Google

For OCP Discussions

Hyperstack for

Datacenter-ready Integrated System *(DC-Stack)*

For conversation at OCP forums

For OCP Discussions

Preface

Based on the current DC-SCM effort, our goal will include:

- Streamline the producer-to-consumer pathway
- Win-win: allow faster delivery of products into Hyperscaled, Enterprise, and Edge datacenters
- Reduce the complexity of providing a common mngmt and security infrastructure into datacenters
- Increase the value-add and diversity of compute, storage, and IO elements that the suppliers may deliver into the products that Hyperscalers may consume
- While driving a standard for the interface to the HPM, limit the impact to the HPM; allow different instances of DC-SCMs for one or many HPM types (either directly or via an Interface Board)

Use the OCP legal framework for multi-party CLA based on OWF

Each participant will contribute a portion:

Spec Chapters
Program management
PoC system
Software, firmware, testbench, ...

For OCP Discussions

Alignment with Other OCP Activities

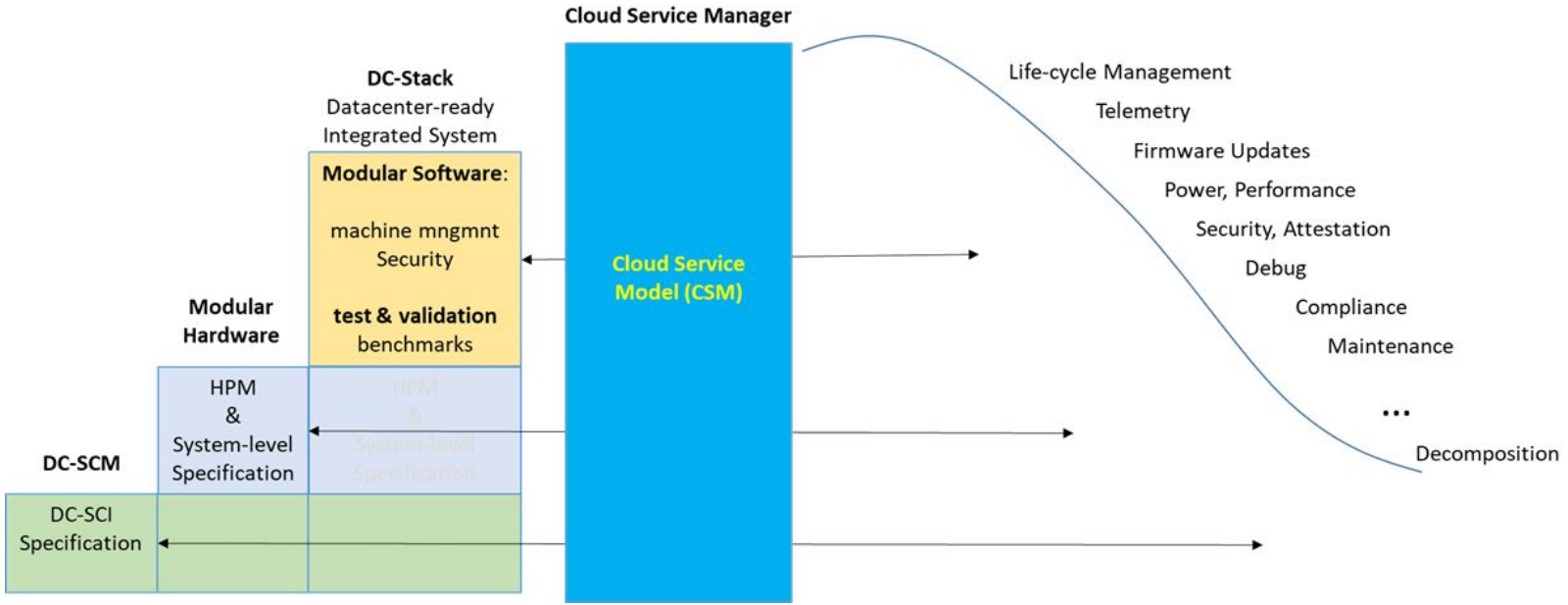
We will align this system-level activity with the foundation we are building within OCP at the module level and deliver an integrated solution for others' contribution at the datacenter level:

- **DC-SCM**: Starting with DC-SCM and DC-SCI specifications (underway)
- **Modular Hardware System**: The system around DC-SCM and HPM (about to start) and extend to Expansion Chassis such as storage and GPU/Accelerators
- **Datacenter-ready Integrated System** (the effort outlined in this document): Add Software and Security apparatus to the Modular Hardware System
- **Open System Firmware** (OSF)
  - Conforms to OSF 1.2 requirements to support owner control, circular economy
- **Security**
  - Implement "Gold" level Security as defined in the Security checklist
- **Test & Validation**: Accommodate Qualification and Certification (an effort has started on this topic within OCP)
- **Benchmarking**: Allow standard benchmarking
- **Cloud Service Model** Initiative: (formed from participants from Google, Microsoft, Facebook, Samsung, and Intel) Deliver the DC-ready Integrated System to the OCP Cloud Service Model (**CSM**) team for datacenter-level life-cycle management (the effort has just been started for 2021)

For OCP Discussions

Alignment with other OCP Activities

The following figure depicts where Datacenter-ready Integrated System (DC-Stack) falls within the continuum from DC-SCM through the datacenter-level Cloud Service Model initiative within OCP.



Hyperstack Vision:

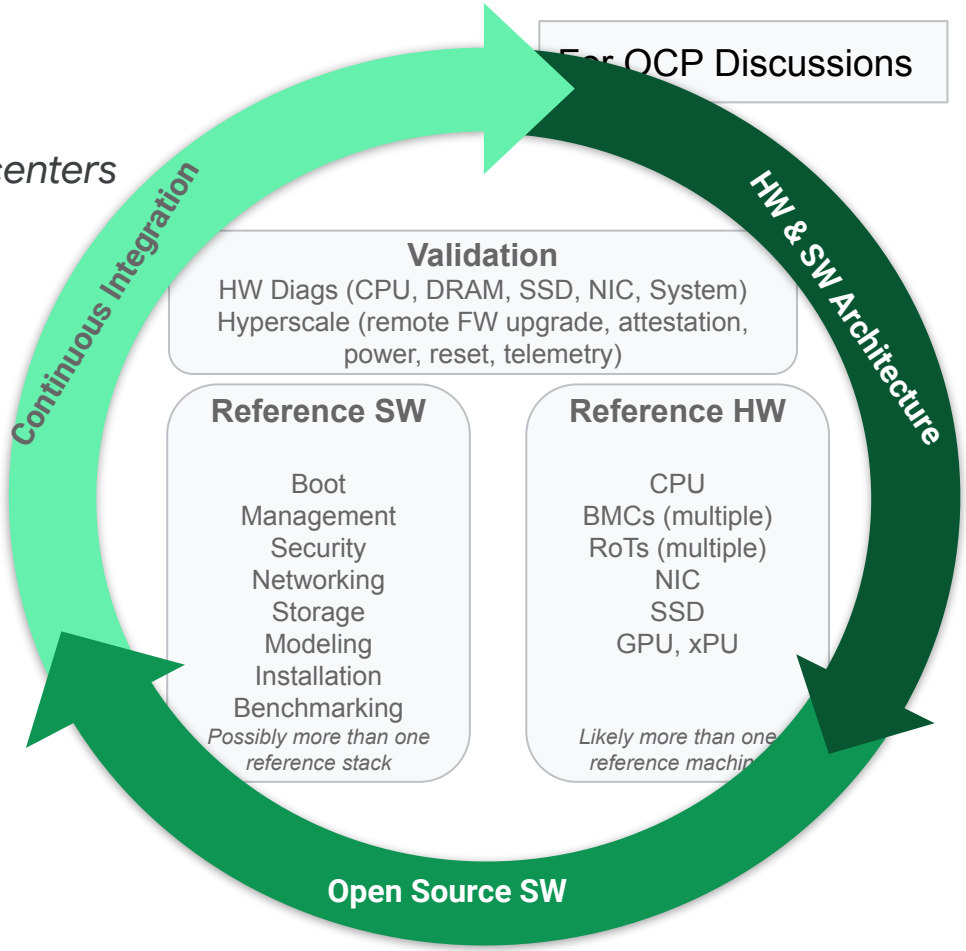
Streamline the producer-to-consumer pathways

Win-win: allow faster delivery of products into datacenters

Open ecosystem
Consumable by hyperscalers, testable by suppliers

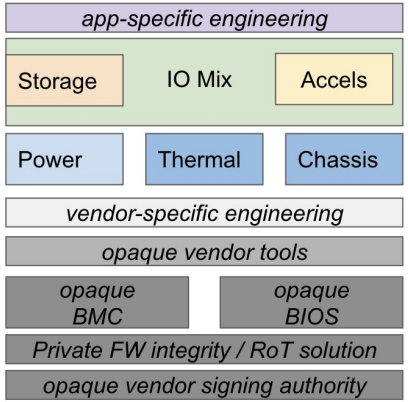
Requirements expressed as:

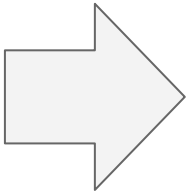
- **Modular hardware**, enabling a vendor to build a base solution for multiple hyperscalers
- **Modular software**, with open-source reference implementation
- **Validation suite** certifying satisfaction of End Customers

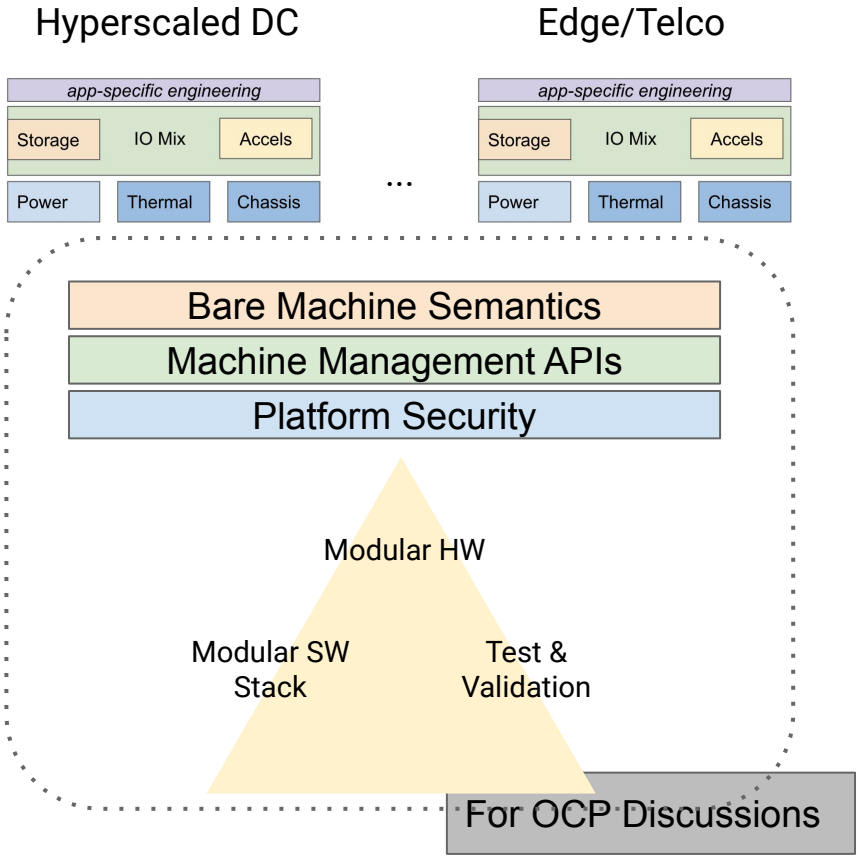


Hyperstack Flywheel

Scaling to Handle Diversity







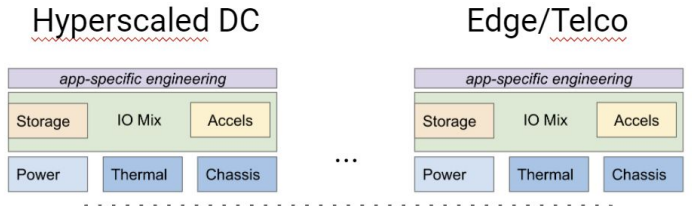
Context: Historically focused on the hyperscaled datacenter optimizations

- Vertically integrated, custom thermal, power, mech, security, machine management, SW stack
- Deliver maximal TCO/cycle at global DC scale, for internal and cloud customers

Scaling Across Internal Workloads, Cloud, Edge, 3P

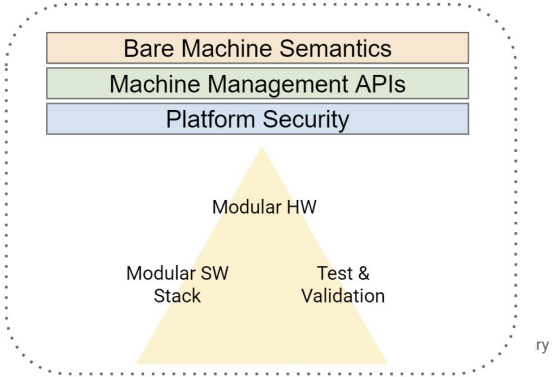
Enable diversity above the hyperscaled-optimized baseline

- Mech/thermal/power for Edge and Enterprize are different
- Different IO mix and flexibility: front vs back IO, disaggregation, etc..
- Machine size: mission-critical 8S/16S, DC 2S/1 (large) S, Edge 1 (small) S



Conquer with common baseline of requirements & reference implementation

- Bare Machine -- separate the customer from platform management
- Platform Security -- firmware integrity & control, physical protection of data confidentiality
- Machine Management -- telemetry & actuation for inventory and repairs



Google

ry

For OCP Discussions

Goals

- Develop hardware and software specifications for the **Datacenter-ready Integrated System**
  - Contribute the *Base Specification* to OCP

- Develop hardware and software **specifications** with key contributors
  - Develop a reference *Design Specification*
  - Common core: bare metal first, platform security, and machine management
  - Develop a reference *Product* for contribution

- Fully **engage OCP**
  - Take advantage of the OCP legal (e.g., CLA, JDA, OWFa 1.0, …) for multi-party collaboration
  - Develop reference hardware based on modular pieces, as a datacenter-ready integrated system
  - Drive the adoption of Hyperstack software to validate the reference implementation

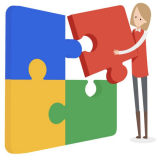
For OCP Discussions

Technical & Execution Details

Datacenter-ready Integrated System

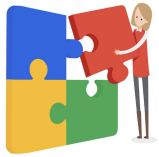
For OCP Discussions

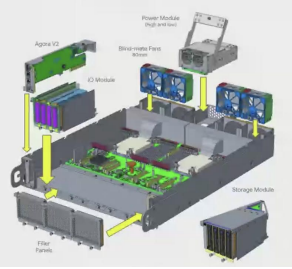
- Datacenter-ready Integrated System for Edge, Private Cloud, and Large Datacenters
  - HW, FW, SW, management, at-scale debug, security, and test & validation
- Built on successes within OCP efforts such as modular DC-SCM+HPM and OAM/OAI
  - Articulate one complete modular system (*The Base Specification*) for each solution category
  - Allow variations at each module (multiple *Design Specifications* based on the Base Spec)
  - Work with suppliers to build *products* (PCBA, Chassis, etc. based on Design Specs)
- Modular System (DC-SCM + HPM + DC-MIO + Modular Power)
  - DC-SCM (BMC, RoT, CPLD)
  - HPM (CPU/Memory/IO Slots)
  - Representative firmware for RoT and BMC (refer to the software strategy slide)
  - DC-MIO: Spec, cable/adapter prototypes
- Rack-level specifications (DC requirements: Mechanical, Power, Cooling, Weight, EMI, Acoustic, …)
- Rack Manager Interface
- Contribute a reference design
  - Mechanicals (new enclosure which fits Open Rack and 1RU/2RU Blades)
  - Generic motherboard requirements (not secret sauce!)
  - Contribute the Base Specification to OCP (generic system)
  - Suppliers will contribute Design Specifications and build Products



Hyperstack Hardware Modules:

Logical Blocks overlaid on Physical Blocks for a *Datacenter-ready Integrated System (**DC-Stack**)*





DC Environmental Requirements

Mechanical | Power/Cooling | EMI/Acoustic | At-scale Debug | Physical Security | Management

DC-SCM
OoB Control

BMC

RoT

Partners go here
Internal

e.g., HPE's iLo
Dell's iDRAC
Lenovo
… others

HPM (baseboard)
Compute

Security & Control sidebands
DC-SCI

USB / I3C / 1xPCIe

CPUs or GPUs, TPUs, xPUs

DIMMs

Partners go here

OCP tracks go here

1S, 2S, 4S, 8S CPUs
Xeon, EPYC, ARM64, …

xPU Expansion Chassis
… others

DC-MIO (modular IO)
Interconnect

Form Factors (details here)

Cables & Interfaces
DC-XPI

IO & Accelerators

NVMe requirements

RoT requirements

OCP tracks go here

SSD
IB NICs

Accelerators
… others

SmartNIC
Dataplane Control

SmartNIC

Partners go here
SmartNIC

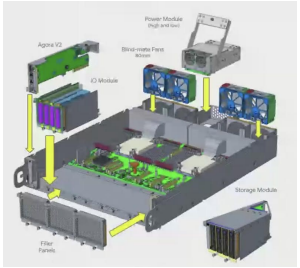
e.g., Blue Field, Stingray,
AWS Nitro,
MSFT FPGA,
… others

For OCP Discussions

Modular Hardware

Reference designs for three different architectural instances

1S/2S Server



Rack
Power
Cooling
Multi-Blade/Instance Chassis
    HPM (CPUs + Memory) per Blade
Multiple DC-SCMs or Multi-Host DC-SCM
DC-MIO
IO Module/Cage (IO Slots)
Multiple SmartNICs or Multi-Host SmartNIC

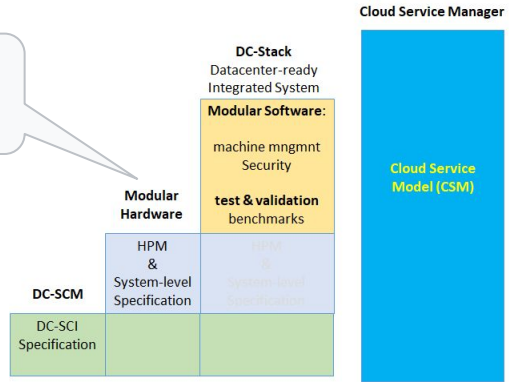
Multi-CPU Server

Rack
Power
Cooling
Single Instance Chassis
    Multi-HPM (CPUs + Memory)
IO Module/Cage (IO Slots)
Single-Host DC-SCMS
DC-MIO
Single-host SmartNIC

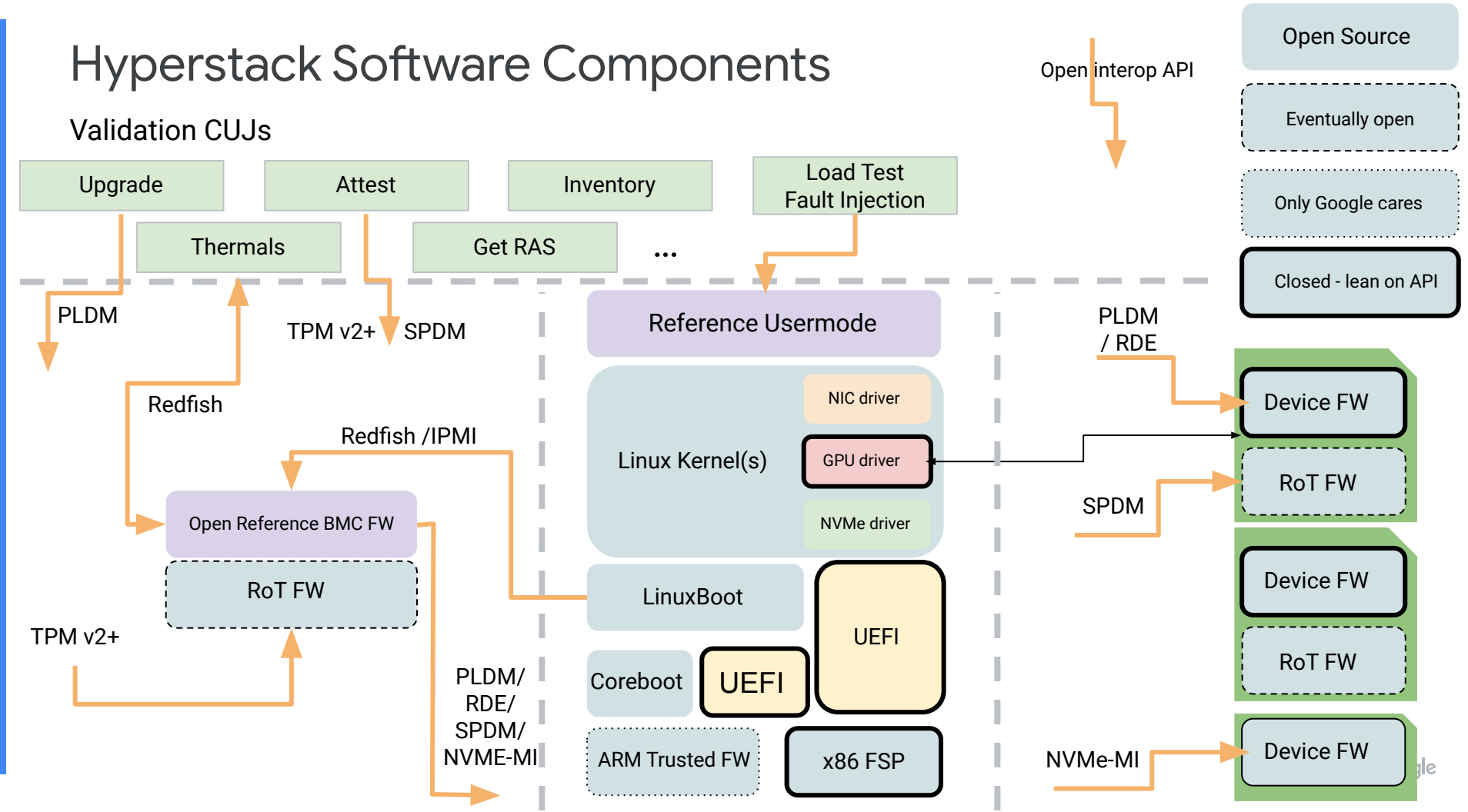
GPU/Acc. or Storage Expansion Chassis

Expansion Chassis, Power, Cooling
Head Node (Server)
Interconnect (cables, retimer?)
Out-of-band Management
BMC (or a variant of DC-SCM)

The Hardware Portion



Hyperstack Software Components



Hyperstack Compliance Suite

For OCP Discussions

Upgrade	Remote Firmware Upgrade/Downgrade (PLDM)					
Thermals/Power	Voltage/Current	Fans/Liquid Cooling	ASPM State Test	Reset/Reboot Tests		
Get RAS	Memory Error Reporting	Soft Repair Support	PCIe AER Support	Platform MCE->BMC	SEL Persistence/BMC Resiliency	
Attest/Security	Verify Locked/RO Firmware	Disk Lock/Unlock Encryption	Remote Attestation Running vs. Static	BIOS/BMC “Fuzzer”	CPU Feature Checks (BootGuard, AMD HVB, disable DCI/Tap, etc)	DXE Inventory/Checks
Inventory	Redfish BMC Inventory (CIM)	FRU Reporting Compliance (CIM)	SMBIOS Device/CPU/Slot/Chassis/DIMM Checks	IPMI Power/PnP/CPU Interlink Checks		
Error Resilience/ Recovery	Kernel Panic/kexec/kdump	PCIE Error Injection	DIMM Error Injection	Firmware Attestation Recovery		
Performance & Load Test	Core Freq/Thermal	Memory Bandwidth/Latency	PCIe Performance	DMI Performance	CPU Interconnect Bandwidth/Latency	Storage Performance
	Network Performance	CPU Performance				
Platform Features	Power Stepping	PCIe Feature Set	RTC Jitter/Wander	Multi-OS Boot	NV Boot Count	Kexec compliance test

Hyperstack Program Workstreams

For OCP Discussions

Reference Software Modules

Boot Stack:
coreboot,
LinuxBoot
U-root

Vendor &
community
adoption

OOB Stack:
OpenBMC

NVME-MI
PLDM
OOB RAS

Security Stack:
Rafter

SPDM, TPM
(AMD PSP)
(Intel S3M)

Modeling, Simulation:
TBD

Google or
Other
Contribs

Test: TBD

Hyperscale
CUJs, Infra,
Tests

Also, SSD & SmartNIC Reference Modules

OCP & Hyperscale Contributors

Ref Modules

DC-SCM
DC-MIO

Ref System

Server
& More

Ref Image

Kernel/Driver
Automation

Validation

Fleet, CI,
System Test

Standards

DMTF

Redfish
SPDM
MCTP

PCI-SIG

Gen5
USB 2.0
sideband

NVMe

NVMe-MI

TCG

TPM 2.0
DICE