
2. [15] What’s next for open hardware standards—intentions, call for participation, Community involvement and coordinate

3. [30] Commercial products in the OCP pipeline

4. [30] Panel Discussion on Telco/Operator sourcing models and ecosystem

5. [30] Updates from community: POCs, deployments, and disaggregation
1. Carrier Grade Open Rack Architecture (CG-OpenRack-19)
CG-OpenRack-19 Achieves OCP Acceptance

A collaborative community focused on redesigning hardware to efficiently support the growing demands of compute infrastructure.

Radisys contributed the Carrier Grade Open Rack concept to OCP in the form of a Rack + Sled interop specification.

DCEngine is a commercially available product family compliant with this specification.

CG-OpenRack-19 Specification

OCP-ACCEPTED™

CG-OpenRack-19 Specification
CG-OpenRack-19 High Level Architecture

Switching

**Usable Compute / Storage Capacity**

Power

Standard 19” Rack

**Vertical 12VDC bus bar in frame mates with power connector located on sled**

**4 x optical fiber ports via blind mate rear connector to sled**
**Physical**
- Suitable for CO retrofit and new telco data center environments
- 19” rack width and standard “RU” spacing for greatest flexibility
- 1000 to 1200mm cabinet depth, supporting GR-3160 floor spacing dimensions

**Content/workload**
- Heterogeneous compute and storage servers

**Management**
- Ethernet based OOB management network connecting all nodes via a TOR management switch
- Optional rack level platform manager

**Networking/Interconnect**
- One or more Ethernet TOR networking switches for I/O aggregation to nodes
- Fiber cables, blind-mate with flexible interconnect mapping.
- Environment, power, seismic & acoustic CO environmental requirements applicable
- Safety and other certification standards also applicable
- NEBS optional (L1/L3)
OCP Design Principles: OpenRack

• **Open**
  Community-driven; Multi-vendor; No lock-in; Fast-moving

• **Efficient**
  Performance optimized for IT data centers; Simple core building blocks; Power and thermal efficiency

• **Scale**
  Web-scale ready; Simple management & maintenance; Mass upgrades

Disaggregates and Normalizes Web-scale Computing
OCP Design Principles: CG-OpenRack-19

- **Open**
  - Open spec and designs starting from OCP baseline
  - Multi-vendor and multi-user collaboration from day one
  - Aligns with existing standard telco and COTS geometries and interfaces
  - Support for heterogeneous and accelerated solutions via standard plug-in cards

- **Efficient**
  - Inherits key OCP principles
  - Performance optimized for CO data center environment
  - Self-contained sleds for thermal and emissions isolation
  - Half-rack sled width well suited for brawny server designs across multiple processor generations

- **Scale**
  - Leverages OCP web-scale principles
  - Standard blind-mate optical interconnect for faster build-out, maintenance and multi-generational upgrading

Brings OCP to Service Providers, Tracking but Decoupled from Web-co driven changes
What does OCP-ACCEPTED™ status mean to me as a Service Provider?

• Break Open the Black Box of Proprietary Infrastructure
• Gain Control and Choice
• Reimagine the Hardware and Software
• Make Solutions More Efficient, Flexible and Scalable
• Customize
• Save $
OCP CG-OpenRack-19 Status and Next Steps

• Framework/Interop Specs
  • Current spec focuses mainly on sled-level interop, which is most critical for supplier ecosystem development; next focus on Rack and Management aspects
  • Updating of specs as new innovations take place in community

• Product Contributions
  • Vendors contributing DCEngine product designs, including rack, compute, and storage sleds
  • See later section in Workshop Agenda for more details

• OCP Events
  • Sessions at Summit (March, Santa Clara): “Delivering Carrier-Grade OCP to Telco Data Centers” and “Hardware Management for Radisys DCEngine Hyperscale Platform”
  • Sessions at this Workshop (May, Austin): Ecosystem and sourcing model focus

• Ecosystem Incubation and Promotion
  • Multi-vendor ecosystem in use in current solutions
  • Expanding to include more options
  • Encouraging new participants to expand market footprint
  • Customers also key part of ecosystem
2. What’s next for open hardware standards
• **Management**
  • See following (subset of) presentation from 2017 OCP Summit

• **Rack**
  • Product contributions for various sized racks
  • Potential area for some basic normalization across solutions – i.e., via framework specs
Hardware Management for CG-OpenRack-19

Suzanne Kelliher, Product Line Manager, Radisys
Nilan Naidoo, Principal Engineer, Radisys
**CG-OpenRack-19 Hardware Management Overview**

- **Create Cohesion Across CG-OpenRack-19 Implementations**
  - Leverage existing HW management standards: IPMI 2.0, DCMI 1.5 and Redfish
  - Each node is independently managed by BMC
    - Includes cooling of shelf containing the node

- **Leverage OCP hardware management premise**
  - Leverage existing HW management standards: IPMI 2.0, DCMI 1.5 and Redfish
  - Each node is independently managed by BMC
    - Includes cooling of shelf containing the node

- **Add Options as Necessary for Simple, Efficient Rack Management**
  - Device Management switch can be used to run Rack Management applications
    - Example, Location Aware Discovery
  - Rack Agent Module provides access to PSU & PDU, and additional physical security features, i.e. door locks

- **Options for Rack Management**
  - Provide basic rack level management using Redfish API based on open sourced Intel® RSD framework
  - Intel® RSD Architecture Compliant

- **Connects to dedicated BMC port on each node, Rack Agent & Management port of other switches**
- **One uplink out of rack provides OOB management access to all devices in rack**
- **Open Linux environment enables Rack Level Management applications**

- **Shelf HW Management provided by Server BMC**
  - FRU Inventory
  - Sensor Data
  - Power on/off/reset
  - Power consumption
  - Boot order control
  - Remote Console (SOL, KVM)
  - Virtual media
  - Front Panel Indicators
  - Interfaces: IPMI 2.0, DCMI 1.5, Redfish

- **Rack Agent provides Ethernet access to PSU & PDU**
  - Abstracts PSU & PDU management standard interface (IPMI, Redfish, SSH CLI)
  - PSU & PDU Inventory
  - Rack level power
Intel® RSD is a logical architecture that disaggregates compute, storage, and network resources

- Introduces the ability to pool these resources for more efficient utilization of assets
- Provides the ability to dynamically compose resources based on workload-specific demands from a set of compute, fabric, storage, and management modules that work together to build a wide range of virtual systems

The design uses four basic pillars:

- POD Manager for multi-rack management
- Pooled system of compute, network, and storage resources are composed based on workload requirements
- Pod-wide storage built on Ethernet-connected storage
- A configurable network fabric of hardware, interconnect with cables and backplane, and management software

Intel RSD based on open industry standard Redfish*

Intel has open sourced reference implementation of following components:

- Pod Manager
- Pooled System Management Engine (PSME)
- Rack Management Module (RMM)
- Validation Test Suite (VTS)

Source code: https://github.com/01org/intelRSD
A key attribute of Intel® RSD management is location-aware discovery
- A mechanism for numbering each component is required

Each Rack has a unique ID
- Configured by operator

RSD defines a 3 level hierarchy for modeling computer systems
- Drawer – maps to a shelf
- Module – logical entity
- Blade – maps to server motherboard

Numbering scheme for blades in a rack:
- `<Drawer Row>.<Drawer Column>.1.<Blade Id>`
Extended RSD PSME reference code to run on Device Management switch
- Extended Chassis and Compute GAMI IPMI interfaces to interact with BMC
- Extended Network Agents to run on Cumulus Linux on Data switches

Added Location Aware Discovery application to discover and determine blade locations
- Monitors switch ports to determine presence/absence of devices in the rack
- Uses Port-to-Device Mapping configuration file to map learned MAC addresses to Blade & Switch location
  - MAC -> Port -> Location
- To overcome limited visibility of blade inventory through IPMI, uses a configured server device tree file for each Product Id
  - Server device tree file describes list of components (CPU, Memory, Drives, etc.)

PSME Interfaces to Location Aware Discovery application through API
- Retrieves BMC parameters
- PSME will use contents of device tree file to fill in information not accessible via IPMI
- Listens for device state changes
Rack Agent Architecture

- Rack Agent module consists of a Controller module following I/O:
  - I2C interface to interface to PMBus
  - Ethernet Interfaces for uplink to device management switch
  - Serial console for debugging & initial setup
  - GPIO signals to monitor PSUs and Circuit Breakers on PDU
  - Other sensors required to monitor health of the module
  - OpenBMC is a good fit

- PSU/PDU Management
  - Presence & Inventory info of PDU & PSU
  - PSU Input and Output Voltage/Current
  - PDU Circuit Breakers
  - Temperature
  - Fan speed & status
• **Discovery**
  • Chassis
  • Computer systems
  • Managers

• **Server Information**
  • Server identification and asset info
  • Host Network MAC addresses
  • Local storage
  • Power supply and fans
  • State and Status

• **Common Manageability**
  • Change boot order / device
  • Reboot / power cycle server
  • Power usage and thresholds
  • Temperature

• **BMC Infrastructure**
  • View / configure BMC network settings

• **Access and Notification**
  • Subscribe/publish event model
• To provide cohesion across CG-OpenRack-19 implementations

• We are considering contributing the location aware discovery application and Intel® RSD enhancements
  • It enables basic hardware management of rack using Redfish

• Please join us on in the Radisys booth to see DCEngine and see a demonstration of this work.
3. Commercial products in the OCP pipeline
• Radisys
• ADLink
• Others – roundtable; general call for inputs, for CG-OpenRack or any OCP
• **DCEngine – NFVi for Hyperscale DCs & COs**
  - Pragmatic NFV and OCP deployment initiative of carrier networks
  - Ready for full NEBs

• **Carrier-grade Environmental**
  - Seismic rack
  - Extended operating temperature
  - Certified w/ EMC, EMI and CO safety requirements
  - High capacity cooling while minimizing noise
  - -48V and 400V DC power options
Inspirited first mainly due to docs readiness; Accepted contribution to follow

- **Racks**
  - 42U DCEngine Rack
    - DCE-RACK-V2-3-MM01
    - DCE-RACK-V2-3-MM02
  - 16U DCEngine Rack
    - DCE-16U-V2-3-MM01

- **Sleds**
  - ½ Wide Compute Sled
    - DCE-CSLED-V2-3-001
    - DCE-CSLED-V2-3-002
  - Full Wide Storage Sled
    - DCE-SSLED-V2-3-001
    - DCE-SSLED-V2-3-002
• **Rack Core**
  - 600mm & 800mm wide rack options
  - Power \( \rightarrow \) 110/208VAC 3ph & 230/400VAC PDU
    - 3 PSU shelves provides 12 x 2500W PSU’s
  - Management Switches (x2)
    - Switch #1 : Connects 1G to each server BMC
    - Switch #2 : Connects 1G to each server CPU
  - Data Switches
    - 1 or 2 switches (up to 3.2 Tbps each)
    - 40G uplinks to spine switch, 10G downlink to each server
    - Option for 100G uplinks & 25G downlinks (v2.3)

• **Standard Configurations**
  - Balanced : 8x Compute (16 sleds) + 8x Storage
  - Storage : 16x Storage Shelves
DCEngine 16RU Rack Core

- **16U Rack Core**
  - 600mm wide x 1000mm deep
  - Single phase AC power
    - PSU shelf with 4 x 2500W units
  - Management Switches
    - Switch #1: Connects 1G to each server BMC
  - Data Switches
    - 1 or 2 switches (3.2Tbps each)
    - 10/40/100G uplinks, 10/25G downlinks to sleds

- **Standard Configurations**
  - 4x compute shelves + 2x storage shelves
DCEngine Sleds

• **Half width compute sled**
  - 2 x dual socket server boards per sled
  - 2 x E5-2600 v4 series CPU per server
  - 16 DIMMs per server (16GB, 32GB, 64GB)
  - 512GB SSD boot flash per server
  - 2 x 2TB SSD per server
  - 10G, n x 10G & 25G NIC options

• **Full width storage sled**
  - 1 x dual socket server board
  - 2 x E5-2600-v4 series CPU per server
  - 16 DIMMs per server (16GB, 32GB, 64GB)
  - 16 x 3.5” SAS drives (160TB)
  - 512GB boot flash, 2 x 2TB SSD
4. Panel Discussion on Telco/Operator sourcing models and ecosystem

NOTE: The panel discussion did not take place at this meeting – it was deemed to be meaty enough to have a separate session, timing TBD.
Panel discussion on Operator Sourcing Models & Ecosystem

• **Suggested Topics:**
  - Do operators want to source from a single integrator or individually from component providers – and at what level of granularity?
  - Do operators want to negotiate directly with ODMs? Silicon providers? Is price negotiation separate (more disaggregated) from procurement/deployment?
  - Expectations on margins (and cost reductions) over time – for initial POCs/deployments, small deployments, large deployments
  - Which is the bigger driver: Opex or Capex? Can a new architecture win with higher initial Capex but lower Opex and TCO?
  - What projects (and what part of network) is the best candidate for change? Are there different procurement orgs for different areas – e.g., Access/Edge/Core/Cloud?
  - Who makes technology choices – and at what level (silicon/component, boards, sleds, racks, etc.)?
  - How do tech choices translate to projects and deployment? (e.g., science projects vs. deployments)
5. Updates from community: POCs, deployments, and disaggregation
How to stand up a 600 node bare metal Mesos cluster... in two weeks

Craig Neth
Distinguished Engineer -- Architecture & Infrastructure
Verizon Labs

- PAAS services – Logging, Monitoring, “External” networking, Storage
- HW – Radisys DCEngine w/ 4x switches, 10x storage sleds, 10x compute sleds (~50 CPU sockets + ~1PB storage)
- SW – CoreOS, Cumulus, Ansible, Mesosphere, EMC ECS/ScaleIO

See details in presentation here:
OCP in CORD POCs

- **CORD** – Central Office Rearchitected as a Data center
  - ONF/ON.Lab

- **Flavors of CORD**
  - R-CORD – Residential (PON)
  - M-CORD – Mobile access (4G/5G)
  - E-CORD – Enterprise (Wavelength Services)

- All use a common infrastructure
  - Edge compute on OCP based systems

- Multiple POCs at carriers globally

- Partnerships with hardware
  - “Whitebox” Open-OLT & Micro-OLT
    - per AT&T contributed OCP specs
  - Traditional vendors like Calix
R-CORD Software to Hardware Mapping

Head Nodes

Compute Nodes

XOS
CORD 1.0

ONOS
CORD 1.0

vRouter
CORD 1.0

OpenStack
OpenStack Kilo

vSG
Pre-built Ubuntu image

vCPE
Docker image

OVS
Open Virtual Switch

vOLT
Open Stack

64bit Ubuntu

System Front

System Back

ONT, Splitter, CPE (WiFi)

Fully functional multicast in POC
R-CORD Control and Data Plane Mapping

R-CORD Functional Blocks
Running on OCP CG-Openrack Compute
R-CORD Example Rack Configs

42U 10K-Sub Micro-OLT GPON, 100GE

42U 10K-Sub Micro-OLT XGS-PON, 100GE
- Edge compute is very important in 5G
  - Very low latency doesn’t allow for backhaul of all traffic
  - Hardened OCP is key
- Several POCs in tier 1 carriers beginning
- M-CORD is still nascent but carriers are interested because it meets 5G needs
• CenturyLink used a Radisys OCP POD for MEF16 POC

• Ciena Blue Planet Service Orchestrator and two Domain Controllers from Ciena and RAD

• Original plan was to use compute and storage sleds but the compute sleds provided enough capacity that the entire POC was run on one sled

• Won best demo of show
Thank You