NVIDIA DGX-1
THE ESSENTIAL INSTRUMENT FOR AI RESEARCH

The Challenges of Building a Platform for AI

Data scientists depend on computing performance to gain insights and innovate faster, using the power of deep learning and analytics. GPU technology offers a faster path to AI, but building a platform goes well beyond deploying a server and GPUs.

AI and deep learning can require a substantial commitment in software engineering. Such an investment could delay your project by months as you integrate a complex stack of components and software, including frameworks, libraries, and drivers. Once deployed, additional time and resources are continually needed as you wait for the ever-evolving open source software to stabilize. You'll also be waiting to optimize your infrastructure for performance and incur administrative costs that increase as the system scales.

The Fastest Path to Deep Learning

Inspired by the demands of AI and data science, NVIDIA® DGX-1™ fast-tracks your AI initiative with a solution that works right out-of-the-box so that you can gain insights in hours instead of months. With DGX-1 you can simply plug in, power up, and get to work, thanks to the integrated NVIDIA deep learning software stack. In addition to leveraging the Ubuntu Linux Host OS, popular among developers, DGX-1 also supports Red Hat for organizations that require seamless integration within their existing enterprise IT management tools. Now you can start deep learning training in as little as a day, instead of spending months integrating, configuring, and troubleshooting hardware and software.

Effortless Productivity

DGX-1 removes the burden of continually optimizing your deep learning software and delivers a ready-to-use, optimized software stack that can save you hundreds of thousands of dollars. It includes access to today's most popular deep learning frameworks, including NVIDIA DIGITS™ deep learning training application, third-party accelerated solutions, the NVIDIA Deep Learning SDK (e.g. cuDNN, cuBLAS, NCCL), NVIDIA CUDA® toolkit and Docker Engine Utility for NVIDIA GPU.
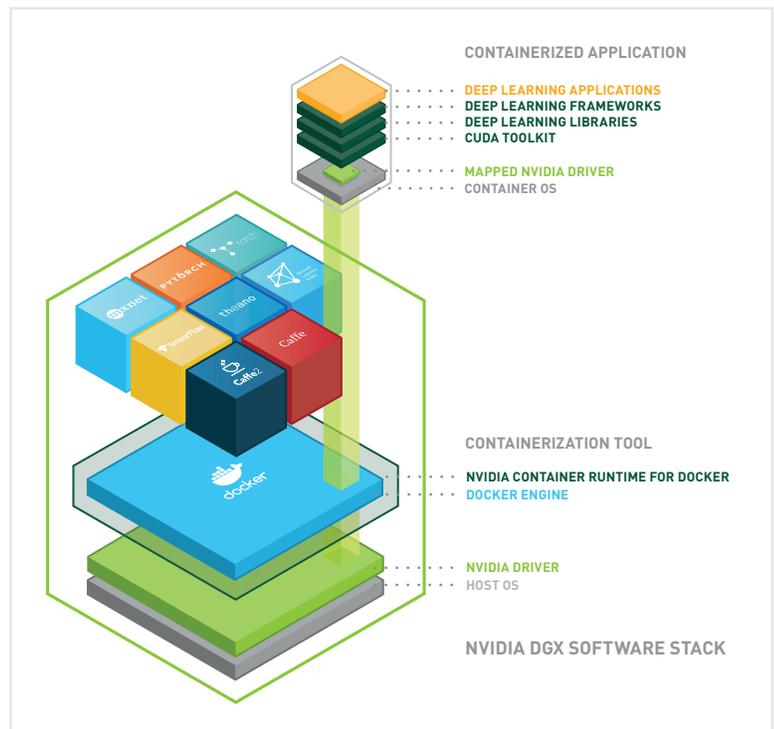

SYSTEM SPECIFICATIONS

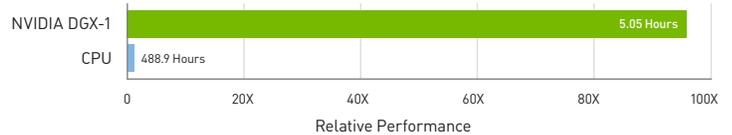|	
GPUs	**8X NVIDIA® Tesla® V100**
Performance (Mixed Precision)	**1 petaFLOPS**
GPU Memory	**256 GB total system**
CPU	**Dual 20-Core Intel Xeon E5-2698 v4 2.2 GHz**
NVIDIA CUDA Cores	**40,960**
NVIDIA Tensor Cores (on Tesla V100 based systems)	**5,120**
Power Requirements	**3,500 W**
System Memory	**512 GB 2,133 MHz DDR4 RDIMM**
Storage	**4X 1.92 TB SSD RAID 0**
Network	**Dual 10 GbE, 4 IB EDR**
Operating System	**Canonical Ubuntu, Red Hat Enterprise Linux**
System Weight	**134 lbs**
System Dimensions	**866 D x 444 W x 131 H (mm)**
Packing Dimensions	**1,180 D x 730 W x 284 H (mm)**
Operating Temperature Range	**5–35 °C**

Revolutionary AI Performance

While many solutions offer GPU-accelerated performance, only DGX-1 unlocks the full potential of the latest NVIDIA GPUs, including innovations like next generation NVIDIA NVLink™ and new Tensor Core architecture. With its performance-engineered deep learning software stack, DGX-1 delivers up to three times faster training speed than other GPU-based systems. With the computing capacity of 140 servers in a single system that integrates the latest NVIDIA GPU technology with the world's most advanced deep learning software stack, you can take advantage of revolutionary performance to gain insights faster than ever.

Access to AI Expertise

With DGX-1, you get access to NVIDIA's AI expertise, enterprise-grade support, extensive training, and field-proven capabilities that can jump-start your work for faster insights. Our dedicated team is ready to get you started with prescriptive guidance, design expertise, and access to our fully-optimized DGX software stack. You get 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.

CONTAINERIZED APPLICATION

- DEEP LEARNING APPLICATIONS
- DEEP LEARNING FRAMEWORKS
- DEEP LEARNING LIBRARIES
- CUDA TOOLKIT
- MAPPED NVIDIA DRIVER
- CONTAINER OS

CONTAINERIZATION TOOL

- NVIDIA CONTAINER RUNTIME FOR DOCKER
- DOCKER ENGINE

- NVIDIA DRIVER
- HOST OS

NVIDIA DGX SOFTWARE STACK

NVIDIA DGX-1 Delivers 96X Faster Deep Learning Training

NVIDIA DGX-1 5.05 Hours
CPU 488.9 Hours

| 0 | 20X | 40X | 60X | 80X | 100X |

Relative Performance

Workload: ResNet-50, BS=256, 90 epochs to solution | CPU: dual Xeon Platinum 8180 | GPU: 8x NVIDIA Tesla V100 32GB

For more information, visit **www.nvidia.com/dgx-1**

NVIDIA