White Paper

The Rise of the DPU: Reconciling Resource Needs of Applications with Requirement for Scalable Network Functions

Sponsored by: Dell Technologies

Brad Casemore
September 2022

SITUATION OVERVIEW

Digital transformation has altered enterprise perceptions of how IT infrastructure relates to their business. Whereas IT infrastructure was formerly viewed as a cost center and a form of tactical overhead, organizations embracing digitization now understand that truly digital infrastructure is the strategic foundation for long-term organizational success.

Consequently, most organizations are modernizing their IT infrastructure. They are identifying and deploying the appropriate datacenter and edge platforms and systems to serve as their foundation for digital transformation, not only architecturally but operationally. This modernized, digital IT infrastructure must be capable of supporting cloud-native workloads as well as traditional applications.

One of the salient challenges in modernizing IT infrastructure involves ensuring that the new architecture can address the full range of requirements while also delivering operational simplicity. To complicate matters further, the benefit of operational simplicity must not come at the cost of sacrificing control and security.

Another challenge, which has accompanied the rise of virtualization and will certainly grow in complexity as enterprises embrace cloud-native containerization, involves the virtualization of network and security functions. While insertion and chaining of these functions are challenges in their own right, an added complication is that the processing of network and security functions and services claims valuable CPU cycles and resources that should be dedicated to business workloads. While important to the performance and protection of business workloads, network and security functions are secondary supporting elements that should serve the purposes of business workloads rather than detract from their effectiveness. What's more, network and security services have their particular resource requirements that must scale elastically in distributed infrastructure environments to support the ebbs and flows of digital business.

The solution obviously cannot involve eschewing or otherwise eliminating these network and security functions, which can individually and collectively consume a growing percentage of CPU resources. Indeed, virtualized and containerized network and security services are essential to digital infrastructure and must be accorded the resources they require to provide resilient and robust connectivity and strong security to the business. So what is the right architectural answer to meet the challenge of prioritizing business workloads for better performance and optimization while also ensuring that those workloads receive adequate support from integral network and security services?

September 2022, IDC #US49640322

Given the dimensions and profundity of the challenge, it is readily apparent that the traditional CPU-centric architecture cannot address the burgeoning requirements of the cloud era. An integrated, streamlined, and elastically scalable offload element is required to relieve overburdened CPUs, freeing processor and memory resources to serve application workloads.

In adding an offload complement to this modern IT architecture, we must ensure that it offers certain capabilities and characteristics, including:

- Hardware-based mitigation capabilities to minimize the impact of any low-level security vulnerabilities in the central execution environment
- Provision for a consistent operating model across distributed processing, addressing service quality issues that can be difficult to troubleshoot and remediate quickly
- Built-in hardware-based isolation, for separation of network and security functions as well as for management and control functions, protecting against attacks that might compromise the operating system or workload execution environment

A NEW APPROACH TO MODERN INFRASTRUCTURE ARCHITECTURE

The data processing unit (DPU), also called a SmartNIC, has risen to the secure scale-out challenges mandated by distributed IT architectures, which often feature disaggregation predicated on the principles and capabilities of compute and network virtualization.

In essence, a DPU is a programable system on a chip (SoC) device, with hardware acceleration and a CPU complex capable of processing data. Today the DPU, or SmartNIC (PCIe form factor), can be plugged into a server to support a range of processing offload functions.

Benefits of DPU Technology

In that DPUs are designed to operate independently of the CPU, the architectural result is that CPUs are aware of the presence of DPUs but do not control them. Introducing an abstraction layer to the operational state of the architectural platform, DPUs boot up using their own independent microcode or firmware or a lightweight hypervisor that treats its CPU subsystem like a virtual machine (VM) running in reduced-privilege mode.

Consequently, the DPU controls access to physical resources such as network interfaces, through which sensitive data can be accessed. Any payload executed on the CPU, including on the kernel itself, that must gain access to those resources must go through function-offload interfaces, presented in virtualized form to the operating system environment running on the CPU.

This architecture bifurcation allows the DPU to assume direct execution of network and security functions and services.

Benefits of this modern approach to infrastructure architecture include the following:

- **Improved and uniform resource utilization for business workloads.** Offloading of network and security services to the DPU ensures that CPU processing resources are fully available to enhance the performance of business workloads. Even the most advanced server platforms, with two or more host CPUs per system, can experience workload performance degradation as a result of resource constraints due to the parallel processing needs associated with infrastructure services, including network and security services. In offloading the compute-intensive network and security, as well as the management and control plane functionality, to the DPU and off the host CPU, business workloads receive optimal processing resources.
- **Protected architectural environment.** Security is enhanced through the ability to execute network and security services on the DPU, which ensures that payload code is isolated from other components executing on the host processor. Similarly, execution of management functions, in a separate physical device, provides physical isolation between management and payload components.
- **Ability to support offload of additional functions and payloads.** A long-term benefit of using DPUs is that they provide a foundational management fabric that spans physical, virtual, and containerized compute architectures, allowing for a CPU-agnostic approach to platform offload. DPUs can be particularly beneficial in shared cloud infrastructure environments where multitenancy and workload isolation are required to host a heterogenous mix of traditional and cloud-native workloads.
- **Automated extension of network policy.** The DPU is a natural extension for SDN architectures and network-virtual overlays, which can define policy for the creation and extension of specific network and security services that require dynamic offload for elastic scale, particularly useful during spikes in network demand and traffic and in providing mitigation or protection against network-security incidents.
- **Improved network performance and observability.** Offloading of network and security services to the DPU enables the use of hardware acceleration to enhance the performance of virtual networking and security functions such as overlay networking, load balancing, and firewalling. Furthermore, network visibility and analytics functions implemented on the DPU provide comprehensive visibility for all traffic flows directly on the SmartNIC without any complex network taps or spans, simplifying security operations and micro-segmentation planning.

DELL TECHNOLOGIES' SMARTDPU SOFTWARE SOLUTIONS

In response to the rise of the DPU, VMware announced that its Project Monterey would take an ecosystem approach, supporting DPUs from several providers such as Intel, NVIDIA, and Pensando Systems, integrating server offerings such as Dell Technologies.

VMware's Project Monterey has been architected so that network functions and processes such as distributed switching, routing, and network observability and analytics are offloaded to the SmartNIC. The result is that resource allocations and traffic bypass the server CPU. Project Monterey will also facilitate installation of a version of VMware's ESXi and NSX on the DPU itself, offloading the attendant CPU resources from the server to the DPUs. This not only frees CPU resources for business workloads but also provides an extra layer of security and network performance enhancements.

Dell Technologies and VMware have worked together to engineer and develop a range of modern IT architectures and infrastructure to support enterprise cloud initiatives and digital transformation. In the realm of DPUs, Dell Technologies has partnered with VMware in relation to Project Monterey, resulting in an integrated DPU solution portfolio that extends across hardware, operating system, and platform management. Dell also provides extensive global services and support for its DPU solutions.

In key respects, Dell Technologies' work with VMware on Project Monterey can be viewed as the latest development in Dell's Open Networking portfolio. Dell's Open Networking portfolio was first established a few years ago through network disaggregation, which offered a choice of several modern NOSs that allowed customers to decouple selection of network software (the network operating system) from their choice of Dell network hardware. This gave customers the flexibility to select an NOS that was best suited to their environment and requirements, often allowing for another NOS to be adopted as requirements evolved. Dell later extended its Open Networking to the edge, where Dell offered a choice of SD-WAN software overlays that could be run on Dell edge platforms. More recently, Dell has developed and released its own version of SONiC, the highly scalable open source NOS first developed by Microsoft and used extensively in Microsoft's own datacenters as well as in the datacenters of other cloud providers, service providers, and large enterprises. Now Dell is further extending its Open Networking to DPUs/SmartNICs supported by VMware's Project Monterey, giving customers choice and flexibility in their adoption of DPU technology to meet varying use cases and requirements.

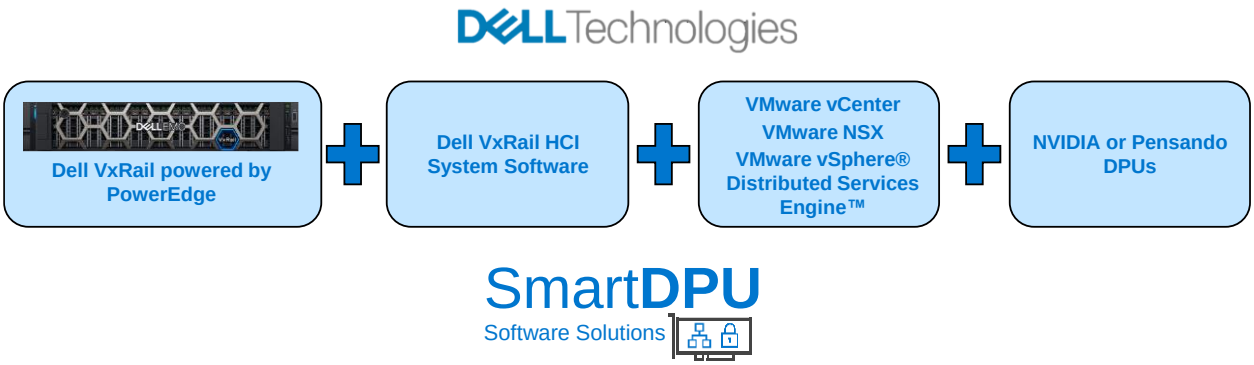
SmartDPU Software Solutions by Dell Technologies also represents an evolution of IT infrastructure disaggregation through the incorporation of DPU technology into an extensible platform, with the initial release built to support Dell VxRail. As the only jointly engineered HCI solution with VMware, the platform is designed to enable enterprises to effectively transition to a modernized architecture for IT infrastructure, inclusive of automated technologies provided by an ecosystem of DPU suppliers.

The architectural components (see Figure 1) of the Dell Technologies SmartDPU Software Solutions include the following:

- **Dell Technologies platform and server support.** This includes VxRail E660F running on the Dell PowerEdge R650, VxRail P670N running on the Dell PowerEdge R750, and VxRail V670F running on the Dell PowerEdge R750.
- **Dell software support.** Software support is inclusive of VxRail HCI System Software, the Integrated Dell Remote Access Controller (iDRAC), and OpenManage Integration for VMware vSphere (OMIVV).
- **VMware software support.** VMware support is provided for vSphere ESXi 8.0 on included DPUs and vCenter 8.0, the vSphere Distributed Services Engine, and VMware's NSX.
- **DPU support.** Initial DPU support is provided for the Nvidia Mellanox BlueField-2 (25GbE FH only) and for AMD/Pensando DSC-25/DSC-100 (25GbE and 100GbE, both FH only).

FIGURE 1

The Architectural Components of Dell Technologies SmartDPU Software Solutions



Source: Dell Technologies, 2022

Benefits

Organizations can realize the following benefits with the SmartDPU Software Solutions by Dell Technologies:

- Optimize performance for demanding applications, as well as provide support for accelerated network infrastructure services and data processing functions, using DPU-offload capabilities and hardware acceleration engines.
- Reduce operational complexity across architectures through the use of a common environment and operational model, providing for automated and uniform operational processes, enhanced network observability, and full life-cycle management from Day 0 and Day 1 through Day 2/N.
- Utilize Dell's global support and services, extending from presales consulting through post-deployment service and support.

CHALLENGES/OPPORTUNITIES

DPU architectures represent a relatively new technological phenomenon, especially for enterprise customers, and IT buyers and practitioners might be inclined to approach the technology cautiously. Given the technology's relative nascence and unfamiliarity, prospective enterprise customers will need to be educated about use cases, business value, and technology and resource requirements.

Another notable challenge for enterprise customers could relate to role-based demarcations of responsibility, especially considering that the DPU is where various traditional operational disciplines — server and virtualization admins and network teams — all come together. Network personnel will have to work with server/virtualization teams to ensure a shared understanding of what needs to be offloaded and when and how offload should be performed. As automated offload policies must be defined and implemented, skills upgrades might also be required.

By extension, the foregoing also represents challenges for any vendor providing DPU-based technology solutions to enterprise customers. Nonetheless, the benefits of deploying the technology, as discussed in the preceding sections of the document, are compelling, and most organizations will find that the challenges can be overcome.

CONCLUSION

The growing wave of virtualized and containerized applications has strained the capacity of CPU-centric architecture. Many organizations have recognized that an integrated, streamlined, and elastically scalable offload element is required to relieve overburdened CPUs, thus ensuring that processor and memory resources are allocated to meet the dynamic needs of application workloads.

The data processing unit, also called a SmartNIC, has emerged to meet the secure scale-out challenges mandated by distributed IT architectures, which often feature disaggregation predicated on the principles and capabilities of compute and network virtualization. The benefits that accrue from the use of DPU technology include consistent and uniform resource utilization for business workloads, a protected architectural environment for scalable processing, a capacity to support additional functions and payloads (such as accelerated networking and improved network observability), and the ability to define and enforce automated extensions of network policy via the DPU.

If Dell Technologies can successfully surmount the challenges noted in this paper, its SmartDPU Software Solutions, building on the flexibility and choice of Dell's Open Networking portfolio, appears well placed to help a wide range of customers derive value across an expanding array of use cases.

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

140 Kendrick Street
Building B
Needham, MA 02494
USA
508.872.8200
Twitter: @IDC
blogs.idc.com
www.idc.com

Copyright Notice

External Publication of IDC Information and Data – Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Vice President or Country Manager. A draft of the proposed document should accompany any such request. IDC reserves the right to deny approval of external usage for any reason.

Copyright 2022 IDC. Reproduction without written permission is completely forbidden.