



NVIDIA DGX SYSTEMS

Demystify AI with World-Class Expertise

Artificial Intelligence (AI) is enabling business transformation at a rapidly increasing rate, delivering innovation that can optimize operations, improve customer experiences, and speed up outcomes. While AI is expected to create an estimated \$16 trillion of GDP growth by 2025, the path to ROI with AI can still be elusive for many organizations.

NVIDIA® DGX™ systems give you direct access to world-class expertise that can jumpstart your initiative with field-proven experience and enterprise-grade support.



GETTING STARTED,
MADE SIMPLER



PROVEN RESULTS,
FASTER



OPERATIONAL
PEACE OF MIND

Getting Started, Made Simpler

If you're new to AI platforms, finding the right balance of data center resources, combined with engineering an optimized software stack and ensuring a productive data science workflow, can be challenging and delay time to insights.

With NVIDIA DGX, you have direct access to a uniquely-skilled team that's prepared to get you started using prescriptive guidance, design expertise, and access to our fully-optimized AI stack. DGX customers also benefit from exclusive team collaboration features that help data scientists enjoy effortless productivity without wasting time on non-value added efforts.

PRIVATE CANCER CENTER	HEALTHCARE PROVIDER	AI-AS-A-SERVICE COMPANY
<p>50% Reduction in time to install 8 DGX-1 units</p>	<p>90 Minutes Unboxing to power-up</p>	<p>3 Days Expert training time for smooth custom Docker experience</p>

Many organizations have dramatically shortened their time-to-insights by tapping into the expertise of their DGX team—getting started in hours, instead of months. For example, the world’s oldest and largest private cancer center had eight DGX-1 units up and running within eight hours, a 50% time reduction in installation. With the power and speed of their new platform, they’ve surpassed competitors, receiving results from AI experimentation on tumor segmentation, digital pathology, and more—faster than ever before.

Another healthcare provider went from unboxing to powering-up in 90 minutes and an AI-powered product-as-service company benefited from three days of expert training on their Linux environment, helping to ensure a smoother experience with their Docker containers.

Proven Results, Faster

Deploying an AI platform can be complex and time-consuming, including the effort required to build a platform and engineer a software stack that’s optimized for maximum performance.

Our global team of DGX AI experts has the most extensive track record of field-proven deployments, backed by the world’s most advanced AI infrastructure, NVIDIA DGX SATURNV. This is a massive DGX-based AI infrastructure comprised of over 1500 systems, offering a valuable proving ground for ensuring that we can help you maximize performance for the fastest insights from data.

Thousands of NVIDIA AI experts, combined with a \$17 billion investment in GPU-accelerated computing, gives us the unique ability to help ensure your AI success—from concept to production. Every DGX solution architect has access to proven best practices and valuable lessons-learned, so that you can get results faster and focus on innovation, instead of systems integration and troubleshooting.

“Delivering near real-time data science is a game changer when it comes to making sense of our network data. Before, we had access to terabytes of data daily with no efficient way to gain insights. Now, every time we look at the data, we see something new that we can take immediate advantage of. We are able to build models to predict high-surge Wi-Fi usage, offload access points swiftly, and streamline operations to save millions of dollars.”

- **Jared Jitter**, Director of Wireless Engineering, Charter Communications

GLOBAL TECHNOLOGY FIRM IN DIGITAL MEDIA	WORLD-LEADING MEDICAL RESEARCH CENTER	LEADING TELECOM SERVICE PROVIDER COMPANY
<p>1.5X</p> <p>Boost in images per second on DGX-optimized TensorFlow</p>	<p>4X</p> <p>Spike in images per second using cuDNN best practices</p>	<p>10X</p> <p>Faster machine learning model training for network health monitoring</p>

Many organizations realize faster time to insights working directly with DGX experts. A global technology firm specializing in digital media saw a 1.5X boost in training on DGX-optimized TensorFlow. Compared with 1680 images per second on their home-grown TensorFlow software stack, they achieved 2600 images per second on the DGX-optimized stack, using ResNet50.

A world-leading medical research center saw 4X faster training, going from 619 images per second to 2600 images per second, leveraging best practices from the NVIDIA software team, which enabled the speedup needed for life-saving tasks.

In the telecom field, a leading service provider saw a 10X speedup on machine learning model training for network health monitoring. Working with one of our NVIDIA solution architects, they're now able to predict the future of their network operations, increase customer satisfaction, and positively impact their bottom line.

Operational Peace of Mind

Moving AI from prototype to production at scale can be challenging if your IT team isn't equipped to support the shift. With DGX, you can help insulate your IT operations from unstable code with an IT-proven solution, backed by enterprise-grade support, and a team of experts who can help ensure your mission-critical AI applications stay up and running.

NVIDIA Enterprise Support provides DGX customers with rapid response and timely issue resolution so that you can get the fastest path to answers. We provide access to remote hardware support, including diagnostic tools that bring systems back to full function. We also offer onsite hardware support and replacement, including advanced return merchandise authorization (RMA), for a next-business-day turnaround.

LEADING AI CONSULTING COMPANY	HEALTHCARE PROVIDER	CONVERSATIONAL AI PLATFORM COMPANY
<p>45X</p> <p>Quicker root cause analysis for framework performance issue</p>	<p>4X</p> <p>Faster training using model parallelism, instead of data parallelism</p>	<p>3 Weeks</p> <p>From misbehaving application to concrete solution</p>

Many DGX system customers have reaped the benefits of having a partner who can give them the fastest path to issue resolution. In one case, a leading AI consulting company experienced a 45X faster root cause analysis through NVIDIA Enterprise Support. For a deep learning framework performance issue, bugs were identified in open source (MXNet) in two days and fully addressed within two weeks. Similar scenarios have historically impacted developers for up to three months.

A client in the healthcare provider space, servicing over three million patients in 45 countries, worked with NVIDIA Enterprise Support to enable 4X faster training. They used model parallelism in place of data parallelism for multi-GPU training with TensorFlow Keras. The operation was completed within two weeks and required minimal code change and no model changes.

In another example, a conversational AI, or chatbot, solution company worked with NVIDIA Enterprise Support to address a resource exhaustion issue on their TensorFlow-based model. The lead engineer indicated that the bug was seriously affecting their business and putting them at risk of losing a considerable amount of money. To fix the situation, we provided them with an application workaround that resolved their complex application issue in the TensorFlow container—within three weeks.

NVIDIA DGX Expert Services and Benefits

SERVICE	BENEFIT
Best practices for implementing data science workflow.	Get best practices for petabyte scale data management, including data ingestion, indexing, transcoding, labeling, curation, and exporting for AI training.
AI infrastructure management software stack and documentation.	Gain access to the NVIDIA data center management stack. Learn best practices with tried and tested methods of deployment for optimized multi-node GPU clusters for AI and HPC environments.
Technical deep dive into infrastructure management.	Learn how to leverage the NVIDIA data center management stack to achieve peak datacenter utilization and performance, similar to the NVIDIA SATURNV supercomputer.
Consultation on maximizing GPU utilization in the data center.	Discover leading-edge performance tuning techniques from NVIDIA experts, covering topics such as configuring Kubernetes for GPUs, utilizing Volta Tensor cores and DALI to improve training, and tips for profiling and debugging for performance.
Assistance with AI workflow management.	See reference architectures for AI workflow management. Learn how to design and run large-scale reproducible multi-step workflows and perform end-to-end tracking of model versions, jobs, experiments, and workflows.
Multi-node AI training for improved cluster utilization.	Explore how to effectively parallelize deep learning training from NVIDIA Deep Learning Institute certified experts.

AI-powered business transformation can be challenging, but NVIDIA DGX offers direct access to AI experts who bring years of field-proven experience and expertise so that you can deliver business-impacting results much sooner.

Embark on your AI journey with NVIDIA DGX to stay ahead of competitors, transform your enterprise and deliver a differentiated experience for your customers. To learn more about NVIDIA DGX Systems, visit www.nvidia.com/dgx