



NVIDIA DGX-1 ARTIFICIAL INTELLIGENCE SYSTEM

The World's First AI Supercomputer in a Box

Get faster training, larger models, and more accurate results from deep learning with the NVIDIA® DGX-1™. This is the world's first purpose-built system for deep learning and AI-accelerated analytics, with performance equal to 250 conventional servers. It comes fully integrated with hardware, deep learning software, development tools, and accelerated analytics applications. Immediately shorten data processing time, visualize more data, accelerate deep learning frameworks, and design more sophisticated neural networks.

Iterate and Innovate Faster

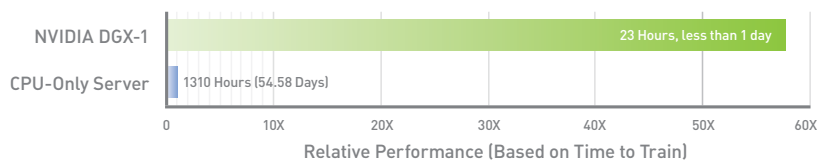
High-performance training accelerates your productivity giving you faster insight and time to market.



SYSTEM SPECIFICATIONS

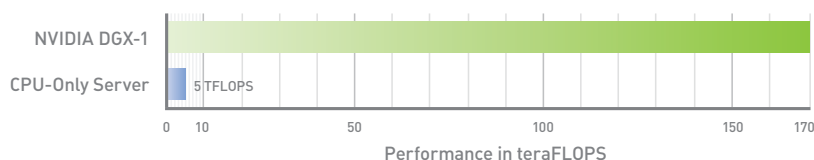
GPUs	8x Tesla P100
TFLOPS (GPU FP16 / CPU FP32)	170/3
GPU Memory	16 GB per GPU
CPU	Dual 20-core Intel® Xeon® E5-2698 v4 2.2 GHz
NVIDIA CUDA® Cores	28672
System Memory	512 GB 2133 MHz DDR4
Storage	4x 1.92 TB SSD RAID 0
Network	Dual 10 GbE, 4 IB EDR
Software	Ubuntu Server Linux OS DGX-1 Recommended GPU Driver
System Weight	134 lbs
System Dimensions	866 D x 444 W x 131 H (mm)
Packing Dimensions	1180 D x 730 W x 284 H (mm)
Maximum Power Requirements	3200W
Operating Temperature Range	10 - 30°C

NVIDIA DGX-1 Delivers 58X Faster Training



Caffe benchmark with VGG-D network, training 1.28M images with 70 epochs | CPU servers uses 2x Xeon E5-2699v4 CPUs

NVIDIA DGX-1 Delivers 34X More Performance



CPU is dual socket Intel Xeon E5-2699v4. 170TF is half precision or FP16

Computing for Infinite Opportunities

The NVIDIA DGX-1 is the first system built with NVIDIA Pascal™-powered Tesla® P100 accelerators. The NVIDIA NVLink™ implementation delivers a massive increase in GPU memory capacity, giving you a system that can learn, see, and simulate our world.

Analyze. Visualize. AI-Accelerate

The NVIDIA DGX-1 software stack includes major deep learning frameworks, the NVIDIA DIGITS™ GPU training system, the NVIDIA Deep Learning SDK (e.g. CuDNN, NCCL), NVIDIA Docker, GPU drivers, and NVIDIA CUDA® for rapidly designing deep neural networks (DNN). It's the ideal stack for accelerating popular analytics and visualization software.

This powerful system includes access to cloud management services for container creation and deployment, system updates, and an application repository. This software, running on Pascal-powered Tesla GPUs, lets applications run 12X faster than previous GPU-accelerated solutions.

Turn Data into Knowledge

The innovative NVIDIA DGX-1 system lets you uncover patterns in large data sets to reveal new knowledge and insights in hours or minutes.

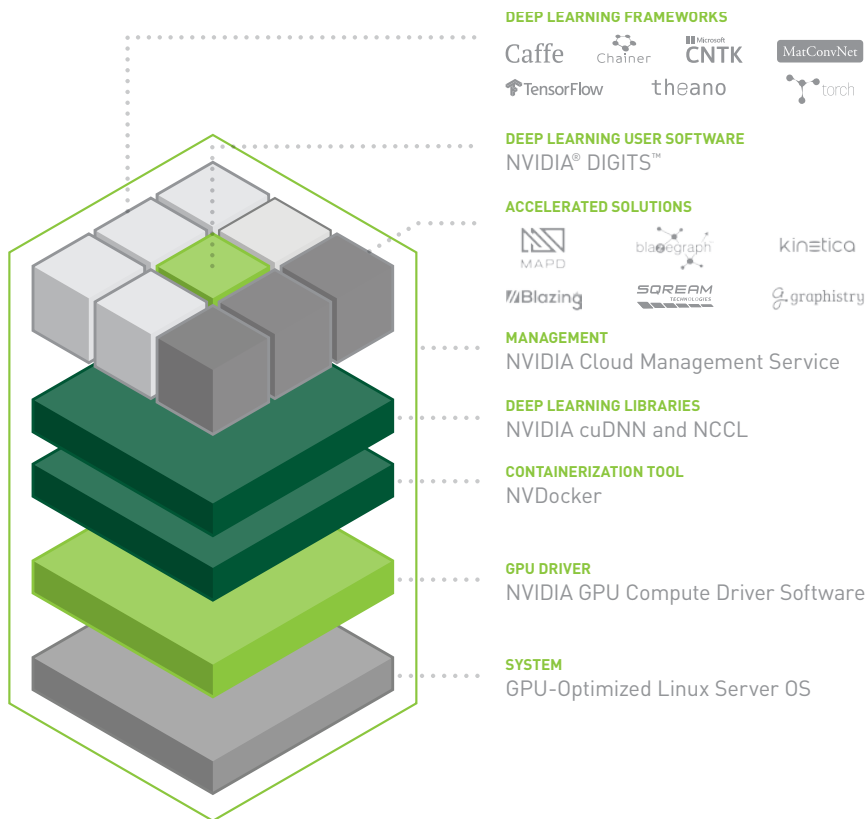
Stay Ahead of the Competition

NVIDIA DGX-1 is engineered with groundbreaking technologies that deliver the fastest solutions for your deep learning training and AI-accelerated analytics workloads.

Maximize Your Investment

Hardware and software support gives you access to NVIDIA deep learning expertise and includes cloud management services, software upgrades and updates, and priority resolution of your critical issues.

NVIDIA DGX-1 Software Stack



For more information on NVIDIA DGX-1, visit www.nvidia.com/dgx1

© 2016 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. DEC16

