



# ORNL

ORNL trusted Microway—and was rewarded with 2 DGX-2 systems installed, operational, & running benchmarks within 4hrs of the first crate being opened



Talk to an Expert:

+1-508-746-7341

www.microway.com  
wespeakhpc@microway.com

## World's First Public Deployment of DGX-2H

The U.S. Department of Energy's Oak Ridge National Laboratory (ORNL) is home to the world's fastest & most powerful supercomputer, Summit. Even so researchers at ORNL had need for specialized infrastructure for machine learning tasks with extremely high GPU density—as well as a space for helping to test projects before running them on Summit. They needed a complementary infrastructure to Summit's unique scale. The Compute and Data Environment for Science (CADES) team sought a groundbreaking new architecture, and they worked with HPC & AI specialist Microway to ready, install, deliver, and deploy it.

### THE DEPLOYMENT

#### ▶ 1 DGX-2 AI SYSTEM

NVIDIA® DGX-2™ delivers 2 petaFLOPS of deep learning performance in a single system—combining the power of 16 fully interconnected NVIDIA® V100 GPUs with the outstanding ease of use of the NVIDIA DGX platform and services. DGX-2 delivers world record AI performance.

#### ▶ 1 DGX-2H AI SYSTEM

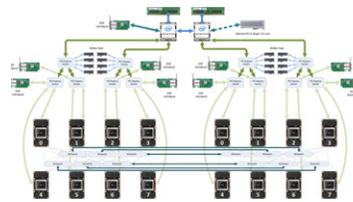
DGX-2H is a DGX-2 system in overdrive, and ORNL is a launch customer for the DGX-2H platform. Higher-clocked NVIDIA V100 GPUs enable unheard of throughput for any GPU workload. The system also includes high-clock speed Intel Xeon CPUs to further augment performance.

#### ▶ NVIDIA NVSWITCH

Unlike other servers, DGX-2 systems allow any two GPUs and all GPUs to talk at full 300GB/s bandwidth via NVSwitch technology. This kind of "all-to-all" internal switch fabric enables large scale 16-GPU AI training. Without it communication would take extra hops: dramatically extending runtimes.

#### ▶ MELLANOX INFINIBAND

The systems scale to multi-node 16-GPU computing via InfiniBand. A multi-node architecture allows researchers to understand how their models scale by adding additional dense-GPU nodes—particularly helpful when the DGX-2s are used to test models before running on Summit.



Thanks to the unique features of the DGX-2 systems and their successful installation at ORNL, the ORNL research teams have been able to expand existing projects throughout the Lab and enable new ones. Researchers in the Center for Molecular Biophysics and Geographic Data Science were first to see results.

"I get requests for access to these systems often, from both researchers and students from all over the lab who want to learn on the best hardware around. These requests cover the full range of use cases and the DGX-2s never fail to impress. As word of mouth, combined with outreach, ramps up, I only see usage of these systems increasing," said Chris Layton, Linux Systems Engineer for the new deployment.

The new DGX-2 systems have already provided unexpected capacities and insights to the ORNL team, and the researchers expect that this will continue into the future.

## Oak Ridge National Laboratory

Oak Ridge National Laboratory hosts the premier supercomputing site in the world. Its computing infrastructure focuses on research and simulation. Oak Ridge is also the largest DOE lab—with thousands of researchers and an over \$1 billion budget. It supports missions in national security, energy, and the environment.

# Case Study

**“Microway was able to, via their installation crew, make the integration of the DGX-2 into the CADES environment a smooth process...This was done under a deadline and they met all the timelines flawlessly.”**

**- Chris Layton, Linux Systems Engineer for ORNL’s Compute and Data Environment for Science (CADES) team**

## THE PROCESS

The ORNL team decided to trust AI & HPC specialist Microway with their deployment. They were rewarded with the two DGX-2 systems physically installed, up and running, and doing benchmark testing within 4 hours of the first crate being opened.



## ABOUT MICROWAY

Microway designs and builds hardware solutions for the intersection of AI and HPC. These include clusters, servers, quiet workstations designed for bleeding-edge computational performance. Microway also delivers the data-planes that keep up with these advanced workloads—with a complete array of storage and network offerings.

Since 1982, customers have trusted Microway to design and deliver them unique and superior hardware—enabling them to remain at the forefront of supercomputing and solve the world’s toughest challenges.

Microway’s strategic partners include NVIDIA, Intel, AMD, Mellanox, DDN, and NetApp. Classified as a small business, woman owned and operated, Microway’s CSA Schedule is GS-35F-0431N.



- ▶ Researchers at ORNL investigated NVIDIA’s new DGX-2 when seeking a new machine learning & data-intensive computing environment—because of the system’s unique scale, density, & architectural design,
- ▶ ORNL reached out to NVIDIA to learn about DGX-2 & confirm that it would meet their specific needs. NVIDIA in turn pointed ORNL to Microway as one of the only organizations authorized to sell & install DGX-2 at launch.
- ▶ The ORNL admin and Microway teams constantly collaborated via, phone, email, and web meetings in the weeks prior to ensure a smooth deployment.
- ▶ Onsite Installation day neared. Both teams ensured ample preparation was made to ensure the room, network, contacts, cooling, & system admins were all ready for the installation and launch of the DGX-2 systems.
- ▶ The day for installation of the new systems arrived. Late in October 2018, Microway installed one DGX-2 and an even more powerful DGX-2H—the world’s first—with upgraded CPUs & higher-clocked NVIDIA V100 GPUs.
- ▶ Others at ORNL have taken note of the new systems. In 2019 another DGX-2 deployment with a third DGX-2 system was delivered to ORNL.



# Vyasa Analytics

How an AI analytics firm grew by adding on-prem resources to save money, create new products, and improve R&D

## Cultivating a Deep Learning Edge GPUs + DGX-1

Talk to an Expert:

+1-508-746-7341

www.microway.com  
wespeakhpc@microway.com



**What kind of HPC resources are most cost effective for the development and delivery of AI-based analytics products? While some entrepreneurs would look to cloud options alone, deep learning analytics firm Vyasa Analytics finds that the best solution includes top-of-the-line on-site hardware. For an enterprise like Vyasa, building cutting-edge software serves as a primary differentiator, but for their flagship offerings to operate efficiently, blazing-fast hardware infrastructure proved mandatory.**

### THE DEPLOYMENT

#### ► NVIDIA DGX-1 AI SYSTEM

DGX-1 offers unparalleled deep learning training performance from 8 NVIDIA V100 GPUs. It can complete a training workload more than twice as fast as other 8 GPU servers & eclipses the speed of a CPU-only server: 96X the throughput. "We call the DGX 'beast mode'" says Vyasa's CEO.

#### ► MICROWAY TESLA GPU SERVERS

NumberSmasher Tesla GPU servers with 4 NVIDIA V100 GPUs were deployed for development. This infrastructure proved out the first product concepts & helped grow the company. Vyasa's team tests out new ideas, develops for newmarkets, and smaller than DGX-1 scale-runs here today.

#### ► NVIDIA V100 GPU ACCELERATORS

NVIDIA V100 GPUs are massively parallel accelerators designed to crunch through data-intensive applications. Each GPU offers 125 TFLOPS of deep learning performance & 32GB of memory. Vyasa's applications require this massive throughput: just one code analyzes 70,000,000 articles/day.

#### ► INTEL DATA CENTER SSDS

Intel Data Center SSDs provide high-bandwidth, low-latency access to data in Vyasa's infrastructure. They provide high-performance scratch space and can consistently serve up the bulk data sets and training data to the GPUs.



Rather than perform simplistic analytics on small sections of siloed data, Vyasa Analytics deploys **deep learning** to tease insight from the reams of data across the entire breadth of an organization. They call their cutting-edge software "Cortex."

Deep learning allows a company like Vyasa to create products that analyze beyond a simple matching search: they provide results in new shapes and varieties, across

diverse data types, & don't even require an operator to know the relationship or structure they are looking for in order to tease relationships out. These skills augment researchers & users to perform their jobs better and stimulate new discoveries.

Backed by new hardware, Vyasa has been freed to make more & more expansive products available. The company has used its deployment to analyze the entirety of the PubMed database and now offers a product that surfaces new research trends—before they are named. They have created new products for the legal field as well.

## Vyasa Analytics

Massachusetts-based Vyasa Analytics provides a deep learning analytics platform to customers in many markets, including life sciences. Vyasa's capabilities also provide value to those in the legal, competitive intelligence, and other fields—by analyzing anything from patent filings to newsfeeds to troves of images.

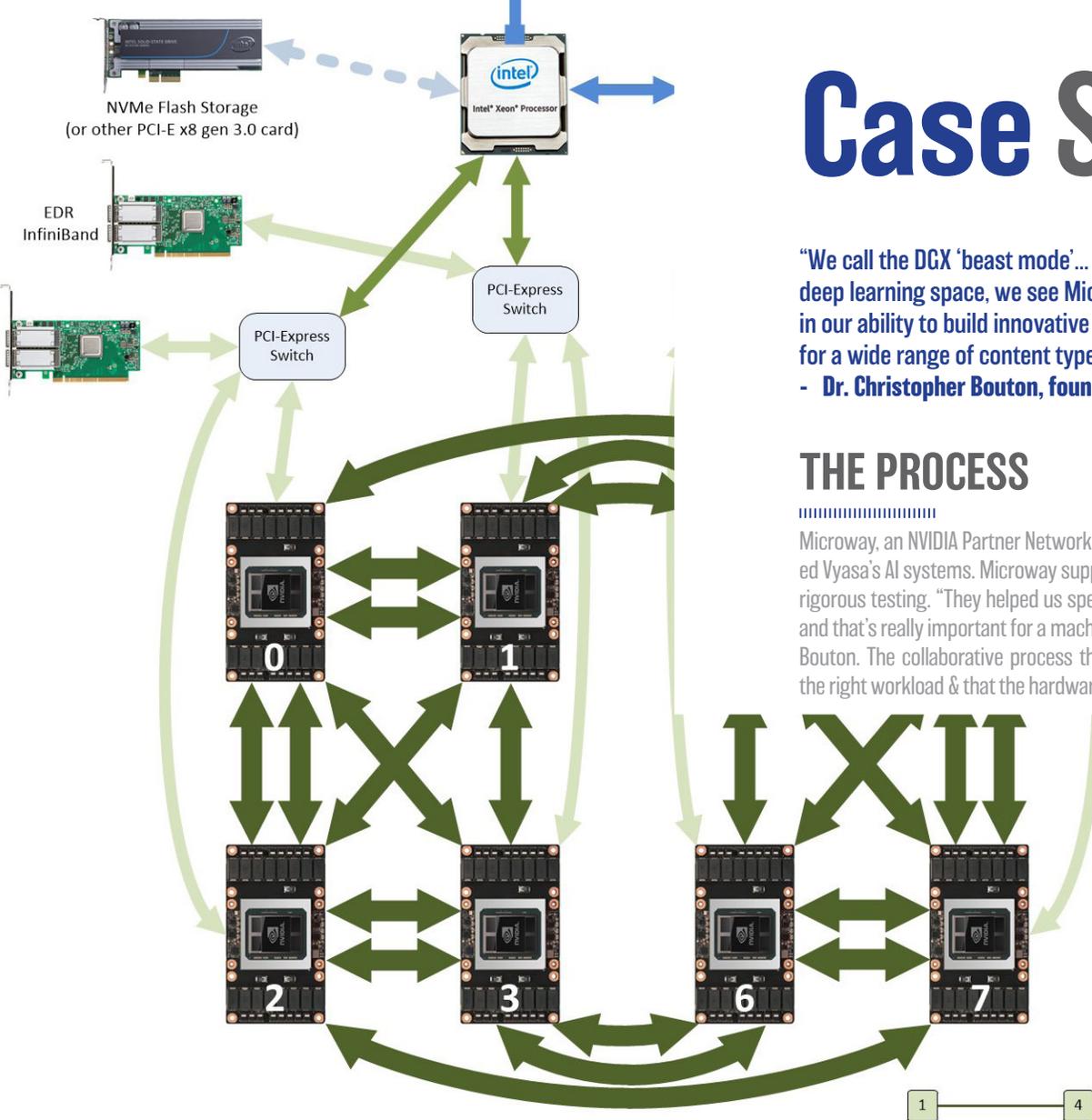
# Case Study

“We call the DGX ‘beast mode’... As a company working in the deep learning space, we see Microway and NVIDIA as key partners in our ability to build innovative novel deep learning algorithms for a wide range of content types”

- Dr. Christopher Bouton, founder and CEO of Vyasa Analytics

## THE PROCESS

Microway, an NVIDIA Partner Network HPC Partner of the Year, built & integrated Vyasa’s AI systems. Microway supported Vyasa from specification through rigorous testing. “They helped us spec the systems & architect the systems, and that’s really important for a machine of this size and complexity” said CEO Bouton. The collaborative process the right architectures were matched to the right workload & that the hardware delivered bleeding-edge performance.



## ABOUT MICROWAY

Microway designs and builds hardware solutions for the intersection of AI and HPC. These include clusters, servers, quiet workstations designed for bleeding-edge computational performance. Microway also delivers the data-planes that keep up with these advanced workloads—with a complete array of storage and network offerings.

Since 1982, customers have trusted Microway to design and deliver them unique and superior hardware—enabling them to remain at the forefront of supercomputing and solve the world’s toughest challenges.

Microway’s strategic partners include NVIDIA, Intel, AMD, Mellanox, DDN, and NetApp. Classified as a small business, woman owned and operated, Microway’s CSA Schedule is GS-35F-0431N.



- ▶ Early on, Vyasa Analytics relied solely on the cloud for GPU capabilities. However, this strategy proved a drag on operations. “If you do everything on cloud instances, then that can get very costly very quickly”
- ▶ Vyasa’s CEO, a successful entrepreneur with a Ph.D. in Neuroscience, did the math. “The modeling shows that within six to eight months you save the full cost of your own hardware...just by savings in cloud costs.”
- ▶ The company settled on an on-prem hardware purchase to complement cloud instances. They reached out to Microway, who assisted them in previous growth companies, to collaboratively design the deployment.
- ▶ Microway worked with Vyasa to find the best solution for their application and their business, enabling a stepwise process of infrastructure deployments and effective investment.
- ▶ A Microway NumberSmasher Tesla GPU Server was delivered first, acting as a development platform to prove out the potential of many new projects. It was up and running applications within 24 hours of delivery.
- ▶ After receiving a \$1.8 million loan from MassDevelopment’s Emerging Technology Fund, Vyasa added a more powerful DGX-1 system to their deployment.



# Oregon State

How a major research university transformed its GPU resources - and how it drove "what's next" for science across the campus



Talk to an Expert:

+1-508-746-7341

www.microway.com  
wespeakhpc@microway.com

## Campus-Wide AI Transformation with DGX-2 At Scale

As the Oregon State University College of Engineering developed a strategy for provisioning and providing high-power GPU computational capabilities, it worked with industry leaders NVIDIA® and Microway, embracing the opportunity to set the institution apart with a cutting-edge computational solution. The investment in the equivalent of six new supercomputers has already dramatically improved the school's educational and research capabilities, helping the machine learning and artificial intelligence group to grow into an increasingly important presence.

### THE CLUSTER

#### ▶ 6 DGX-2 AI SYSTEMS

Each DGX-2 AI system delivers world record AI performance of up to 2 petaFLOPS: an entire supercomputer worth of computational horsepower. There are 16 fully-connected NVIDIA® V100 GPUs per DGX-2— far greater density than available in most GPU servers available today.

#### ▶ NVSWITCH AND NVLINK

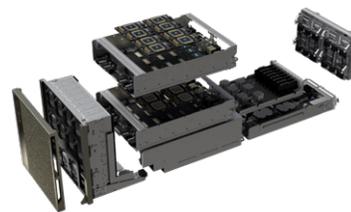
New technologies support increased application scale & allow for more users to utilize each system. The DGX-2 solves common scaling predicaments by leveraging the NVSwitch scalable fabric architecture & NVLink™ GPU-GPU interconnect for high-speed GPU-to-GPU inside each appliance.

#### ▶ NGC CONTAINERS

NVIDIA GPU Cloud Containers dramatically simplify the user experience & administrative ease of the new hardware. NGC containers encapsulate GPU software for deep learning & HPC applications as well as pre-trained models, training scripts, SDKs, for rapid development & deployment.

#### ▶ MELLANOX EDR INFINIBAND

100Gb EDR InfiniBand is used to bridge between DGX-2s for jobs that require multiple AI supercomputers. The improved fabric infrastructure stitches together systems far more tightly, with higher bandwidth and lower latency, than in past computational resources on campus.



"Artificial intelligence, machine learning, parallel programming: those are all really hot items right now," says Todd Shechter, Director of IT at Oregon State University College of Engineering. We are seeing a lot of interest in the undergraduate curriculum space." Students, staff, and faculty all agreed that the University should fully embrace this emerging trend.

For Oregon State, the new DGX-2s optimize functionality & flexibility. They handle single & double-precision workloads, and crunch data in essentially any form. They can be tied together in a cohesive computing unit when large jobs are required, yet easily partitioned for smaller projects. "We have a hugely broad set of users who will be making use of the investment, and DGX-2 was the best solution to take care of all of their needs," emphasized Shechter. The tools are used by electrical engineers & computer scientists, as well as mechanical engineers, civil engineers, nuclear engineers, biologists, chemists, and others.

## Oregon State University

Oregon State University's College of Engineering comprises about a third of the University, with about 10,000 students, staff, and faculty. It's an enormous and diverse community of users. Throughout the past several years, interest has steadily grown in an emerging force in science and engineering: GPU computing.

# Case Study

“Microway was really great at helping us through the nitty-gritty details... I’ve had faculty members who have run their simulations on existing hardware, and then run it on DGX hardware, and the difference just blows your mind away”

- **Todd Shechter, Director of Information Technology, Oregon State University College of Engineering**

## THE PROCESS

Microway, an NVIDIA Partner Network HPC Partner of the Year, installed and integrated the deployment at Oregon State. The teams at Microway and Oregon State collaborated on extensive preparatory work. Microway experts then worked onsite to rack the systems, integrate software, and burn the cluster in, sharing their knowledge with Oregon State IT along the way.



## ABOUT MICROWAY

Microway designs and builds hardware solutions for the intersection of AI and HPC. These include clusters, servers, quiet workstations designed for bleeding-edge computational performance. Microway also delivers the data-planes that keep up with these advanced workloads—with a complete array of storage and network offerings.

Since 1982, customers have trusted Microway to design and deliver them unique and superior hardware—enabling them to remain at the forefront of supercomputing and solve the world’s toughest challenges.

Microway’s strategic partners include NVIDIA, Intel, AMD, Mellanox, DDN, and NetApp. Classified as a small business, woman owned and operated, Microway’s CSA Schedule is GS-35F-0431N.



- ▶ Faculty and administrators gathered together to define the characteristics of an ideal campus-wide computing resource. The solution had to scale, yet it also had to represent a dramatic leap in capability.
- ▶ In part because of docker images that have NVIDIA’s containerized software, in part because of technical support, but mostly for the unmatched computational horsepower, the team selected NVIDIA DGX-2™
- ▶ After factoring in myriad use cases from researchers, enrollment figures for undergrad tech courses, & an analysis of overuse of existing infrastructure, OSU calculated an upscale to 6 DGX-2s was warranted.
- ▶ The proposed increase in scale was not only about capacity planning—it was also focused on unlocking new possibilities. The Oregon State team moved forward confidently—with Microway delivering the systems.
- ▶ The teams collaborated to get the details right. Higher amperage power was readied at the racks, installation paths cleared, and space made available at the bottom of racks for the new DGX-2 equipment.
- ▶ Microway experts arrived onsite to rack the systems, integrate cluster software, properly burn the cluster in, and guide Oregon State IT in knowledge transfer about the systems.