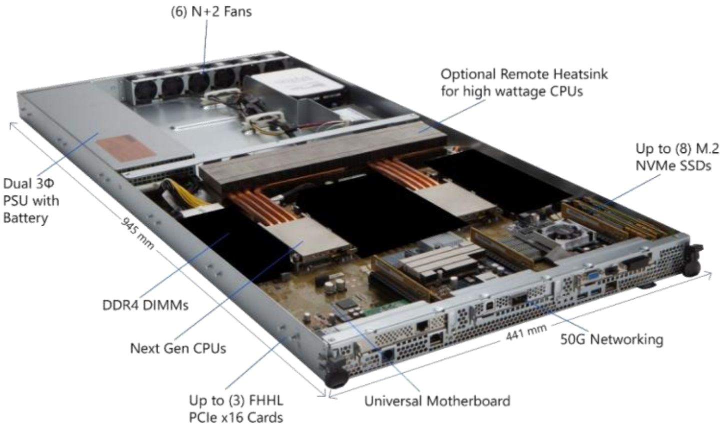
http://www.opencompute.org/wiki/Server/ProjectOlympus

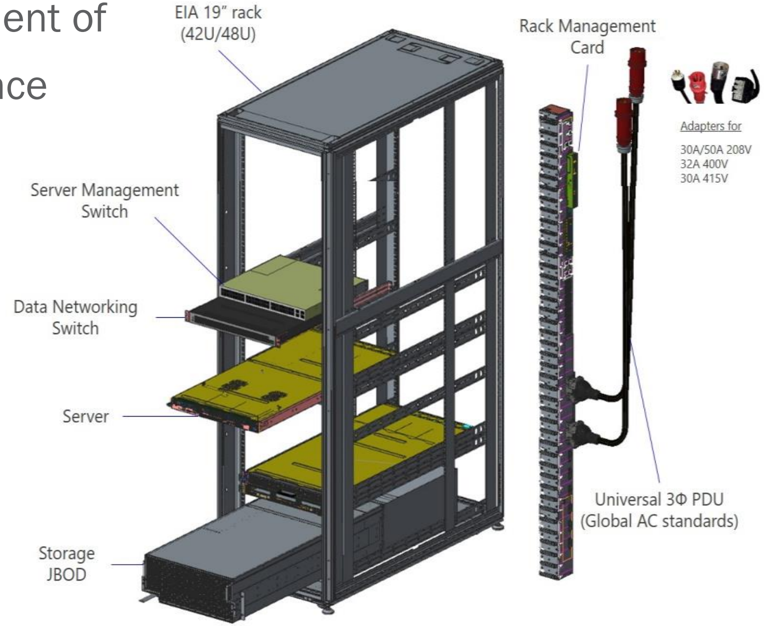
Microsoft

PROJECT OLYMPUS BASE SPECIFICATION

PROJECT OLYMPUS MODULAR ARCHITECTURE

Establishes a baseline for cloud-scale standard deployment of datacenter management, power, cooling, and performance

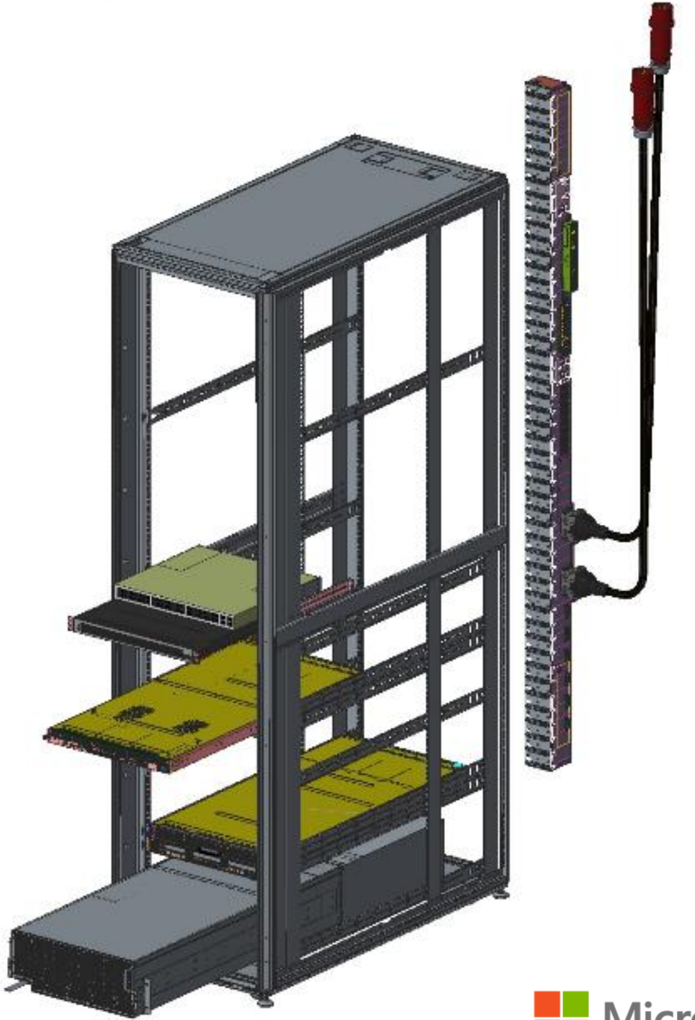




Microsoft

PROJECT OLYMPUS RACK

The Structure



19” EIA RACK STRUCTURE

NO-FRILLS HIGH CAPACITY RACKS

- EIA-310 Standard
  - 600mm wide, 1200mm deep
  - 42U & 48U options
  - 3,000 lbs. capacity
- Features
  - Integrates with Project Olympus modules
  - Front & rear locking door, sidewall options
  - Three EIA rails for standard equipment
  - Baffles and air blocking panel options

RITTAL

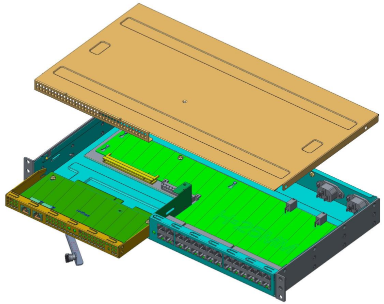
Microsoft

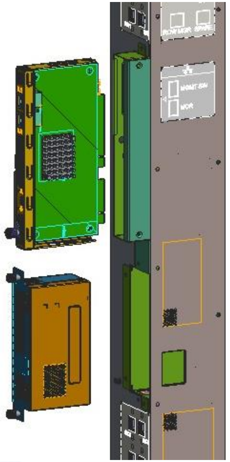
RACK POWER & MANAGEMENT

NO-FRILLS HIGH CAPACITY RACKS

Universal Power Cord

Management

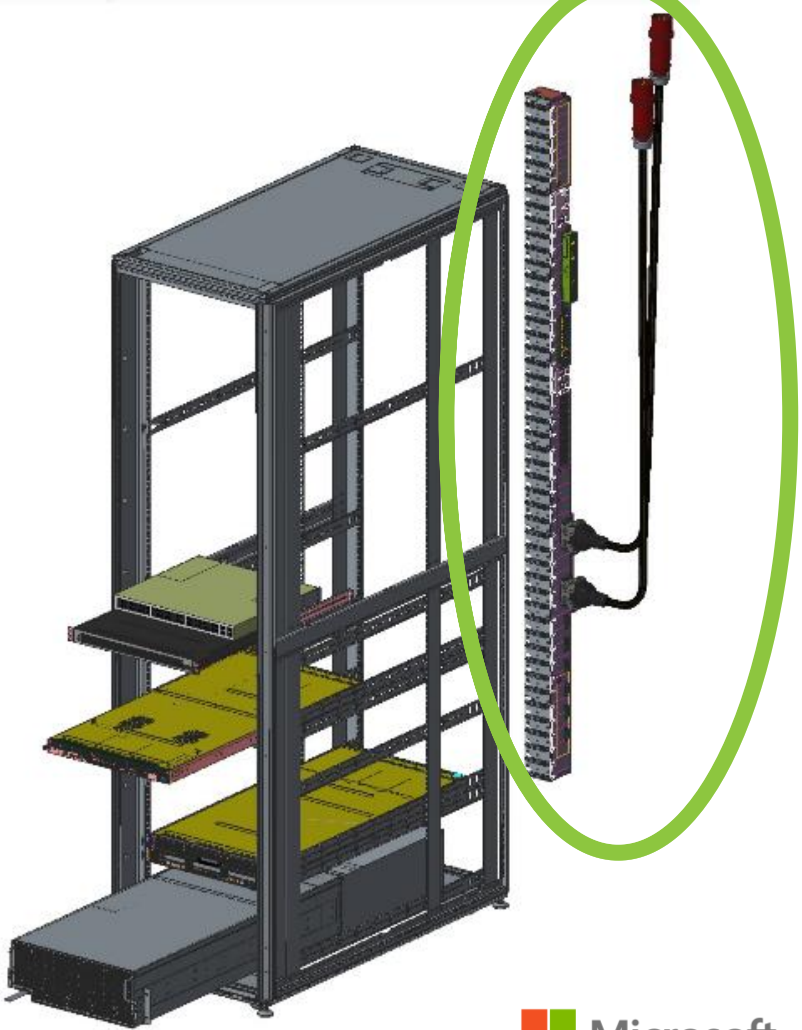




Microsoft

PROJECT OLYMPUS

Power



Microsoft

UNIVERSAL POWER

A/C POWER ADAPTED TO YOUR DATACENTER

- Supports high availability data centers
  - Dual-feed, three-phase A/C
  - Rack power monitoring and capping
- Universal Power Distribution
  - Distributes power and management
  - Integrated Rack Manager
  - Supports blind-mate servers
- A/C power cord adaptation
  - Racks never change, only A/C cord
    - 208V-30A, 208V-50A, 415V-30A, 400V-32A

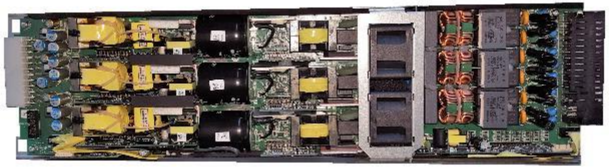
ingrasys®

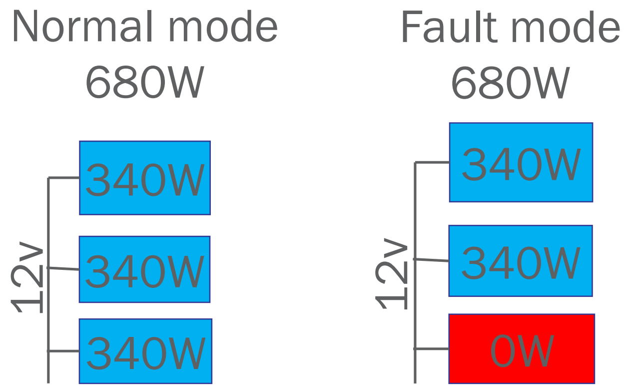
Microsoft

BLADE POWER SUPPLY

OPTIMIZED FOR EFFICIENCY, RELIABILITY, HYPERSCALE DATACENTERS

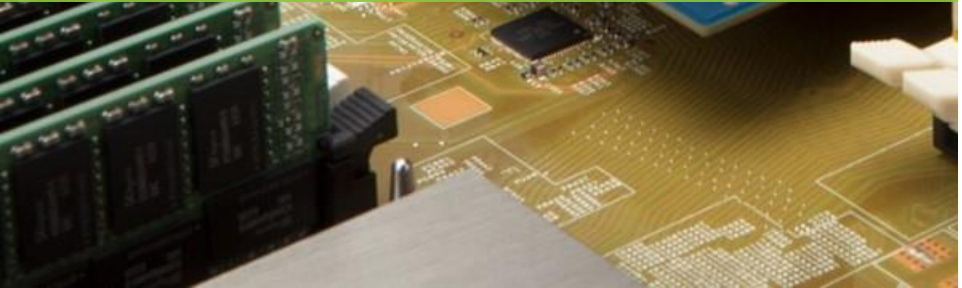
- Three x 340W PSUs Fully Integrated
  - Three-phase balanced AC power
  - 680W N+1 (1020W total)
  - Dual-feed auto-selection (IVS)
- Fault Mode Resiliency
  - AC feed failure, automatic fail over
  - PSU failure caps power if necessary
  - N+1 HA => no repair on failure
  - Double fault will be extremely rare

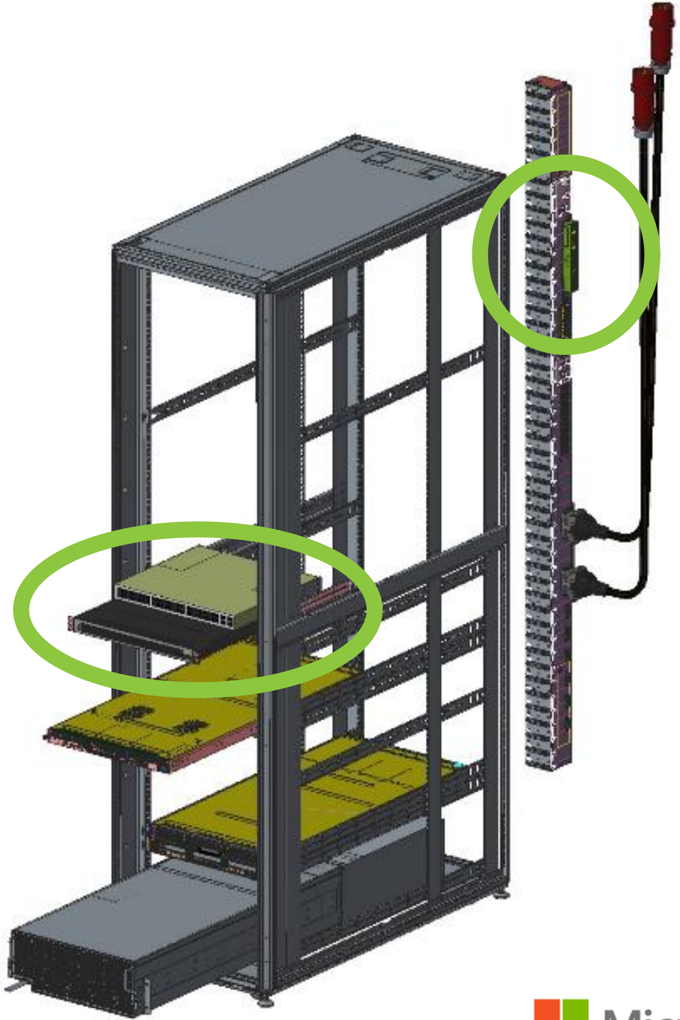




Microsoft

PROJECT OLYMPUS RACK MANAGEMENT



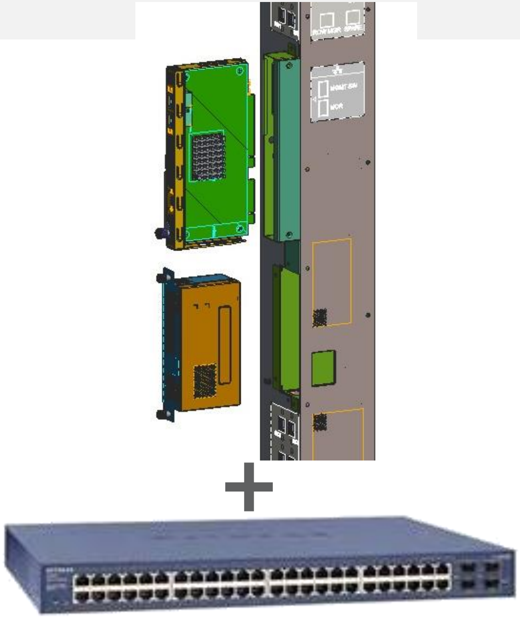


Microsoft

RACK MANAGEMENT

DRIVING UP DATACENTER UTILIZATION

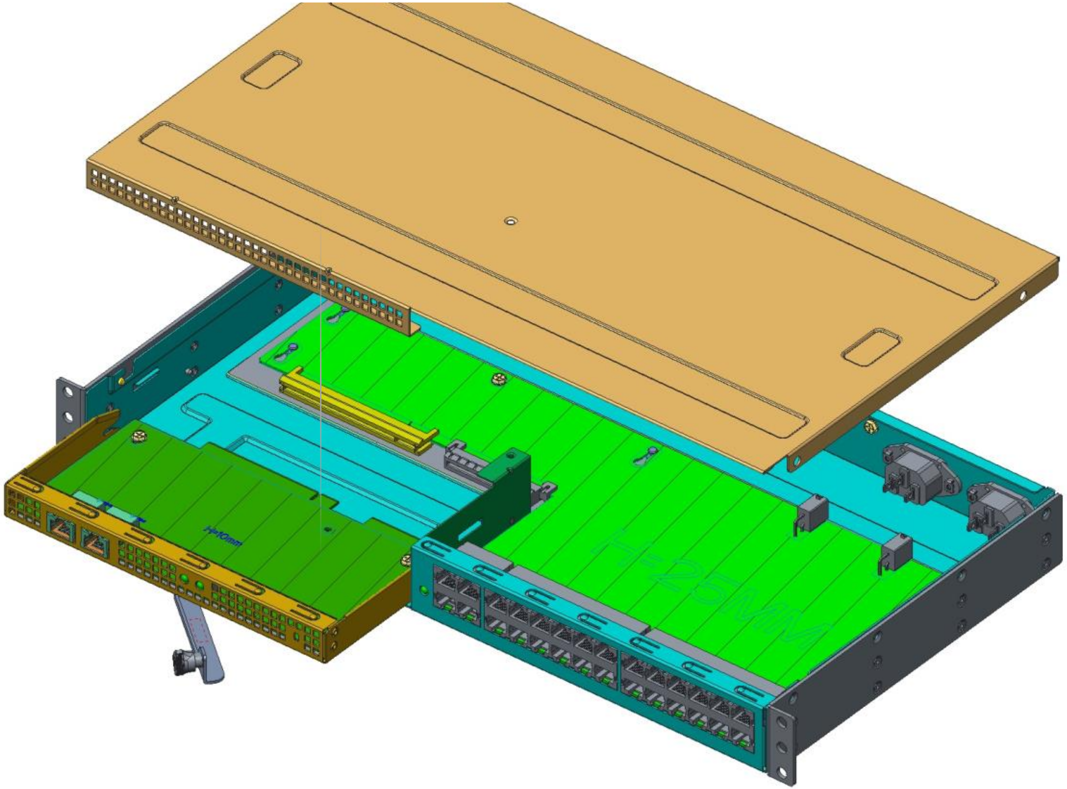
- Rack Management
  - Restful API I/F or Redfish via Ethernet
  - Rack Manager (RM) ARM CPU
  - Integrated into PDU
- Blade Management flexible to your needs
  - GbE I/F to each blade's BMC
  - NCSI enabled, cable to OCP Mezz Carrier
  - KVM enabled on motherboard
- Standalone, 1U rack mount version
  - For hardware that does not use the PDU



Microsoft

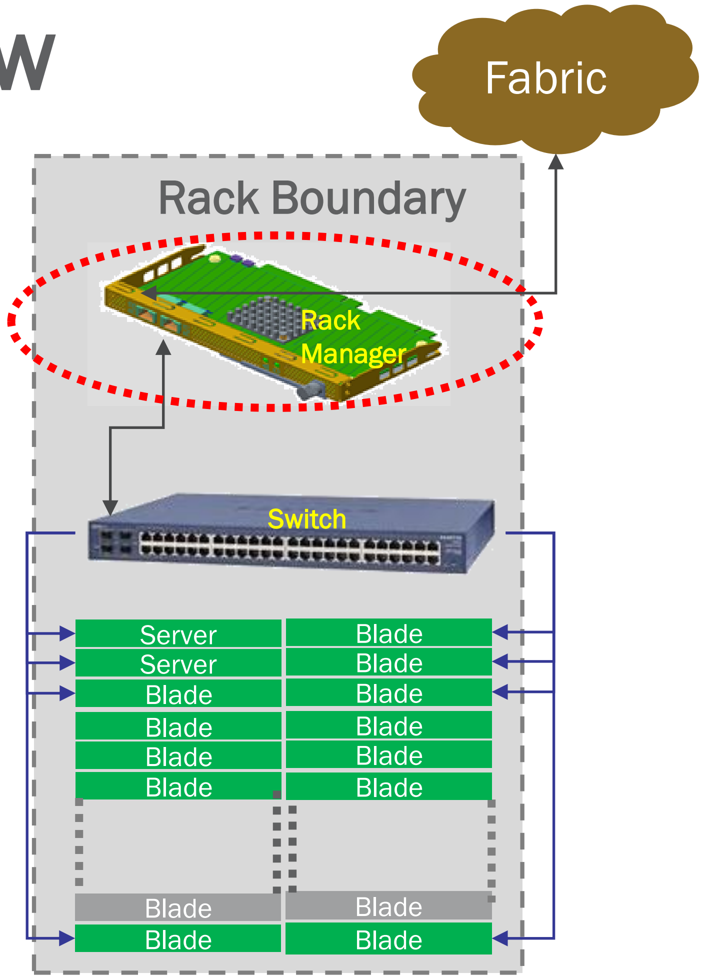
STANDALONE RACK MANAGER

- 1U Rack Manager Assembly
- Reuse of Rack Manager
- Supports Non-Olympus Rack Management
- Support Row Management
- Redundant AC Input
- Single SW image



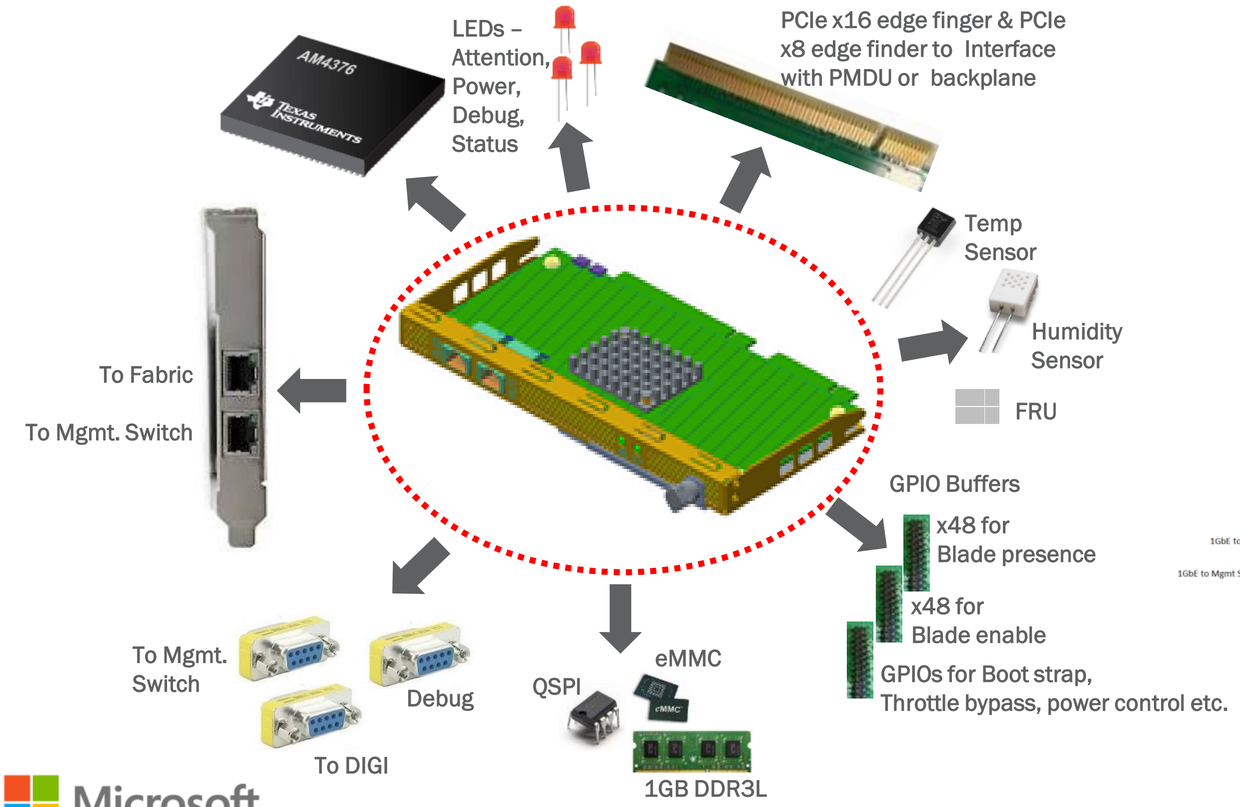
Rack Manager Overview

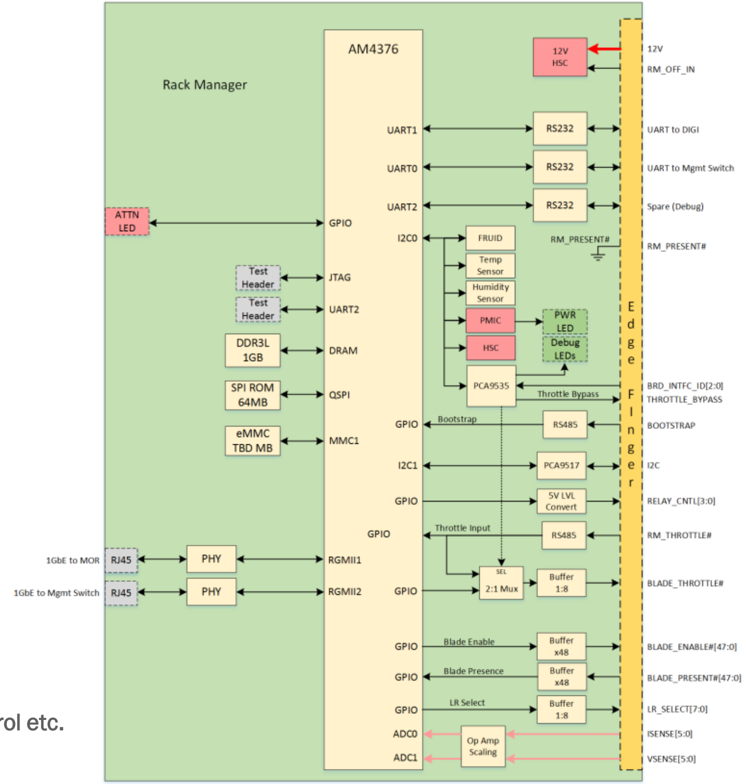
- Rack Manager Functions
  - Power Management
  - Out-of-band Server Management
  - Protocol abstraction, IPMI, Redifsh.
- RM Instances in Cloud
  - Discrete (Consumes 1U)
  - Integrated into PMDU
- Communication
  - Network
  - TTY Console
- Hardware Signaling
  - Server Presence
  - Server On/Off
  - Server Throttle
- Security
  - Secure Boot
  - BMC authentication
- Power Metering & Control
- Remote Debug
  - Kernel Debug
  - HW Debug (JTAG / ITP)
  - Logs / Telemetry
- Remote Media
- Out of band FW update and recovery
  - UEFI, CPLD, FPGA, PSU



Microsoft

Rack Manager Block Diagram





Microsoft

RACK MANAGER ARCHITECTURE

Auth at Interface:

- Ethernet / REST
- CLI / TTY

Privilege:

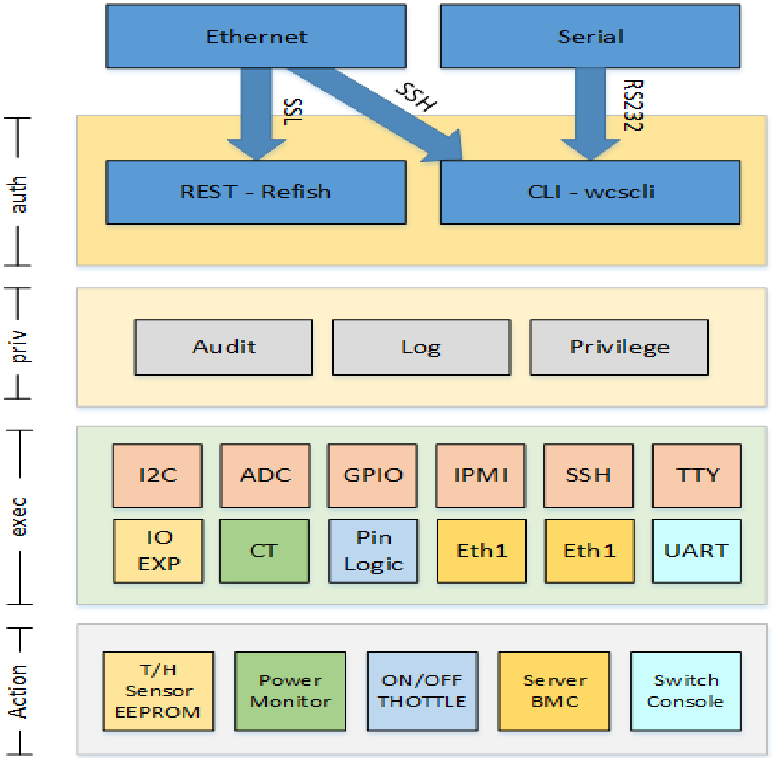
- Permissions token
- Log / Audit

Execution

- Task execution
- Logical to Physical

Action

- Hardware Control



Microsoft

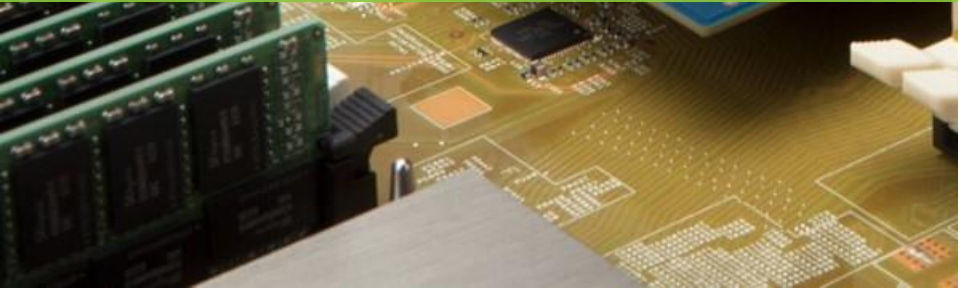
RACK MANAGEMENT LESSONS LEARNED

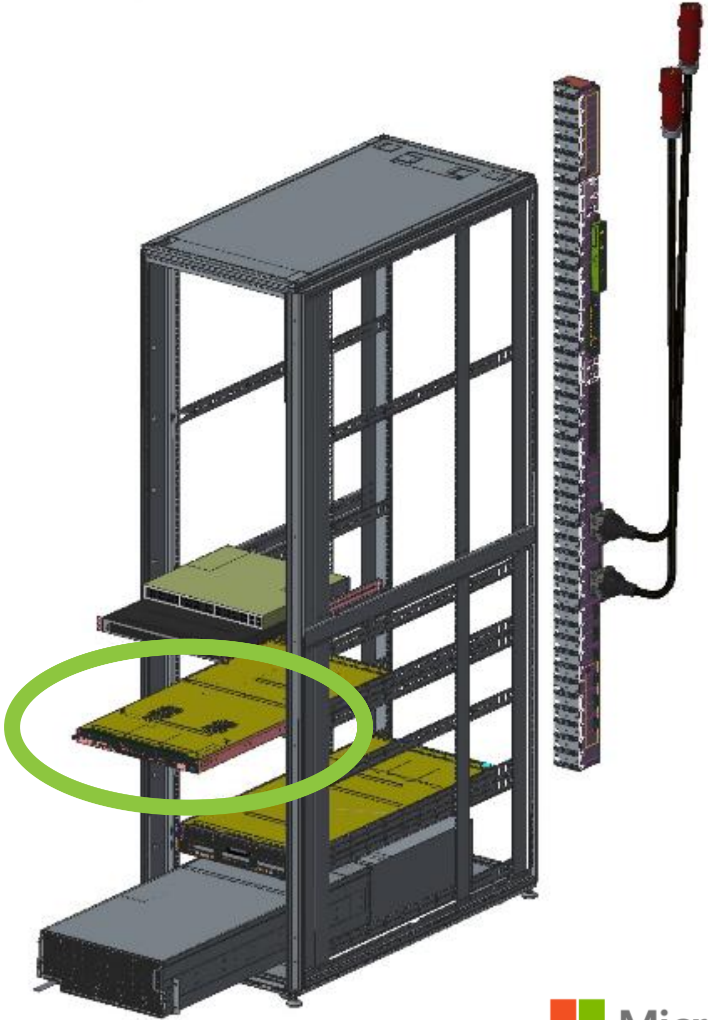
- ARM vs. X86
- Linux vs. Windows
- GbE vs. RS232
- Rack vs. 12U Chassis
- Redfish vs. custom API

RACK MANAGEMENT WHAT'S NEW

- Rack/Row/DC level Dynamic Power Capping
- Presence and Slot-ID
- Remote JTAG debug
- OoB FW update
- Odata client for auto-discovery

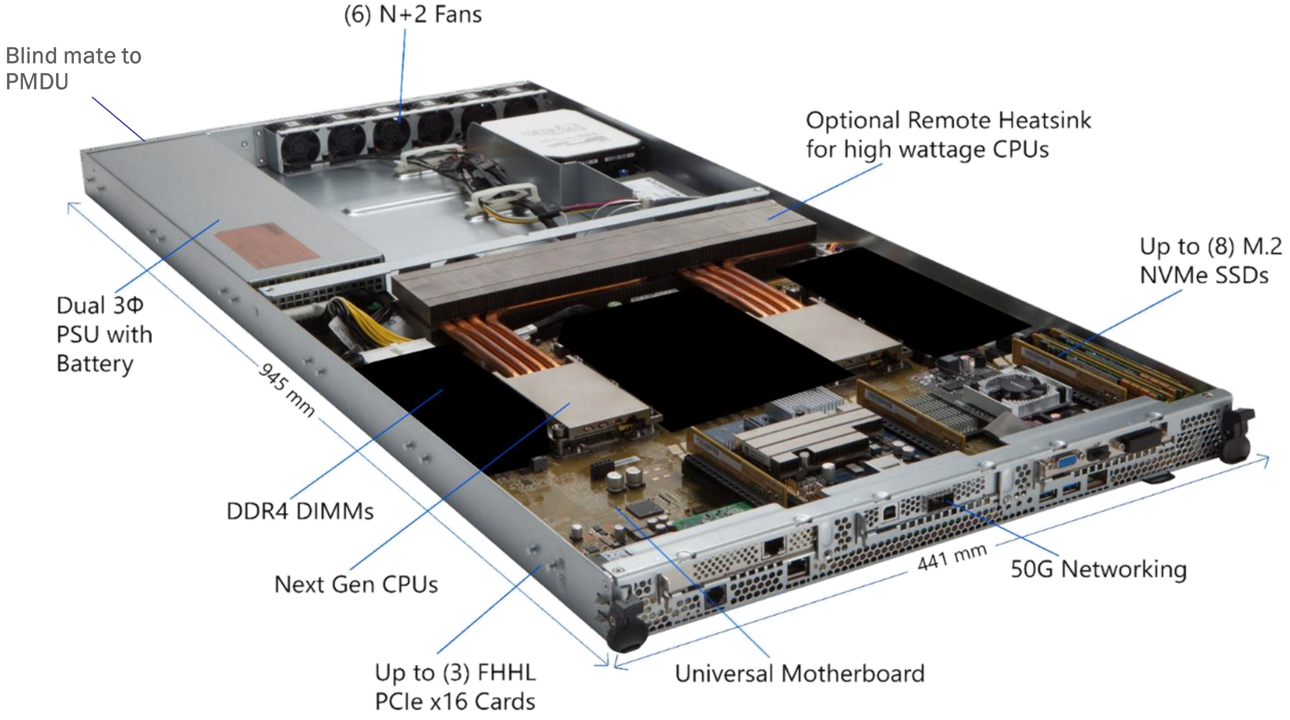
PROJECT OLYMPUS
Servers





Microsoft

PROJECT OLYMPUS 1U Universal Server
(Base Specification)



Microsoft

Servers

(Several Design Specifications)

PROJECT OLYMPUS SERVERS

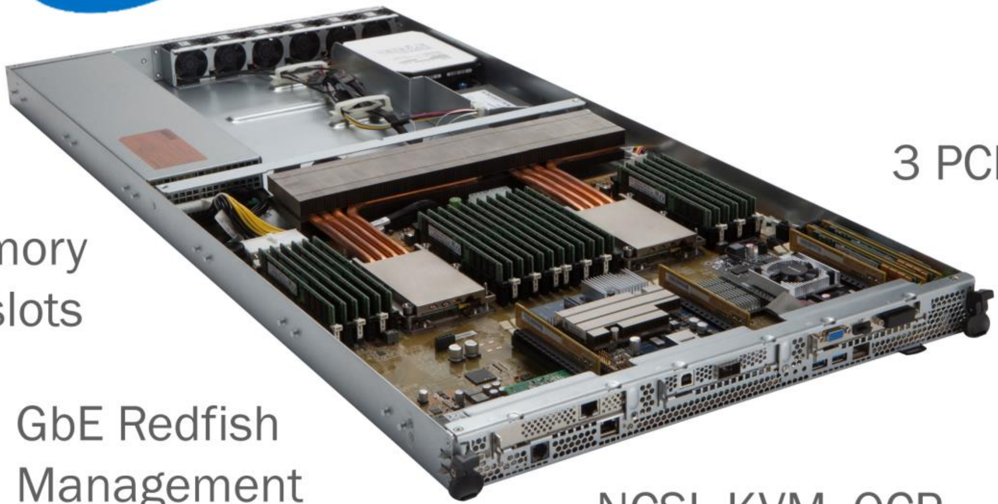
- Collaborative work with Industry CPU Leaders
- US1-XSP: Project Olympus 1U Server Motherboard, based on a dual-socket SXP CPU Platform from **Intel**
- US1-EPYC: A Motherboard and Server Design around dual-socket x86 CPU on EPYC Platform from **AMD**
- US1-THX2: A Motherboard and Server Design around dual-socket ARM64 ThunderX2 SoC from **Cavium**
- US1-Centriq: A Motherboard and Server Design around single-socket ARM64 Centriq™ SoC from **Qualcomm**

Microsoft

PROJECT OLYMPUS
Intel XSP-based 1U Universal Server (US1-XSP)

US1-Xeon

Intel® Xeon® Scalable Platform



3 PCI-E x16 FHHL slots

DDR4 memory
24 DIMM slots

GbE Redfish
Management

Up to 16 M.2
CloudSSD

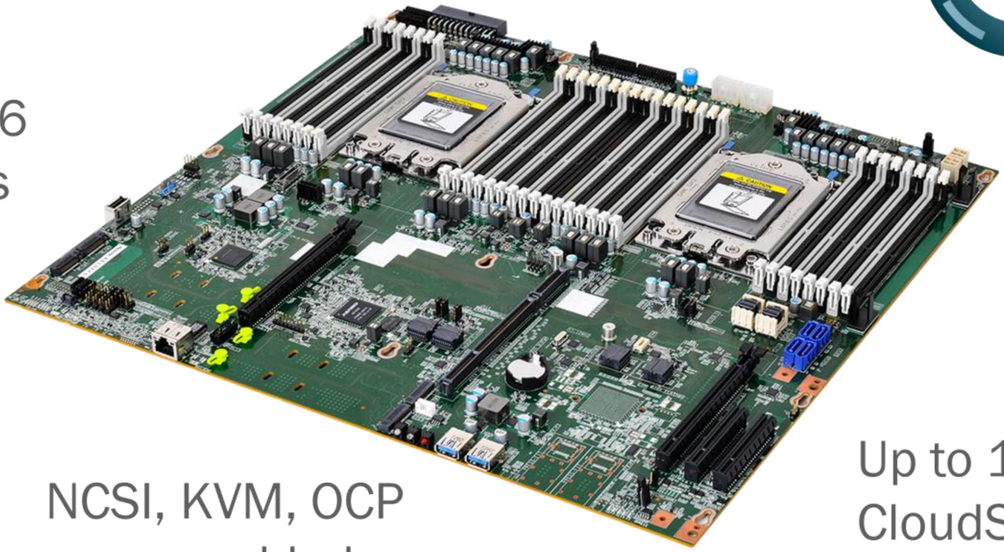
NCSI, KVM, OCP
mezz enabled

PROJECT OLYMPUS
AMD EPYC-based 1U, 2S Server (US1-EPYC)

US1-EPYC

AMD

EPYC



3 PCI-E x16 FHHL slots

GbE Redfish Management

NCSI, KVM, OCP mezz enabled

DDR4 Memory 32 DIMM Slots

Up to 16 M.2 CloudSSD

Microsoft

PROJECT OLYMPUS
Cavium THX2-based 1U, 2S Server (US1-THX2)

CAVIUM

Two THX2 SoCs

3 PCI-E x16
FHHL slots

US1-THX2

DDR4 memory
24 DIMM slots

Up to 16 M.2
CloudSSD

Microsoft

OTHER MOTHERBOARDS

DROP IN OR ADAPT TO PROJECT OLYMPUS UNIVERSAL SERVER CHASSIS

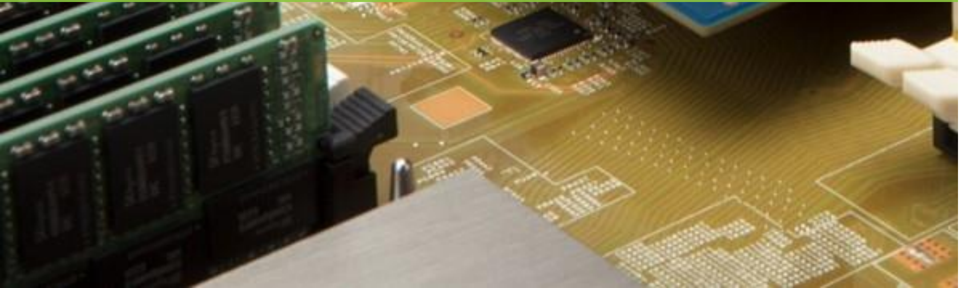
One ARM64 SoC

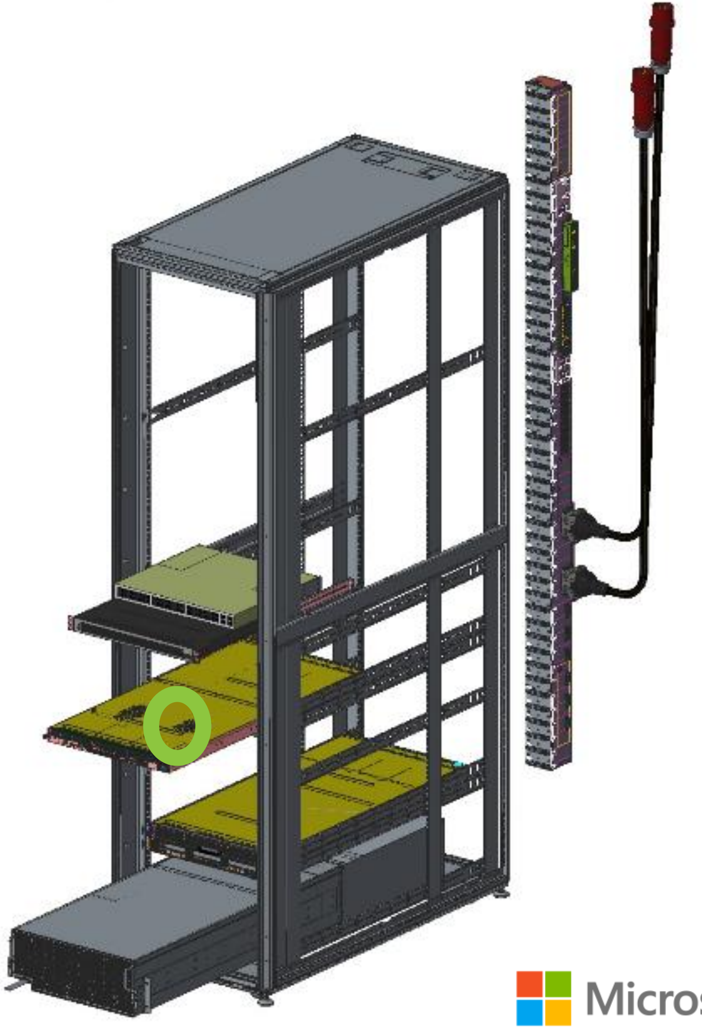
Various
PCIe Riser Options

US1-Centriq

DDR4 Memory
12 DIMM Slots

PROJECT OLYMPUS
System Firmware

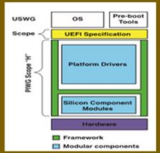




Microsoft

CLOUD UEFI

Mostly written in C.
High code re-use.

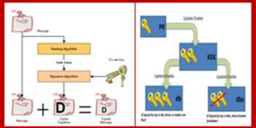


Emphasis on
Specifications.
Standards compliance.

Better platform
scaling. For e.g. removes
shadow ROM limits.

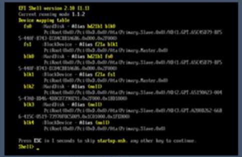
Storage.
GPT removes 2.2 TB
MBR restriction.

CPU Architecture
independent. Platform
design flexibility.



Secure boot solves
“trust” related system
integration challenges.

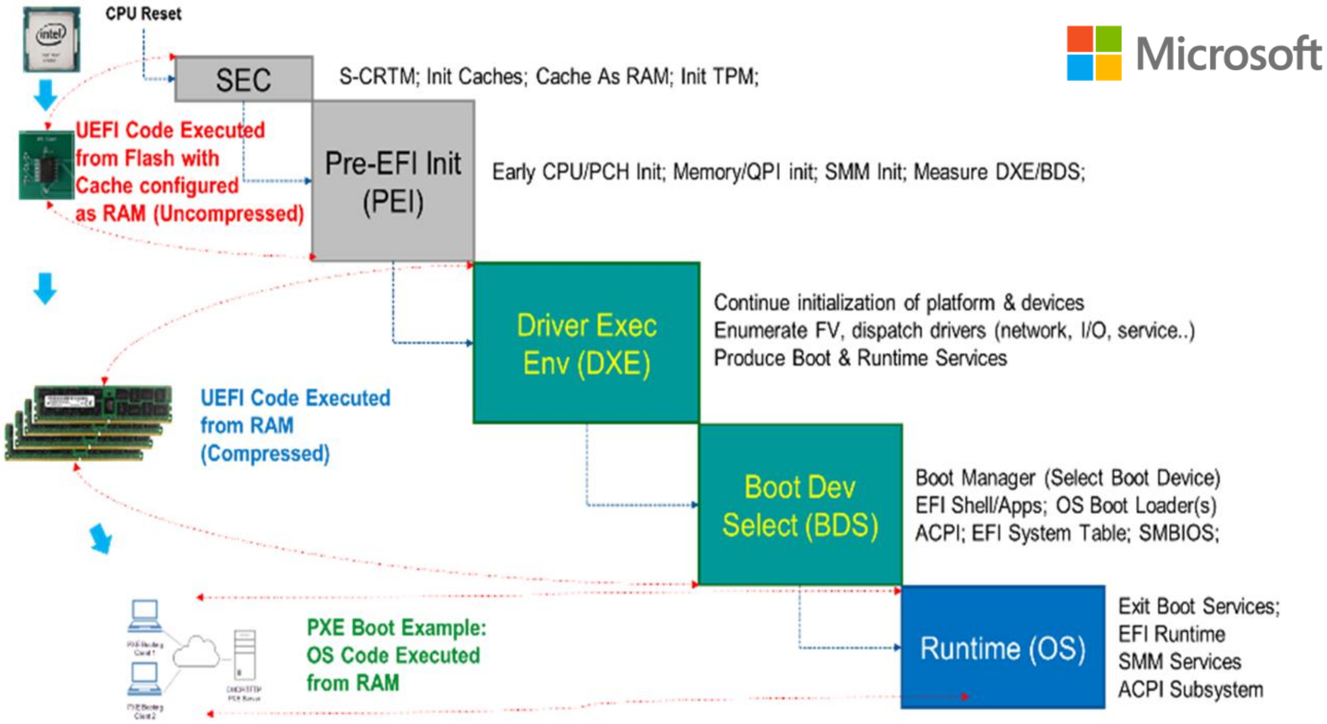
Pre-boot Networking.
Ipv4, Ipv6, PXE,
VLAN, iSCSI etc.



UEFI shell improves
pre-boot testing &
diagnostics experience.

Microsoft

CLOUD UEFI BOOT FLOW



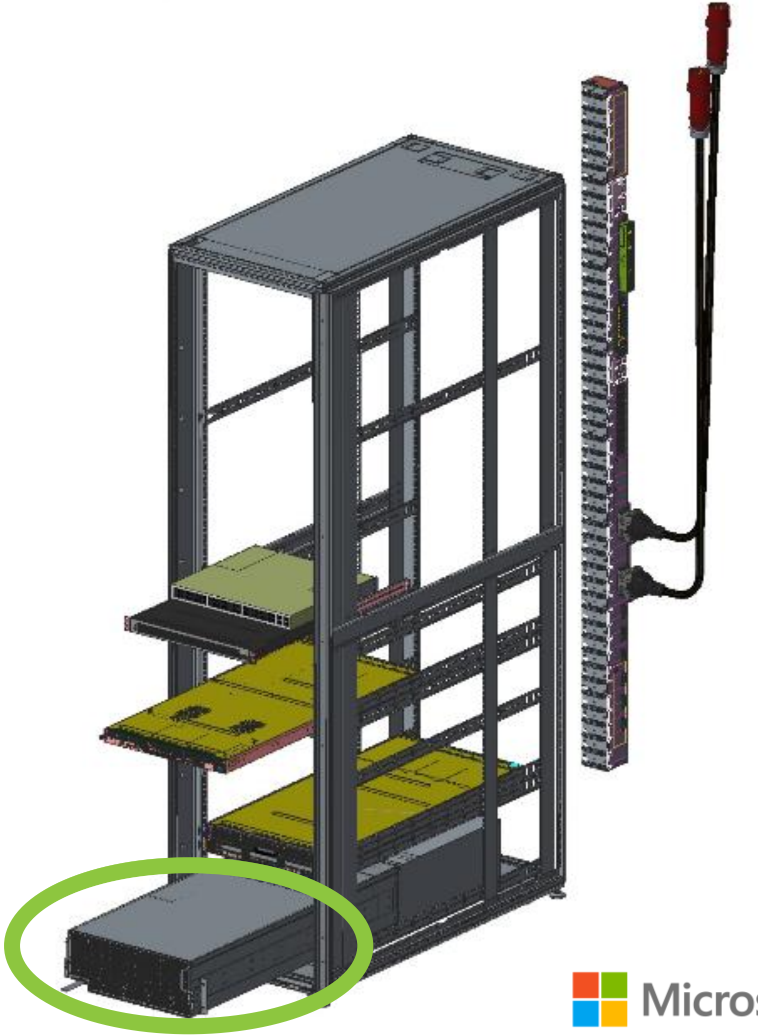
- CPU Fetches UEFI Code
- Cache as RAM
- CPU Init
- Interconnect Init
- PCH Init
- Memory Init
- MP Init
- PCI Init
- Platform Init
- RAS Features
- Standards Compliance
- Boot Target Selection
- Windows OS

Platform Discovery → Publish Resource Map → Runtime Services → OS Boot

SALIENT UEFI FEATURES

- Standards
  - UEFI 2.6+
  - PI 1.3+
  - SMBIOS 3.0
  - ACPI 6.1
  - TCG
  - TPM 2.0
  - PCI 3.0
  - USB 3.0 / USB 2.0
- Security
- Serviceability
  - IPMI
- Boot Target variety
  - Network, M.2, SATA, etc.
- Binary consolidation

PROJECT OLYMPUS
Expansion Chassis



Microsoft

Expansion Chassis

PROJECT OLYMPUS EXPANSION CHASSIS

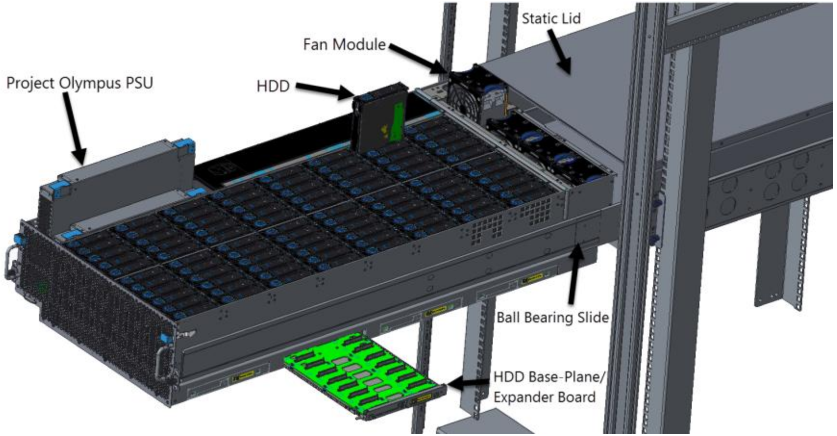
- DX-88 JBOD with 88 HDDs
- FX-16 JBOF with 16 SSD Carriers
- HGX-1 Hyperscale GPU Accelerator

Expansion Chassis

DX-88 JBOD

- 4U JBOD
  - 88 Hot-plug HDDs
  - N+1 Dual Rotor Fans
  - N+N Redundant, 1650W 3-Phase, Hot-Swap PSUs
  - Drawer design – slide out for service
- Robust Feature Set
  - Integrates into Project Olympus Infrastructure through Universal PDU
  - Runs on OpenBMC
    - Gathers HDD temps and component status info
  - Individual HDD on/off to minimize NTF

Quanta



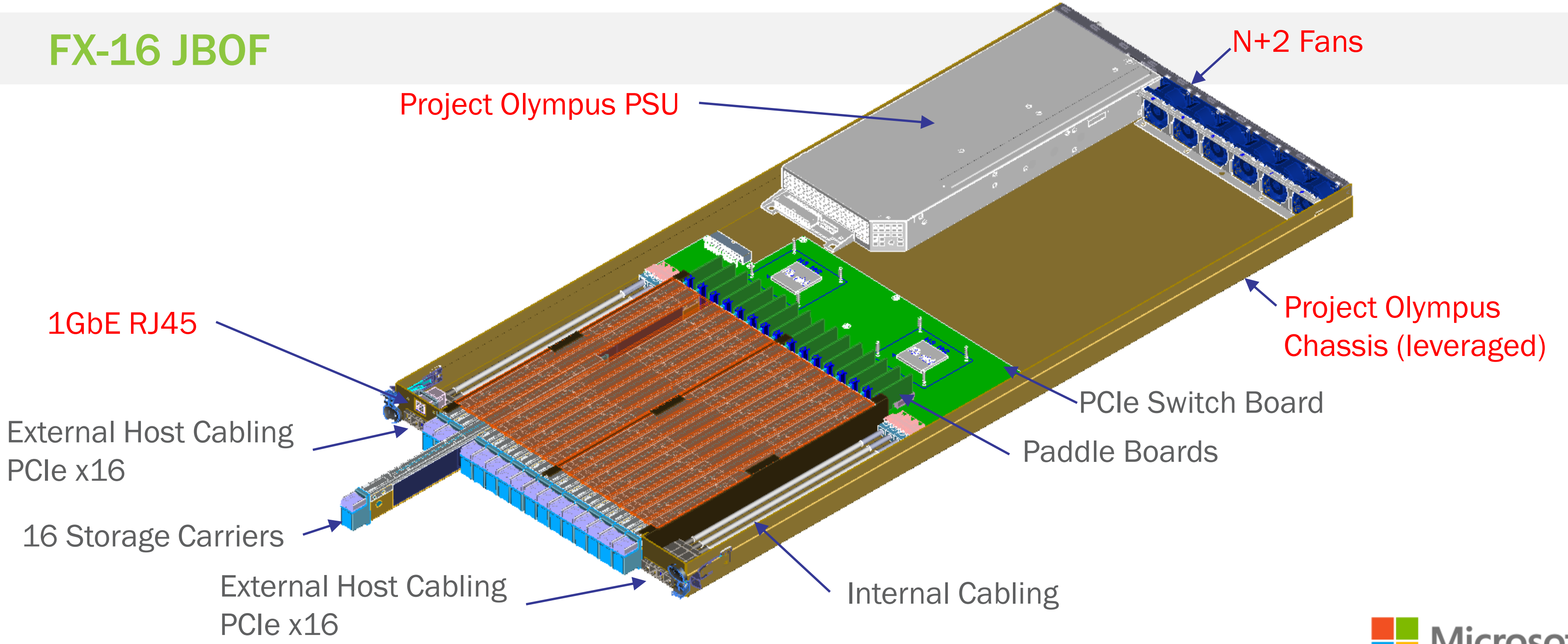
ARTESYN
EMBEDDED TECHNOLOGIES

Microsoft

OPEN Compute Project

Expansion Chassis

FX-16 JBOF



Microsoft

Industry-standard Accelerated CLOUD COMPUTING

PROJECT OLYMPUS HYPERSCALE GPU ACCELERATOR (HGX-1)

- Configurable and Flexible Accelerators
  - 8 x NVIDIA V100_SXM2 & NVLink
  - 8 x GPGPUs in PCIe Card Form Factor
- Expandable to Scale UP
  - From one to four Chassis
  - Internal PCIe Fabric Interconnect
- Scale Out via InfiniBand Fabric
- Host Head Node Options
  - 2S Project Olympus Server
  - 1S, 2S, 4S Server Head Nodes (eight x16 PCIe Links)
  - Up to 16 Head Nodes (sixteen x8 PCIe Links)

Summary

Cloud-scale, Cloud-speed!

- Azure servers diverse set of customers
- Efficiency is key at all levels

- At Cloud-scale, we need everybody's help!

- Early Contribution to OCP, encouraging feedback
  - Modular Architecture
  - Detailed specification
  - Promote Independent Collaboration and Interoperable Components

Microsoft

HOW TO DOWNLOAD

PROJECT OLYMPUS V1.0 CONTRIBUTIONS

http://www.opencompute.org/wiki/Server/ProjectOlympus

Microsoft

PROJECT OLYMPUS V1.0 CONTRIBUTIONS (NOV 2017)

Servers:

- Project Olympus Chassis Mechanical Base Specification
  - Design Package: http://files.opencompute.org/oc/public.php?service=files&t=09991252d3ecff77aa88027c68d075e9&download
- Project Olympus 1U Server Mechanical Specification
  - Design Package: http://files.opencompute.org/oc/public.php?service=files&t=a5dc716cb134a1c51905cc3411644b24&download
- Project Olympus 2U Server Mechanical Specification
  - Design Package: http://files.opencompute.org/oc/public.php?service=files&t=621dc0ddd20bf6bb287ccfa32a40279d&download
- Project Olympus Universal Motherboard Base Specification
- Project Olympus Intel XSP Motherboard Specification
  - Design Package: http://files.opencompute.org/oc/public.php?service=files&t=fa1435ef86ea3ab831bc1fc229e709e6&download
- Project Olympus Intel BIOS Specification
- Project Olympus AMD EPYC Motherboard Specification
- Project Olympus Cavium ThunderX2 ARM64 Motherboard Specification

Expansion Chassis:

- Project Olympus HGX-1 PCIe Expansion Chassis Specification
- Project Olympus DX-88 JBOD Specification
- Project Olympus DX-88 Power Supply Specification

Rack:

- Project Olympus Rack Base Specification
- Project Olympus Air Blocker Specification

Management:

- Project Olympus Universal Power and Management Distribution Unit Specification (PMDU)
- Project Olympus Rack Manager Specification
- Project Olympus Standalone Rack Manager Specification

Power:

- Project Olympus Power Cord Specification
- Project Olympus Server Power Supply Specification
- Project Olympus Power Supply Software Interface Specification
- Project Olympus PDU Specification

Microsoft

MARCH 2017 OCP SUMMIT

- Microsoft Project Olympus Overview - [video], [slides]
- Microsoft Project Olympus Servers - [video], [slides]
- Microsoft Project Olympus High Density Flash - [video], [slides]
- Microsoft Project Olympus Storage JBOD - [video], [slides]
- Microsoft Project Olympus Hyperscale GPU Accelerator Chassis (HGX-1) - [video], [slides]
- Microsoft Project Olympus Rack Management - [video], [slides]
- Microsoft Power capping in Project Olympus - [video], [slides]

GET INVOLVED (PROJECT OLYMPUS SUB-GROUP)

- Project Olympus is Microsoft's next generation rack-level solution that is open-sourced through Open Compute Project. (Video Introduction)
- The charter of Project Olympus sub-group within OCP Server Project is to enable the OCP community to further explore, invent, collaborate, enhance, and produce great solutions for customers using Project Olympus modular building blocks.
- Initially introduced in November 2016 and with V1.0 contribution in November 2017, Project Olympus addresses several cloud workloads for Microsoft Azure.
- Since inception, Project Olympus has attracted a large group of partners such as compute silicon providers, ODMs, OEMs, and component manufacturers. It is becoming the de facto open-compute standard for cloud workloads.
- Project Olympus base specification defines a modular architecture with clear internal and external interfaces. Hardware modules include Rack, Universal PDU, Rack Manager, 1U/2U Server and mechanical Enclosures, Power Supply, Universal Motherboard, PCIe Riser Boards, and Expansion Modules for storage and accelerators; while, Software/Firmware components include RESTful API, Rack Manager Software/Firmware interface, BMC Firmware, System BIOS/UEFI Firmware, and Software APIs.
- Project Olympus comprises these Hardware and Software Modules to realize a holistic rack architecture; however, individual Modules are applicable to Racks, Chassis, Rack Managers, PDUs, PSUs, Blades and Motherboards from other architectures such as OCP Open Rack, Open Rack 19, Scorpio, 19” EIA Rack, Rack-mount Servers, and Tower Servers.
- While, Microsoft is planning Design and Product implementations based on Project Olympus specification, we encourage the community to use as-is, buy/sell as-is, modify to use or sell, and to provide feedback on any of these software or hardware Modules.
- Find other material on Project Olympus Sub-group at Wiki Page.
- To get on the sub-group mailing list, go to here.

OPEN
Compute Project