



ITMO UNIVERSITY

Erasmus Mundus Master Program PERCCOM.

ITMO University

Cluster, Grids and Cloud Computing



**UNIVERSITÉ
DE LORRAINE**



Open your mind. LUT.

Lappeenranta **University of Technology**



**Erasmus
Mundus**



**LULEÅ
UNIVERSITY
OF TECHNOLOGY**

REPORT

Globus Data Service



Prepared by: Mukhammadjon Jalolov

Index:

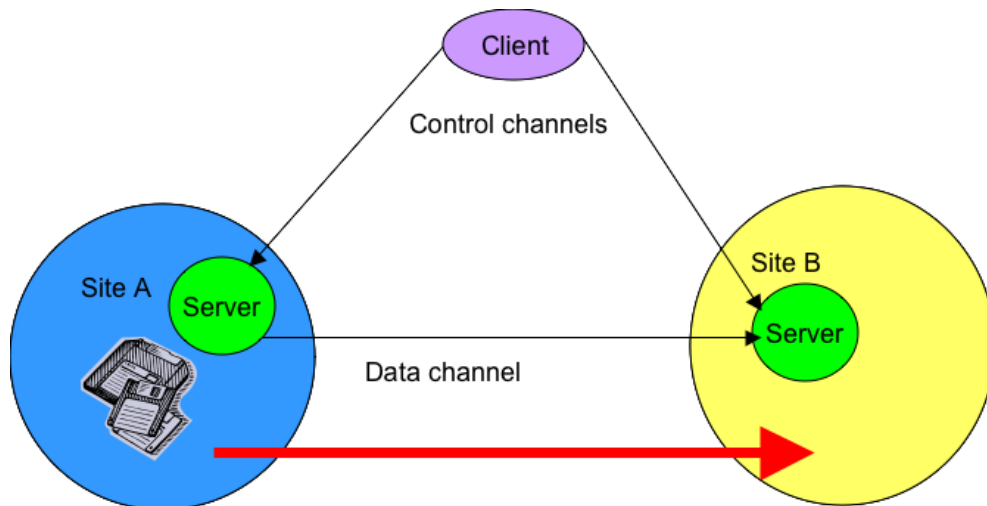
1. Background
2. Globus differences from other data services.
3. Globus key terms
4. Pricing
5. Globus file transfer
6. Data sharing possibility with Globus
7. Deploying on my machine
8. Conclusion
9. References:

Background

Globus is a connected set of services for research data management. These services are designed with a goal of providing researchers with tools to make their work easier. Globus transfer service helps move data where needed, and the added sharing capability (available through Globus subscriptions) gives the ability to share data with others, without moving it off storage. The technology underlying this service is GridFTP, developed by Ian Foster, Carl Kesselman, and Steve Tuecke. (Ian Foster and Steve Tuecke co-lead the Globus project, based at the Computation Institute at the University of Chicago and Argonne National Laboratory.)



After analyzing Globus, I figured out that it uses GridFTP protocol which is an extension of the standard File Transfer Protocol (FTP) for high-speed, reliable, and secure data transfer. Due to the fact that GridFTP provides a more reliable and high performance file transfer (compared to protocols such as SCP or rsync), Globus enables the transmission of very large files. Moreover, GridFTP also addresses the problem of incompatibility between storage and access systems.



Globus file sharing allows individual researchers to simply and securely share their data with others, freeing HPC system administrators from routine requests for access. Subscriptions offer a range of services that are vital to running a successful data service in a university, laboratory, or computing facility. Subscriptions can include managed endpoints, file sharing, data publication and discovery, enhanced support, branded sites, a management console, and other advanced features. Subscriptions also allow Globus Plus to be enabled for individual users, giving them additional transfer and sharing capabilities from their personal computers.

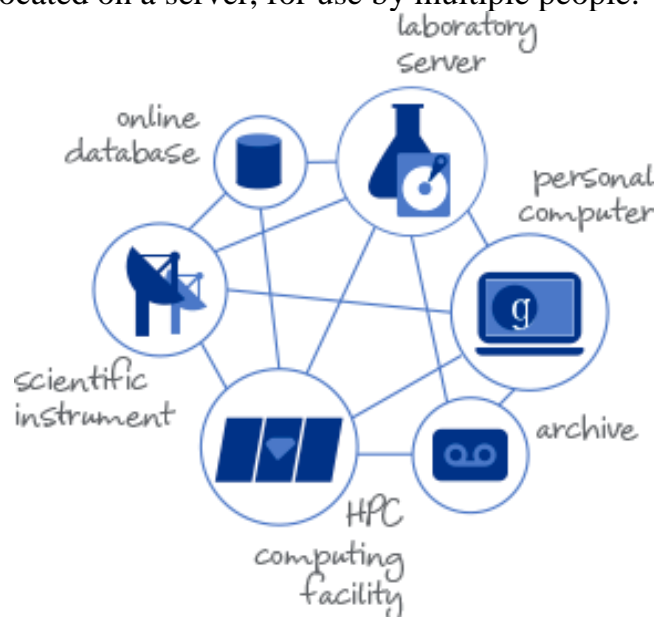
Globus differences from other data services.

Globus offers easy, fast, secure, and reliable transfer between two end nodes. Nowadays, the largest laboratories and universities around the world exploit Globus to get big data in a quick way where needed, and this allows those researchers to focus on productive tasks rather than involving in IT issues. All storage tasks are done by the Globus. Globus deploys cloud. However, Globus is not a cloud storage service, which means that subscribers don't have to pay for storing their data. With the use of Globus data transferring and sharing to and from the storage options, researchers can reach effectiveness. A campus computing cluster, a lab server, a supercomputer, or laptop can be easily turned into endpoints. The Globus services are purpose-built to meet the needs and requirements of the research community.



Globus key terms

In Globus service the usage of the term “endpoint” is described as the different locations where data is moved to or from using the Globus transfer, sync, and sharing service. There are two types of endpoints which are either personal on a user's personal computer or multiuser located on a server, for use by multiple people.



In order to move data from a cluster on a personal computer, setting up Globus Connect Server on a cluster and installation of Globus Connect Personal is required, which creates two endpoints in each location. In technical terms, the Endpoint is a logical

address for a GridFTP server, similar to a domain name for a web server. Data is transferred between Globus endpoints. The software installed on each endpoints have two different types:

Globus Connect Personal – is targeted for individual users – a client for communicating with other GridFTP servers, via the local computer using Globus. Installing Globus Connect Personal on a computer creates an endpoint that can be used to transfer data.

Globus Connect Server is purposed for multiuser environments – a Linux package that sets up a GridFTP server for use with Globus. Once installed on a server, those with access to the server can move data to and from this location.

Pricing

Although Globus file transfer is free to use for non-profit research and educational purposes, an account need to be in a necessary level to use with full potential. Subscriptions offer researchers access to enhanced features such as file sharing and data publication, and provide system administrators with a comprehensive view into how storage systems are being utilized. Subscription helps make Globus self-sustaining, offering high-quality data management infrastructure to all researchers at the same time. The table represents the types of user subscriptions for Globus service.

Features <small>(click ⓘ for description)</small>	Basic - Free	Starter Subscription	Standard Subscription
Transfer level ⓘ	Unlimited	Unlimited	Unlimited
User level ⓘ	Unlimited	Unlimited	Unlimited
Managed endpoints ⓘ	None	1	Unlimited
Management console ⓘ	–	✓	✓
Usage reports ⓘ	–	✓	✓
Support for Globus Connect, Web, CLI ⓘ	–	✓	✓
Shared endpoints ⓘ	–	✓	✓
Globus Plus users ⓘ	–	–	✓
Data publication ⓘ	–	–	✓
Application integration support ⓘ	–	–	✓
HTTPS support <i>(coming soon)</i> ⓘ	–	–	✓
Support service level ⓘ	–	Monday-Friday, 9am-5pm Central; 1-business day response	
Named support contacts ⓘ	–	1	5
Pricing	Free	Contact us for subscription pricing	

Globus file transfer

Globus provides a secure, unified interface to research data. Use Globus to “fire and forget” high-performance data transfers between systems within and across organizations.

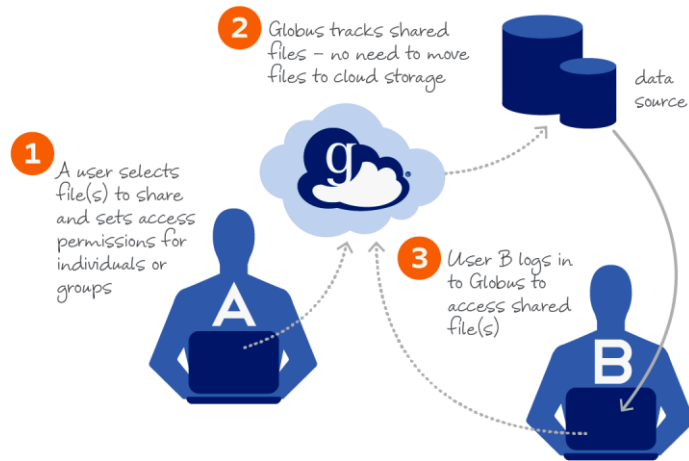


Research often requires sophisticated data management capabilities across systems and institutions. Globus is intended to provide these capabilities. Laptops, supercomputers, tape archives, cloud storage, HPC clusters, and scientific instruments are some of the systems that can be connected to Globus.

Most importantly, Globus allows use a web browser or command line interface to submit transfer and synchronization requests, optionally choosing encryption. With this 'fire and forget' model one can concentrate on research while Globus handles the mundane (but important) details of successful large-scale data transfers. Data is transferred directly between the source and destination systems and during this time Globus tunes performance parameters, maintains security, monitors progress, and validates correctness. A transfer status can be checked at any time via the Globus activity page and will receive email when the transfer completes. If a network or system involved in the transfer goes down, Globus automatically resumes the transfer when the component comes back online. If an issue requires action from a user, such as an expired credential or exceeded disk quota, Globus resumes the transfer after the problem is fixed. If a transfer has not made progress after a period of time (usually 3 days), the transfer will expire and a user is notified. Moreover, in order to transfer valuable information, Globus offers sharing, publication, platform, and management capabilities.

Data sharing possibility with Globus

Another mostly used service within the Globus is easy and secure data sharing with the people who need it — without requiring to create temporary accounts or transfer data to an external storage system.



Almost all file sharing services require copying valuable data to an external storage system that they manage—usually hosted in the cloud. This sharing model is reasonable for a few megabytes of data, but can be cumbersome, slow, and expensive for big research data. On the other hand, in Globus the data is not moved to the storage for sharing it. Any storage system that has a Globus connection can be easily configured to allow secure data sharing directly by users of the system. Once configured, directory paths can be select to be securely shared with offsite collaborators and grant them read-only or read-write access. All collaborators receive an email with a link to the shared directory paths, and can then use Globus to transfer data from/to storage system directly, without taking a detour through a third-party cloud storage provider.

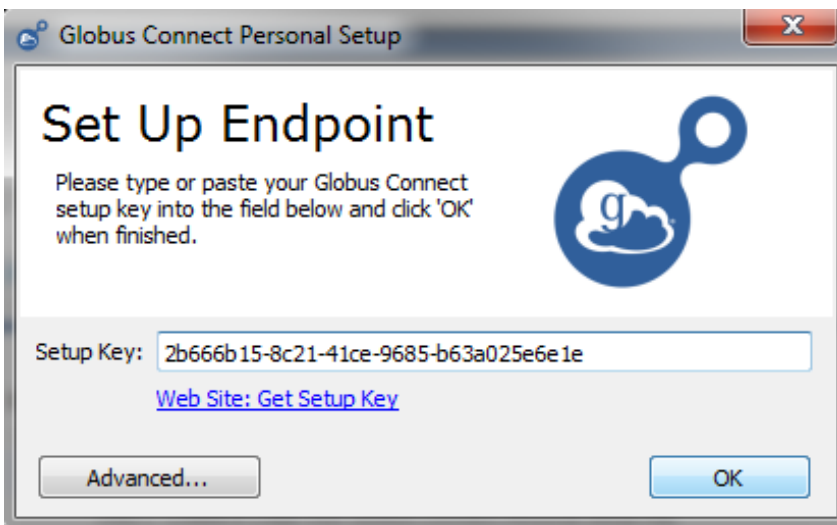
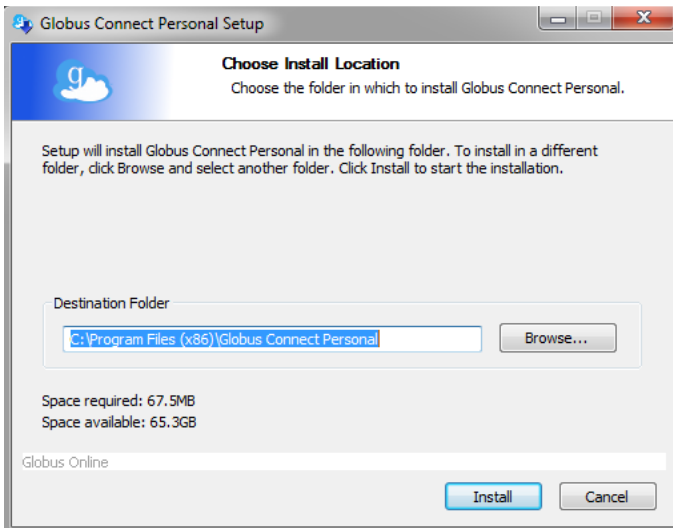
Deploying on my machine

In order to start using the Globus service, the installation of its software is required. Hence, I first downloaded the app signing up into the service.

Manage Endpoints

The screenshot shows the 'Manage Endpoints' interface in the Globus web application. At the top, there is a navigation bar with 'Endpoint List' and two buttons: 'add Globus Connect Personal endpoint' and 'add Globus Connect Server endpoint'. Below this is a modal window titled 'Add Globus Connect Personal Endpoint'. The modal has a close button (X) in the top right corner. The main content area is divided into two steps. Step 1, 'Create & Copy Your Globus Connect Personal Setup Key', is completed, indicated by a large green checkmark and the text 'My Laptop created'. Below this, the 'Setup Key' is displayed as '2b666b15-8c21-41ce-9685-b63a025e6e1e' with a 'copy' button next to it. Step 2, 'Download & Install Globus Connect Personal', is currently active. It includes the instruction 'Click one of the buttons below to download and install Globus Connect Personal for your operating system.' and three buttons: 'for Mac OS X', 'for Linux', and 'for Windows'. At the bottom of the modal, there is a note: 'Once downloaded, run the installer. When prompted, paste in the Setup Key to complete the installation.'

Following the shown steps, I installed the Globus Connect software. After successful configuration, I was able to create my endpoint and give an access to my selected directory which can be shared and performs as a location for my transfers. All the screenshots of the main important stages is shown in the following page.



Conclusion

In conclusion, I would like to highlight that using Globus data service enhances collaboration with other researchers, without requiring them to have accounts on the system. After installation of the Globus Connect software, I was able to experiment with transferring my files on the endpoint. Moreover, the fact that my data is not duplicated on the cloud storage makes it more secure which I appreciate and essential requirement these days. Finally, the service itself monitors all transfers, informing users with notifications.

References:

<https://www.globus.org/subscriptions>

<https://www.globus.org/globus-connect-server>

<https://docs.globus.org/how-to/>

<https://fasterdata.es.net/data-transfer-tools/globus/>

<https://www.globus.org/globus-connect-personal>

<https://docs.globus.org/how-to/get-started/>

<https://docs.globus.org/how-to/share-files/>

<https://www.globus.org/subscriptions>