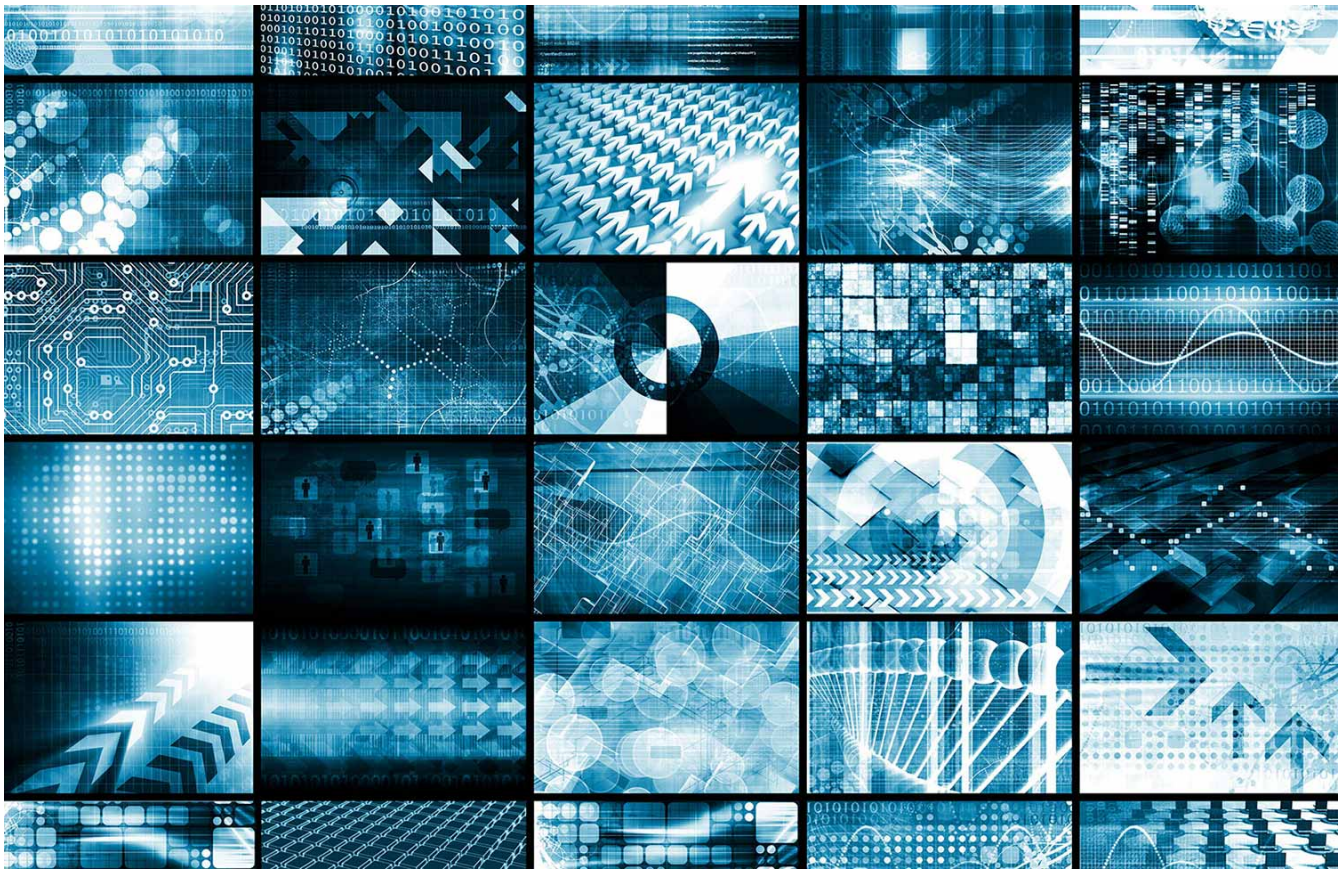
UC San Diego
UC San Diego News Center

November 12, 2015 | By Jan Zverina

SDSC's 'Comet' Supercomputer Blazing Trails via Science Gateways

Web-based portals attract researchers across numerous domains



Just six months after coming online, *Comet,* the new petascale supercomputer at the San Diego Supercomputer Center (SDSC) at the University of California, San Diego, is already blazing new paths of discovery, thanks in part to its role as a primary resource for an assortment of science gateways that provide scientists across many research domains with easy access to its computing power.

Simply described, science gateways provide web browser access to applications and data used by specific research communities. Gateways make it possible to run the available applications on supercomputers such as *Come*t, so results come quickly, even with large data

sets. Browser access offered by gateways allows researchers to focus on their scientific problem without having to learn the details of how supercomputers work and how to access and organize the data needed.

“It’s possible to support gateways across many disciplines because of the variety of hardware and support for complex, customized software environments on *Comet*,” said Nancy Wilkins-Diehr, an associate director of SDSC and co-director of Extended Collaborative Support Services for the National Science Foundation’s XSEDE (eXtreme Science and Engineering Discovery Environment) program, an advanced collection of integrated digital resources and services that include *Comet* as a national resource for U.S. academic researchers. “This is a great benefit to researchers who value the ease of use of high end resources via such gateways.”

For the most recent quarter ended September 30, there were 3,310 gateway users across all XSEDE systems, according to data compiled by Wilkins-Diehr. There were 64,377 research jobs run by all gateways across all systems during the quarter, and 86 percent of them were run on either *Comet* or SDSC’s data-intensive *Gordon* supercomputer.

“That’s a notable level of usage for a new machine,” said SDSC Deputy Director Shawn Strande, who also is *Comet’s* program manager. “We anticipate that *Comet* will reach an active research community of more than 10,000 users, mostly via gateways. Our goal for *Comet* is to speed up as many researchers as possible, rather than supporting a handful of heroic calculations, so we configured it to serve as one of the most productive HPC systems available to the academic research community, just as its predecessor, *Trestles*, was.”

In recent years the most popular science gateway in XSEDE has been CIPRES, which stands for CyberInfrastructure for Phylogenetic RESearch (*see sidebar*). Typically, about 200 CIPRES jobs are running simultaneously on *Comet* and another 100 on *Gordon*.

“The scheduling policy on *Comet* allows us to make big gains in efficiency because we can use anywhere between one and 24 cores on each node,” said Mark Miller, a bioinformatics researcher with SDSC and principal investigator of the CIPRES gateway. “When you are running 200 small jobs 24/7, those savings really add up in a hurry.”

Currently, 30 science gateways are available via XSEDE’s resources, each one designed to address the computational needs of a particular community such as computational chemistry, phylogenetics, or the neurosciences. SDSC itself has delivered 77 percent of all gateway cycles since the start of the XSEDE project in 2011.

Supported by an NSF grant worth almost $24 million including hardware and operating funds, *Comet* is designed to meet the needs of what is often referred to as the 'long tail' of science – the idea that the large number of modest-sized computationally-based research projects represent, in aggregate, a tremendous amount of research that can yield scientific advances and discovery. A video about *Comet* can be viewed here.

Some of the science gateways now accessible on *Comet* and other selected XSEDE HPC resources include:

CIPRES: The CIPRES science gateway was created as a portal under the NSF-funded Cyberinfrastructure for Phylogenetic Research (CIPRES) project and began using supercomputers at the end of 2009. In 2013 SDSC received a $1.5 million award from the NSF to extend the project to make supercomputer access simpler and more flexible for phylogenetics researchers. "Access to supercomputers is a key part of modern evolutionary science, where evolutionary relationships are explored by comparing DNA sequence information between species," said SDSC's Miller.

To date, the CIPRES science gateway has supported more than 14,000 users conducting phylogenetic studies involving species in every branch of the Tree of Life. It is the most popular science gateway in XSEDE, with 49% of all XSEDE users running via CIPRES during the third quarter of 2015. The gateway is used by researchers on six continents, and their results have appeared in more than 1,800 scientific publications since 2010, including *Cell*, *Nature*, and *PNAS*.

Neuroscience Gateway (NSG): The Neuroscience Gateway eliminates most administrative and technical barriers facing neuroscientists who need to use high-performance computing resources for large modeling projects and other computationally intensive tasks, such as analysis of neuroimaging data.

Last month, the NSF and the United Kingdom's Biotechnology and Biological Sciences Research Council (BBSRC) awarded funding for the next phase of the NSG. The project will contribute to the national BRAIN initiative announced by the Obama administration in 2013 to advance researchers' understanding of the human brain. That collaborative project, funded under three separate awards, is between UC San Diego, Yale University, and University College of London. Amit Majumdar, division director for SDSC's Data Enabled Scientific Computing group, is a principal investigator in the project.

The NSG is a single access point for powerful simulators and data analysis tools widely used by neuroscientists. Its web-based interface simplifies the tasks of uploading models or data, specifying job parameters, monitoring job status, and storing and retrieving output data. NSG has logged 4.5 million core hours on XSEDE supercomputers, serving more than 270 neuroscientists since 2013.

MP-Complete: This gateway is a collaboration with the Materials Project to crowd-source novel crystal structures for first principles computations. The Materials Project is an open science database of computed materials data funded by the Department of Energy. Shyue Ping Ong, an assistant professor at UC San Diego, Gerbrand Ceder, a professor at UC Berkeley, and Kristin Persson, an assistant professor at UC Berkeley are the principal investigators of the project. This gateway is available only on *Comet*.

"The Materials Project provides a Google-like database of material properties to serve the materials science, computational physics, and chemistry research communities," said Wilkins-Diehr. "Thanks to advances in codes and computing, this gateway contains software, algorithms, databases, and web tools for researchers to create and suggest novel materials for data to be calculated using first principles methods, while reducing duplication of calculations on identical materials. Researchers can also data-mine scientific trends in materials properties."

SEAGrid: The SEAGrid science gateway, funded by the NSF's SciGaP project, serves researchers and teachers in the areas of computational chemistry, molecular dynamics, structural dynamics, and fluid dynamics. Currently, SDSC's *Comet* cluster is the workhorse for these communities. Led by Sudhakar Pamidighantam, a member of the Science Gateways Group in Research Technologies division at Indiana University, SEAGrid has been in operation since 2005 and serves more than 600 scientists and students under more than 300 projects. "SEAGrid has supported and enabled more than 120 publications," said project leader Pamidighantam. "Last year alone, and with about 3 million compute hours via XSEDE, it enabled execution of almost 25,000

MEDIA CONTACT

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