Publication date:
27 July 2021
Author:
Vladimir Galabov

Open Computing is for Everyone and is Here to Stay

A blueprint for how the open source community is driving global innovation

Commissioned by:
inspur

Brought to you by Informa Tech

Contents

Catalyst	2
Open computing is born	3
Adoption takes off	9
Latest open compute initiatives	13
Conclusion and recommendations	16
Appendix	18

© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.

Catalyst

Over the last 20 years, the world has become more and more digital and the data center (DC) has become key to business success, especially when it comes to the skyrocketing e-commerce, internet content, and cloud services industries. The larger the scale, the bigger the impact simple design improvements have, be that in the physical or IT aspects of a data center. It is no surprise then that, as companies like Facebook, Amazon, and Alibaba became leaders in digital industry, the re-engineering of their DC equipment became a top priority. Three design considerations stand out: power efficiency, manageability, and ease of deployment.

In 2009, Facebook began an initiative to design its own DC IT and physical equipment to create the world's most energy-efficient data center that would enable hyperscale computing at the lowest cost possible. Two years later, it shared its designs and results with the public, demonstrating a data center that was 38% more energy efficient with a 24% reduction in operational expenditure (opex). The Open Compute Project (OCP) was born, in partnership with Intel, Rackspace, and Goldman Sachs, with the aim of establishing a collaborative hardware design process mirroring the open source software ecosystem.

Building on this trend, six months after the launch of the OCP, in October 2011, China's three hyperscale cloud service providers (cloud SPs) – Baidu, Alibaba, and Tencent – launched the "Scorpio Project" with support from Intel, to develop and test integrated racks of servers. In 2014, the Open Data Center Committee (ODCC) was established, extending Scorpio's work from all-in-one rack servers, to include micro-modular data centers.

The efforts to expand open hardware development continued in 2016 with the inception of the Open19 Foundation, started by LinkedIn, HPE, and Vapor IO. Open19's aim was to develop DC equipment that is well suited to colocation and edge location deployment. In 2019, LinkedIn contributed the Open19 designs to the OCP in the first step toward convergence of open hardware development models. Most recently, the two open hardware development efforts are converging with the OCP, expanding its collaboration with the ODCC.

Today, members of the open computing community include IT equipment and physical DC equipment vendors, cloud and communication service providers, enterprises, colocation providers, and system integrators (SIs), as well as semiconductor manufacturers. These companies aim to work collaboratively and openly share ideas, hardware specifications, and other intellectual property to evolve the IT industry.

This report will discuss (1) why open computing is an important technology trend in the evolution of DC IT and physical infrastructure that should be considered when organizations plan future DC investment, and (2) how the latest innovations by the open computing community are making open compute hardware well suited to a broader set of workloads, market segments, and deployment scenarios.

© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.

Open computing is born

OCP and the world of collaborative engineering

Facebook designed DC hardware for energy efficiency and sustainability

The Open Rack specification was the first hardware design released by the OCP and is based on the work Facebook started in 2009. The specification includes a re-engineered rack and power delivery architecture that, according to the OCP, offers an efficient, scalable alternative to the widely adopted EIA-310 rack, often referred to as the traditional 19-inch rack. Open Rack has four key features that make it efficient to deploy, support, and operate.

- Each individual server or storage node does not have an individual power supply. Power is supplied by a pair of bus bars located in the rear of the rack. The bus bars are supplied with either 12V or the latest 48V DC by a power shelf at the base of the rack. According to Facebook's Matt Corddry, to maximize efficiency a design principle was "convert the power the minimum number of times required" to avoid losing 2–5% of power with each conversion.[1] The Open Rack design brings unregulated 480V three-phase power straight into the rack where a power shelf converts it to 12V that goes straight to the server/storage motherboard. By comparison, other DC designs convert power three or four times: from 480 to 208, into an uninterruptible power supply (UPS), back out of the UPS, into a power distribution unit, and into a server power supply. According to Facebook, this results in 21–27% cumulative power loss. By converting only once, it saw only 7.5% power loss[2] – an improvement of nearly 20%.

- The Open Rack IT equipment enclosures are wider and taller to accommodate better airflow. The taller enclosures also allow better efficiency of airflow through the use of 60mm or 80mm fans as opposed to 40mm fans. According to Corddry, the bigger fan blades turn slower and use less energy to move an equal volume of air. As a result, fans in the Open Rack servers consume much less energy: approximately 6 watts, compared to up to 80 watts in a traditional server design.[3,4]

[1] Corddry, M., "How Facebook threatens HP, Cisco, and more with its 'vanity free' servers," *Ars Technica*, https://arstechnica.com, accessed April 5, 2021

[2] Graner, A., *Why Open Hardware?*, Open Compute Project, https://www.socallinuxexpo.org, accessed April 5, 2021

[3] Corddry, M., *Interview at Open Compute Project Summit 2015*, March 10, 2015, https://www.youtube.com/watch?v=dwk4-bg0c4A, accessed April 5, 2021

[4] M. Corddry, *Demonstration of Windmill Server*, July 31, 2012, https://www.youtube.com/watch?v=ckNzwqhDS60, accessed April 5, 2021

© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.

- Cables and interconnects are made from the front of the rack for easier access when servicing. IT equipment is hot-pluggable, which aids servicing. As a result, service personnel can avoid working at the rear of the rack, often known as “the hot aisle.”

- Component faults are also identifiable from the front of the rack and route service procedures do not require tools.

Overall, the physical and IT equipment redesign offers capital expenditure (capex) savings of 45%, when compared to a traditional approach, according to Facebook. It also achieved a 38% improvement in energy efficiency and a 24% reduction in opex.[2]

Since the conception of the Open Rack design, the issue of data center sustainability has increased in importance. Apart from reducing the power consumption of the equipment, two key sustainability design principles Facebook implemented are:

- Minimizing the use of nonrecyclable components

- Heating office spaces with server waste heat.

Then Microsoft contributed its Open Cloud Server, designed for software-defined environments

In early 2014, Microsoft contributed in-house-designed DC equipment to the OCP, called Open Cloud Server (OCS). The design builds upon a key software-defined data center principle – through effective workload distribution policies and practices, computation can be moved dynamically from one device to another. As a result, a high level of hardware redundancy and availability and 24x7 support is not needed – if a server fails, workloads can be re-routed to another server. Microsoft concluded the following:

- In a traditional DC deployment, IT equipment is designed for high reliability, requiring several layers of redundancy to ensure the infrastructure is always available – typically 99.999% or more uptime. Applications require that the hardware on which they are running is persistently available, may have dedicated redundant servers, and typically lack the capability to shift their workloads to alternate servers in the event of a failure. Failure of any component in the stack results in application downtime. Consequently, each layer of the stack requires multiple levels of redundancy – backup power feeds and supplies, batteries and generators, backup storage, backup cooling, and backup network connections.

- In a hyperscale cloud DC environment, the sheer scale of operations dictates that at any given time numerous components will be in a failed state. To keep a service up and running in this scenario, software needs to provide the resiliency. Properly architected applications can instantly have their workloads shifted to alternate hardware – even a different data center – in the event of a component failure or configuration error. Hardware availability of 99.9% or less is acceptable. As a result, the servicing model can move from a 24x7 schedule to an 8 am–5 pm, five days a week schedule, thus reducing the cost of support and maintenance operations.

Microsoft followed the principle of software providing resiliency and moved to a model where its cloud services software was designed to be software resilient, enabling hardware designs that do

© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.

not require high levels of redundancy and availability, and offering significant efficiencies and cost savings.

Microsoft's OCS hardware design, now used for all of its key cloud services, including Bing, Office 365, and Microsoft Azure, had two key design goals – low acquisition costs and lower opex. Microsoft wanted the OCS design to be optimized for managing and operating an installed base of more than 1 million servers across a global footprint of data centers.

The OCS equipment fits in an EIA-310 19-inch rack; each server or JBOD shelf is designed to be 1U height and half width in a density-optimized blade design, where the server and the JBOD shelf share power management and a signal backplane. The result of the design is an extremely dense rack. This inevitably reduced airflow.

In addition to the enclosure size and design, Microsoft worked on standardizing key components like the Ethernet adapter, management module, and software suites.

Microsoft estimated that the OCS design offers cost savings of 40% when compared to traditional IT equipment. The company achieved 15% improvement in energy efficiency. These values are both lower than those Facebook reported for the Open Rack design. Where OCS shines is in its operational savings. Microsoft achieved an up to 50% improvement in deployment and service times and up to 75% improvement in operational agility vs. traditional rack servers.

LinkedIn followed with Open19 server

A non-hyperscale DC alternative

LinkedIn developed an alternative to the Open Rack and OCS designs, addressing two of the features that it considered shortfalls of Facebook's and Microsoft's contributions. LinkedIn's Open19 servers were designs to fit 19-inch racks because they found that Facebook's 21-inch Open Rack was not available in enough colocation facilities, LinkedIn's primary server deployment location. The Open19 gear also does not require huge scale to reap benefits for the end user. In fact, the design favors a dispersed server installed base with no IT staff at the location of deployment. The design targets were:

- Reduce rack-scale capex by 50%
- Reduce integration time by 3–5x.

To achieve these, LinkedIn used two blind mate connectors – one networking and one for power – with snap-on cables at the back of the Open19 racks. These connectors provide 100GE network connectivity and deliver 400W of power to each server, called "bricks" by the Open19 engineers. This cabling system enables easy server deployment because the servers do not need to be cabled to a power distribution unit (PDU) or a switch. This cabling system also enables partially toolless installation. In a conversation with Yuval Bachar, the lead architect of the Open19 project, we learned that LinkedIn's long-term goal is for Open19 servers to be so easy to install that a courier would both be able to ship it to the colocation DC and plug it into the existing Open19 rack.

© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.

LinkedIn estimated that the Open19 design offers cost savings of 35–40%, when compared to traditional IT equipment. Unlike Microsoft and Facebook, they did not communicate any energy savings. LinkedIn observed a staggering 7–8x improvement in deployment time and found the design will enable them to keep the number of service personnel low. It took the LinkedIn team 75–90 minutes to install a rack and a further 15 minutes to plug in 96 servers; a total of 100 minutes for physical and IT equipment per rack with two technicians.

Open19 is formally a part of the Linux Foundation. With Microsoft's acquisition of LinkedIn in late 2016, we anticipated Open19 designs to be contributed to the OCP and the start of a co-development process. In 2018, LinkedIn formally joined the OCP and in 2019 it contributed the Open19 designs to the OCP. LinkedIn eventually moved away from the Open19 foundation with future equipment development falling on the shoulders of the Linux foundation, Equinix and Cisco. In a recent interview with Yuval Bachar, Open19 Foundation Fellow, we learned that an update of the Open19 server and physical infrastructure is coming later this year. Going forward, we expect the teams behind the two OCP designs and the Open19 foundation to co-develop server and storage component standards like open Ethernet adapters, co-processors, and SSD designs.

ODCC – cost savings through standardization

An Alibaba, Tencent, and Baidu lead project encompassing IT and physical DC equipment

Within six months of the inception of the OCP, the three Chinese hyperscale cloud services provider (cloud SP), Alibaba, Tencent, and Baidu set up their own open equipment design project, called Project Scorpio. In 2012, China's Academy of Information and Communication Technology – under the Ministry of Industry and Information Technology (MIIT) – joined the effort, resulting in the foundation of the Open Data Center Committee (ODCC). The ODCC working groups cover not just equipment but also physical data center designs. The Server Working Group develops rack and density-optimized blade server designs, including servers with co-processors for AI. The Data Center Working Group develops modular and containerized data center designs. A network group works on Ethernet switch and adapter designs. The Testing and Certification Group verifies whether the other two groups are working in accordance with technical specifications set by members and partners.

Similar to the OCP and Open19 projects, the ODCC is focusing on cost saving through standardization, with an emphasis on achieving this through the development of rack-scale and high-density servers. The Scorpion rack-scale server adopts centralized power supply and heat dissipation, which can reduce the total cost of ownership (TCO) by 10% and make the power usage effectiveness (PUE) of the data centers equipped with Scorpion rack-scale servers less than 1.3, or even below 1.2, which significantly improves energy efficiency. In 2017, the ODCC reported that its 8-node and 4-node 4U server designs were 8% and 4% more energy efficient than a traditional 2U rack server. At the scale of Baidu, Alibaba, and Tencent this can result in significant cost savings.

ODCC welcomes telecoms with OTII Initiative

The Open Telecom IT Infrastructure (OTII) project, now governed by the ODCC, was initiated in late 2017 and counts Chinese cloud cloud SP, JD.com, and a large number of original design manufacturers (ODMs) as members. It aims to standardize open-developed server for 5G and comms SP edge computing applications.

© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.

In June 2019, an OTII server became a recommended open solution for the Open Radio Access Network (O-RAN). Later, OTII started to focus on developing 1U servers for specific O-RAN scenarios that require low scalability.

In conversations with comms SPs over the past five years, we have learned that there is interest in benefiting from hyperscale cloud SP innovation and best practices. Additionally, telecom companies we have interviewed highlighted capex and opex reduction and energy consumption as key reasons why they are considering open designs. The OTII initiative is a step in the right direction in expanding the open computing ecosystem.

Table 1: Comparing the open computing organizations

	Open Compute Project (OCP)		Open19	Open Data Center Committee (ODCC)
	OpenRack	**OCS**		
Mission	Redesign DC hardware for efficiency		Customizable, flexible, economical, and open	Openness, innovation, cooperation, win-win
Main problems solved	Energy efficiency, maintenance	Hardware, deployment & operations costs	Ease of deployment, minimum personnel	Standardization, compute density
Server market value (2020)	~$8bn	~$3bn	~$0.1bn	~$3bn
End users	Facebook, Rackspace, Yahoo Japan, OVH, AT&T, Verizon	Microsoft	LinkedIn Equinix	Alibaba, Baidu, Tencent, Meituan, Jingdong
Ecosystem	>200 members		9 members	>100 members

Source: Omdia

Convergence of open standards is coming

Cross-membership is increasing and will drive co-development

Over time companies have joined all three open computing organizations improving the cooperation between them. Baidu, Microsoft, and Facebook have recently announced closer partnerships in the development of new equipment designs. This new partnership shows a willingness of two major regions to work together to drive even larger economies of scale.

Convergence can happen at many levels, be it the API or software layer, the architecture layer, or simply multiple vendors that provide interoperability in the market. A common misconception is that the industry has to reach complete hardware or implementation convergence. We think that at a hardware level a balance between standardization and differentiation will occur based on workload differences and natural competition. This is a central benefit of the collaboration between working groups.

© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.

The many are stronger than the few

The long-term success of the open computing ecosystem is dependent on the contribution of all members. The origins of the organizations described above are centered on the engineering proficiency and needs of a few big end users. The growth of open computing is dependent on the contribution of all members, big and small, and the initiative of both end users and vendors.

While there are differences in implementation between the OCP, Open19, and ODCC, we are seeing the best ideas converging and becoming core themes in all three organizations.

We should also not understate the importance of vendors that are members of all three initiatives and help to drive cooperation in the market. Notably, Inspur is one of the few IT equipment vendors involved in the OCP, Open19, and ODCC efforts, developing and shipping servers based on all three standards.

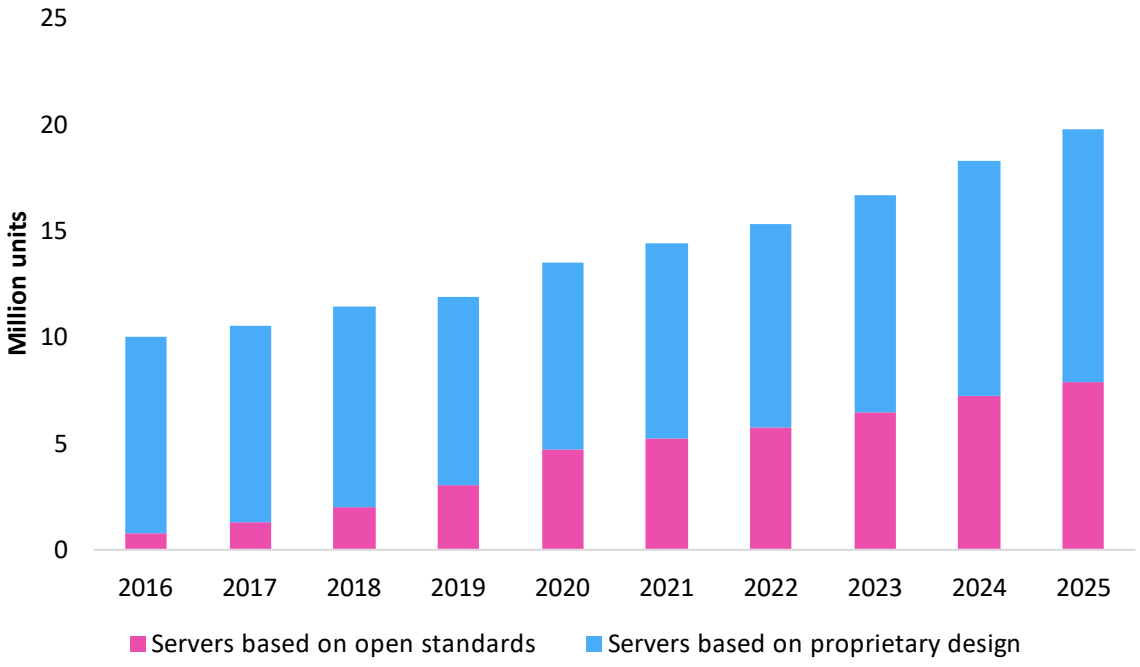
© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.

Adoption takes off

The increasing appeal of open computing is reflected in the adoption trends

Apart from the cloud SPs that initiated the different open compute equipment development projects like Facebook, Microsoft, LinkedIn, Baidu, Alibaba, and Tencent, a number of tier-2 cloud SPs like Yahoo Japan, comms SPs, enterprises, and governments have deployed open compute equipment from one of the projects listed above. Cumulatively, this is expected to result in 40% of the servers shipped worldwide in 2025 being developed based on an open standard, up from 7% in 2016, as highlighted in Figure 1.

Figure 1: Open computing will make up 40% of worldwide servers shipped by 2025



© 2021 Omdia

Source: Omdia

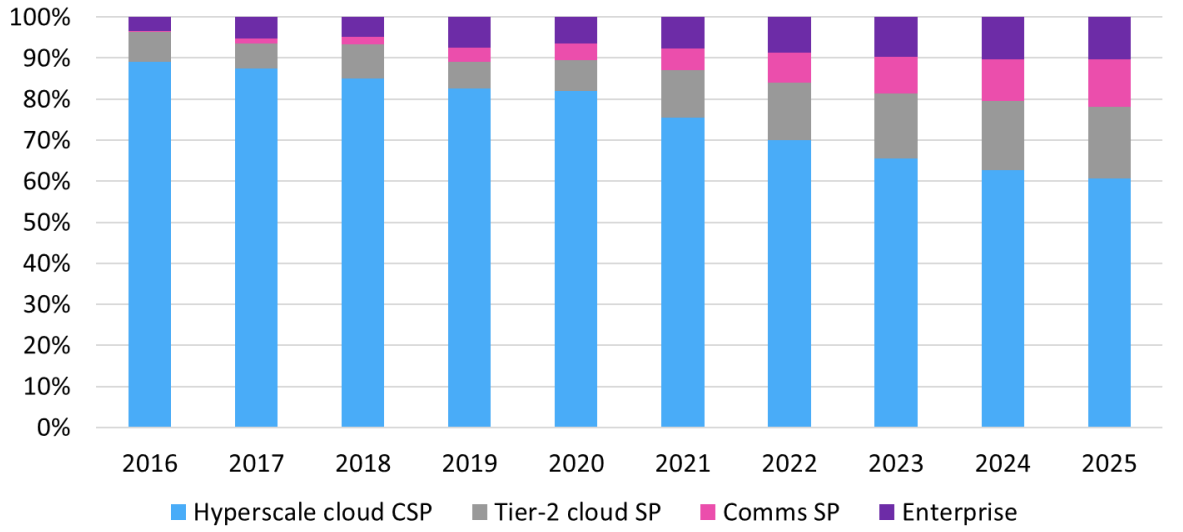
© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.

Comms service providers are the next wave

Since 2017 Omdia has run annual surveys with vendors shipping open compute hardware and companies that have adopted equipment based on open standards that were not a founding member of a standard body. We found that the earliest adopters of servers developed based on open standards were tier-2 cloud SPs like Rackspace, Salesforce, Yahoo Japan, and, in China, Meituan and Jingdong, and other hyperscale cloud SPs like Amazon and Google. We also saw governments in the US, China, and a few European countries adopt IT equipment based on open standards due to specific initiatives to decrease energy consumption of government-owned equipment and an aim to copy hyperscale cloud SP innovation. The latter was also a major factor for tier-2 cloud SPs to deploy equipment based on the OCP, Open19, and ODCC standards.

When we worked on our very first survey, we saw a large number of test and proof-of-concept work at communication service providers (comms SP). Over the following years, many of these turned into production deployments making the comms SP segment a big potential market for IT equipment developed based on open standards. Investment in 5G networks proved to be a catalyst for central office and disaggregated cell site gateway deployments. Figure 2 shows the deployment of the OCP accepted and inspired equipment by non-board member companies (i.e., companies that were not founding members of the OCP).

Figure 2: Servers based on open standards by end-user segment



Source: Omdia

In a recent interview with a large comms SP we learned that servers developed based on open standards appealed to the company as its demand for server computation grew and as it sought to centralize server purchasing with a focus on power and cost efficiency. Trial deployments of OTII and

© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.

ODCC servers demonstrated lower TCO compared to traditional rack servers. The company observed lower power consumption, expedited delivery times due to standardization and consistency of operations and management.

Out with the share resource blade servers and in with the open compute

The comms SP hopes to weed out traditional shared resource blade servers from its installed base, primarily because of their high cost. The comms SP believes shared resource blade servers have higher TCO due to smaller-scale production and a lack of standardization between manufacturers. The service provider also found shared resource blade servers to lack node scalability. In many of its data centers the power density of each rack was relatively low, requiring power-efficient servers. The comms SP could not fully populate blade enclosures, as that would bring the total power consumption above each rack limit.

The comms SP also encountered a lack of management consistency due to blade server vendors having their own management systems. This created a management challenge in large-scale deployments as it could not easily manage the fleet centrally.

The comms SP concluded that OTII, ODCC, and OCP equipment have better node scalability and provided flexibility when matching server configurations to individual application requirements. A high degree of uniformity in the server installed base also made management easier while enabling high server density per rack and lower energy consumption.

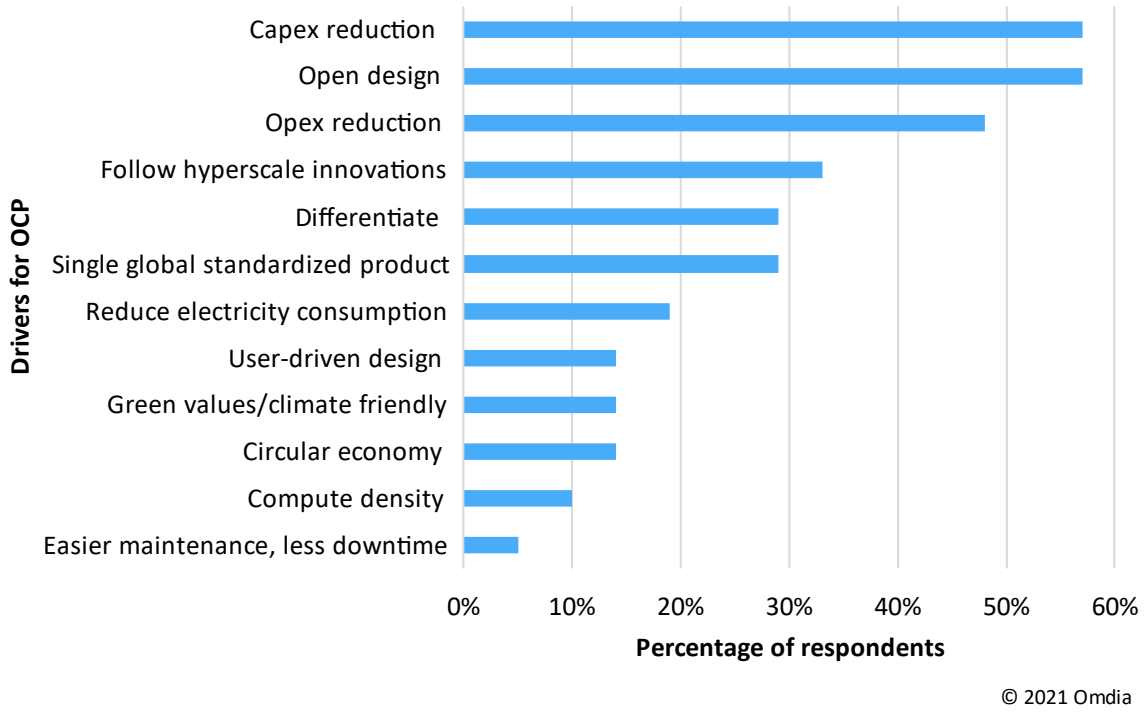
Capex and opex reduction are the main reasons for adoption

In Figure 3 we summarize the results of our 2020 OCP pulse survey. The DC IT equipment buyers we interviewed shared much of the comms SP's sentiment, pointing out capex and opex reduction as the main reasons for adopting open compute equipment.

They also highlighted how they see value in utilizing the innovation the hyperscale companies are contributing via the open designs. They recognized that deploying open compute equipment would provide differentiation from their competitors.

© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.

Figure 3: 2020 OCP survey results: Top reasons for adopting OCP-certified products



Source: Omdia

More enterprise verticals are following suit

Our annual interview pointed to financial enterprises, gaming companies, e-commerce, and energy sectors experimenting with and deploying at-scale servers based on open standards. High performance compute (HPC) was cited across several enterprise verticals as an important growth driver. These enterprise verticals will also benefit from having workloads reside both in the cloud and on premises, and having a physical architecture that mirrors large-scale cloud deployments is therefore critical. This includes the need for acceleration and offload for AI and fast networking offload to improve latency.

Healthcare and manufacturing companies were highlighted as next-wave deployers. Automotive and industrial companies were also expected to ramp up use of servers developed based on open standards. We expect many of those to be deployed at the edge as connected cars and IoT devices need low latency compute. Interviews indicated that proof-of-concept deployments (POCs) have started.

© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.

Latest open compute initiatives

Rack scale is key to many end users

One of the key objectives of the open computing initiatives is optimizing the installation and operation of servers by the rack. Furthermore, the ability to configure the compute, storage, acceleration, and networking to meet the needs of the workload and do this at scale is fundamental.

One of the most important areas of open source contribution has been the development of the OCP Open Rack specification. Open Rack 2.0 has helped solve some fundamental problems by increasing compute density and increasing energy efficiency, which has shown substantive return on investment to the internet and telecom companies that have deployed. With Open Rack 3.0, which will incorporate additional features such as 48V DC power supply and liquid cooling, and by increasing the rack height from 410U to 440U, we expect the capability to add needed AI GPUs while maintaining server density, a change that will match the needs of the ever-evolving workloads.

ODCC has also devoted itself to developing rack-scale server specifications from Scorpio 1.0 to Scorpio 3.0. And its open computing architecture is seeing similar returns. Increasing the number of server nodes from 16 to a full allocation of 32 optimizes the infrastructure cost of the power supply unit (PSU) and fans and can result in an energy saving of 10–20%.

In an interview with Inspur, which was a strategic partner in building one of Baidu's data centers, we learned that Scorpio 3.0 enabled the data center to maintain a PUE of less than 1.3 on the adoption of the architecture. We also found that by delivering rack scale infrastructure to Baidu, Inspur might have set a global record by installing 10,000 nodes in eight hours.

Rack management extends the savings to scale

The ability to provide an open rack management infrastructure that enables various implementations and a modular ecosystem was a major catalyst for the market.

Rack management extends management beyond the GPU server to components in the cabinet, such as top-of-rack switches, PDUs, and fans, etc., and the standards are applicable from the data center to the edge deployments, which is critical.

The Open Rack Management Controller (OpenRMC) working group works in close cooperation with the Desktop Management Task Force (DMTF) organization, which has established a clear strategy and roadmap for vendors and operators to:

© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.

- Establish a specification for northbound interfaces, which allows the equipment resource profile to present itself to the remote client for management
- Establish a specification for southbound interfaces, allowing transparency of the rack resources and control of the bottom elements in the rack
- Establish safety management and align with DC operations and business resource scheduling.

The fundamental value proposition is clear – unified management will make it easier for end users to manage and maintain equipment among various manufacturers and greatly reduce the system management cost.

The OpenRMC is a blueprint for how collaboration can work. The initiative was led by Inspur and helped influence vendors to open their code to the industry. Contributors include Facebook, Microsoft, Google, HPE, Dell, Intel, and Wiwynn.

Open compute standards turn to AI

New software technologies that use ML and AI techniques have driven the market for specialized processors capable of high degrees of parallelism. Co-processor options are multiplying, giving customers choice with a long-term roadmap developing. Additionally, more vendors are offering servers with co-processors, providing choice of server form factors and power envelopes. Virtualization software also continues to add features that make it possible for customers to increase the utilization of servers shipped with a co-processor.

As AI evolves, different suppliers produce new AI accelerators, but due to the technical challenges and design complexities of current proprietary AI hardware systems, it generally takes about 6–12 months to integrate them into systems. This delay prevents quick adoption of new competitive AI accelerators. It seems well timed that the OCP would kick off the open accelerator module (OAM) project. The OAM design specification defines the mezzanine form factor and common specifications for a compute accelerator module. In contrast with a PCIe add-in card form factor, the mezzanine module form factor of OAM facilitates scalability across accelerators by simplifying the system solution when interconnecting high-speed communication links among modules. Facebook, Microsoft, and Baidu are leading the charge with first OAM shipments landing in the latter's data center in late 2019. Engineers from the three cloud SPs authored an initial OAM design and collaborated on the final specification with other cloud SPs (Google, Alibaba, and Tencent), semiconductor vendors (NVIDIA, Intel, AMD, etc.) and server vendors (Inspur, Penguin Computing, IBM, Lenovo, etc.).

Open computing at the edge

The development of new devices and software technologies in response to a growing requirement to improve business processes, relieve humans of repeatable tasks, and make life more interesting is accelerating the global computing demand. At the same time, the nature of devices and applications is changing where the collection and real-time processing of data are becoming increasingly important. As a result, latency and bandwidth are becoming key performance determinants and are

© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.

driving the need for better telecom networks and more computing power to be placed closer to end users and machines.

With the number of servers deployed at the edge projected to double over the next five years, the open computing community has ramped up their efforts to provide IT equipment designed for deployment at the edge. Servers from the OCP, OTII, and Open19 are already being tested by enterprises and comms SPs.

Open Rack designs are optimized for the telecom industry with the Open Edge server

The OCP Open Edge server design, for example, can fit in a 600mm-deep 19-inch rack, which is commonplace across existing base station sites. At maximum density, end users can place 5x1U half-width Open Edge servers in a 3U enclosure that provides power and management to the server. The server was designed with telecom industry requirements, such as electromagnetic shielding and seismic tolerance, in mind. We've learned of a large number of POCs with North American and European comms SPs, but a number of commercial deployments have already been announced including American mobile network operator US Cellular, which will be running its virtualized radio access network on Open Edge servers.

While the Open Edge server was designed with the telecom industry in mind, enterprises have also explored the design. MIRIS, a Norwegian real estate and technology firm, deployed Open Edge servers to support delivery of Smart City services in business parks and residential areas. MIRIS has built edge data centers in over 20 urban locations in Norway since 2019, and is currently exploring a wider rollout across the Nordic region.

OTII offers an extra compact edge server option

The OTII-based edge server design has been shown at 19-inch width, 2U height, and a shallow 430mm depth, which is more than half of the traditional server depth. This allows servers not only to be deployed directly on traditional racks but also directly on the wall, which simplifies edge deployments. Chinese comms SPs kicked off field trials of the OTII edge server for virtualized radio access networks, content delivery, and virtual customer premises equipment (vCPE). Given that comms SPs expect to slash radio access network (RAN) opex by 53% and capex by 30% by virtualizing, with savings in power consumption, site rental fees, and onsite management, we expect the OTII-based edge server to be a hit.

This is just the beginning

We expect edge-optimized designs to continue coming from the open computing organization. One interesting new development is the Open Edge Flexi Cabinet Outdoor (FCOB), which enables a mini-DC-style deployment outdoors. With an operations air temperature of -40°C to +50°C it can withstand most Asian, North American, and European climate conditions. We expect similar outdoor server enclosure designs from other members of the open computing ecosystem.

© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.

Conclusion and recommendations

While it has been a decade in the making, it is apparent that there is an open computing ecosystem that is thriving, growing, and continues to be a force of innovation. The open compute infrastructure started from the largest hyperscale cloud SPs on the planet and is now reaching into the tier-2 cloud and comms SP, and enterprise sectors. The largest compute farms needed to change fundamentals of cooling and power, and be at the leading edge of AI. The developed technology is now bearing fruit beyond its roots and is gaining traction in much smaller deployments where the design fundamentals are just as relevant. Expect to see data centers in a box or prefabricated modular data centers become a new norm at the edge, and inside will be the technology contributed from the OCP, ODCC, and Open19.

The OCP, ODCC, and Open19 as industry groups have been born from different ideas, market needs, and geographic differences. It is fortunate for the industry that there are companies that are contributing in all the groups and serving a function of needed coherency. OEMs like Inspur are showing that they can deliver products with similar technology to each in a quick and efficient manner and are helping to guide the industry with contributions in hardware but, more importantly, software that allows the differences to become fluid.

We recommend end users:

- Spend time understanding their cross-business computing needs and the shortfalls of their present operational practices. Bear in mind that the compute environment needs to support the entire lifecycle of existing and future applications – from early prototyping to production.
- Assess the open computing options discussed in this report as they could reduce capex and opex by a third or more. Keep in mind that, depending on the solution, to realize these savings you do not need to deploy equipment on the huge scale that Facebook, Microsoft, or Alibaba do.
- Many open computing servers are already optimized for popular software options like Red Hat Linux and VMware's vSphere.
- Many SIs already offer open computing options. You don't have to always work directly with the vendor. Most open computing organizations will have a list of SI partners.
- Utilize the open computing organizations as partners and provide improvement recommendations and feedback. Collaboration is in their DNA.

© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.

© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.

Appendix

Methodology

The Technology team at Omdia is the leading source of information, insight and analytics in critical areas that shape today's technology ecosystem—from materials and components, to devices and equipment, to end markets and consumers. Businesses and governments in more than 150 countries around the globe rely on the deep market insight we provide from over 300 industry analysts in technology sectors spanning IT, telecom, media, industrial, automotive, electronics, solar and more. What sets Omdia's Cloud and Data Center Research Practice apart is our team of technical, experienced analysts, and our end-to-end coverage of the industry.

- The 10 lead analysts that are the main contacts for our clients all have been in the industry for over a decade, have a technical background and a strategic mindset. This gives us the confidence to say that we have the experience, training and skills needed to effectively help you connect the dots, see all perspectives, and stay ahead of disruption. The 10 lead analysts are supported by a large team of primary and secondary research experts, data scientists and specialists. We're also more global and diverse than ever, located in 4 countries, across 6 time zones, communicating in 10 languages.
- Our unique data points/perspective include the physical aspects of the data center (IT enclosures, backup power, cooling, modular DC construction) and a detailed view of DC IT (servers, storage, networking, operations and development software), including who is shipping and who is buying DC equipment, as well as, where they are placing it (centralized vs. edge location). We've also built a comprehensive view of the cloud and colocation services ecosystem including service provider investment in IT equipment and DC buildout. Our rapidly growing portfolio of primary research is also helping us provide end-user perspectives and what is impacting their purchasing decisions. In fact, it is hard to list all viewpoints our team can provide.

The quantitative data used for the completion of this report is based on Omdia's Data Center Server Intelligence Service. Our server forecast model is based on more than 20 years of actual worldwide yearly growth in computing capacity shipped, the rate of servers becoming multi-tenant—running virtual machines or software containers—and increases in computing capacity per server. We also reconcile three quantitative data aspect – (1) detailed server shipments provided by the most impactful vendors in the market, (2) detailed server CPU shipments and inventory data provided by the most impactful semiconductor makers, and (3) end-user level data from the largest cloud SPs and comms SPs on how many servers they bought. Our forecast is done at the form factor and end user segment level to ensure accurate forecasting of servers based on open computing standards. Omdia then projects these trends out to CY25 to derive the forecast.

The qualitative data used in this report comes from our 2020 OCP Adoption Survey, presented at the 2020 Virtual Summit. We concluded a series of in-depth interviewed with key industry leaders across North America, Europe and Asia Pacific on the topic of open compute adoption drivers and barriers. Omdia has been the official research partner for the OCP since 2018 and will be updating our forecast again at the 2021 Global Summit in November. The OCP interview data was supplemented with an in-depth interview with an ODCC member, the comms SP detailed on page 10-11 of this report.

© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.

Author

Vladimir Galabov
Director, Cloud and Data Center Research Practice
askananalyst@omdia.com

© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.

Get in touch

www.omdia.com
askananalyst@omdia.com

Omdia consulting

Omdia is a market-leading data, research, and consulting business focused on helping digital service providers, technology companies, and enterprise decision-makers thrive in the connected digital economy. Through our global base of analysts, we offer expert analysis and strategic insight across the IT, telecoms, and media industries.

We create business advantage for our customers by providing actionable insight to support business planning, product development, and go-to-market initiatives.

Our unique combination of authoritative data, market analysis, and vertical industry expertise is designed to empower decision-making, helping our clients profit from new technologies and capitalize on evolving business models.

Omdia is part of Informa Tech, a B2B information services business serving the technology, media, and telecoms sector. The Informa group is listed on the London Stock Exchange.

We hope that this analysis will help you make informed and imaginative business decisions. If you have further requirements, Omdia's consulting team may be able to help your company identify future trends and opportunities.

About Inspur

Inspur Electronic Information Industry Co., LTD is a leading provider of data center infrastructure, cloud computing, and AI solutions, ranking among the world's top three server manufacturers. Through engineering and innovation, Inspur delivers cutting-edge computing hardware design and extensive product offerings to address important technology arenas like open computing, cloud data center, AI, and deep learning. Performance-optimized and purpose-built, our world-class solutions empower customers to tackle specific workloads and real-world challenges. To learn more, please go to www.inspursystems.com.

© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.

Copyright notice and disclaimer

The Omdia research, data and information referenced herein (the "Omdia Materials") are the copyrighted property of Informa Tech and its subsidiaries or affiliates (together "Informa Tech") or its third party data providers and represent data, research, opinions, or viewpoints published by Informa Tech, and are not representations of fact.

The Omdia Materials reflect information and opinions from the original publication date and not from the date of this document. The information and opinions expressed in the Omdia Materials are subject to change without notice and Informa Tech does not have any duty or responsibility to update the Omdia Materials or this publication as a result.

Omdia Materials are delivered on an "as-is" and "as-available" basis. No representation or warranty, express or implied, is made as to the fairness, accuracy, completeness, or correctness of the information, opinions, and conclusions contained in Omdia Materials.

To the maximum extent permitted by law, Informa Tech and its affiliates, officers, directors, employees, agents, and third party data providers disclaim any liability (including, without limitation, any liability arising from fault or negligence) as to the accuracy or completeness or use of the Omdia Materials. Informa Tech will not, under any circumstance whatsoever, be liable for any trading, investment, commercial, or other decisions based on or made in reliance of the Omdia Materials.

© 2021 Omdia. All rights reserved. Unauthorized reproduction prohibited.