New Globus Online service to use the grid for high-performance file transfer

November 22 2010, By Vas Vasiliadis

A new tool that will provide secure file–transfer service to manage large–scale data was introduced Nov. 18 at an international conference in New Orleans. Globus Online, which will use a cloud–based system rather than complex, custom information technology infrastructure, was displayed at SC10, an international conference for high–performance computing, networking, storage and analysis.

The team that created Globus is housed at the Computation Institute, a joint initiative between the University of Chicago and Argonne National Laboratory, and includes collaborators from the University of Southern California's Information Sciences Institute and Lawrence Berkeley National Laboratory.

Globus Online automates the mundane but error–prone and time–consuming activity of moving files across wide area networks. Users can make their file–transfer request, and Globus Online will manage the entire operation, monitor performance, retry failed transfers and recover from faults automatically whenever possible so that users focus on their research.

The National Energy Research Scientific Computing Center was an early adopter of Globus Online and is now recommending the service to its users as a secure, fault–tolerant file–transfer solution.

"We are excited to offer a simplified, yet reliable data movement method to our users," said David Skinner, group leader at NERSC. "As

the collection of Globus Online endpoints grows, our users will be using the highest–performing wide area network–tuned systems with simplicity."

Globus Online works across and between any servers that have the GridFTP software installed, including systems at Department of Energy facilities such as NERSC, Oak Ridge, and Argonne; large–scale Globus based cyberinfrastructure in the United States and abroad, such as TeraGrid, Open Science Grid, European Grid Infrastructure, and Australian Research Collaboration Service; as well as many campus grids. Users can rapidly access this international pool of resources using Web, command line and Representation State Transfer interfaces, without installing any software.

"We are in an era of explosive data growth, where requirements for reliable data movement are more demanding, and distributed environments are more complex. Globus Online goes a long way toward helping users navigate this 'perfect data storm' without expensive, custom–built solutions," said Ian Foster, director of the Computation Institute. "Building on our 15–year heritage with the Globus Toolkit, we continue to challenge ourselves to deliver innovative solutions for large–scale distributed computing."

Commonly used file transfer mechanisms, such as the secure copy command found on most systems, require complex configuration for optimal performance and frequent attention from the user to deal with transient faults — perhaps the most frustrating aspect of distributed data management. Globus Online transparently handles faults by re–trying failed transfers and provides detailed logs for users and service operators to understand the reasons behind any failed transfers.

Managing security across multiple domains has traditionally been a challenge for grid users. Globus Online manages multiple security

credentials automatically on behalf of users. When a certificate expires, for example, Globus Online suspends the file transfer, notifies the user and automatically resumes when a valid credential is received.

In its current release, Globus Online allows users to move files between systems with GridFTP installed. Upcoming releases will add support for HTTP, opening up the service to any file system with access to a Web server, as well as InCommon, allowing members of many academic and research institutions to access the service using their existing campus logins.

Provided by University of Chicago