Cornell News Highlights

Contact: Paul Redfern, red@cac.cornell.edu

OpenStack Cloud Implementation Toolkit Under Development by XSEDE and Aristotle

Two NSF-funded projects have joined forces to develop a new toolkit and service designed to help university systems administrators implement cloud computing systems on their campuses. "Combining the consulting strengths of the XSEDE Cyberinfrastructure Resource Integration (CRI) group with Aristotle Cloud Federation project expertise in OpenStack cloud implementation is a win-win for U.S. campus cyberinfrastructure administrators interested in adding clouds to their campus research resources," said John Towns, principal investigator and project director for XSEDE.

The toolkit and service are expected to be available in 2020. The plan is after an initial phone consultation, XSEDE staff would assist campus systems administrators onsite, if necessary, with the configuration of their OpenStack cloud and ensure that they have the knowledge they need to maintain it. There would be no fee for XSEDE staff to travel to U.S. campuses and assist local admins. Campuses or researchers interested in learning more about this opportunity should contact Rich Knepper at *rich.knepper@cornell.edu.*

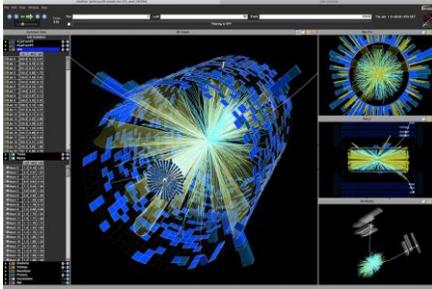### Dartmouth Joins Aristotle Cloud Federation to Explore Federated Cloud Computing Model

Dartmouth joined the Aristotle Cloud Federation, a group of universities that are sharing cloud computing resources to reduce time to science. Dartmouth joins project leader Cornell University and co-PIs University at Buffalo and UC Santa Barbara, in exploring this computing model, which the three schools built under an NSF Office of Advanced Cyberinfrastructure award. "We're excited about this opportunity," says Mitchel Davis, vice president and chief information officer at Dartmouth. "A federated cloud model has the potential to facilitate resource-sharing between campuses and is emerging as an important consideration in cyberinfrastructure planning." Davis says he hopes to provide Dartmouth researchers with access to cost-effective, on-demand systems, software, and data in a hybrid fashion that will let them choose from private, public, and federated cloud offerings. Universities interested in joining the Aristotle Cloud Federation should contact *help@federatedcloud.org*.

Cornell Builds Faster Infrastructure for Research

Thanks to a $376,714 NSF grant, Cornell Information Technologies developed 100 gigabit per second networking at Cornell, bringing the next generation of networking connectivity directly to the University's faculty and student researchers. "Providing the research community with high performance connectivity is extremely important to the university's mission," says CAC director and VP/CIO David Lifka.

Cornell Part of $25M NSF Effort to Untangle Future Physics Data

Laboratory of Elementary Particle Physics director Peter Wittich and his team, including CAC sr. research associate Steve Lantz, are developing software to handle the tsunami of data expected from the CERN LHC upgrade. The Cornell effort is part of the Institute for Research and Innovation in Software for High-Energy Physics (IRIS-HEP) project. Wittich said his team will focus on improving the performance of a particular part of the data analysis tool box in anticipation of the high-luminosity era of the LHC.

Aristotle Cloud Federation Investigating Open Cloud Marketplace

Cornell and its University at Buffalo and UC Santa Barbara partners received a $995,000 supplemental award from the NSF to investigate the feasibility of integrating public cloud cost metrics into the Aristotle Cloud Federation portal. Testing has begun with select Aristotle science use case apps and standard benchmarks. Public cloud providers AWS, Google, and Microsoft have provided credits for this investigation.

Cloud Usage Continues to Grow at Cornell; GPU's Now Available

Cornell faculty, staff, and student researchers continue to shift workloads where appropriate to public clouds and to Cornell's Red Cloud, the University's on premise IAAS platform that now includes GPUs.
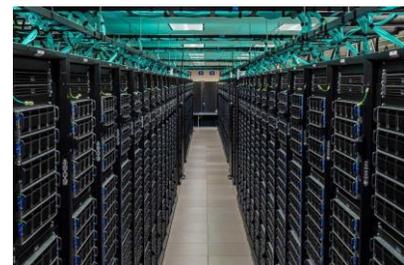

For example, Cornell Institute for Social and Economic Research scientists have acquired over 250 Red Cloud subscriptions. Social sciences computing can be an ideal fit for the cloud. The availability of large memory instances, such as 28 core instances with up to 240GB RAM on Red Cloud, is also helpful. Most recently, 4 Tesla NVIDIA V100 GPUs and 4 Telsa T4 GPUs were added to Red Cloud.

CAC staff Kim Burlingame and Ben Trumbore built 5 ready-to-use cloud images for the researchers: a Windows image with IBMSPSS, MATLAB, Mathematica, R, Rscript, RStudio, SAS, and Stata/MP and 4 custom Linux images that include Python, Miniconda, PostgreSQL, DataGrip, Kate editor, QIME2, and PICRUSt.

Cornell Developing Training for Frontera

Cornell CAC staff members are developing online, asynchronous training modules for the $60 million National Science Foundation-funded Frontera supercomputer that was launched on September 3, 2019 by the Texas Advanced Computing Center (TACC). As an operations partner in the Frontera project, Cornell is authoring content for online courses using its proven Cornell Virtual Workshop (CVW) platform, as well as adapting content created by the rest of the Frontera team.

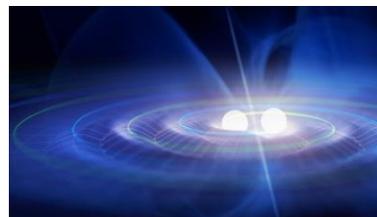**SC19 Keynote Speaker is Dr. Steven Squyres, Cornell Alumnus and Former Professor**

Between 2004 and 2019, mankind spent more than 5,600 cumulative days exploring Mars, yet the principle investigator on those missions never needed a spacesuit. That's because Spirit and Opportunity, NASA's two Mars rovers, were guided by the computing power that made it possible for Dr. Steven Squyres to oversee their findings from as much as 250 miles away.

Dr Squyres, Cornell alumnus and former James A. Weeks Professor of Physical Sciences, will be the keynote speaker for the SC19 supercomputing conference on Tuesday, November 19, 2019 at 830 a.m. in Denver. While he is best known to the public as the "face and voice" who chronicled Spirit and Opportunity's extended missions—each of which lasted years beyond initial expectations—Dr. Squyres played his most mission-critical role behind the scenes, helping guide the project science as the two rovers examined the terrain of Mars for signs of water or life.

CAC's Brazier is Tech Lead for Scalable Cyberinfrastructure Institute for Multi-Messenger Astrophysics

CAC is among ten collaborators awarded a $2.8 million grant from the NSF to develop the concept for a Scalable Cyberinfrastructure Institute for Multi-Messenger Astrophysics. The goal of the institute, called SCIMMA, is to develop algorithms, databases and computing and networking cyberinfrastructure to help scientists interpret multi-messenger observations such as gravitational waves and particles.

The code for SCIMMA will be developed and tested on CAC's NSF-funded Aristotle project – a federated cloud that allows multiple institutions to aggregate, share, and analyze very large datasets.

Cornell High School Programming Contest: STEM in Action

The Computer Science Department at Cornell and Cornell Tech, the University's campus located in Manhattan, invite teams of high school students to compete in the annual Cornell University High School Programming contest. More than 180 students competed in this year's event. An all-girls contest was added in February to encourage girls to participate in the general coding competition in April.

Alessandra Visca felt more confident returning for the April event. "I wanted to see how I would do against all the guys," joked Visca, a senior at Webster Schroeder High School. "It's growing every year," said Robbert van Renesse, computer science research professor, who coordinates the event. At Cornell Tech twice as many girls competed as last year. In Ithaca, Greg Morrisett, dean and vice provost of Cornell Tech, spoke to the competitors, encouraging them to pursue computer science. "We need computer scientists desperately," said Morrisett, citing U.S. Bureau of Labor Statistics showing that 70 percent of STEM jobs created in the next 10 years are expected to be in computer and information science.

New and Updated XSEDE Online Training Topics Available

New and updated Cornell Virtual Workshop training topics are available at the XSEDE user portal including *Code Optimization, Introduction to Wrangler, Message Passing Interface, Publishing Jetstream Virtual Machines, Science Gateway Roadmap,* and *Stampede2 Environment.* Cornell Virtual Workshops are freely available at all times to the entire scientific community of researchers, HPC practitioners, students, and educators at *https://portal.xsede.org/online-training/.*

Cornell's online training is developed in collaboration with CI experts at leading universities and centers, including the Texas Advanced Computing Center, the Indiana University Pervasive Technology Institute, and UC Berkeley.

XSEDE Software Toolkits Available for Campus Systems

Providing computational resources to researchers can be difficult, especially in terms of getting the necessary expertise to meet research needs. The NSF-funded XSEDE project provides access not only to resources but also to software toolkits that can be implemented on campus resources with minimal difficulty. XSEDE also offers consulting to help implement local resources that look like the supercomputing systems available at national centers. Rich Knepper and his XSEDE team, including CAC staff, are building an XSEDE Community Software Repository (CSR) that allows schools to implement a range of software, including cloud solutions, to support research computation needs and create a community of practice among implementing institutions. Visit the XSEDE CSR at *https://software.xsede.org/.*

Aristotle REU Students Make Meaningful Contributions to Science

Sixteen NSF-funded Research Experience for Undergraduates (REU) students have made meaningful contributions to Aristotle Cloud Federation science use case projects to date at Cornell and at UCSB.

The Aristotle project is grateful to the NSF for their support, as are the students. Student projects have included modeling, data analysis, and the enhancement of job duration and serverless computing tools, and containerization of scientific codes for portability to Aristotle, NSF clouds, and public clouds. Domain-specific experiences included building code to down-resolve radio astronomy search data, processing over 10 terabytes of complex datasets produced by weather models on multiple computing platforms, and creating visualizations to present data in a way that a lay audience could understand.