

Deep learning teams to get monster power in a box

April 6 2016, by Nancy Owano



Nvidia on Tuesday announced its DGX-1, a rather formidable, deep learning supercomputer. Deep learning is an already familiar concept; it is fairly well known that deep learning and AI signify modern-day keys for people who are searching for patterns, insights and answers for what's to come.

Nvidia said this in announcing its supercomputer: "Data scientists and artificial intelligence (AI) [researchers](#) require accuracy, simplicity, and speed for [deep learning](#) success.

Faster training and iteration ultimately means faster innovation and

faster time to market."

Nvidia is promoting the DGX-as the world's first purpose-built system for deep learning with fully integrated hardware and software, "a deep learning supercomputer in a box."

Engadget said there was "an insane amount of [computing power](#) (170 teraflops in a [single](#) machine)." Devin Coldewey in *TechCrunch* said the machine was an enclosure for an 8-GPU supercomputing cluster with "built-in neural network training software."

It is said to be the equivalent of 250 servers in a box. Agam Shah, IDG News Service, said, according to Nvidia, "The DGX-1 supercomputer can deliver the computing power of 250 two-socket servers in a desktop box." Writing in *iTWire*, Stephen Withers, a senior member of the Australian Computer Society, made similar observations about the DGX-1 edge. "He said, "because of the diminishing returns from adding nodes, it requires more than 250 Xeon servers to match the speed of the [DGX-1](#)."

Attention is being given to the power delivered via eight Tesla P100 GPUs.

Agam Shah explained: "Ironically, the DGX-1 runs on two Intel Xeon chips, though Nvidia didn't provide exact CPU details. But it's the other components like the computer's GPUs that provide the serious horsepower. At the center of DGX-1 are eight Tesla P100 graphics processing units, which are based on the company's new Pascal GPU [architecture](#)."

The Tesla P100 design is to put more power behind the deep learning technique.

MIT Technology Review's Tom Simonite also took time to reflect on this: "Deep learning involves passing data through large collections of crudely simulated neurons. The P100 could help deliver more breakthroughs by making it possible for computer scientists to feed more data to their artificial neural networks or to create larger collections of virtual [neurons](#)."

Cooling system? According to the IDG News Service report, as a small system which draws a lot of power, it will need a special cooling system to dissipate the heat, said Jim McGregor, principal analyst at Tirias Research.

TechCrunch took a step back to reflect on the significance of this launch. "GPUs are already [massively](#) parallelized, having to handle huge amounts of data under extremely strict time constraints, so they're a great match for supercomputing rigs. 8 Teslas in parallel is nothing to sneeze at (they produce 170 teraflops), and while you could rent time on a cloud cluster with more raw power, there's a lot to be said for running your own hardware in-house. (Though you'll probably be relying on Nvidia for troubleshooting and maintenance.)"

Nvidia introduced the system Tuesday. What is next? The DGX-1 will be available in the US in June, Nvidia's announcement said, "General availability for the NVIDIA DGX-1 deep learning system in the United States is in June, and in other regions beginning in the third quarter direct from NVIDIA and select systems [integrators](#)."

More information: nvidianews.nvidia.com/news/nvidia-announcing-supercomputer

Citation: Deep learning teams to get monster power in a box (2016, April 6) retrieved 25 April 2024 from <https://techxplore.com/news/2016-04-deep-teams-monster-power.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.