Journal of Physics: Conference Series **664** (2015) 052004 doi:10.1088/1742-6596/664/5/052004

Experience of public procurement of Open Compute servers

Olof Bärring, Marco Guerri, Eric Bonfillou, Liviu Valsan, Alexandru Grigore, Vincent Dore, Alain Gentit, Benoît Clement, Anthony Grossir

IT department CERN CH-1211 Geneva 23 Switzerland

Olof.Barring@cern.ch, Marco.Guerri@cern.ch, Eric.Bonfillou@cern.ch, Liviu.Valsan@cern.ch, Alexandru.Grigore@cern.ch, Vincent.Dore@cern.ch, Alain.Gentit@cern.ch, Benoit.Clement@cern.ch, Anthony.Grossir@cern.ch

Abstract. The Open Compute Project, OCP (http://www.opencompute.org/), was launched by Facebook in 2011 with the objective of building efficient computing infrastructures at the lowest possible cost. The technologies are released as open hardware, with the goal to develop servers and data centres following the model traditionally associated with open source software projects. In 2013 CERN acquired a few OCP servers in order to compare performance and power consumption with standard hardware. The conclusions were that there are sufficient savings to motivate an attempt to procure a large scale installation. One objective is to evaluate if the OCP market is sufficiently mature and broad enough to meet the constraints of a public procurement. This paper summarizes this procurement, which started in September 2014 and involved the Request for information (RFI) to qualify bidders and Request for Tender (RFT).

1. Introduction

The Open Compute Project (OCP) was launched by Facebook in 2011 with the mission to design and enable the delivery of the most efficient server, storage and data centre hardware designs for scalable computing. The OCP foundation is a non-profit corporation with Facebook, Intel and Rackspace as founding members and the board of directors currently (May 2015) also includes Microsoft, Arista Networks and Goldman Sachs. Specifications and mechanical drawings are developed in multiple areas including server and storage design, networking, hardware management, certifications and data centre technology. Hardware specifications are licensed by one or more contributors under the Open Web Foundation Agreement [1]. An alternative more restricted (GPL like) license agreement is also under discussion [2].

Section 2 gives the motivation why it was decided to have a closer look at OCP and what are its potential benefits to our infrastructure. Two OCP twin servers were acquired for testing and benchmarking. The tests and results from this first acquisition are described in Section 3.

Based on the encouraging test results a larger procurement of OCP hardware started at the end of summer 2014 with the goal to attempt a public competitive tender for a medium-sized installation. The objectives for this project were to:

- probe the maturity of the OCP market with respect to multiple offerings from different suppliers and their presence in the CERN member-states
- establish if a larger volume allows for a comparative or better pricing to the standard servers
- get experience with the hardware itself, its integration and interoperability with our data centre (power distribution and cooling) and tools

The public procurement project is described in Section 4 followed by the results and conclusions, which also summarizes the experience so far.

2. Motivation

Our main motivations for having a closer look at OCP in 2013 and gain some practical experience were:

2.1. Electrical power efficiency

Due to limits on the electrical power available in the CERN IT data centre, power efficiency is an important technical requirement and the cost for electrical power is routinely factored into the contract adjudications. OCP hardware is designed for power efficiency and therefore provides an interesting alternative to standard servers and storage. Furthermore the standardization also allows to benefit from further optimizations such as rack-level AC/DC conversion without risks of vendor lock-in.

2.2. Cost

Standardization of interfaces (firmware and tools) together with simplified and vanity-free mechanical and electronic designs are factors that would presumably reduce the manufacturing costs, which coupled with the potential market volume generated by Facebook and other hyper-scale web companies, may also be of benefit for smaller end-customers.

The underlying assumption, however, is that there actually exists a market with standardized items. In reality the large volume may actually be entirely custom built, or in the best case “OCP inspired” and therefore not directly available to smaller customers.

2.3. Serviceability

One of the stated goals of OCP is the design of solutions engineered for maximum operational efficiency and serviceability. In a post from 2013 [3] the total repair time is reduced with a factor 2-5 depending on the type of failure.

3. First tests with OCP hardware

In the summer of 2013 two Hyve-1500 twin servers, based on OCP design retrofitted to a 19-inch rack form-factor, were acquired. The intention was to run our standard benchmarks and compare performance and power consumption to a standard server based on the Intel S2600JF platform [4] with a similar configuration:

	OCP server, Hyve-1500	**Standard server, S2600JF**
Chassis	1.5U Twin system	2U Quad system
PSU	1	2 (1+1 redundant)
CPU	2x E5-2650 (Sandy Bridge)	2x E5-2650 (Sandy Bridge)
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

Table 1: main features of OCP and reference systems

For the comparison the HEP-SPEC06 [5] was used. It is the agreed benchmark for computing performance within the High Energy Physics (HEP) community. The results shown in Figure 1 indicate a 25% gain in the power consumption per performance (VA/HEP-SPEC06).

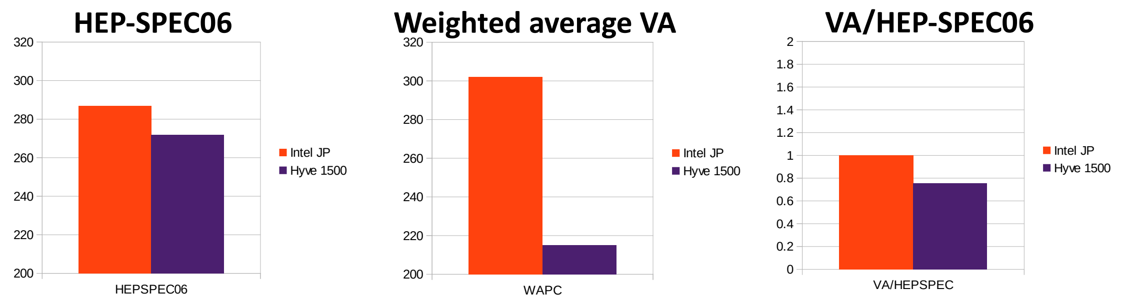

Figure 1: benchmark results comparison for performance and power consumption

Apart from the benchmarking, the OCP systems were also evaluated to a set of standard technical requirements in our tender specifications. The most important shortcomings found were:

- Lack of redundant local storage, since only one 2.5” drive is supported by the Hyve-1500 (Figure 2)
- Lack of virtual KVM over IP
- Non-redundant Power Supply

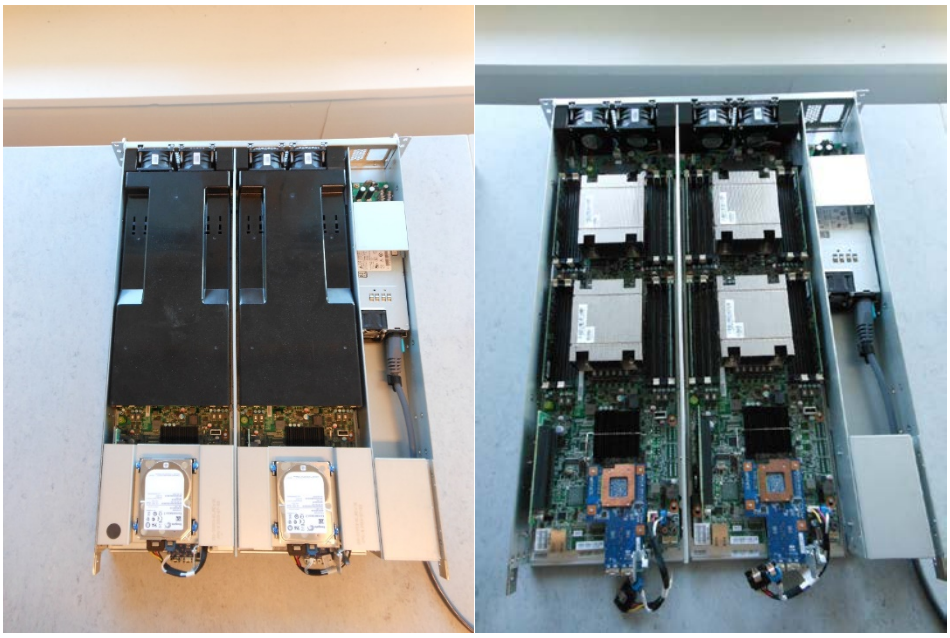

Figure 2: Hyve-1500 twin system

None of those shortcomings were real showstoppers and have been addressed in recent designs. Encouraged by the potential electrical power efficiency it was therefore decided to go ahead with a larger project.

4. Competitive tender for OCP hardware

Encouraged by the initial results from the small OCP server acquisition described in the previous section, a larger project was launched for the supply of about 6 racks of hardware with the objectives to:

- assess the efficiency expected from centralized AC/DC conversion and rack level DC current distribution
- probe the OCP market and supplier ecosystem

- compare and assess potential cost benefits for a reasonable volume of OCP server to a similar acquisition of standard servers

4.1. Qualifying bidders

The OCP Foundation maintains a list of "Solution Providers" [7] with capabilities and experience to supply OCP solutions. The list currently has 6 members:

- Hyve Solutions
- AMAX
- Penguin Computing
- QCT (Quanta Cloud Technology)
- ITOCHU Techno-Solutions Corporation
- StackVelocity

The CERN procurement rules stipulate [8] that the "Country of Origin" shall be a CERN member state[1], where Country of Origin is defined as the country in which the supplies are manufactured or undergo the last major transformation by the contractor or its sub-contractor(s). Some of the companies listed as Solution Providers have assembly facilities in CERN member states.

Apart from documented experience with OCP and capabilities to delivery such solutions, an OCP Solution Provider must attain Gold or Platinum membership and pay an annual program fee. The total annual fee is in the order of 1% of the annual turn-over for a small-medium sized integrator that will typically bid for a CERN tender so, unless its other customers also procure OCP hardware, the additional cost will need to be spread onto the unit price.

The compromise was to qualify Solution Providers fulfilling the Country of Origin criteria as well as integrators either in a consortium with a Solution Provider or with documented experience in the assembly of OCP solutions.

4.2. Technical specification

There is a plethora of published OCP designs and the challenge is to find a subset that is sufficiently broad to allow for competition while fulfilling the technical requirements for the intended usage:

1. Rack design
   a. Facebook Open Rack v1 Mechanical Specification (revision 1.8), March 2013
   b. Facebook Open Rack Hardware v1 Specification, January 2013
   c. Facebook Power Shelf v1 Specification
   d. Power Supply Hardware v1 Specification
2. Server design
   a. Facebook Server Intel Motherboard v2, April 11th 2012
   b. Facebook Server AMD Motherboard v2, April 11th 2012
   c. Facebook Server Intel Motherboard v3, January 1st 2014

The requirements on processing performance, memory capacity and connectivity were devised to match our requirements for standard servers. Due to the limited number of mechanical designs supporting a redundant configuration with 3.5-inch disk drives, some compromises were necessary for the local storage capacity on the servers. However, as the revised requirements for local storage open for configurations with solid state drives they turned out to be useful in our tenders for standard servers.

The CERN IT main machine room has a fixed rack infrastructure with cold air containment. The racks are arranged in rows of 15 racks per cold aisle with an average electrical power density of

[1] The CERN Member States are currently Austria, Belgium, Bulgaria, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Israel, Italy, Netherlands, Norway, Poland, Portugal, Slovakia, Spain, Sweden, Switzerland and United Kingdom. In addition: Serbia is an Associate Member State in the pre-stage to Membership and Romania is Candidate for Accession

approximately 3.5 kilo-Watt per rack. The procurement units are therefore servers, storage arrays and network switches, which are mounted and cabled in the existing rack infrastructure. However, for OCP the procurement unit is a fully assembled and cabled rack with servers, storage expansion units and Top-of-the-Rack (ToR) network switches. Some adaptations to the infrastructure and procurement are therefore necessary: the assembled racks will be commissioned in a refurbished area of the machine room allowing for approximately 7 kilo-Watt power density. The two ToR switches, one 10 Gigabit Ethernet switch and one 1 Gigabit Ethernet switch for management, are not part of the supply but instead procured separately by CERN and shipped to the successful contractor for the final rack assembly.

Despite being an active area, the certification of the compliance with OCP designs and interfaces is not mature. There are currently two certification flavours:

- OCP Ready: a maintained software toolkit for self-testing
- OCP Certified: a comprehensive certification executed by certification labs

The OCP Ready toolkit is a subset of the OCP Certified certification. As it is not clear what are the benefits of either, the specification was kept open for both by specifying that the rack, system unit, server enclosure and storage expansion unit shall comply with one of the “OCP Certified” or “OCP Ready” certifications.

4.3. Adjudication criteria

The tender specified that the offer should provide CPU performance capacity of 40’000 HEP-SPEC06, and exactly 12 JBODs with in total about 2 petabytes of storage based on 6 terabyte hard disk drives certified for continuous (24/7) operation. It was required that the proposed servers and JBODs shall be assembled in exactly 6 identically populated OCP racks with a power consumption not exceeding 7 kilo-Watts per rack (under full load).

5. Result and conclusions

The tender described in the previous section was launched at the end of 2014. Given the early days for the OCP market, and the rather complex technical specification, bidders were allowed 10 weeks, compared to the usual 4 weeks, for preparing their offers. At the time of the CHEP 2015 conference the contract adjudication had been completed, and purchase orders placed, but equipment not yet delivered.

The cheapest offer was found to be technically conforming and prices can therefore be compared to the ones from the latest tender for standard (non-OCP) servers adjudicated in September 2014. In both tenders the platforms were based on the latest (dual-socket) Intel® Xeon® Processor E5-2600 v3 (Haswell) family. This allows to determine any potential commercial benefits for end-customers. Care has to be taken, however, when comparing the prices:

- **Volume**: the tender volumes are significantly different: in the latest server tender the capacity used for the adjudication was 250’000 HEP-SPEC06 compared to 40’000 for the OCP tender
- **Chronology**: there were about 6 months in between the two tenders so market prices for some components, e.g. the CPUs, would normally be lower at the time of the more recent OCP tender
- **Adjudication unit**: in the OCP tender the fully assembled Rack is the adjudication unit whereas it is the individual servers in the standard server tender. The OCP tender also included 12 disk arrays (JBODs)

The chronologically more recent pricing for the OCP capacity in the second item is expected to partly compensate for the smaller volume. The two latter items can also be corrected because the offers in the OCP tender also specified the pricing for the individual servers.

The relative cost difference for the server only:

$$\frac{Cost\ per\ \text{HEP-SPEC06}, OCP\ server}{Cost\ per\ \text{HEP-SPEC06}, standard\ server} = 1.06$$

The relative cost for assembled Racks (minus the JBODs)

$$\frac{Cost\ per\ \text{HEP-SPEC06}, assembled\ OCP\ rack}{Cost\ per\ \text{HEP-SPEC06}, standard\ server} = 1.30$$

The comparisons are subjected to differences in the HEP-SPEC06 measurement where the values used are the ones provided by the bidders. For the standard server delivery the bidder's value was about 5% higher than the average measured once the full delivery had been commissioned. A similar comparison for the OCP offer is only possible once the OCP supply has been delivered to CERN.

The price difference for the assembled OCP rack is amplified due to the limit imposed on the power density per rack (see Section 4.3.), which implies that less than half of the rack is populated with equipment. A denser solution would yield a better price ratio.

In summary: similar pricing for a server unit compared to a tender for standard servers. The comparison can only be completed once the full delivery has been benchmarked and the actual aggregated performance has been measured. It will also allow for a broader understanding of the other cost benefits such as the electrical power efficiency and simplified serviceability.

References

[1] Open Web Foundation specification agreement, http://www.openwebfoundation.org/legal/the-owf-1-0-agreements/owfa-1-0
[2] See OCP post from 24 of February 2014, http://www.opencompute.org/blog/request-for-comment-ocp-hardware-license-agreement/
[3] See OCP post from 24 of October 2013, http://www.opencompute.org/blog/facebooks-perspective-on-serviceability-and-operational-efficiency/
[4] Intel S2600JF platform http://ark.intel.com/products/56303/Intel-Server-Board-S2600JF
[5] HEP-SPEC06 benchmark, http://w3.hepix.org/benchmarks/doku.php/
[6] Open Compute at CERN, HEPiX spring workshop 2014 at LAPP Annecy-le-Vieux, http://indico.cern.ch/event/274555/session/12/contribution/22
[7] OCP solution providers, http://www.opencompute.org/about/open-compute-project-solution-providers/
[8] How to do Business with CERN, http://procurement.web.cern.ch/how-to-do-business-with-cern