

Campus Bridging

Call for Proposals: Pilot sites to deploy and evaluate XSEDE infrastructure

The Campus Bridging team of the National Science Foundation's Extreme Science and Engineering Discovery Environment (XSEDE) project asks for proposals from campuses to become early adopters and evaluators of XSEDE architecture components. Campus Bridging seeks two to four campuses that are well-experienced (and have current, active users) with the recently-ended TeraGrid program and one to three campuses that are not as well-experienced with TeraGrid to pilot the XSEDE Global Federated File System (GFFS) and the features it enables.

Specifically:

- Users can include (export) in the GFFS resources — such as a directory structure or cluster — on campus or research group machines. Users, including collaborators at other institutions, can then map the GFFS into their local filesystem using a GFFS-aware FUSE (Filesystem in Userspace) driver and securely access remote resources as files and directories in their local environment. In other words, authorized clients can access (create, read, write, delete) locally exported resources from anywhere in XSEDE using the GFFS. The result is a globally shared file system that spans the centers, campuses, and research groups. A critical point is the use of FUSE (<http://fuse.sourceforge.net/>) — which allows a “normal user” without systems administrator privileges to install the GFFS software and map the filesystem her/himself.
- Two collaborating research groups might map their PBS/SGE controlled clusters into XSEDE and create a shared grid queue in the global namespace. The research groups and other selected collaborators can then use the shared grid queue. The grid queue can schedule jobs on their private clusters or on private clusters and an XSEDE resource where they have an allocation. The creator of the queue controls access to this new, shared grid queue. Jobs running via the shared queue can access data and executables via the GFFS.

The Campus Bridging effort within XSEDE offers consulting and assistance on XSEDE and related middleware and cyberinfrastructure architecture; it does not offer money for those campuses selected to participate in the pilot project. Campus Bridging representatives will visit each campus selected to participate in a pilot project. They will offer training and hands-on help with installing the GFFS client (Windows, Linux, and MacOS), how to define and run jobs from the desktop, how to use and access the GFFS, how to include users' own data resources in the GFFS, and how to manage access control to those resources.

The Campus Bridging team will expect thorough feedback, good and bad, on XSEDE architecture elements deployed on their campuses. Campus IT staff and researchers on the selected campuses will be exposed to the latest XSEDE-recommended infrastructure, and they will play a vital part in shaping future updates and deployments of that infrastructure.

Every campus is different, but in selecting proposals, strongest consideration will be given to campuses that have:

- active XSEDE Campus Champions
- an established strategic plan that includes research computing
- a significant level of a shared computing resource, which might be shared in a closer engagement with XSEDE in the future
- a specific need to share data between users at multiple institutions and preferably some reasonable estimate of overall campus requirements for data movement to and from XSEDE resources.

Proposals should be no longer than three pages, including at least one paragraph each on the intellectual merit and broader impacts of the proposal — similar to NSF proposal requirements. In addition, each proposal should include a facilities statement for the campus detailing the size and scope of its computing resources, as well as NSF-style bios for at least two people who will be the responsible parties on the campus.

Early Adopter proposals are to be submitted to **campusbridging@xsede.org** from December 1-9, 2011. Selected proposals will be notified by Monday, December 19, 2011.

A sample proposal follows this CFP.

FOR MORE INFORMATION

- **XSEDE:** <https://www.xsede.org>
- **Campus Bridging:** <https://www.xsede.org>



— S A M P L E —

Proposal for XSEDE Campus Bridging Early Adopter Program

Bringing Indiana University data to XSEDE resources

OVERVIEW – Implementation of GFFS at Indiana University

Data management for researchers is a task that is frequently characterized as difficult to orchestrate effectively and perilous to execute poorly. Indiana University proposes implementation of the Global Federated Filesystem (GFFS) for research users at IU, as well as on the Quarry Gateway Web service, making it possible for users to develop and deploy code on the gateways in a new, more seamless fashion. IU also will provide support and consultation to research users to make data on their systems available to XSEDE resources, as well as on IU local systems and with collaborators at other institutions.

Research is built upon the collection and organization of data, and data loss is catastrophic to any research project. Most researchers tend to re-use solutions they have relied upon in the past and pull together solutions they are familiar with and understand to move data from collection through cataloging, analysis, and archival stages. When researchers collaborate across multiple sites, they find these difficulties increase, with additional layers of authentication, filesystem types, and security controls. The Global Federated Filesystem makes it possible for data from one location to be easily shared on many systems. It creates a single namespace for all files and makes use of the implementation of Filesystem in Userspace (FUSE), which allows unprivileged users on systems to access filesystems in a familiar way without extensive permissions and installation of network filesystem services on resources. IU will work with local researchers to make the GFFS installers available and provide consulting to export filesystems proximate to the researchers – on desktops and in laboratories – to local resources as well as national resources.

GFFS provides a number of useful functions that interact well with the resources provided at IU. One of the primary use cases described by the GFFS team is the ability to export local data and code to XSEDE resources. By providing support to IU researchers for the GFFS installer, IU will make it easier for users to access their data on XSEDE resources and share it with XSEDE users and collaborators at other institutions. IU's gateway hosting service frequently is used by users who are developing and implementing gateway code. Keeping local development and science gateway machines in sync can be a problem due to different operating system choices, maintenance issues, and organizational concerns. The GFFS team has made a strong case for GFFS as a development tool that allows rapid development on code, which is then exported via GFFS and implemented in place on a given resource. While this can be a computational resource, it also is possible to deploy code in this way to a science gateway. IU will: make GFFS available in the default software selection for gateway web service hosts and assist gateway hosting service users with GFFS installation for their own machines, make training and documentation available, and provide support for users making use of GFFS on the gateway web service.

INTELLECTUAL MERIT

By making GFFS available to IU users, IU will be able to bring experiences of a number of users who are familiar with national-scale infrastructure — with TeraGrid and XSEDE systems — and who are active developers of gateway systems, as well as code running on national-scale resources. The experiences of these users will provide significant input for the level of training and documentation appropriate to users who are familiar with these types of resources. In addition, these users have seen many different solutions for file transfer and frequently make use of some kind of metadata services that will take advantage of advanced uses of GFFS, allowing these user experiences to provide input on future needs for GFFS development.

The IU team has been involved with the TeraGrid and XSEDE projects and is familiar with the software involved in national cyberinfrastructure for a number of years, providing a solid background of experience with software comparable to GFFS and with research users who make use of similar software.

BROADER IMPACT

Use of GFFS by researchers at IU will allow for improved sharing of data and utilization of data on XSEDE resources, accelerating the advancement of research projects and increasing the usability of XSEDE resources. Facilitation of collaboration will improve research by allowing greater participation both between and within multiple fields of science, allowing researchers to tackle broader and more complex projects easily. The IU team plans to work with researchers who are carrying out work on both local resources and national cyberinfrastructure, facilitating collaborations within the IU campus, to XSEDE, and with other institutions.

FACILITIES STATEMENT

IU has a number of high performance computing systems that can take advantage of files exported via GFFS and, most importantly, a robust network infrastructure that allows the transfer of files to GFFS to take place quickly and seamlessly, both on campus and to national cyberinfrastructure. Resources and networks are briefly detailed below.

High performance computing (HPC) systems

IU has the following production high performance computing systems:

- Big Red, an IBM e1350 distributed shared-memory cluster with 4096 processor cores, 6 TB total memory capacity, and a peak theoretical processing capability of 40.96 TFLOPS
- Quarry, an IBM e1350 distributed shared-memory cluster with 1120 processor cores, 1.2 TB total memory capacity, and a peak theoretical capability of 8.96 TFLOPS
- Mason, an HP distributed shared-memory cluster with 512 processor cores, 8 TB total memory capacity, and a peak theoretical capability of 3 TFLOPS

Facilities for handling sensitive data

IU has established appropriate administrative, technical, and physical controls to protect data in accordance with the HIPAA security rule. Electronic Personal Health Information may be stored on all of the HPC and storage facilities described in this document with the exception of the FutureGrid systems.

Networking

Services include routed Internet Protocol (IP) and switched Ethernet, Gigabit Ethernet (GigE), and 10GigE. Research Technologies (RT) leverages single nonblocking backplane data center switches from Force10 and Cisco for interconnectivity between Indianapolis and Bloomington, as well as network access for computing resources. IU provides I-Light project dark fiber assets for exclusive research use, connecting resources across both campuses. RT lights these dark fiber paths with four native 10GigE links providing 40Gb total throughput between the campuses. Additionally, the RT network peers with XSEDE (two 10GigE) and FutureGrid (one 10GigE) connections, providing access to nationally connected computing resources. The RT network connects to the Indiana Gigapop at 10GigE for commodity Internet access and access to national research networks, Internet2 and National LambdaRail (NLR).

PARTICIPANTS

Craig A. Stewart is the executive director of the Pervasive Technology Institute (PTI), IU's flagship initiative for advanced information technology research, development, and delivery in support of research, scholarship, and artistic performances. Stewart is associate dean for research technologies and leads the Research Technologies Division of University Information Technology Services.

Richard J. Knepper is the manager of Research Technology Core Services, providing infrastructure services to researchers at IU and their collaborators that allow software project management, development, and deployment, and he assists in the support of the use of research cyberinfrastructure at IU.

Therese Miller is the manager of Collaboration and Engagement for Research Technologies, managing outreach and engagement to researchers at IU and the community at large.

Submitted by Indiana University
Proposal contact: Dr. Craig Stewart
stewart@iu.edu
812-855-4810
December 1, 2011