Open Compute at CERN

HEPiX

21/05/2014

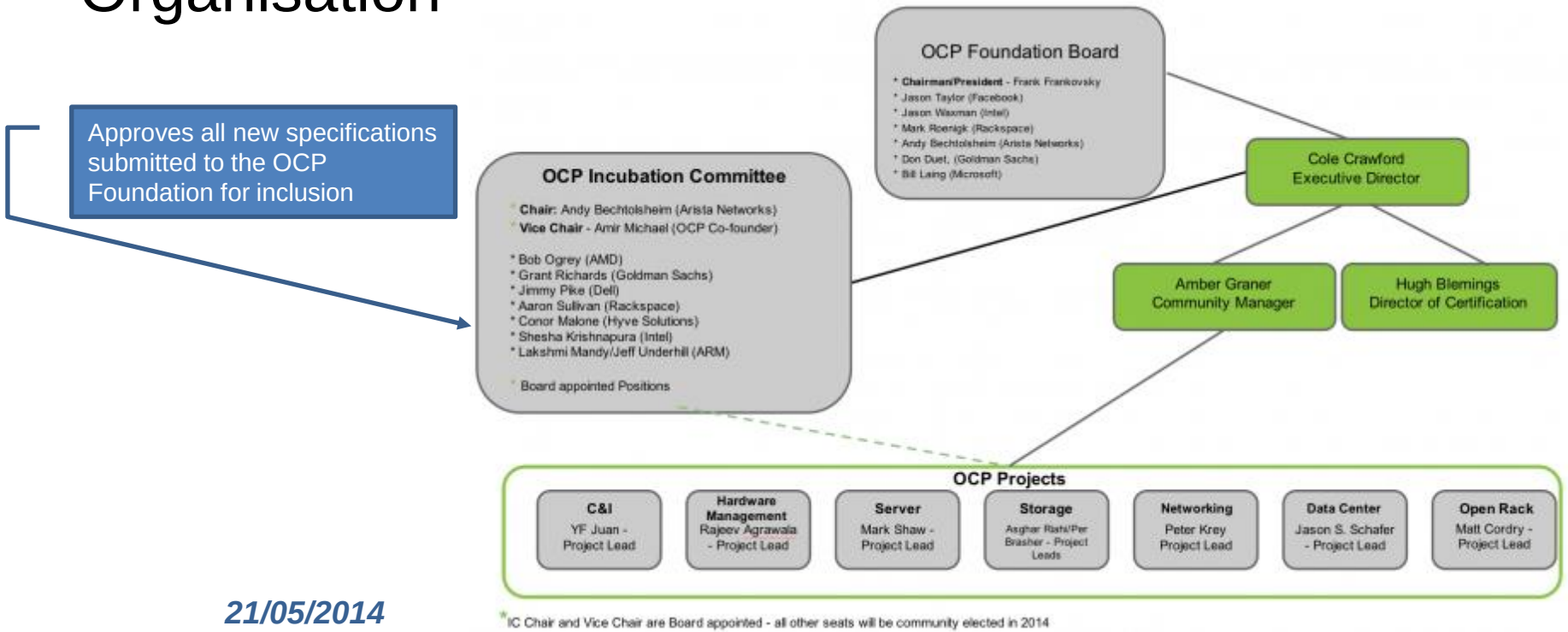
Olof Bärring, Marco Guerri – CERN IT

Outline

- What is Open Compute Project (OCP)
- Why OCP
- Initial OCP tests at CERN
- Benefits & issues
- Open Rack
- Plans
- Conclusions

What is OCP

- www.opencompute.org
- Started by Facebook in 2011 when building their first own Data Centre
  - *Design and enable the delivery of the most efficient server, storage and data centre hardware designs for scalable computing*
  - Open hardware design http://www.opencompute.org/wiki/Motherboard/SpecsAndDesigns
- Organisation



Four corporate membership
- Community (free)
- Silver
- Gold
- Platinum

All but community assume a fee or work time contribution

Why OCP

- Our main motivations for a closer look are potential benefits from
  - Energy efficiency
  - Platform and infrastructure standardization
  - Economy of scale generated from existing customers
  - Ease serviceability

"38% more efficient"

"24% less expensive to build"

	Other webi-class servers					OCP v1 Repairs					OCP v2 Repairs				
Primary Repair Types	Pre-repair activities minutes	Part swap duration minutes	Additional Steps minutes	Post-Repair activities minutes	Total Repair Time	Pre-repair activities minutes	Part swap duration minutes	Additional Steps minutes	Post-Repair activities minutes	Total Repair Time	Pre-repair activities minutes	Part swap duration minutes	Additional Steps minutes	Post-Repair activities minutes	Total Repair Time
1 Hard drive (nonraid)	2:00	3:00	-	2:00	7:00	0:00	0:58	-	0:19	1:18	0	0:51	-	:19	1:11
4 DIMM (offline)	2:00	3:00	-	2:00	7:00	0:00	0:49	-	0:19	1:09	0	1:00	-	:19	1:19
6 CPU (offline)	2:00	7:00	-	2:00	11:00	2:00	4:00	-	0:19	6:19	2:00	2:15	-	:19	4:34
8 Motherboard	2:00	20:00	20:00	2:00	20:00	2:30	10:24	2:30	0:19	15:44	2:00	9:13	-	:19	11:33
10 PSU (swappable)	2:00	5:00	-	2:00	9:00	0:00	0:39	-	0:19	0:58	0	1:45	-	:19	2:04
24 Fan	2:00	2:00	-	2:00	6:00	0:00	1:24	-	0:19	1:43	0	2:24	-	:19	2:44
AVG total repair duration	2:00	6:67	20:00	2:00	14:00	1:15	3:04	2:30	0:19	4:20	2:00	3:33	-	:19	4:30

www.opencompute.org/blog/facebooks-perspective-on-serviceability-and-operational-efficiency

CERN Meyrin	April 2014
Tasks	**Nb. interventions**
disk	127
mainboard	5
memory	12
PSU	18
RAID controller	1
BBU	3
Total interventions	**172**

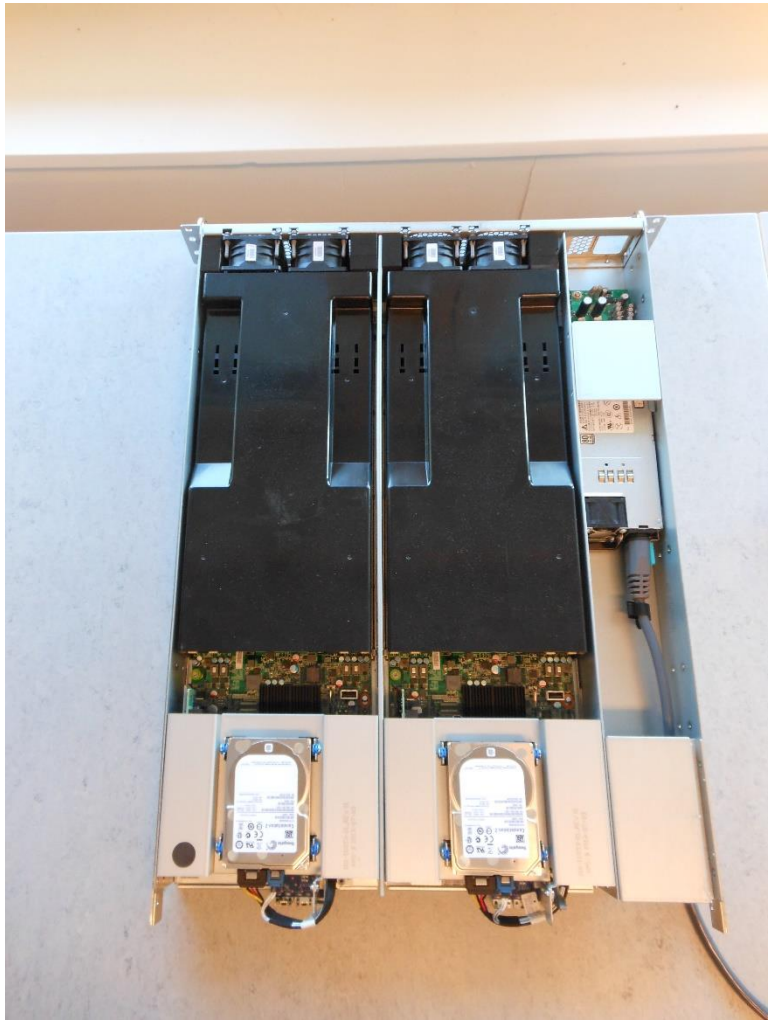
Initial OCP tests at CERN

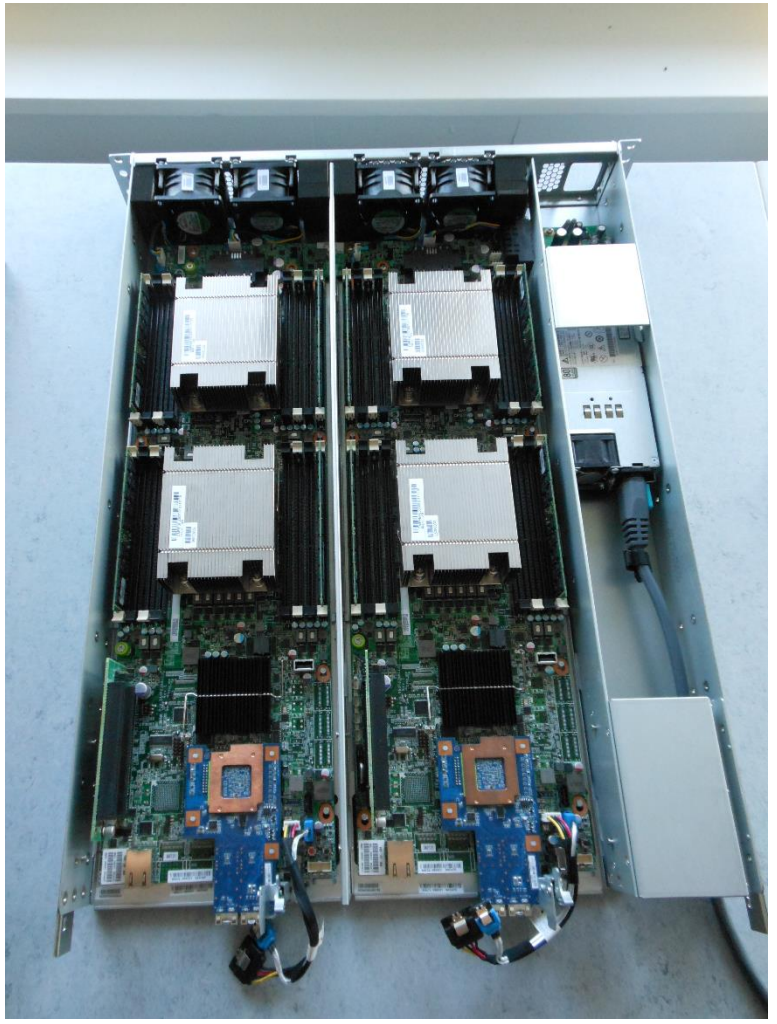
- Summer 2013: acquired two Hyve-1500(*)
  - OCP twin server in 1.5U enclosure for 19” rack
  - Comparable spec to recent deliveries

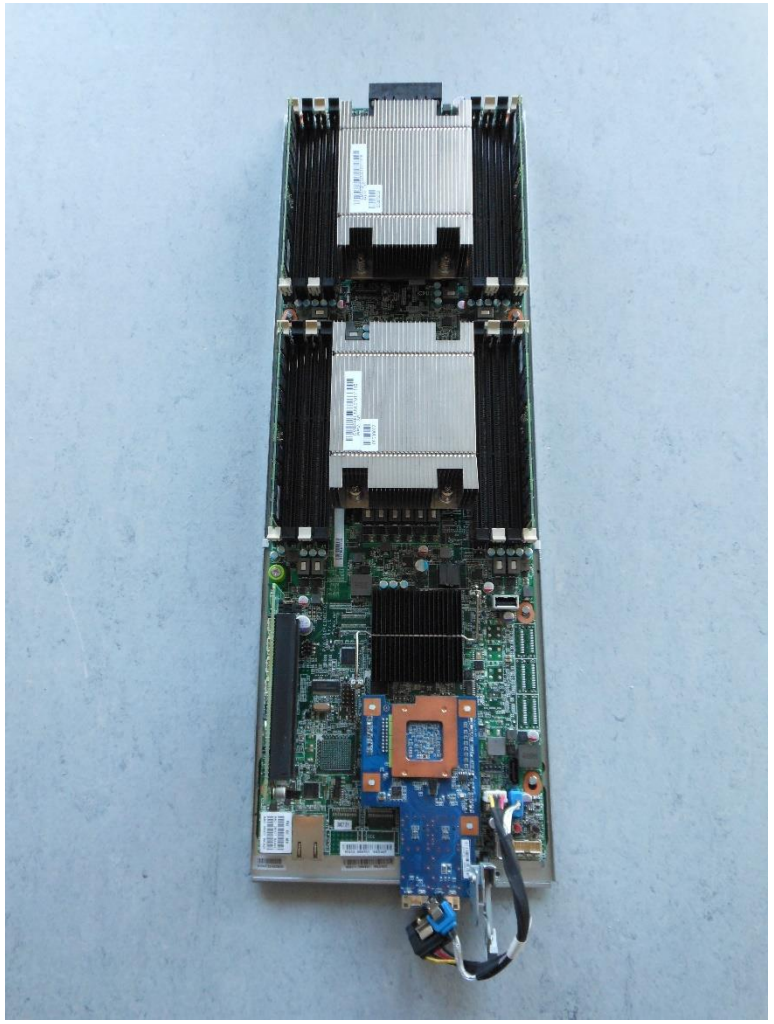
	Hyve-1500	**S2600JF (Intel Jefferson Pass)**
Chassis	1.5U Twin system	2U Quad system
PSU	1	2 (1+1 redundant)
CPU	2x E5-2650 (SnB)	2x E5-2650 (SnB)
Memory	64GB (8x 8GB)	64GB (8x 8GB)
Local storage	1x 2.5” 1TB HDD or 1x 2.5” 480GB SSD	2x 2TB HDD
Network	1GbE + 10GbE mezzanine	1GbE + 10GbE mezzanine

(*) http://hyvesolutions.com/resources/docs/2013Hyve1500Datasheet.pdf

Hyve-1500 twin



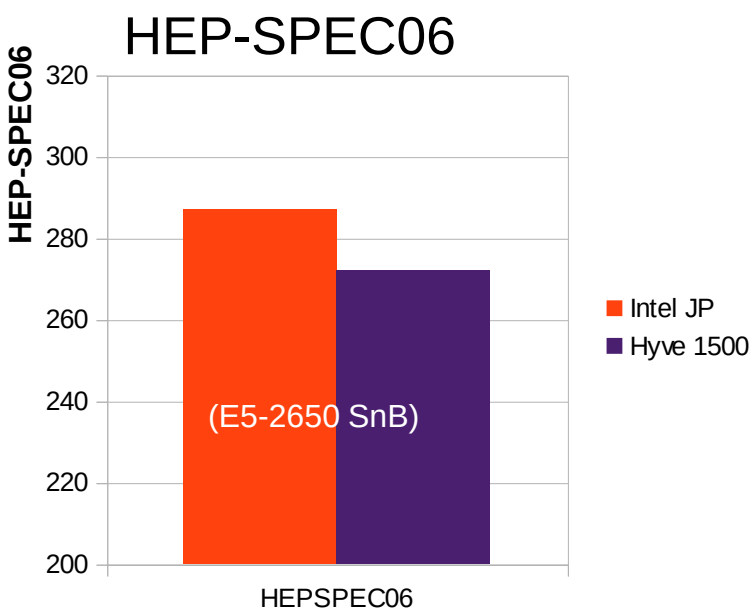


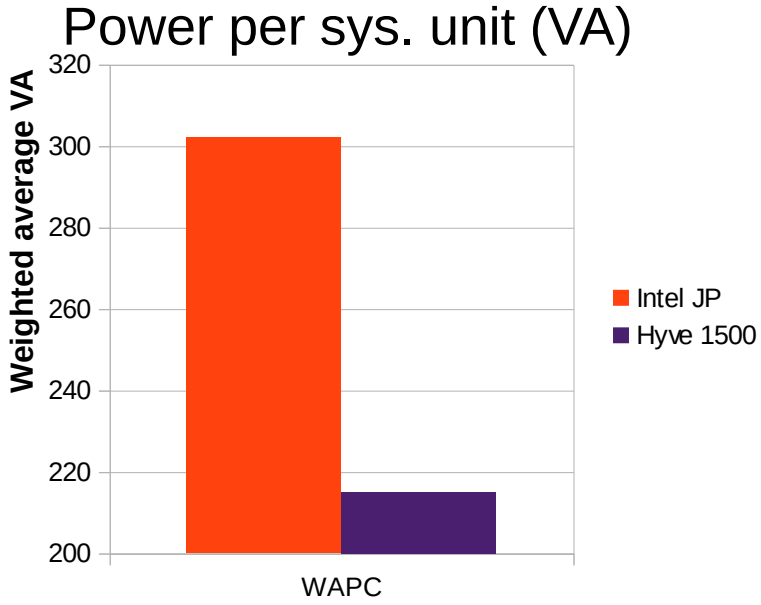


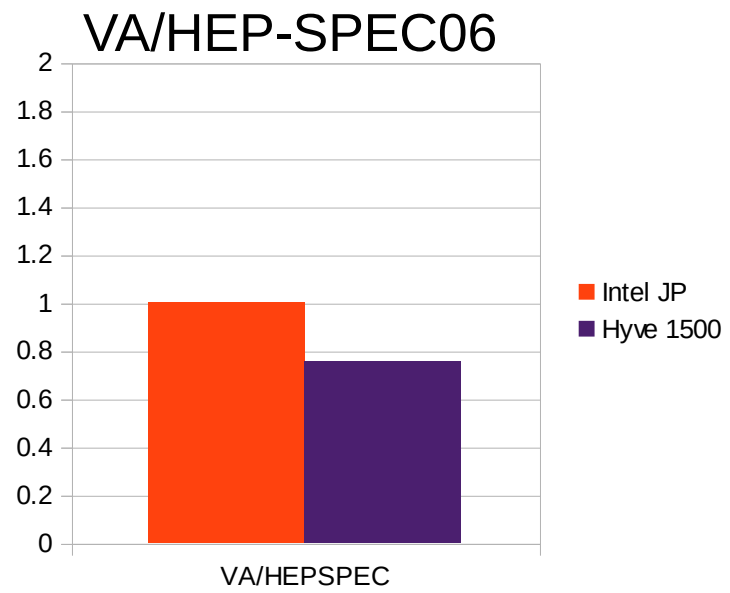
Hardware features

- Hyve-1500
  - Single 2.5” drive possible
    - 1xTB
  - Console: on debug header
  - Single PSU for two blades
  - No BMC!
    - C6xx, Management Engine Firmware
    - New set of drivers/binaries for Intel Management Engine and DCMI
- S2600JF
  - Three 3.5” drive bays
    - 2x 2TB
  - Console: iKVM (requires hw key)
  - Redundant (1+1) PSUs
  - IPMI

Benchmarking results





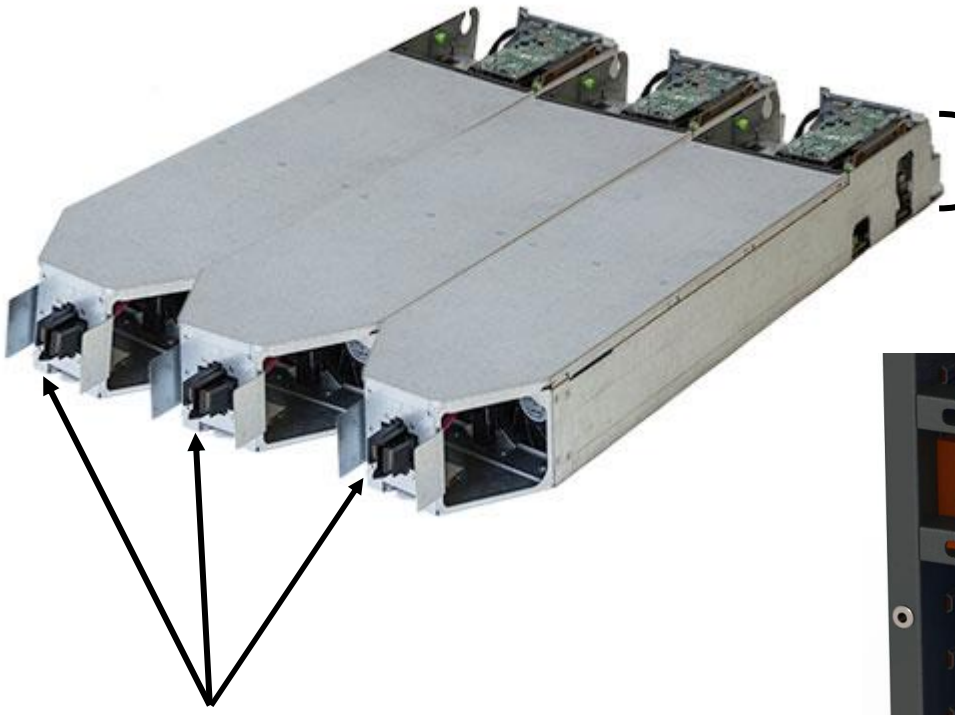


- Common settings
  - Hyper-threading and Turbo enabled
  - Power saving options disabled
  - Weighted power average: 80% loaded / 20% idle
- ~5% less performance but 25% power gain

Benefits & issues

- Power consumption
  - 25% from platform but more gain expected from OCP Open Rack with DC distribution
- Standard design
  - Ideally competing offers must be technically identical. Good to know what you pay for
  - But contract manufacturing is not cheap in small volumes
    - Economy of scale assumes “large” scale
    - Can we benefit from supplier ecosystem around Facebook, Microsoft, …?
- Manageability and redundancy
  - BMC is back in Intel mainboard v3.0 spec (Jan 2014)
  - The single PSU issue disappears with OCP Open Rack DC distribution (next slide)
  - Single non-redundant HDD (or SSD)

Open Rack



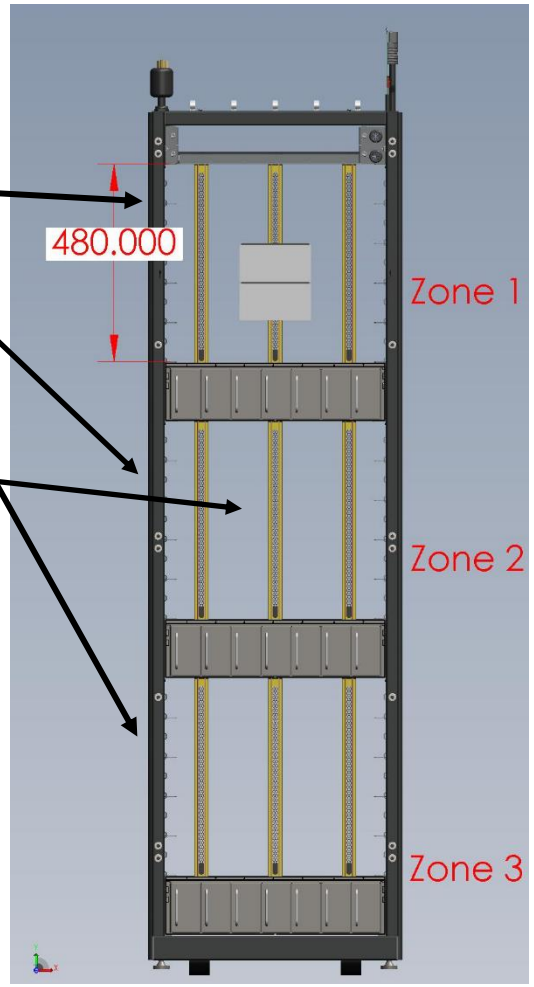
2 OpenU

DC power connector

3x 13xOpenU zones:

- 1 OpenU = 48mm
- 3xOpenU power zone
- 10xOpenU “innovation” zone

3 bus bars for DC 12V



Plans

- Acquire a small part of 2015 capacity in a few OCP Open Racks?

- Procurement considerations
  - Unit is a fully populated, cabled and tested rack
  - Finding and qualifying bidders
    - OCP certification – still early days
      - Understand what “OCP Ready”, “OCP certified” means
    - Complex rack level cabling (is it part of certification?)
    - Top of Rack switch part of the “unit” (customer selected non-OCP)
  - Tender specification
    - OCP is a collection of specifications – select a comprehensive sub-set in order to be able to compare offers

Conclusions

- Open Compute Project is an interesting new direction with
  - Potential far-reaching impact for industry and data centres
  - A constantly growing provider community and private customer space
- Encouraging results from our initial tests with two twin systems
  - Sufficiently interesting to motivate launching a project for larger deployment
- Public procurement challenges
  - So far no public procurement has been attempted (to my knowledge)
  - Finding and qualifying bidders (in CERN member states) will be a challenge
  - Specifying the tender will likely be quite different from what we are used to
- Any other site with OCP experience or interested in collaborating?