April 07, 2014 | By Jan Zverina

SDSC Enables Large-Scale Data Sharing Using Globus

The San Diego Supercomputer Center (SDSC) at the University of California, San Diego, has implemented a new feature of the <u>Globus</u> software that will allow researchers using the Center's computational and storage resources to easily and securely access and share large data sets with colleagues.

In the era of "Big Data"-based science, accessing and sharing of data plays a key role for scientific collaboration and research. Among SDSC users there is a need to share datasets, which can be large, with collaborators who may not have accounts on SDSC resources. The new Globus feature addresses this need.

Described as a "dropbox for science", Globus is already widely used by resource providers and users who need a secure and reliable way to transfer files. SDSC is the first supercomputer center in the National Science Foundation's <u>XSEDE (eXtreme Science and Engineering Discovery Environment)</u> program to offer the new and unique Globus sharing service.

While SDSC has been offering file transfer capability via Globus to users for several years, the Center is now providing a number of Globus Plus accounts via a <u>Globus Provider plan</u> to selected users free of charge so that they can allow their collaborators, including those who don't have an account on SDSC clusters, to access (read and write to their shared file space) data on SDSC resources.

SDSC staff will issue these accounts based on researchers' needs for sharing data with their collaborators, such as if they are part of a larger collaboration where data sharing becomes crucial. Separately, researchers will be able to purchase a Globus Plus account from Globus directly, with subscriptions currently priced at $7/month or $70/year.

"Integrating the Globus sharing capability into SDSC's widely used data-intensive computing and storage systems that include *Gordon*, *Trestles*, and *Data Oasis* is important because it allows researchers and resource providers to hand off the challenges of data sharing and

movement to a hosted service that manages the entire process, while also monitoring performance and providing status reports," said Amit Majumdar, director of SDSC's Data Enabled Scientific Computing division.

"Big data has become an integral part of the research landscape, and with that comes the challenge of extracting meaningful value from those massive data sets," said SDSC Director Michael Norman. "That process is often done through multi-site collaborations. With SDSC at the forefront of big data management and expertise, enabling Globus sharing on our high-performance compute and storage systems lets scientists focus on their research, and not be distracted by challenges associated with sharing data or having to seek time-consuming IT help. I view Globus data sharing as a way to reach a broader audience of researchers beyond those who do the simulations."

Rick Wagner, manager of SDSC's HPC Systems group, and Mahidhar Tatineni, manager of SDSC's User Services group, have been working with Globus staff to install Globus software on SDSC's GridFTP servers and test its various features. Based on their experience, they expect SDSC users to rapidly adopt the software for data sharing because of its ease of use. SDSC users from domain sciences such as genomics, economics, and astrophysics are already starting to use Globus to share research data with their collaborators.

"We are excited to see SDSC become the first XSEDE resource provider to offer Globus sharing, and we will work with the SDSC team to increase adoption of the service and facilitate enhanced scientific collaboration among their users," said Steve Tuecke, Globus project co-lead. "As an early Globus Provider plan subscriber, we appreciate SDSC's support in helping Globus become a self-sustaining service for all researchers."

To start using the Globus sharing feature, users who hold a Globus Plus account at SDSC need to follow the instructions provided here.

Full details on the sharing service are provided here.