

Raghib Hussain: Hi. My name is Raghib Hussain, and I am President of Marvell Products & Technologies. Today, I'm going to talk about how hyper-scale cloud data centers are evolving, driven by the needs of data infrastructure, and how Marvell is best positioned to be the leading semiconductor company for the data infrastructure.

So, if you look at cloud data center. When the cloud data center started, their focus was to implement a set of functionality. They needed to implement it fast and roll out their application and services to their customers. This is why they took whatever equipment was available. They took a bunch of standard OEM servers, along with all kind of OEM appliances for networking, security, and server load balancing. [inaudible] it then was the first generation of cloud data center. It served the need. It was designed to provide the functionality, but it had all kinds of challenges related to scalability, performance and efficiency.

So as the cloud business grew, the hyper-scale cloud vendors needed to scale rapidly to service the need of their customers. They needed to make it more efficient so that they could deploy more and more data centers. So at that time OEM appliances that they chose in the first phase were not designed for cloud scale data center. A DCO was not good with them. This is why hyper-scale cloud vendors move to ODM based server that were designed according to their specification, improving their overall TCO.

They also adopted virtualization and moved network services from OEM appliances to software based implementation. It not only gave them ability to scale their overall network, to make it cloud scale, but it also gave them control over their network services and security, giving them a lot more flexibility to deploy newer services on the go. However, with the rising number of customers and increasing amount of data, performance and efficiency was still a big challenge.

The cloud had another unique characteristics that was different from traditional server workload. Traditional servers were designed to process applications with the characteristics of less compute, intensive processing and data mainly controlled by the application. These applications generally create data and manage data in their own format. For example, web servers, web browsers, emails, or word processing application, they all manage and create their own data. This allowed CPU to process a structured data, which was easy to handle.

The data center compute is the processing that not only involves complex math function, but deals with a large volume of data movement. Data center compute is a unique characteristics of cloud workloads. For example, things like, security with your processing, storage optimization algorithms like compression and a recovery in coding, and especially AI and ML. All these applications had one thing in common, they are all complex math functions. In addition, they all require high volume of data movement in and out of the system to get the most value out of it. Several CPUs are designed for application centric compute, cannot handle the overall complex data center compute tasks very efficiently.

So, as the cloud vendors were looking to make it more efficient and improve performance, and especially with the increase of volume of data as computer

requirement of these data centric processing or increasing. The third phase of the hyper-scale cloud, it started to deploy accelerators for all kinds of data center compute. They use readily available TPUs, FPDAs and accelerators. These accelerators help remove performance bottlenecks for data center compute. However, they were not most efficient implementation. They were still not optimized for the target application and not built according to application specification. In addition, it was not that efficient for the TCU.

During the last 18 months, the shift to cloud has accelerated, driven by COVID. All these things were already happening, the workloads were already moving to the cloud, but the pandemic has really accelerated that option of the cloud. Whether it is e-commerce, video conferencing, whether you were streaming, telemedicine or online gaming. All these applications have moved to the cloud and driving a lot more data in the cloud infrastructure. We see more and more remote users and hence the growth of cloud computing.

With this growth of cloud computing, the shift of workload to cloud is also causing that type of processing needed in the cloud. And with the increase of volume of data, the overall need of data center compute is increasing. That trend, once again, was already into play with more and more data centric services being deployed in cloud. And it has been accelerated. As time goes by, the amount of data is increasing, the amount of data centric compute needed in the cloud. The recent shift in this trend has only accelerated the need of data center compute, and that is why cloud architectures are looking into finding newer ways to improve the overall data centric processing.

In order to handle the data centric processing at cloud scale, and do it in an efficient manner, hyper-scale cloud are implementing a completely new data center architecture. This architecture is designed to not only remove performance bottlenecks at every level, but also getting more efficient solution for the best TCO. To start with, hyper-scale cloud vendors are improving the fundamental building block of data center [inaudible]. This time, they are not only optimizing the server at the system architecture level, but also building server processor, according to their own application needs. This allows them to put the right amount of compute balance with the right amount of CaaS specific to their application need, and the IO to handle bandwidth requirement of the data centric applications. This gives birth to all server processors.

In addition to augment the processing, they are also complimenting application server processor with the data-centric processing engines, like a smart links and DPUs. This is special purpose data centric engine, or DPUs, allow them to process various data-centric tasks like security, where you're processing computational storage, AI and network services, in a most efficient manner, design specifically for their workload needs.

However, to achieve the best efficiency, instead of using the off the shelf FPDAs and GPUs, the cloud vendors are choosing DPUs designed specifically for this need and also, custom application specific processors. To remove the bottleneck of the network they're implementing a new network architectures with upgraded network switches to handle higher bandwidth and also upgrading interconnect and optical connectivity.

Finally, to scale that cloud data center beyond a single data center and implementing the reliability and redundancy, they are connecting data center with multiple data center with the optical interconnect between the data center. This not only gives them much better connectivity and bandwidth, but also reduce the overall cost of implementation. This new data center architecture allows the cloud vendor to achieve the most optimized TCO for the data center and with the cloud scale performance and efficiency.

Over the last four years, Marvell has been focused to bring the best of products to serve all the needs of the data infrastructure. First, we have standard products to address the needs of growing security, storage, networking, and data centric processing. We have liquid iOS smart deck, that are the leading smart deck solutions, shipping in volume production and deployed at cloud data centers.

We have, 'do it yourself' storage solution. This SSD solution helps cloud data center to implement, virtualize and desegregated this solution. In addition, they also provide additional functionality needed for computational storage. We have liquid security to provide the highest performance, but cloud scale security. So this is the highest performance virtualized security solution available with the highest level of security. Next we have high-performance data center switches and 'PAM-4' optical interconnect. This allows hyper-scale data center to achieve the highest bandwidth and cloud scale networking. Then we also have, for the enter data center connectivity, we have 'Colors' and 'Colors Z Two', which is coherent optical interconnect. This allows the scaling of the cloud networking beyond the single data center. And finally we have experienced custom Silicon team and business model to partner with the hyper-scale cloud customers and serve their customer Silicon needs, for all server processor, as well as various kind of EPU's and accelerators.

Today, Marvell is providing products and solution based on leading new technology, but this didn't happen by coincidence or chance. Marvell has been investing for the last several years to become a leader in process technology and IP. As a result, today, we have a state of the art IP in the latest process node and leveraging it not only for our standard products, but also for custom Silicon. That gives a good opportunity for the cloud guys to implement their own application specific solutions.

In addition, we are driving Silicon and system level innovation. Our Silicon team is experienced in implementing most sophisticated packages and driving new trends like multi chip modules, 2D/3D packages, and co-packaged optical interconnect. All this ability to customize Silicon, along with our flexible business model, puts us in the best position to partner with hyper-scale cloud vendors.

So if you look at a high level, Marvell is very well aligned to be the best partner for hyper-scale cloud data center. We have high-end data center switches and for optical interconnect. We have expertise and business model for custom server processors, as well as, we have not only experienced expertise and product for the off the shelf DPUs and a smart link, but we also have a partner model for custom Silicon and 'ASIC' to implement a specific solution for the specific needs of the hyper-scale cloud vendor.

We have computationally storage and do it yourself SLD. And that also gives our customer a way to work with us, partner with us and implement their own custom variant solution to achieve the best storage solution for their hyper-scale cloud solutions. We have data center interconnect, coherent optical modules and solution. This allows our hyper-scale cloud partner to scale beyond a single data center. And then finally our highest performance security solution allows them to implement the highest performance security and enable their customers to have the highest level of trust in working with them.

So in summary, cloud scale infrastructure requires, optimize and build to specification solution. This solution needs to be designed according to the needs of the data centric compute and the applications that are being deployed at cloud scale. Marvell has complete product portfolio and latest process node IP needed for the data sector. In addition, Marvell has flexible custom Silicon model to be the best partner for hyper-scale cloud vendors. With that, I thank you for attending today's session. Thank you.