



OPEN
Compute Project

OCP Technology day

@Facebook, Aug 30, 2016

Goals & topics of the day

Would my software work
on your hardware?

“An OCP Technology Day is meant to showcase software solutions on OCP (Open) Hardware. Through these events, we move up the stack and continue to build on the strength of open hardware as we showcase how communities come together to consume, collaborate, and contribute to the infrastructure ecosystem.”



Agenda

GlusterFS, Scale out storage - Richard Wareing, Facebook

Open CloudServer vNext - Mark Shaw, Microsoft

Openstack RDO - Redhat

MAAS / Juju - Canonical

Disks for Data Center - Lawrence Ying, Google

Announcement of Facebook Open Lab

opening of a lab space in their Menlo Park office for vendors to test their software on open hardware—hardware that is open sourced through OCP—and for potential adopters to test-drive their software solutions.

<http://www.opencompute.org/blog/facebook-opens-new-lab-space-helping-the-open-compute-project-community-move-up-stack/>

<https://code.facebook.com/posts/1155412364497262>

“We built a lab space specifically for vendors to test their software on open hardware. We hope this will help advance the industry even faster and make the idea of using open hardware more appealing to smaller companies that may not have the resources to develop or rely on custom solutions”



Michael Liberte

Bill Carter, CTO at OCP

Explained the new certification process and OCP Inspired/Accepted.

The foundation is working on a reference architecture, currently listing the orderable SKUs -> should be done by next month.

“Make the OCP product more easily consumed”

“Would like the community to take our products and build solutions”

GlusterFS on OCP

Richard Wareing, Production engineer at Facebook

Numbers

Datasets of GBs to many PBs

100s millions of files

10s Billions of file operations a day

1000s of Bricks

100s of thousands of clients per cluster

Motivations:

- problem of pace of vendors
- lead time when ordering hardware
- configuration took days
- long response time to bugs
- rejected customization
- custom/complex cabling
- different voltage in OCP datacenter
- higher cost of traditional solutions

"our customers expect bug fixes same day, same hour"



GlusterFS on OCP

Use cases:

- Archival
- Backing data store for large scale applications
- Being the glue between large scale systems
- Anything that doesn't fit into other specialized storage (ex: Haystack, Video backend...)

Test setup

6 racks, 54 hosts

OCP Winterfell server + Openvault KNOX

Gluster 3.6.3_fb with erasure coding

File size 100-500MB

2 bricks per server

multi region tests across 3 datacenters

All IPV6

8 cores Xeon

24GB memory

1x10G Nic (no dedicated network for storage, don't have issues of latency, packet loss etc...)

1x HW RAID

2x15 disk RAID 6

OpenVault drive sled : 4TB x 30 HDDs (~100TB per host)

Why RAID?

- Data protection and availability
- bit rot, patrol read (for free in the controller)
- single disk failure > hot swap

want Gluster to focus on bigger problems, single disk failure is managed by the RAID controller

OCP: half to 2/3 the cost of past solutions

- future erasure coding will reduce the cost by half or so

Microsoft vNext

Open source motherboard (CPU up to 200W, 32 DIMMs, 3 x16 PCIe, 16 SSDs)

EIA 19" rack

Front cabling

Independent chassis

Power supply in each server chassis

Management node in the PDU based on ARM + Linux

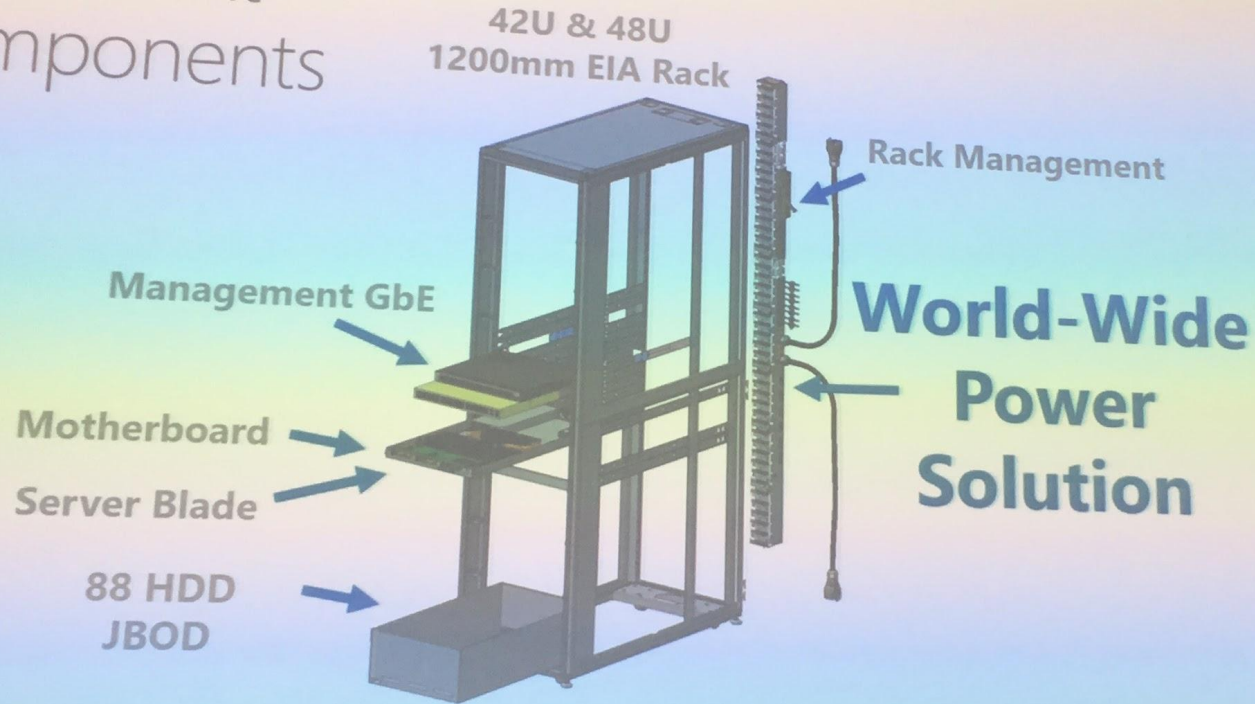
Redfish based OOB management

Expected sometimes in 2017 for review/collaboration

Not deployed at Microsoft yet.

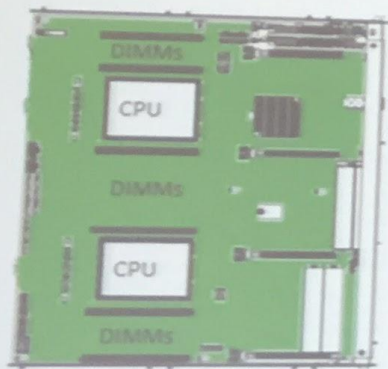
At very early stage of design.

OCS vNext – Components

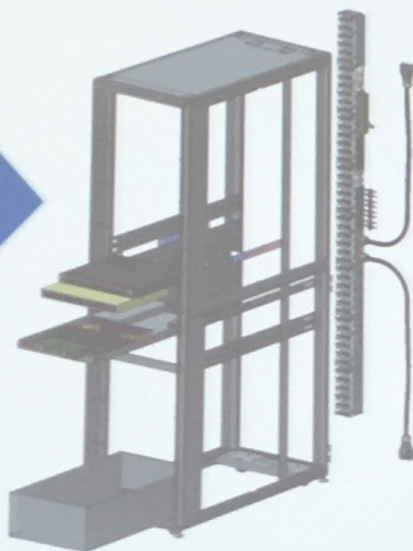


Choose the components you need

Motherboard



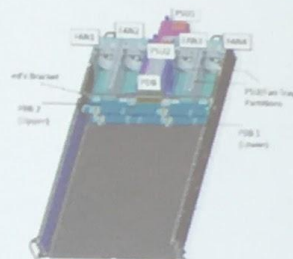
OCS vNext



Open Rack



EIA 19"



**Other
19"/21"**



Redhat

open hardware and open software are complimentary

- open hardware: cut off long tail of maintenance, buy replacement from anyone
- Open software: greatest amount of innovation
- similarities of development and certification models

Openstack installation on OCP in collaboration with Facebook

spec:

- 1x Intel 10GBE
- 10 nodes
- 2x intel broad well CPU
- 256G of RAM
- 80G SSD

Canonical

MAAS + Juju

Similar deployments as Redhat

Demo live using the OrangeBox + Wedge

Capable of deploying Linux Snappy on the Wedge + FBoss (some manual steps though because of ONIE not compatible with MAAS)

Google

Google is using custom hard drives

Low level commands in the drives, for instance to enable additional security beyond encryption

At very large scale, you can deal with getting drives of “about 6TB” (5.9 to 6.1TB), as long they are averaging 6TB

Proposing to work on many research topics: Change geometry of drives, protocols, remove unnecessary sensors/protection, different connectivity