

November 19, 2019 | By Jan Zverina

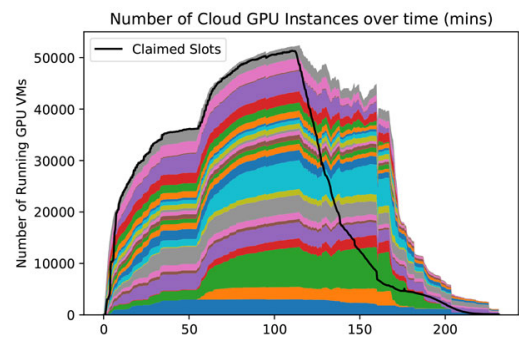
SDSC, Wisconsin University IceCube Center Conduct GPU Cloudburst Experiment

Multi-messenger astrophysics run has implications for other research domains

The San Diego Supercomputer Center (SDSC), an Organized Research Unit of UC San Diego; and the Wisconsin IceCube Particle Astrophysics Center (WIPAC) at the University of Wisconsin–Madison successfully completed a computational experiment as part of a multi-institution collaboration that marshalled all globally available for sale GPUs (graphics processing units) across Amazon Web Services, Microsoft Azure, and the Google Cloud Platform.

In all, some 51,500 GPU processors were used during the approximately two-hour experiment conducted on November 16 and funded under a [National Science Foundation EAGER grant](#). The experiment used simulations from the [IceCube Neutrino Observatory](#), an array of some 5,160 optical sensors deep within a cubic kilometer of ice at the South Pole. In 2017, researchers at the NSF-funded observatory [found the first evidence of a source of high-energy cosmic neutrinos](#) – subatomic particles that can emerge from their sources and pass through the universe unscathed, traveling for billions of light years to Earth from some of the most extreme environments in the universe.

The experiment – completed just prior to the opening of the International Conference for High Performance Computing, Networking, Storage, and Analysis (SC19) in Denver, CO – was coordinated by Frank Würthwein, SDSC Lead for High-Throughput Computing, and Benedikt Riedel, Computing Manager for the IceCube Neutrino Observatory and Global Computing Coordinator at [WIPAC](#).



The chart shows the time evolution of the burst over the course of ~200 minutes. The black line is the number of GPUs used for science, peaking at 51,500 GPUs. Each color shows the number of GPUs purchased in a region of a cloud provider. The steep rise indicates the burst capability of the infrastructure to support short but intense computation for science. Image: Igor Sfiligoi, SDSC/UC San Diego

Igor Sfiligoi, SDSC's lead scientific software developer for high-throughput computing, and David Schultz, a production software manager with IceCube, conducted the actual run.

"We focused this GPU cloud burst in the area of multi-messenger astrophysics, which is based on the observation and analysis of what we call 'messenger' signals, in this case neutrinos," said Würthwein, also a physics professor at UC San Diego and executive director of the [Open Science Grid \(OSG\)](#), a multi-disciplinary research partnership specializing in high-throughput computational services funded by the NSF and the U.S. Department of Energy.

"The NSF chose multi messenger astronomy as one of its [10 Big Ideas](#) to focus on during the next few years," said Würthwein. "We now have instruments that can measure gravitational waves, neutrinos, and various forms of light to see the most violent events in the universe. We're only starting to understand the physics behind such energetic celestial phenomena that can reach Earth from deepest space."

Exascale Extrapolations

The net result was a peak of about 51k GPUs of various kinds, with an aggregate peak of about 350 PFLOP32s (according to NVIDIA specifications), according to Sfiligoi.

"For comparison, the Number 1 TOP100 HPC system, [Summit](#), ([based at Oak Ridge National Laboratory](#)) has a nominal performance of about 400 PFLOP32s. So, at peak, our cloud-based cluster provided almost 90% of the performance of *Summit*, at least for the purpose of IceCube simulations.

The relatively short time span of the experiment showed the ability to conduct a massive amount of data processing within a very short period – an advantage for research projects that must meet a tight deadline. Francis Halzen, principal investigator for IceCube, a Distinguished Professor at the University of Wisconsin-Madison, and director of the university's Institute for Elementary Particle Physics, foresaw this several years ago.

"We have initiated an effort to improve the calibration of the instrument that will result in sensitivity improved by an estimated factor of four," wrote Halzen. "We can apply this improvement to 10 years of archived data, thus obtaining the equivalent of 40 years of current IceCube data."

"We conducted this experiment with three goals in mind," said IceCube's Riedel. "One obvious goal was to produce simulations that will be used to do science with IceCube for multi-messenger astrophysics. But we also wanted to understand the readiness of our

cyberinfrastructure for bursting into future Exascale-class facilities such as Argonne’s *Aurora* or Oak Ridge’s *Frontier*, when they become available. And more generally, we sought to determine how much GPU capacity can be bought today for an hour or so GPU burst in the commercial cloud.”

“This was a social experiment as well,” added Würthwein. “We scavenged up all available GPUs on demand across 28 cloud regions across three continents – North America, Europe, and Asia. The results of this experiment tell us that we can elastically burst to very large scales of GPUs using the cloud, given that exascale computers don’t exist now but may soon be used in the coming years. The demo also shows such bursting of massive data, is suitable for a wide range of challenges across astronomy and other sciences. To the extent that the elasticity is there, we believe that this can be applied across all of scientific research to get results quickly.”

HTCondor was used to integrate all purchased GPUs into a single resource pool that IceCube submitted their workflows to from their home base in Wisconsin. This was accomplished by aggregating resources in each cloud region, and then aggregating those aggregators into a single global pool at SDSC.

“This is very similar to the production infrastructure that OSG operates for IceCube to aggregate dozens of ‘on-prem’ clusters into a single global resource pool across the U.S., Canada, and Europe,” said Sfiligoi.

An additional experiment to reach even higher scales is likely to be made sometime around the Christmas and New Year holidays, when commercial GPU use is traditionally lower, and therefore availability of such GPUs for scientific research is greater.

MEDIA CONTACT

Jan Zverina, 858-534-5111, jzverina@sdsc.edu

Madeleine O’Keefe, WIPAC Communications, 608-515-3831 mokeefe@icecube.wisc.edu

UC San Diego’s [Studio Ten 300](#) offers radio and television connections for media interviews with our faculty, which can be coordinated via studio@ucsd.edu. To connect with a UC San Diego faculty expert on relevant issues and trending news stories, visit <https://ucsdnews.ucsd.edu/media-resources/faculty-experts>.