Cornell University
Center for Advanced Computing

Cornell News Highlights
Fall/Winter 2016
Contact: Paul Redfern - red@cac.cornell.edu

Cornell Leads New NSF Federated Cloud Project

Cornell University will lead a five-year, $5 million project sponsored by the National Science Foundation to build a federated cloud comprised of Data Infrastructure Building Blocks (DIBBs) designed to support scientists and engineers requiring flexible workflows and analysis tools for large-scale data sets.


The federated cloud will be deployed at Cornell University (CU), the University at Buffalo (UB), and the University of California, Santa Barbara (UCSB) and shared by seven science teams with over forty global collaborators. David Lifka, Director of the Cornell University Center for Advanced Computing (CAC) will lead the project with colleagues Tom Furlani, Director of the UB Center for Computational Research, and Rich Wolski, Professor of Computer Science at UCSB. Initial users of the cloud federation—earth and atmospheric sciences, finance, chemistry, astronomy, civil engineering, genomics, and food science—were selected based on the diversity of their data analysis requirements and cloud usage modalities. Their use cases will demonstrate the value of sharing resources and data across institutional boundaries.

CAC to Design Data Management System for New NSF NANOGrav Physics Frontiers Center (PFC)

The NANOGrav PFC will address a transformational challenge in astrophysics: the detection of low-frequency gravitational waves using the unique capabilities of the Green Bank and Arecibo telescopes.
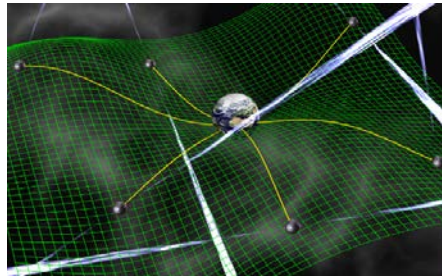
Cornell personnel on the project include astronomers James Cordes (Co-PI), Rachel Bean, and Shamibrata Chatterjee and CAC computational scientist Adam Brazier. Brazier is overseeing the NANOGrav cyberinfrastructure design and is also designing and maintaining the data management system. This Center, led by University of Wisconsin-Milwaukee PI Zavier Siemens, will enable the study of super-massive black hole binaries and stochastic backgrounds from cosmic strings and the earliest stages of the universe.

Cornell Simulations Support State-of-the-Art Planning for US Water Systems
Patrick Reed, Civil and Environmental Engineering, relies on a high-performance computing cluster maintained by CAC to simulate decision analytics for water management systems that take into account complex factors such as energy systems, ecosystem services, expanding population, and climate change. In July, he reviewed the Southern California Metropolitan Water District's water management plans and provided recommendations for the future. New planning frameworks and software are essential to model the complex issues facing US water systems.
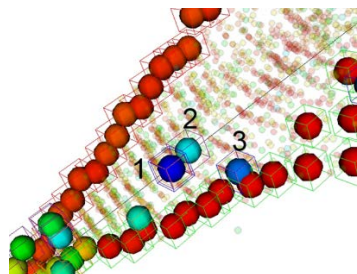
There are currently more than 15,000 users of the visual, model-based tools and software that Reed and his team have created. Reed appreciates CAC's computational and systems expertise. "CAC is the best support group in the United States. They help make my work possible."

**Cornell Virtual Workshop Users Surpass 100,000; New Cornell Job Runner Service[SM] Available**
Over 100,000 learners have accessed Cornell Virtual Workshop modules since the online learning platform was launched by CAC. The latest platform enhancement is called Cornell Job Runner Service (CJRS). It provides learners immediate access to a compute environment that looks and feels like a typical HPC login node or batch job, allowing incorporation of on-demand learning experiences interspersed with general training content. The web service and SLURM that supports the Cornell Job Runner Service are running as Red Cloud instances. CAC is the training lead for XSEDE and is developing a new generation of Cornell Virtual Workshops for XSEDE and for Jetstream, a new NSF-funded science and engineering cloud computing system.

Annual Cornell IT Conference Panel Discusses Role of Women in Computer Science
The fourth annual IT@Cornell Conference, held June 25 in Bailey Hall, was attended by hundreds of IT and research professionals from Cornell and neighboring colleges and universities. Irene Qualters, division director of advanced cyberinfrastructure at the National Science Foundation, led a panel of Cornell faculty in the discussion of the role of women in computer science and IT, and why they are in the minority. The all-female panel included Eva Tardos, the Jacob Gould Schurman Professor of Computer Science; Marjolein van de Meulen, the Swanson Professor and the James M. and Marsha McCormick Chair of Biomedical Engineering; and Perrine Pepiot, assistant professor of mechanical and aerospace engineering. The conclusion: Women need to help other women, men need to get on board, and employers need to expand their "comfort zone," recognizing that diverse voices can bring in new ways of thinking. Van de Meulen said that men and women must educate themselves about racial and gender unconscious bias in selection processes in every level of the STEM pipeline. More women will be attracted to STEM majors and professions, she added, if we can show that engineering makes things that help people. "The way to get diversity in engineering is not to market it as 'cool,' but to show how impactful it is," she said.

**Fennie Group Expands HPC Cluster at CAC for Materials Research**
Craig Fennie's Research Group has expanded their high-performance computing cluster to 1,112 cores to accelerate fundamental and technological innovation using a novel first-principles approach to materials discovery.
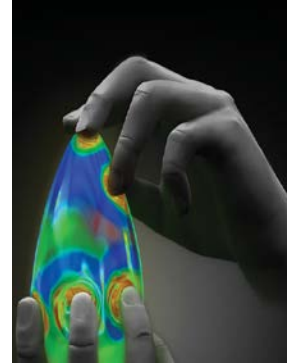
One application of their research is to develop materials with coupled electrical and magnetic states, which opens the possibility of digital data storage being controlled electronically rather than magnetically. Their ultimate goal is to make Materials-by-Design a reality. CAC houses and maintains the Fennie cluster and over 30 other systems for Cornell University faculty and staff. Fennie is a MacArthur Fellow.

Desjardins to Use High Fidelity Simulations to Develop New Turbulence Models
Olivier Desjardins, Mechanical and Aerospace Engineering, recently won a NSF award to develop multiphase turbulence models based on the kinetic theory for granular flows. Areas of application include engines, industrial furnaces, fluidized beds for fossil and renewable energy conversion, and chemical reactors.
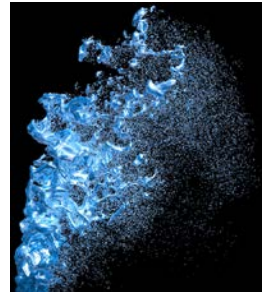
The collaborating PIs plan to offer a full implementation of their models and code in ANSYS and Fluent as well as public OpenFOAM platforms. The goal is to develop and validate a model that is generalizable to industry. CAC operates a high-performance computing cluster for the Desjardins research group. This award by the Fluid Dynamics Program of the CBET Division is co-funded by CIF 21 Software Reuse Venture Fund Program of the CISE/ACI Division.

Cornell Gates Hall Dedicated; Ground Broken for Cornell Tech
On the heels of opening Gates Hall, its new Computing and Information Science Building in Ithaca, Cornell is now taking its most daring step yet, building a $2 billion graduate campus on New York City's Roosevelt Island that will offer startup-focused curricula in business and computer science in the hopes of finding the next Zuckerberg. Called Cornell Tech, it will cement Cornell's identity as an innovation powerhouse—and stamp its bearish imprint on New York City, where a fifth of its 250,000 alums already live. Cornell Tech started classes at Google's offices in 2013 with plans to move to the Roosevelt Island campus in a few years, when it's complete.

Entrepreneurship is a big part of the future of Cornell. "The entrepreneurial passion is something you can't teach," admits Dan Huttenlocher, dean of Cornell Tech, speaking from Google's offices overlooking Manhattan's Chelsea neighborhood. "But I think an Ivy League university is a tremendous place to get the skills to make you much more likely to succeed in your entrepreneurial endeavors." – edited from *Forbes*

Cornell Deploys One of First Private HIPAA Clouds in Academe
Weill Cornell Medical College (WCMC) deployed a HIPAA eligible internal cloud called Red Cloud Secure, with 192-cores spanning two sites: one at CAC in Ithaca, the other at WCMC in New York City. The cloud is built on the Eucalyptus platform and is compatible with Amazon Web Services. Applications will include mobile health databases, genetic sequencing pipelines, and large memory metabolomics data analysis using MetaboAnalyst.

CAC systems staff developed the Red Cloud Secure platform and assisted WCMC in its deployment. CAC HIPAA cloud consulting is available to other universities, public agencies, and industry through the Center's Partner Program.

Cornell Engineering Programs Capture Top 10 Rankings
Cornell graduate programs were ranked top 10 in the following 2016 US News categories: Biological Engineering, Civil Engineering, Computer Engineering, Electrical Engineering, Industrial Engineering Materials Engineering, and Mechanical Engineering.

Research Repository arXiv Hits 1 Million Submissions; Growing Rapidly in New Science Fields
It all started with an electronic bulletin board: one computer on one scientist's desk. Now, more than two decades later, arXiv, is a driving force in scientific communication, it draws in thousands of researchers every day. As an open-access service, it allows scientists from disciplines encompassing physics, statistics, computer science, and others to share research before it's formally published. One million papers have now been uploaded to the repository. "arXiv is the go-to source, the core of an ecosystem, and it fills critical research needs for researchers around the world every day." CAC consultants have worked on software development for arXiv.

Research, Innovation allows Corning to Thrive, says CEO
Corning's commitment to innovation and its ability to reinvent itself have allowed the company to thrive for 163 years in an environment that is increasingly difficult for businesses to remain relevant long term, said Wendell Weeks, chairman and CEO of Corning Inc. Weeks delivered Cornell's 27th annual Lewis B. Durland Memorial lecture.


Weeks showed one of Corning's innovations, flexible glass, and outlined four strategies that allow companies to thrive in what he called "creative destruction:" understanding why your institution is here, organizing staff to deliver that core message, collaborating with others who share your vision, and having a clear framework for guidance. This approach helped Corning develop one of its key products, Gorilla Glass. "Most great things are done in groups, teams, institutions, cultures, and great universities," Weeks said. "We celebrate the individual, but actually that's never the true story." Corning is a member of CAC's Partner Program.

CAC to Develop Cloud Training for U.S. Science and Engineering Community
As a collaborator on the NSF-funded Jetstream program, CAC will be developing cloud computing training for the national community. Jetstream – led by Indiana University's Pervasive Technology Institute – will add cloud-based computation to the national cyberinfrastructure. "As an early cloud adopter and provider with Red Cloud, Cornell CAC staff will leverage their cloud computing expertise to develop and deploy on-demand Jetstream training," says Susan Mehringer, assistant director, CAC. "The first Cornell Virtual Workshop module will introduce users to Jetstream capabilities and teach them how to optimally use, create, and archive cloud services and virtual machines (VMs)," she explained. Subsequent modules will focus on using remote desktops to access Jetstream, XSEDE, and systems in the eXtreme Digital (XD) program; deploying biology and earth science applications on Jetstream; and, publishing, archiving, and curating VM images to ensure scientific reproducibility."

A Fast Track for Data between Ithaca and New York City

A spin-off from research aimed at improving the performance of the Internet will give Cornell researchers a faster way to transfer large files between the Ithaca campus and Weill Cornell Medical College (WCMC) in New York City starting with massive files of genomics data. If the project, dubbed the Cornell Open Science Network (COSciN), is successful it could help researchers everywhere, so it is being supported by the National Science Foundation, from a special fund developed to upgrade the cyberinfrastructure of university research. PIs are Nate Foster and Hakim Weatherspoon of computer science, and Jason Mezey, a computational biologist. They are partnered with Cornell Information and WCMC information technologies.
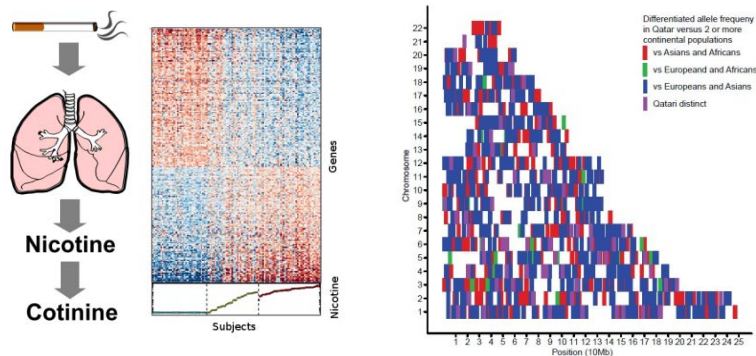
Research projects will include determining the genetic ancestry of individuals, identifying genes important for diseases, and comparing cells from the lungs of smokers and nonsmokers. Huge files of genomics data are becoming common because the cost and complexity of sequencing a genome is rapidly decreasing.