Everybody's Talking About Titan

OLCF Staff Writer

Twenty-four presenters at GTC shared Titan GPU success stories

Kicking off this year's GPU Technology Conference (GTC), an annual GPU developer conference hosted by NVIDIA, the visual computing technology company's CEO and cofounder Jen-Hsun Huang held up a two-GPU, 5,760-core graphics card smaller than a shoebox: the GeForce GTX Titan Z.

NVIDIA's Titan Z follows the company's GE GTX Titan GPU. The name is no coincidence.

An Oak Ridge Leadership Computing Facility (OLCF) vendor partner, NVIDIA is offering scientists, artists, engineers, and gamers—anyone who needs computing muscle for heavy-duty simulation—a slice of the power that NVIDIA GPUs provide to the OLCF's Titan supercomputer located at Oak Ridge National Laboratory (ORNL).

GPU Technology Conference keynote address
Image credit: NVIDIA

Sporting 18,688 NVIDIA GPUs and operating at up to 27 petaflops, the Cray XK7, hybrid CPU–GPU Titan is the nation's most powerful supercomputer. Its users are crunching quadrillions of calculations per second, storing petabytes of information, and pushing the limits of simulation ranging from quantum mechanical to cosmological scales.

Leading up to last year's GTC, OLCF was confident in Titan's ability to accelerate scientific discovery. But how would users respond to the GPUs, a relatively new architecture for high-performance computing (HPC)?

"The expectation at the time was that it was going to be a difficult machine to program," said Fernanda Foertter, OLCF HPC user support specialist. "We were still fighting the perception that the machine would require a lot of time and effort to use."

But after a year of success stories using Titan's GPUs—including application speedups up to seven times what would be possible on an Cray XE6 CPU-only architecture—OLCF wasn't going into the conference with an elephant in the room. Rather, staff and users had a mammoth list of science achieved, lessons learned, and advice for programming and operating a GPU system.

The OLCF-organized series "Extreme Scale Supercomputing with the Titan Supercomputer," chaired by Jack Wells, director of science, brought together Titan users in 24 conference sessions amounting to 540 minutes.

"Titan is the biggest GPU cluster on the planet," Wells said. "A lot of our users come to the conference to learn more about GPUs and what other people are doing with GPUs because they have this great resource they can take advantage of."

Foertter led a talk on accelerating research and development using Titan, in which she described the transition that many OLCF users made to hybrid CPU–GPU computing, increasing Titan's GPU usage over the last year.

"It was a review of the success we've had since the last GTC," Foertter said. "Our users were overwhelmingly positive about Titan and reflected what we've found, which is that the GPUs have been worth it to people and have allowed

them to do more science."

Users like James Phillips of the University of Illinois at Urbana-Champaign (UIUC) and Dag Lohmann of KatRisk talked about the research value in using Titan's GPUs to run at petascale and cut time-to-solution.

Phillips discussed the scientific background for petascale molecular dynamics simulations of the HIV-1 capsid and presented early results from GPU-accelerated simulations of the complete atomic structure of the HIV-1 capsid carried out on Titan and the National Center for Supercomputing Applications' Cray XE6/XK7 Blue Waters.

Including up to 64 million atoms, they are the only all-atom simulations of the complete HIV-1 capsid and could provide information for drug design. The simulations, which were conducted by Juan Perilla and Klaus Schulten of UIUC, were performed using the molecular dynamics program NAMD and the VMD visualization and analysis program.

While Phillips, Perilla, and Schulten need GPUs to compute the science of the very small, Lohmann, CEO and cofounder of KatRisk, talked about how the small, Berkeley, California, company uses Titan to simulate global flood models.

KatRisk develops models on the company's NVIDIA GTX Titan GPU nodes and then scales their models to Titan to run regional and global flood simulations in high resolution (10-meter resolution of US flooding and 90-meter resolution of global flooding). Lohmann said the company's models will provide probabilities for the occurrence of damaging extreme flood events to insurance companies, government agencies, businesses, and home owners.

Lohmann said a single NVIDIA GPU card can perform around 10 times faster than a top-of-the-line multicore CPU, which makes KatRisk's models possible in a competitive timeframe.

"Also, having access to thousands of nodes on Titan compared to our in-house nodes cuts 8 to 12 months of our computing time," Lohmann said.

To describe how users can benefit from OLCF's team of computational scientists who have worked to improve the tools needed to run efficiently on Titan's architecture, OLCF's Oscar Hernandez and Matthew Baker discussed ORNL's efforts to design future programming models that make the most of hybrid systems like Titan.

Baker's work demonstrates that the communication library OpenSHMEM (optimized for small and medium-sized, one-sided communications) and the OpenACC accelerator directives (designed to provide easy access to accelerators) can work together efficiently to significantly speed up applications running on hybrid systems.

With assistance from HPC software provider CAPS, Baker ported the SSCA3 Synthetic Aperture RADAR and image recognition benchmark to show that OpenSHMEM/OpenACC works at scale. Individual accelerated kernels were 20 times faster than the sequential kernel, and the overall application was twice as fast using GPUs.

"Matt was even invited to repeat the talk at another event, so this demonstrates this is something important to the research community and industry," Hernandez said.

Jim Rogers, director of operations for ORNL's National Center for Computational Sciences, also presented on the operational impact and efficiency gains of Titan's GPUs.

As OLCF continues to help pioneer the use of GPUs for HPC, Foertter said she hopes the conference series "showed users that there isn't such a big learning curve. We want to let people know our users have been successful and reached new milestones by using the GPUs." —*Katie Elyce Jones*