



ACCELERATED COMPUTE INFRASTRUCTURE FOR ENTERPRISE IT

Addressing the Challenges of Modern Enterprises

Creative and technical professionals across industries face increasingly complex problems as they produce more data and create higher-quality content faster than ever before. These challenges are compounded exponentially as globally distributed teams continue to work remotely. Data centers must now provide the graphics and compute power that professionals need to tackle multiple visual computing workloads spanning the enterprise, from rendering and engineering simulation, to interactive graphics on virtual workstations. To drive visual computing workflows, professionals require incredibly powerful computing solutions that support the latest technologies and are capable of powering diverse applications and multiple workloads from the data center. Subsequently, the IT teams supporting these professionals require a secure, scalable, and easy-to-manage solution that delivers the performance of a physical workstation, enables real-time collaboration, and offers the flexibility to work from anywhere.

IT Challenges for Modern Applications

- > Modern applications are resource hungry. AI training requires a vast number of compute cycles, and AI inference often demands real-time response. Data analytics can involve huge transfers of data across multiple systems. And product designers need more graphics capabilities than ever. Traditional servers with only CPUs and commodity networking aren't able to provide the compute power needed to run these applications.
- > Modern applications also frequently require a cloud-native architecture, with many microservices organized into complex workflows. This creates challenges for management, scalability, security, and visibility.
- > Companies often end up deploying these applications into one-off, single-purpose clusters or in the cloud. This results in operational overhead, and these silos don't always adhere to IT standards for visibility, security, and governance.

NVIDIA EGX Platform for Enterprise IT

The NVIDIA EGX platform provides a way for customers to run both traditional and modern applications on a single high-performance, cost-effective, and scalable infrastructure. It brings together compute and graphics acceleration, high-speed secure networking, and enterprise-grade management in the leading enterprise data center servers, built and sold by NVIDIA partners. This platform supports a vast collection of accelerated applications that enables users to become productive immediately. It can also be easily integrated into existing industry-standard IT and DevOps frameworks for ease of management, deployment, operation, and monitoring.

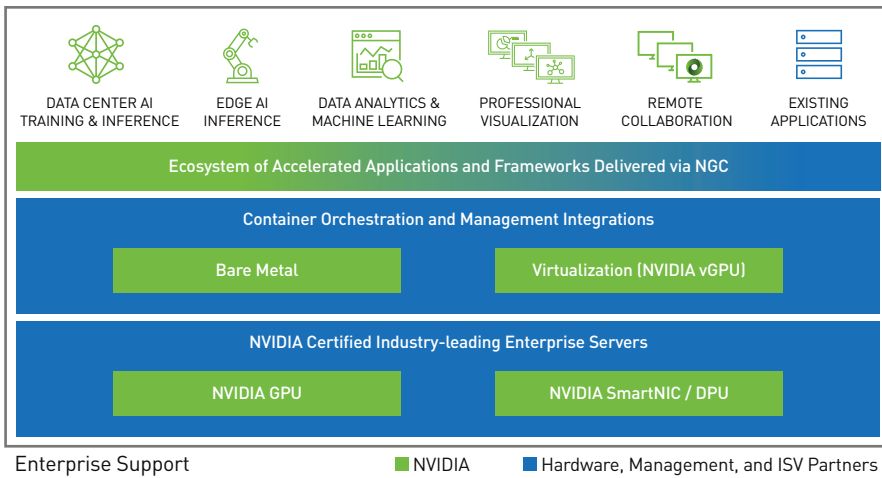
Run diverse applications-both traditional and modern-on a single high-performance, secure, cost-effective, and scalable infrastructure.

KEY TECHNOLOGIES

- > NVIDIA GPUs
- > NVIDIA® adapters, switches, and cables
- > NVIDIA virtual GPU (vGPU)
- > NVIDIA EGX™ stack
- > NVIDIA NGC™ catalog

BENEFITS

- > **Cost Savings:** NVIDIA's continuous full-stack innovation across hardware and software means customers can realize ROI much faster. And by running modern and traditional applications on a shared pool of resources, computing silos are eliminated and fewer systems are needed, reducing costs.
- > **Enterprise-Grade infrastructure:** IT can purchase mainstream servers that are optimally configured for accelerated computing.
- > **Unified architecture:** A single architecture accelerates modern applications across many types of workloads. Developers in all industries can become productive quickly with frameworks, optimized libraries, SDKs, integrated compilers, pre-trained models, inference optimizers, and other software.



Key Elements of the EGX Platform for Enterprise IT

GPUs for Compute Acceleration

NVIDIA's portfolio of GPUs deliver record-setting performance on a diversity of workloads: data center training and inference, edge inference, data analytics, professional visualization, remote collaboration, and more. They're built on a unified architecture, supported by a vast array of software libraries, programming language integrations, and a community of more than 2 million developers.

NVIDIA-Certified Systems

NVIDIA works with partners to build preconfigured servers with this platform for a variety of use cases. These servers are backed by enterprise-grade support, including direct access to NVIDIA experts, minimizing system downtime and maximizing user productivity. NVIDIA-Certified Systems™ are optimally designed to run modern applications in enterprise data centers.

Networking Solutions for I/O Acceleration

NVIDIA offers a comprehensive networking fabric solution with adapters and switches for various target use cases, along with other key hardware and software technologies. This enables data to be moved right to where it needs to be quickly and securely—from storage to memory to processor—greatly reducing bottlenecks in compute processing pipelines.

Enterprise Management Integrations

The platform is designed to be managed using standard data center and DevOps tools and frameworks. NVIDIA vGPU software combines the management and security benefits of server and desktop virtualization with the performance benefits of GPU acceleration. The NVIDIA EGX stack enables the automated use of GPUs and network acceleration in a Kubernetes cluster.

Ecosystem of Accelerated Applications

NVIDIA supports the largest collection of software to help users benefit from accelerated computing quickly in a wide variety of areas. The NVIDIA NGC catalog offers libraries, software development kits (SDKs), and toolkits for writing accelerated code; pre-trained models and frameworks for building accelerated applications; and container-based deployment services for putting applications into production. With hundreds of NVIDIA-accelerated commercial applications also available, customers can immediately benefit from systems running the EGX platform.

- > **Future-proofed:** This new data center architecture will help businesses prepare for the future, where the majority of applications will be hardware-accelerated.

NVIDIA-CERTIFIED SYSTEMS

- > Confidently deploy scalable hardware and software solutions that securely and optimally run accelerated workloads.
- > Learn more about accelerated servers at [pny.com/certified-systems](https://www.pny.com/certified-systems)



Learn More

Learn more about accelerating the enterprise at www.pny.com/ai-enterprise