NVIDIA DGX-1
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 GPU's.

AI and deep learning can require a substantial commitment in software engineering. An investment that 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, along with 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 and DGX-1 cloud management services. 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

NVIDIA 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, NVIDIA DIGITS™ deep learning training application, third-party accelerated solutions, the NVIDIA Deep Learning SDK (e.g. cuDNN, cuBLAS, NCCL), CUDA® toolkit, NVIDIA Docker and NVIDIA drivers.

SYSTEM SPECIFICATIONS

GPUs	**8X Tesla V100**	**8X Tesla P100**
Performance (GPU FP16)	**1 petaFLOPS**	**170 teraFLOPS**
GPU Memory	**128 GB total system**	
CPU	**Dual 20-Core Intel Xeon E5-2698 v4 2.2 GHz**	
NVIDIA CUDA® Cores	**40,960**	**28,672**
NVIDIA Tensor Cores (on V100 based systems)	**5,120**	**N/A**
Maximum Power Requirements	**3,200 W**	
System Memory	**512 GB 2,133 MHz DDR4 LRDIMM**	
Storage	**4X 1.92 TB SSD RAID 0**	
Network	**Dual 10 GbE, 4 IB EDR**	
Software	**Ubuntu Linux Host OS See Software Stack for Details**	
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	**10–35 °C**	

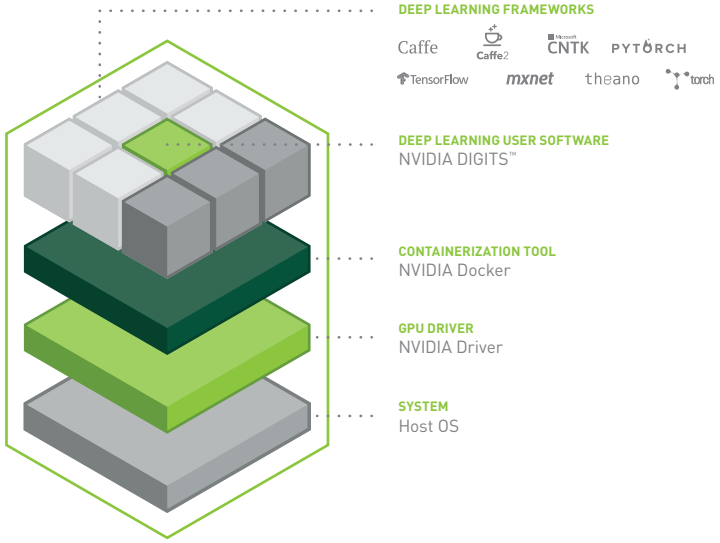
Revolutionary AI Performance

While many solutions offer GPU-accelerated performance, only NVIDIA DGX-1 unlocks the full potential of the latest NVIDIA GPU's like the NVIDIA Tesla® V100, including innovations like next generation 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 25 racks of conventional 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, powered by NVIDIA DGX-1.

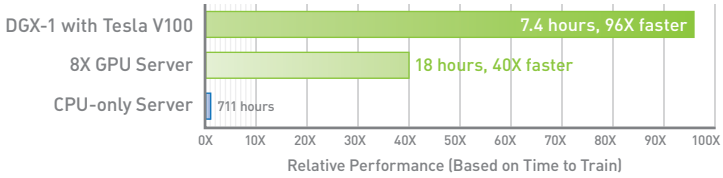
Investment Protection

Your AI initiative is critical to your organization's success, and dependent on a frequently optimized software stack and integrated hardware infrastructure. With today's rapidly evolving open source software and the complexity of libraries, drivers, and hardware, it's good to know that NVIDIA's enterprise grade support and software engineering expertise are behind every DGX-1. This software stack is built on years of R&D, innovation, and deep learning expertise, and maintained by monthly optimized framework releases. Also, NVIDIA's support includes software upgrades and priority resolution of critical issues; you can have peace of mind that your environment is tuned for maximized performance and uptime.

NVIDIA DGX-1 Software Stack



NVIDIA DGX-1 Delivers 96X Faster Training



Workload: ResNet50, 90 epochs to solution | CPU Server: Dual Xeon E5-2699 v4, 2.6GHz

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

© 2017 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo, and Iray are trademarks and/or registered trademarks of NVIDIA Corporation. All company and product names are trademarks or registered trademarks of the respective owners with which they are associated. Features, pricing, availability, and specifications are all subject to change without notice. SEP17

NVIDIA