



POWERING  
UNPRECEDENTED  
DATA CENTER  
TRANSFORMATION  
WITH NVIDIA  
BLUEFIELD



# THE DATA CENTER IS THE NEW UNIT OF COMPUTING

Cloud computing and AI are driving fundamental changes in the architecture of modern data centers. Infrastructure virtualization, which is the foundation of cloud computing, has become mainstream in nearly every data center. Software-defined data centers, despite providing great flexibility and manageability, utilize up to 30 percent of their CPU cores for infrastructure, consuming cores that could otherwise run business applications. With the end of Moore's Law, CPU-centric infrastructures have reached their limits and can't deliver on the stringent performance requirements of emerging technologies, including data science, Internet of Things (IoT), and 5G wireless networks.

The advent of containerized applications and the multi-tenant nature of the cloud are driving exponential growth in east-west communication. As application processing and data velocity scale beyond what a single computing platform can handle, data centers have become the new unit of computing. Lastly, the broad and ever-changing cyberthreat landscape is forcing cloud-minded organizations to continuously assess and adapt their security postures while implementing zero-trust defense initiatives.



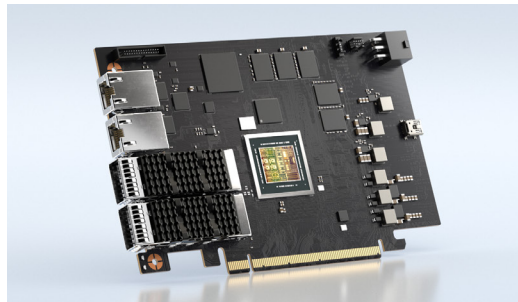
# NVIDIA BLUEFIELD: POWERING UNPRECEDENTED DATA CENTER TRANSFORMATION

The data processing unit (DPU) is a new class of processor that accelerates data center infrastructure processing. The **NVIDIA® BlueField® DPU** is a data center infrastructure on a chip that combines a high-speed networking interface with powerful, software programmable Arm® cores, enabling breakthrough networking, storage, and security performance. Many of the world's top server manufacturers offer or are building systems powered by BlueField DPUs. BlueField offloads, accelerates, and isolates a broad range of software-defined infrastructure services that previously ran on the host CPU, overcoming performance and scalability bottlenecks and eliminating security threats in modern data centers.

BlueField DPUs transform traditional computing environments into secure and accelerated data centers, allowing organizations to efficiently run data-driven, cloud-native applications alongside legacy applications. By decoupling the data center infrastructure from business applications, BlueField DPUs enhance data center security, streamline operations, and reduce total cost of ownership.

The **NVIDIA DOCA™ software framework** enables developers to rapidly create applications and services for the BlueField DPU. Built with a future-proof, API-driven mindset, DOCA makes it easy to leverage DPU hardware accelerators, providing unparalleled data center performance, efficiency, and security. DOCA plays a key role in creating a broad ecosystem of DPU-powered applications in the areas of networking, storage, cybersecurity, manageability, and more.

This paper articulates some of the most prominent DPU applications, demonstrating how BlueField enables the future of data centers.



NVIDIA DPU BlueField

# CLOUD COMPUTING

More than a decade old, the global cloud computing market has experienced rapid growth. According to Gartner<sup>1</sup>, the worldwide infrastructure-as-a-service (IaaS) market expanded during the pandemic, growing 40.7 percent in 2020 to total \$64.3 billion, up from \$45.7 billion in 2019. Providing on-demand compute power, data storage, and network connectivity for organizations of all sizes and industries, cloud computing largely relies on the sharing of resources to achieve coherence and economies of scale. This core principle has guided the world's largest cloud service providers (CSPs) to be the first adopters of DPU technologies, optimizing the data center infrastructure stack for incredible efficiency and scalability.

## Fueling the Trend Toward Software-Defined Data Centers

As more enterprises shift their workloads to the cloud, they aim to bring the same cloud operating model and economics to new, on-premises data centers. NVIDIA BlueField DPUs are ideally positioned to provide best-in-class performance, efficiency, and security for a wide variety of cloud IaaS delivery models.

<sup>1</sup> [www.gartner.com/en/newsroom/press-releases/2021-06-28-gartner-says-worldwide-iaas-public-cloud-services-market-grew-40-7-percent-in-2020](https://www.gartner.com/en/newsroom/press-releases/2021-06-28-gartner-says-worldwide-iaas-public-cloud-services-market-grew-40-7-percent-in-2020)

In virtualized cloud environments, the DPU offloads, accelerates, and isolates the hypervisor software and infrastructure processing from the tenant's applications. This delivers line-rate network performance, storage elasticity, and enhanced security while freeing up CPU cores to run business applications.

## Bridging the Gap Between Bare-Metal and Cloud

Bare-metal clouds have traditionally been hard to provision, secure, and manage. BlueField DPUs provide cloud operations teams with a trusted environment to control their bare-metal servers independently from their tenants. Deploying BlueField has transformed bare-metal clouds into performance-optimized, secure, and agile computing infrastructures. Many enterprises are deploying bare-metal server infrastructures powered by Kubernetes to run modern applications. BlueField DPUs enable a secure and accelerated cloud infrastructure for running cloud-native applications at massive scale.



# SOFTWARE-DEFINED NETWORKING

Modern data center networks are increasingly becoming virtualized and provisioned as a service. These software-defined networks (SDNs) deliver greater flexibility and control, enabling enterprises to easily scale from the premises of one data center to multi- and hybrid-cloud environments.

The longstanding conflict between implementing SDN technologies and hardware-accelerated networking—namely single-root input/output virtualization (SR-IOV)—has primarily been due to fundamental differences between the two: SDN abstracts the underlying network interface card (NIC) hardware and provides a virtualized network device to a virtual machine/container pod. However, SDN also utilizes a considerable amount of CPU processing capacity that could otherwise be used for running business applications. SR-IOV does the exact opposite, providing a direct interface between the virtual machine/container pod and the NIC hardware. It bypasses the host's CPU and operating system, freeing up expensive CPU resources from IO tasks. With that said, data center builders tend to avoid SR-IOV as it eliminates the live migration functionality and may require installing a network driver on the guest operating system. This discord between SDN and SR-IOV

accelerated networking has forced data center architects and network engineers to prioritize one over the other, often at the cost of poor application performance or inflexible system design and higher total cost of ownership (TCO).

## **Best of Both Worlds: Software-Defined, Hardware-Accelerated**

NVIDIA Accelerated Switch and Packet Processing® (ASAP<sup>2</sup>) technology, which is built right into the BlueField DPU, fully offloads and isolates the SDN layer and accelerates the network data path. BlueField's ASAP<sup>2</sup> technology delivers the best of both worlds: the performance and efficiency of bare-metal servers with the flexibility and agility of SDN. The ASAP<sup>2</sup> technology stack provides a range of network acceleration capabilities that gives customers the ability to choose between maximum performance (SR-IOV) or support for legacy environments (VirtIO), all while maintaining the SDN layer.

The ASAP<sup>2</sup> technology is integrated upstream into both the Linux kernel and a range of leading SDN frameworks. Leveraging ASAP<sup>2</sup>, organizations can build distributed firewall capabilities, which enforce network security policies that are orchestrated through the SDN layer.



# COMPOSABLE STORAGE INFRASTRUCTURE

As data volumes continue to increase, enterprises are prioritizing software-defined storage (SDS) technologies to meet their requirements for flexibility, agility, ease of management, and lower costs. SDS enables users and organizations to uncouple or abstract storage resources from the underlying hardware platform for greater efficiency and faster scalability by making storage resources programmable.

While SDS technologies provide obvious advantages over traditional storage infrastructures, they're often limited in the performance they deliver to applications in terms of read/write speed, or input/output operations per second (IOPS), and latency. The degradation in performance is normally due to the software abstraction of the storage hardware. Modern applications, including professional visualization, deep learning, and content delivery networks

(CDN), have stringent performance requirements that SDS solutions can't meet. Therein lies the conflict between what is best for the application and what is most easily composable, maintainable, and cost-effective.

## Elastic Block Storage Made Simple

With NVIDIA BlueField Software-Defined Network Accelerated Processing (BlueField SNAP™) technology, enterprises can use hardware virtualization to achieve all the operational benefits and economics of software-defined storage, with the performance of high-end, directly connected storage. Deploying the DPU allows elastic provisioning of networked block storage to BlueField SNAP, storage becomes virtualized, thinly provisioned, and protected and can be migrated between servers as needed, providing both capex and opex savings.



# CYBERSECURITY

As enterprises adopt cloud and edge computing technologies, cybersecurity remains a top concern for IT and business leaders alike. Modern data center architectures—given the ubiquitous virtualization of their infrastructures—have larger attack surfaces that expose enterprises to cyberthreats. Perimeter cybersecurity solutions that were adequate in the past are no longer equipped to provide comprehensive protection for modern, cloud-native data centers. Multi-tenant environments with distributed, containerized applications make network isolation techniques ineffective against east-west cyberthreats. These types of attacks are increasingly common; once a network has been compromised, the attacker will attempt to move laterally across the network, hopping from one server to another.

## Taking Security to a Whole New Level

NVIDIA BlueField DPUs transform data center security by delivering innovative hardware engines that offload, accelerate, and isolate security across the entire stack for every host. These engines protect data center infrastructures and are geared toward the following functions:

- > Enhancing platform security
- > Accelerating encryption and decryption at line speed
- > Performing stateful packet filtering and enforcing distributed security policies
- > Storing and managing keys in hardware and accelerating public key infrastructure (PKI) exchanges
- > Detecting malicious code and mitigating attacks

BlueField DPUs enable security functions to run fully isolated from the host's CPU and operating system. This isolation is key in making BlueField work best for zero-trust security solutions, as it delivers the needed separation of the security functions from the host, while delivering unmatched performance. In the event a host has been compromised, the separation between the security functions and the compromised host helps stop the attack from spreading further.

BlueField DPUs also act as sensors for **NVIDIA Morpheus**, an open AI cybersecurity framework. Morpheus can receive rich, real-time network telemetry from every DPU-accelerated server in the data center without impacting performance. Integrating the DPU's real-time telemetry into Morpheus brings the world's best AI computing to take on complex network security challenges.



10.20.16.248 -> 10.244.0.59  
SI: none

```
{  
  "who": "32205660 pVycjWHH",  
  "thing": "within there lmXxovGN",  
  "condition": "NhpdtcvS 89234632",  
  "key": "nearly appear n1KLSr8A 41142934 actually QZnPjQyd"  
}
```

# DATA SCIENCE AND AI

The growing prevalence of GPU-accelerated computing for data science and AI workloads increasingly relies on robust and powerful network infrastructures. As application processing scales far beyond a single computer to an entire data center, high-throughput, low-latency network connectivity is key to harnessing the power of GPUs.

## Pushing Today's Compute Architecture Beyond Its Limits

NVIDIA BlueField DPUs bolster a range of hardware acceleration engines for GPU-to-GPU or GPU-to-storage communications.

**GPUDirect®** remote direct-memory access (RDMA) technology is an important BlueField capability that unlocks high-performance network connectivity to feed GPUs with data. GPUDirect RDMA (GDR) allows efficient,

zero-copy data transfers between GPUs while facilitating the hardware engines in the BlueField application-specific integrated circuit (ASIC).

**GPUDirect Storage (GDS)** is a new technology that provides a direct path to local/remote storage—like non-volatile memory express (NVMe) or NVMe over Fabrics (NVMe-oF)—and GPU memory. BlueField enables this direct communication within a distributed environment, when the GPU and storage media are not hosted in the same enclosure. BlueField's GDS increases bandwidth, lowers latency, and increases capacity between storage and GPUs. This is important as dataset sizes no longer fit into system memory and data IO to the GPUs becomes the runtime bottleneck. Enabling a direct path alleviates this bottleneck for scale-out data science and AI workloads and provides incredible application performance.





# VIDEO STREAMING

With the increasing demand for video content around the world, the media and entertainment market is projecting tremendous growth in 4K and 8K ultra-high-definition (UHD) video. Among a wide range of emerging trends in video production are high dynamic ranges and higher frame rates, which are reinventing the quality of video.

Video production studios gearing up for the anticipated demand are gradually moving their proprietary serial digital interface (SDI)-based video production solutions to next-generation Internet Protocol (IP)-based infrastructures. Most of these production houses are adopting commercial-off-the-shelf (COTS) server hardware, which presents various challenges around compliance, scale, and performance, while lacking the economical and operational benefits of virtualized infrastructures and cloud-based solutions.

## Enabling Cloud-Scale Video Solutions

NVIDIA BlueField DPUs transform video streaming in the cloud, data center, and edge, enabling high-performance network infrastructure to deliver HD and UHD video streams at massive scale. Powered by the purpose-built, Society of Motion Picture and Television Engineers (SMPTE)-compliant **NVIDIA Rivermax SDK**, BlueField provides zero-copy, frame-level video transport that reduces CPU cycles by offloading and accelerating video streaming to the DPU and bypassing the system kernel. What's more, BlueField DPUs provide built-in hardware-based video acceleration that delivers superior performance in GPU environments.

For production studios, deploying BlueField DPUs lets them repurpose the same hardware infrastructure for diverse workloads, enabling successful, cost-effective delivery of next-generation UHD video streaming with unprecedented flexibility and agility.



# INTELLIGENT SERVICES AT THE EDGE

5G networks are ushering in a new era in wireless communications that delivers 1,000X the bandwidth and 100X the speed at 1/10th the latency of 4G. 5G, which allows for millions of connected devices per square kilometer, is being deployed as an alternative to WiFi at edge locations like factories and retail stores. These applications demand a new network architecture that's fully software-defined, dynamically reconfigurable, easily deployed, and easily managed to guarantee a specific quality of service.

## Converging AI and 5G Capabilities

**NVIDIA AI-on-5G** is an innovative computing platform that leverages NVIDIA BlueField DPUs with NVIDIA's extensive AI software library and the **NVIDIA Aerial™ SDK** for 5G. The AI-on-5G platform speeds time to deployment and fuels an array of ultra-low-latency enterprise AI projects, including precision robots, automated guided vehicles, and digital twins for product development and manufacturing.

BlueField's role is to help every enterprise deliver AI applications with industry-leading performance and data center security. It's optimized for 5G connectivity and multi-tenant, cloud-native environments, offering software-defined, hardware-accelerated networking, storage, security, and management services at the edge.



# CLOUD-NATIVE SUPERCOMPUTING

Today's most challenging high-performance computing (HPC) and AI workloads rely on the power of supercomputers to transform the world with research and results that benefit science, industry, and society. NVIDIA cloud-native supercomputing takes supercomputing to a whole new level.

**NVIDIA cloud-native supercomputing** combines the power of an HPC cloud with a TOP500-class supercomputer, along with the security and ease of use of cloud computing services, to enable delivery of the most demanding HPC and AI workloads.

## **Bare-Metal Performance with Multi-Tenant Isolation**

The NVIDIA cloud-native supercomputing platform leverages the NVIDIA BlueField DPU architecture with high-speed, low-latency NVIDIA InfiniBand networking to deliver optimal bare-metal performance, while natively supporting multi-node

tenant isolation. Handling tasks such as storage, systems management, and multi-tenant security, the DPU enables native cloud services that let multiple users securely share resources without loss in application performance. By offloading the CPU to focus on processing tasks, systems are able to maintain peak efficiency running HPC and AI workloads, maximizing overall system performance.

HPC and AI communication frameworks and libraries play a critical role in determining application performance. Due to their latency and bandwidth-sensitive nature, offloading the libraries from the host CPU or GPU to the BlueField DPU creates the highest degree of overlap for parallel progression of communication and computation. It also reduces the negative effects of operating system jitter while dramatically increasing application performance—key to enabling the next generation of supercomputing architecture.



# TRANSFORM YOUR DATA CENTER

**NVIDIA BlueField DPUs** are a new class of data center processors, optimized for infrastructure processing and transforming traditional computing environments into secure and accelerated data centers. By offloading, accelerating, and isolating the data center infrastructure, BlueField boosts application performance, enhances data center security, and streamlines IT operations.

[Learn more](#) about BlueField data processing unit

© 2021 NVIDIA Corporation and affiliates. All rights reserved. NVIDIA, the NVIDIA logo, Accelerated Switch and Packet Processing, Aerial, BlueField, BlueField SNAP, DOCA, and GPUDirect are trademarks and/or registered trademarks of NVIDIA Corporation and affiliates in the U.S. and other countries. Other company and product names may be trademarks of the respective companies with which they are associated. All other trademarks are property of their respective owners. DEC21.



**nvidia**