National Research Platform (NRP) Space

To hear about future activities, please join the NRP engagement email list via this link.

If you would like to present or have a suggestion for a session, please write to Dana Brunson

Please join us for the NRP Engagement webinar

Monday, January 27, 2020 at 1 ET - 12 CT - 11 MT - 10 PT

Zoom: https://internet2.zoom.us/j/735965245

Feel free to share this announcement with anyone that may be interested. Calls will typically be on the fourth Monday of the month. To hear about future activities, please join the NRP engagement email list via this link.

Running a 380PFLOP32s GPU burst for Multi-Messenger Astrophysics with IceCube across all available GPUs in the Cloud

Igor Sfiligoi and Frank Würthwein

The IceCube Neutrino Observatory is the National Science Foundations (NSF)'s premier facility to detect neutrinos with energies above approximately 10 GeV and a pillar for NSF's Multi-Messenger Astrophysics (MMA) program, one of NSF's 10 Big Ideas. The detector is located at the geographic South Pole and is designed to detect interactions of neutrinos of astrophysical origin by instrumenting over a gigaton of polar ice with 5160 optical sensors. The sensors are buried between 1450 and 2450 meters below the surface of the South Pole ice sheet. To understand the impact of ice properties on the incoming neutrino detection, and origin, photon propagation simulations on GPUs are used. We report on a few hour GPU burst across Amazon Web Services, Microsoft Azure, and Google Cloud Platform that harvested all available for sale GPUs across the three cloud providers the weekend before SC19, reaching over 51k GPUs total and 380 PFLOP32s. GPU types span the full range of generations from the NVIDIA GRID K520 to the most modern NVIDIA T4 and V100. We report the scale and science performance achieved across all the various GPU types, as well as the science motivation to do so.

Igor Sfiligoi is Lead Scientific Software Developer and Researcher at UCSD/SDSC. He has been active in distributed computing for over 20 years. He has started in real-time systems, moved to local clusters, worked with leadership HPC systems, but spent most of his career in computing spanning continents. For about 10 years, he has been working on one such world-wide system, called glideinWMS, which he brought from the design table to being de-facto standard for many scientific communities. He has recently moved his attention in supporting users on top of Kubernetes clusters and Cloud resources. He has a M.S. in Computer Science equivalent from Universita degli studi di Udine, Italy. He has presented at many workshops and conferences over the years, with several published papers.

Frank Würthwein is the Executive Director of the Open Science Grid, a national cyberinfrastructure to advance the sharing of resources, software, and knowledge, and a physics professor at UC San Diego. He received his Ph.D. from Cornell in 1995. After holding appointments at Caltech and MIT, he joined the UC San Diego faculty in 2003. His research focuses on experimental particle physics and distributed high-throughput computing. His primary physics interests lie in searching for new phenomena at the high energy frontier with the CMS detector at the Large Hadron Collider. His topics of interest include, but are not limited to, the search for dark matter, supersymmetry, and electroweak symmetry breaking. As an experimentalist, he is interested in instrumentation and data analysis. In the last few years, this meant developing, deploying, and now operating a worldwide distributed computing system for high-throughput computing with large data volumes. In 2010, "large" data volumes are measured in Petabytes. By 2025, they are expected to grow to Exabytes.

Past Calls:

Monday, October 28, 2019 at 1 ET - 12 CT - 11 MT - 10 PT

Running Genomics Workflows on the Pacific Research Platform's Nautilus Kubernetes Cluster

Alex Feltus, Ph.D.

Professor|Genetics & Biochemistry @ Clemson University; Co-Founder|Praxis AI

Slides

Recording

Abstract: Our core biological research mission is to discover causal alleles underlying complex trait expression in plants and animals. Active projects include (A) discovery of genetic subsystems driving legume-microbe symbiosis that can be engineered into other plants so they can make their own fertilizer, (B) elucidation of gene expression pattern shifts between normal and disordered brain tissue for better diagnosis of intellectual disability, and (C) detection of tumor specific alterations in kidney and other tumors of relevance to precision medicine. Our scientific instrument is the high performance/throughput computer where we run bioinformatic, machine learning, and network biology workflows on tens to thousands of terabytes of in-house and open source deep DNA sequencing datasets. In recent years, we have wrapped applications in containerized NextFlow workflows and now run data intensive experiments on Kubernetes (K8s) clusters including the PRP Nautilus cluster (we have added a node at Clemson) and the Google Cloud Platform. In this webinar we will (A) present results from a large tumor biomarker screen generated with the Nautilus cluster, (B) describe broadly useful open source genomics workflows (GEMmaker, KINC, and Gene Oracle) with Nautilus-specific usage documentation, (C) outline a grassroots strategy to add nodes to the Nautilus cluster and train people how to use that super-awesome system, and (D) discuss a business model where one can build an elastic K8s cluster for a small virtual organization that can be dynamically linked to larger national compute fabrics platforms via aggregation or federation.

Dr. F. Alex Feltus received a B.Sc. in Biochemistry from Auburn University in 1992, served two years in the Peace Corps, and then completed advanced training in biomedical sciences at Vanderbilt and Emory. Since 2002, he has performed research in bioinformatics, high-performance computing, cyberinfrastructure, network biology, genome assembly, systems genetics, paleogenomics, and bioenergy feedstock genetics. Currently, Feltus is an Professor in Clemson University's Dept. of Genetics & Biochemistry, CEO of Allele Systems LLC, Core Faculty in the CU-MUSC Biomedical Data Science and Informatics (BDSI) program, member of the Center for Human Genetics, and serves on the Internet2 Board of Trustees as well as various "Advance Research Computing" engagement workgroups. Feltus has published numerous scientific articles in peer-reviewed journals, teaches undergrad and PhD students in bioinformatics, biochemistry, and genetics. At present, he is funded by multiple NSF grants and is engaged in tethering together extremely smart people from diverse technical backgrounds in an effort to propel genomics research from the Excel-scale towards the Exascale.