

## SDSC Announces First Round of Triton Research Opportunity (TRO) Grants

*Four Computer Science, Bioengineering Projects Selected for Triton Resource System*

February 10, 2010

Jan Zverina

The San Diego Supercomputer Center (SDSC) at UC San Diego has announced the first round of computer allocations under the Triton Research Opportunities (TRO) program associated with the center's new *Triton Resource*, a data-intensive computing system primarily designed to support UC San Diego and UC researchers.

"These awards will help drive the development of the *Triton Resource* and our new flash memory-based *Dash* system that is part of it, while promoting strong levels of collaboration across our campus and ultimately the entire UC system, especially in data-intensive science, which is collaborative, networked, and data-driven," said Chaitan Baru, SDSC Distinguished Scientist and Chair of the Project Advisory Committee (PAC) consisting of 10 UC San Diego faculty members and researchers who judged the submissions. "We are grateful to the SDSC PAC members for their thoughtful and in-depth reviews of a strong roster of proposals in this inaugural round of the competition." Projects selected from the first round of submissions included researchers from several UC San Diego departments. They are:

*"UCSD Triton-based Resource for Managing Large Spatially-registered Biomedical Images."* Ilya Zaslavsky (SDSC), Mark Ellisman (Neurosciences and Bioengineering).

The project will leverage existing tools and community approaches to create a *Triton*-based resource for managing and serving large images and volumes derived from different types of microscopes at the National Center for Microscopy and Imaging Research, in the UC San Diego's Neurosciences Department.

*"Benchmarking Data Intensive Applications on the Triton Resource."* Chaitan Baru (SDSC), Sriram Krishnan (SDSC), Yannis Papakonstantinou (Computer Science & Engineering).

Using high-resolution topography datasets served by the OpenTopography facility based at SDSC, this project will investigate hybrid implementations of scientific data archives, encompassing shared-nothing parallel database systems and MapReduce/Hadoop-based models. The project will install benchmark datasets, develop a set of benchmark queries, and conduct a number of performance experiments on the *Triton* "fat" or large-memory nodes.

*"MapReduce for the Rest of Us: An Intuitive and Generic Architecture Facilitating Data- Intensive Applications in Metagenomics"* Ilkay Altintas (SDSC), Mark Ellisman (Neurosciences and Bioengineering), Jeffrey Grethe (Center for Research on Biological Systems) and Paul Gilna (Calit2).

The project will develop extensions to the Kepler scientific workflow system to incorporate usage of MapReduce, and deploy this capability on the *Triton Resource*. These workflows will initially be published for users via the CAMERA portal.

*"Processing Massive Graphs on Triton: Performance, Optimizations, and Applications"* Sandeep Gupta (SDSC), Trey Ideker (Bioengineering), Fan Chung (Mathematics, Computer Science & Engineering), and Jeffrey Grethe (Center for Research on Biological Systems).

This project focuses on the development of a graph toolkit called Scalable Graph Framework (SGF), to optimize processing of very large graphs by exploiting the solid state disks and the versatile SMP (vSMP™) capability of SDSC's new *Dash* resource. The SGF will be tested initially in applications in the areas of bioinformatics and neuroinformatics.

The Triton Research Opportunities program provides seed funding for developing collaborative, extramural proposals between SDSC and UC San Diego researchers who leverage the *Triton Resource*. The program consists of a proposal competition for 10 percent of Triton Resource cycles and funds equivalent to \$200,000 per year in the form of SDSC full-time employees working on the collaborative project. Full details of the TRO program can be found at <http://cired.sdsc.edu/tro/troinfo.php>. The deadline for submission of requests for the second round of allocations is March 15, 2010.

The *Triton Resource*, which features some of the most extensive data analysis power available commercially or at any research institution in the country because of its unique large-memory nodes, was formally announced last August. One month later, SDSC unveiled an innovative element of the *Triton Resource* called *Dash* - a super-sized version of a flash memory-based supercomputer. This "system within a system" will help researchers looking for solutions to particularly data-intensive problems that arise in astrophysics, genomics and many other domains of science.

For additional details for SDSC's new Triton Resource, including the Triton Affiliates and Partners Program (TAPP) and information on how to obtain allocations on the system, please also visit <http://tritonresource.sdsc.edu>. Industry users interested in learning more about the Triton Resource should contact Ron Hawkins, SDSC's director of industry relations, at (858) 534-5045 or [rhawkins@sdsc.edu](mailto:rhawkins@sdsc.edu).

## **About SDSC**

As an organized research unit of UC San Diego, SDSC is a national leader in creating and providing cyberinfrastructure for data-intensive research. Cyberinfrastructure refers to an accessible and integrated network of computer-based resources and expertise, focused on accelerating scientific inquiry and discovery. SDSC recently doubled its size to 160,000 square feet with a new, energy-efficient building and data center extension, and is a founding member of TeraGrid, the nation's largest open-access scientific discovery infrastructure.

Media Contact: Jan Zverina, SDSC Communications, 858 534-5111 or [jzverina@sdsc.edu](mailto:jzverina@sdsc.edu), Warren R. Froelich, SDSC Communications, 858 822-3622 or [froelich@sdsc.edu](mailto:froelich@sdsc.edu)