Supermicro servers with NVIDIA technology help service providers enable the Edge

In this interview with Jay Lawrence, general manager of technology enablement at Supermicro, we discuss how service providers can leverage Supermicro's servers powered by NVIDIA technology and how these services can increase processing power at the Edge.

Mobile World Live: Tell me a little bit about Supermicro's evolution in the telecom space.

Lawrence: Supermicro has a great leadership position with the traditional data center, hyperscale, and social media markets, and others. The telecom industry, however, presents a number of new opportunities for a company like Supermicro who has a rich DNA in hardware, a building block ideology to our design, and power-efficient green computing as a central tenet of our business.

On the horizon we have new CPU level horsepower which can be simply summed up as more speed and power, plus we have a new suite of very high-powered NVIDIA GPUs that are very enabling — not just at the Edge but everywhere in the telecommunications super structure from the core to the far edge and we have a play across this entire spectrum.

In conjunction with our partner Ribbon Communications, we are doing a lot of application work in 5G that starts at the core and is, in our first successful program with them, performing transcoding. We are advancing our work with Ribbon to extend to the Edge with a couple of mobile edge compute (MEC) applications. We are also integrating some of the new NVIDIA® Bluefield® data processing units (DPU) into our systems which is an example of some of the great work they've done to enable such applications.

SUPERMICR⊙® **⊗ nVIDIA.**

Mobile World Live: What new applications and services are enabled by increased processing power at the Edge?

Lawrence: What I find exciting about 5G are the applications that haven't even been invented yet that 5G will enable. We are starting to see those types of applications come to life.

Supermicro is actively involved in solutions delivery, not just on the hardware side, but also on the software side because we are partnering with innovative software companies.

We're seeing significant progress on the radio access network (RAN) and in the small cell and millimeter wave (mmWave) domain. We are also seeing progress with applications that incorporate artificial intelligence and augmented reality into 5G.

We have one product, for example, that is integrated with augmented reality and is being used for advanced training. Every time that we get more core technology and more processing power from NVIDIA GPUs, we can get more diverse in our workloads and we can do more with the ecosystem.

Mobile World Live: Can you give me more details on that augmented reality application that is possible because of processing power at the Edge?

Lawrence: Because you have stronger processing capabilities these Augmented Reality (AR) applications become powerful enough for enhanced human-machine engagement which becomes very useful in enhancing the human experience and making it immersive. Take training for example: How is training in the workplace handled today? You watch a video or look at some PowerPoint slides and then take a quiz. How much do you really learn? How much do you retain?

As one example of Supermicro's efforts, we are working on an application with a company called Taqtile that allows a subject matter expert to conduct a task and record it and then a novice can put on an AR headset and they can do the same task immediately and the work product is nearly as good as the expert on the first try plus the novice is learning by doing, not just answering a handful of multiple choice questions. That type of application can't be done without the compute power of the Edge and this is just a single example of the exciting applications that 5G is and will continue to enable.

Mobile World Live: Without getting into too much technical detail, can you tell me how NVIDIA GPUs help with these types of applications?

Lawrence: The Edge is exploding with 100 times or more devices now hanging off the end of a 5G network, to meet several of the key requirements of a 5G network such as latency, availability and bandwidth, it's critical that compute power is pushed to the Edge to inject functionality where it is needed and close to the user. To do this successfully, you need to pick the right tool for the job as the workloads are also getting much more diverse.

For a simple example, an autonomous vehicle has a different workload and data requirement profile in terms of latency than a streaming video application. Streaming video can queue and buffer and needs a wider pipe but it doesn't have the same latency requirement.

Augmented reality is another example of an application that demands a special workload requirement where you need to keep the latency low. There are some reports I've seen that indicated if the latency is above 20 milliseconds in this application people will get disoriented and nauseous because of the timing.

Having the flexibility of these various GPU workloads is very empowering because we can put the right amount of horsepower at the Edge. NVIDIA has given us the tools and guidance to do that. I also think the advent of DPU capability — which includes Network Interface Controller (NIC) cards that integrate an ARM-based processorwith GPU — is another example of good engineering ingenuity that helps us provide our customers with the right tool for the job.

Mobile World Live: How has Supermicro adapted its GPU-powered servers for telco environments?

Lawrence: We've done a number of things. We listened to the market and studied what core requirements we need to have. We realized that we needed robust, redundant DC power supplies. In the hyperscale data center, AC power is readily available but in telco central offices they prefer DC power so AC is not always available so we needed to adapt for that. It's a big deal.

We've created multiple products with the form, fit, and function and workload capability including GPUs to meet a wide array of telecom/5G requirements, and taken a further step to invest in taking these through the process of certification for Network Equipment-Building System (NEBS) Level 3. This is a very important standard to the tier-one telecom providers in the U.S. While it is required here in the states, it is not always required globally. However, passing the NEBS Level 3 tests shows your telecom customers that you are serious, and your products are robust. NEBS Level 3 certification tests are divided into two parts. One part focuses on the passive elements like emissions and whether your product is a good neighbor, and if it can tolerate noise from friends nearby.

The second part of the test includes more aggressive testing that apply a variety of ways to make sure the product has durability and is hardened. For example, our servers have been tested by the "Needle Test" where direct flames are applied to each component in the system to make sure it can't be set on fire. Our servers have also been tested for shock and vibration, simulated earthquake, thermal operation to 55C and particulates and moisture.

NEBS Level 3 is now a part of our roadmap. As we move to the next generation of CPU and GPU technology, we are incorporating NEBS into a growing number of our products. We are making this long term investment in the telco segment.

Another element is our product design and our rich hardware DNA. Because heat builds up in a chassis there has to be good air flow and a good cooling system to dissipate that heat. We have that with our products, and it allows us to support data center-class CPUs and GPUs as well as edge-focused systems.

DC power supply, as I mentioned, is another big deal: our engineers were able to, within a matter of a few days, take a DC supply from another platform, retrofit it, put it in our chassis, and get it approved by Underwriters Laboratories (UL) and NEBS Level 3 in record time. Just a small example of the amazing things we can do in this area.

I think this strategy of looking at the application and building off of the lessons we learned has been beneficial to Supermicro. We give our customers a lot of flexibility, by following the "right tool for the job" approach. And we have this partner ecosystem that is helping us craft what we are taking to market that is both hardware and software, optimized so the customer gets the best experience and the best value for their dollar.

Mobile World Live: How is Supermicro working with its partners such as Ribbon to meet the edge computing needs of service providers?

Lawrence: Ribbon is just one example. We have lots of partners in our ecosystem. It really depends upon the partner and what that partner does and how we engage with them. With Ribbon, it is a software company and we are a hardware company. As they are developing their next-generation software tools we make sure they have the right types of servers available, optimized to the requirements of the software. We know where their software ends up in the topology – whether it is more of a core product like transcoding for cross-network voice calls, or a MEC solution. That determines the class of the server.

Because we take our telco products through NEBS Level 3 certification, we can pick the best configuration. Many of our competitors were not able to or willing to get their hardware certified for NEBS.

Because of this, software companies like Ribbon appreciate that we have NEBS-compliant hardware which is another differentiator.

It is a strategy that is perpetually evolving and adapting. As with anything, if you don't have the whole ecosystem of customers, partners, etc. providing you with guidance it is easy to overlook. As we've evolved with our telecom partners, we have become aligned at the planning level and the go-to market level as well; the net effect is that we are working to market each other's products and solutions. This gives us a bigger section of the market to talk too. Bottom line is the combination of software and hardware and the engineering alliance between Ribbon and Supermicro is a good thing.

Mobile World Live: What's the next step for advanced services at the Edge?

Lawrence: Part of the answer is, who knows? The augmented reality, artificial intelligence and Industry 4.0 applications are coming of age. There will be more distance learning and telemedicine applications – these are all things that we are involved in with our partners. And all of this is because of what 5G is making possible.

From Supermicro's perspective, we want our customers to come to us with problems so we can work with them to solve them. With the improvement of CPU and GPU technology, we have more tools at our disposal to solve those problems.

By looking at the gang of three – customer, hardware and software – we are able to create the best possible customer experience. Even if we aren't the authors of the software, we are involved with it and we want the customers experience to warrant five stars.

I'm excited to see what the next thing is that hasn't been invented yet but is on the near horizon.