



Globus

Niklas Kolbe (niklas.kolbe@gmail.com)

ITMO, St. Petersburg, 2. June 2015

Clouds, Grid and Cluster systems, Andrey Shevel

Content

- ▶ Introduction: Globus Project
- ▶ Globus Online for Handling of Research Data
- ▶ Globus Toolkit for Grid Infrastructure
 - ▶ GridFTP
- ▶ Live Demonstration

Globus Project

- ▶ Started in 1997
- ▶ Computation Institute
 - ▶ The University of Chicago and Argonne National Laboratory
- ▶ To enable Grid Computing
- ▶ Make Big Data portable
 - ▶ mainly to support data-intensive research

Globus Online




- ▶ To share, transfer and publish Research Data
- ▶ File Transfer from Endpoint to Endpoint via Web Interface
 - ▶ Endpoint: campus storage systems, HPC clusters, laptops, Amazon S3 buckets, scientific instruments, etc.
- ▶ Transfer options, e.g. encryption
- ▶ Globus Connect Personal and Server

Transfer Files

Get Globus Connect Personal
Turn your computer into an endpoint.

Endpoint ...




Path

select all | none  up one folder  refresh list 

 Globus Folder


Endpoint ...

Path

select all | none  up one folder  refresh list 

 ITMO File.txt	15 b
 file1.txt	4 b
 file2.txt	4 b

▶ more options

Label This Transfer

This will be displayed in your transfer activity.

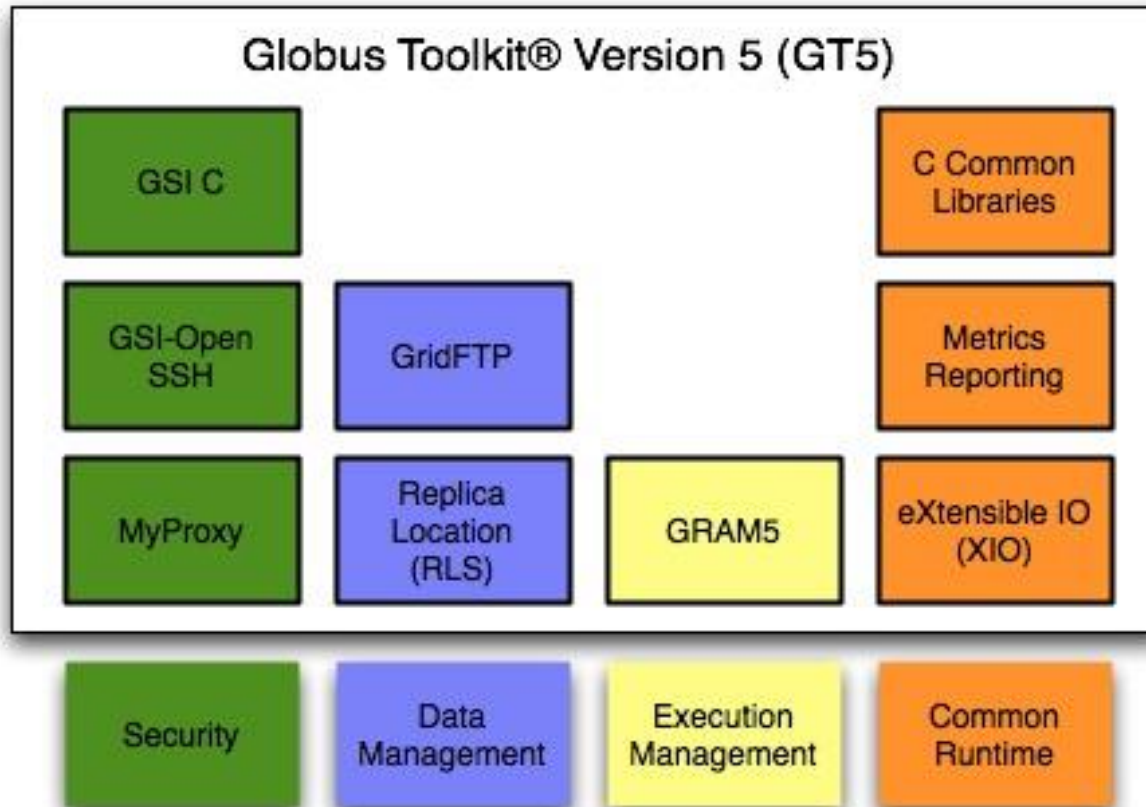
Globus Toolkit

- ▶ Open Source toolkit for Grid Computing
- ▶ Sponsored by
 - ▶ DOE, NSF, DARPA, NASA,
 - ▶ IBM, Microsoft, ...
- ▶ Implementations
 - ▶ European Data Grid (used by LHC)
 - ▶ Grid Physics Network
 - ▶ Network for Earthquake Engineering and Simulation
 - ▶ FusionGrid
 - ▶ ...

Globus Toolkit

- ▶ The toolkit includes software for
 - ▶ Security
 - ▶ Information infrastructure
 - ▶ Resource management
 - ▶ Data management
 - ▶ Communication
 - ▶ Fault detection
 - ▶ Portability
- ▶ Modules can be used independently or together
- ▶ Core components allow to access remote resources

Globus Toolkit



GridFTP

- ▶ Extension of FTP: High-speed, reliable and secure data transfer
- ▶ Enables transmission of very large files
- ▶ Addresses problems of incompatibility between storage and access systems
- ▶ Globus Toolkit provides the most commonly used implementation

GridFTP

► Features

Performance

- supports parallel TCP streams
- supports multi-node transfers

Checkpointing

- server sends restart markers (checkpoint) to the client

Third-party transfers

- separated control and data channels
- transfer of data between two end hosts, mediated by a third host

Security

- Control channel is encrypted by default
- data channel is authenticated by default
- optional integrity protection and encryption.

Globus Online

Live Demonstration

Thank you for your attention!

Any Questions?