XSEDE Capability Delivery Plan
SGW-4 Large file data movement...mediated
by the gateway
Last revised 2016-06-08

Background

Use cases describe community needs, requirements, and recommendations for improvements to cyberinfrastructure "CI" resources and services. A Capability Delivery Plan "CDP" is an executive summary of use case support gaps, of plans to fill those gaps with new or enhanced capabilities, and of existing operational components that already support aspects of a use case.

Use Case Summary

Use case SGW-4 describes how science gateway developers can include a mechanism in their gateways that allows end users to transfer large files (up to 20GB) between their personal systems (desktops or laptops) and XSEDE compute and storage systems. The end users do not have personal allocations on the XSEDE systems, but the science gateway has a gateway allocation. Data transfers should be direct between the end user system(s) and the XSEDE system(s). Use case document: http://hdl.handle.net/2142/43883

CDP Summary

The functionality described in this use case is fully supported by the operational components listed below. The missing pieces are <u>documentation</u> and <u>training</u> for gateway developers.

Gap(s) that we currently plan to address:
- Document and prepare training materials on the use of XSEDE identity management to coordinate user identities between a science gateway and XSEDE.
- Document and prepare training materials on the use of XSEDE/Globus sharing features to provide direct transfer capability between end users and XSEDE storage systems.

Gap(s) that will <u>not</u> be addressed <u>at this time</u>:
- A web portal that science gateway developers use to setup and configure their relationship with XSEDE
- Verification of quality attributes

Time and effort summary:
- A total of seven FTE-weeks of staff effort (documentation research+writing and testing). (3 for SDIACT-245, 4 for SDIACT-246.)

Functionality Gaps

The following plan relies on the Globus groups feature, which has not been fully integrated with XSEDE at this time. (See the entry for Globus groups in the list of system components below.) This plan consequently includes documentation specifically designed to enable science gateway developers to use the groups feature. When XSEDE completes integration with the groups feature, a revised CDP for this use case can be written that takes advantage of any new features, documentation, user support, or interfaces.

1. Document the use of XSEDE identity management for science gateways
(suggested priority: high)

We need to document and prepare training materials on the use of XSEDE identity management to coordinate user identities between a science gateway and XSEDE.

Identity coordination is necessary because the implementation of this use case (described in gap #2 below) relies heavily on authorization by group membership, and group membership requires authenticated identities. In concrete terms, the file access services running on XSEDE systems must be able to identify the end user of the science gateway when the end user connects from their own laptop or desktop. The services then need to be able to recognize the end user's membership in an XSEDE group that was defined by the science gateway. Thus, there must be a way to recognize XSEDE group membership when starting with an identity credential that the end user possesses (as opposed to the science gateway's XSEDE credential, which the user does not--and must not--possess).

Plan: https://software.xsede.org/view/xci-2

2. Document the user of XSEDE/Globus sharing features to provide direct transfer capability between end user and XSEDE storage systems (suggested priority: medium)

We need to document and prepare training materials on the use of XSEDE/Globus system features to accomplish the following things *in the context of developing a science gateway.*
1. Use the XSEDE/Globus groups API or web interface to create a group in Globus.
2. Use the XSEDE/Globus groups API to manage membership of a group.
3. Identify XSEDE storage systems that support Globus sharing.
4. Use the XSEDE/Globus transfer API to create a shared endpoint on an XSEDE storage system.
5. Instruct end users to download and install Globus Connect Personal on Windows, Mac, and/or Linux systems.
6. Put items 1-5 together with XSEDE identity management (gap #1 above) to allow end users to transfer files directly to/from their own systems and XSEDE systems.

Plan: https://software.xsede.org/view/xci-3

3. Web portal for science gateway developers to use to setup and configure their relationship with XSEDE (suggested priority: medium)

XSEDE currently does not provide a self-service web portal for setting up the credentials, registrations, configurations, etc. necessary for a science gateway to use XSEDE's system features. A self-service web portal would automate the communication and configuration currently performed by XSEDE staff members and gateway developers, improving responsiveness for gateway developers, reducing their start-up time, and also enforcing or encouraging best practices and guidelines for how to use the features effectively and securely. *There are no current plans to address this gap.*

4. Verification of quality attributes (suggested priority: low)

Verifying quality attributes requires significant one-time and ongoing testing. XSEDE has decided that the costs of this testing would not bring sufficient benefit. Instead XSEDE will monitor user satisfaction, usage, and available performance metrics and address quality issues when raised by users. *There are no current plans to address this verification gap.*

System Components That Support This Use Case

The following XSEDE operational components currently support this use case.

Component	Supported Functionality
XSEDE User Portal (XUP)	The front-end user interface to the XSEDE system where end users register with XSEDE, manage their user profile information, and request allocations to use XSEDE SP resources.
Globus Auth	Provides the authentication service used by end users to login to XUP or a science gateway and to obtain an XSEDE OAuth2 token that can be used with other XSEDE services, including Globus Transfer. Also provides the ability for end users to link their XSEDE identities with non-XSEDE identities (e.g., InCommon campus identities, DOE and other agency identities, etc.)
XSEDE Kerberos	The repository that stores XSEDE usernames and passwords and authenticates XSEDE identities for Globus Auth
Globus Transfer	The hosted web application accessed by users via a web browser to select the file transfer source and destination and initiate, monitor, or cancel transfers, as well as managing the quality attributes regarding reliability, performance, scalability, and history accessibility. It also supports synchronizing a source file or directory to a destination, Also provides an SSH-accessible command-line interface and a RESTful

	web API for application integration.
	Sharing is a feature of Globus Transfer that allows one end user to share a portion of his/her storage with other end users. All users involved in the share must be known to Globus Auth, but only the user who creates the share needs to have an allocation on the storage system. The sharing feature must be explicitly enabled by the storage system provider.
GridFTP, Globus Connect	Used by XSEDE SPs, campus IT providers, and end users to create endpoints from which and to which Globus Transfer can transfer files.
MyProxy, MyProxy OAuth2	Used by Globus Transfer to acquire tokens for authenticating to source/destination endpoints.
Globus groups	*Groups* is a feature of Globus that allows end users to define and manage user groups, which can then be used to define access control rules for other Globus services, such as Transfer. A common practice i to define a group and then give that group permission to access share storage accessible via Globus Transfer.
	The groups feature has not been fully integrated with XSEDE at this time. This is implicit in the gaps identified in this CDP. The plans for filling these gaps include documentation specifically designed to enabl science gateway developers to use the groups feature.