

Behind the Hype: What are DPUs and Why Do They Matter?

A processing evolution is hitting the data center, and it's bringing improved performance, security, and scalability.



SPONSORED BY



Organizations are placing significant demand on their IT infrastructure to achieve business goals such as innovation and improved user experiences. This puts the pressure on central processing units (CPUs) such as the Intel® Xeon® Scalable processor to not only perform without latency, but also provide efficiencies, security, and scalability. It's also creating increased complexity as businesses seek to modernize applications with a microservices architectural approach.

Data processing units (DPUs) can simplify the delivery of these infrastructure services, and help organizations reap multiple other benefits. The DPU is a server option card that acts as a system on a chip, making intelligent decisions about offloading and accelerating workloads and services including networking, storage, and security.

"Think of the DPU as an impact player, like a quarterback," said Ragav Gopalan, product management leader, VMware. "It partners with the CPU to make more capacity available for workloads."

By offloading and accelerating infrastructure services carried out by the CPU, the DPU helps businesses gain greater performance, security, agility, and efficiencies, and in turn, provide rapid return on investment.

Divide and conquer

Accelerators such as DPUs are not new. Graphical processing units (GPUs) are a common example of a popular accelerator. With their multiple-core processors, GPUs can tackle thousands of computational tasks — making them a good fit for heavy-lift workloads like machine learning models and data analysis.

Yet, a multitude of factors has weighed down processing in the data center. The expanded adoption of hybrid/multiple clouds and

infrastructure-as-code have made the delivery of infrastructure services more complex and time-consuming. In addition, the use of microservices to support distributed applications has increased the burden on CPUs to manage network traffic.

The constant need to meet cybersecurity requirements also introduces latency. For example, network traffic must be passed through a firewall for security inspection before hitting the CPU, then back out through the firewall after processing — thus slowing services delivery.

DPUs are intelligent server cards that serve as the entry point to the network. They decide on the value of the services, and offload duties from the CPU to the DPU, thereby accelerating business application performance.

Yet, the benefits don't stop there.

DPUs improve capacity, security, scalability

"Another advantage is that the DPU expands the server's capacity," said Greg Caetano, integrated systems master engineering architect at Hewlett Packard Enterprise (HPE).

For example, IT managers can offload virtual machines (VMs) from a server's CPU to the DPU. Doing so frees up the CPU for other tasks and to accommodate more workloads for better utilization of capacity. This consolidation reduces total cost of ownership by mitigating the need to buy more hardware — in the data center and at the edge. Also, fewer servers equals less power consumption.

In addition, DPUs make it easier to scale. "DPUs solve problems that microservices introduce, where applications get broken down into distributed services," Gopalan said. "The same pieces of code are running in different parts of the data center, which introduces redundancy and increases traffic volume."

The DPU manages that traffic and ensures the right resources are directed to the right applications or systems.

Another benefit is improved security. IT can assign security functions to the DPU, so that in a sense, the server has its own private firewall, Caetano said. “That functionality is done right there at the server, minimizing traffic that would traditionally have to leave the server to be verified by the firewall.”

IT teams can assign the same levels of security protection to the DPU — such as software isolation, encryption, and Zero Trust principles — to protect each server from firmware and root-based attacks.

All these enhanced performance benefits give the business room to easily scale and grow, while modernizing the data center without adding capital expenditures for processing capacity.

An impactful DPU ecosystem

Gain a fully integrated server solution stack with operational simplicity, thanks to the partnership of HPE, VMware, and Intel — an ecosystem that is built on decades of joint engineering and collaboration.

HPE ProLiant with vSphere® Distributed Services Engine™ powered by Intel® Xeon® Scalable processors provides out-of-the-box functionality that delivers performance, security, and scalability for the business, as well as easy deployment and manageability for IT teams:

- Workload-optimized configurations installed on premises
- Flexible solutions that can package hardware and software in a per-unit model
- Adjustable stacks that meet targeted use cases — such as edge, virtual desktop infrastructure, and data management — as well as cost objectives.

Get a processing power boost

Processing demands will continue to increase right along with the growth of data, the growth of distributed IT infrastructure due to the hybrid workforce, and businesses’ constant need for innovation.

DPUs are ready to meet these demands, providing the extra processing power to offload the burden from CPUs and accelerate workloads.

To learn more, [click here](#).

