

GLOBUSONLINE OVERVIEW

Tanya
Levshina

WHAT IS GLOBUSONLINE (GO)?

- GO is a Software-as-a-Service facility to provide file transfer functionality.
- It will do a third-party transfer on your behalf:
 - Performs transfers of files
 - Retries in case of failures
 - Optimizes gridftp parameters for transfer
 - Concurrency
 - Threading
 - Pipelining
 - Provides CLI, Rest API, WEB UI
 - Handles certificate via MyProxy servers or uses your proxy certificate
 - Provides user support

WHY DO WE(OSG) CARE?

- Big VOs, with dedicated storage and staff, have their own transfer services (FTS, PhedEx, SAM-Grid, ...)
- A VO that relies on OSG public storage encounters common problems:
 1. What SE is available for my VO?
 2. How much space can I use?
 3. What is the most efficient way to transfer significant amount of data (GBs - TBs) from on site to another?
 4. What is the best way to handle errors and retries?
 5. How to monitor the transfer progress?
- GO may help us with #3-5

GO GLOSSARY

■ EndPoints

- Logical Name
- Transfer servers
- Activation Status
- MyProxy Server

■ Transfer Task

- Task ID
- Status
- Start Time
- End Time
- Event Log
- Sync level
- Subtasks List
 - Subtask ID
 - Status
 - Start Time
 - End Time
 - Event Log

HOW CAN I USE IT?

- Sunny day scenario:

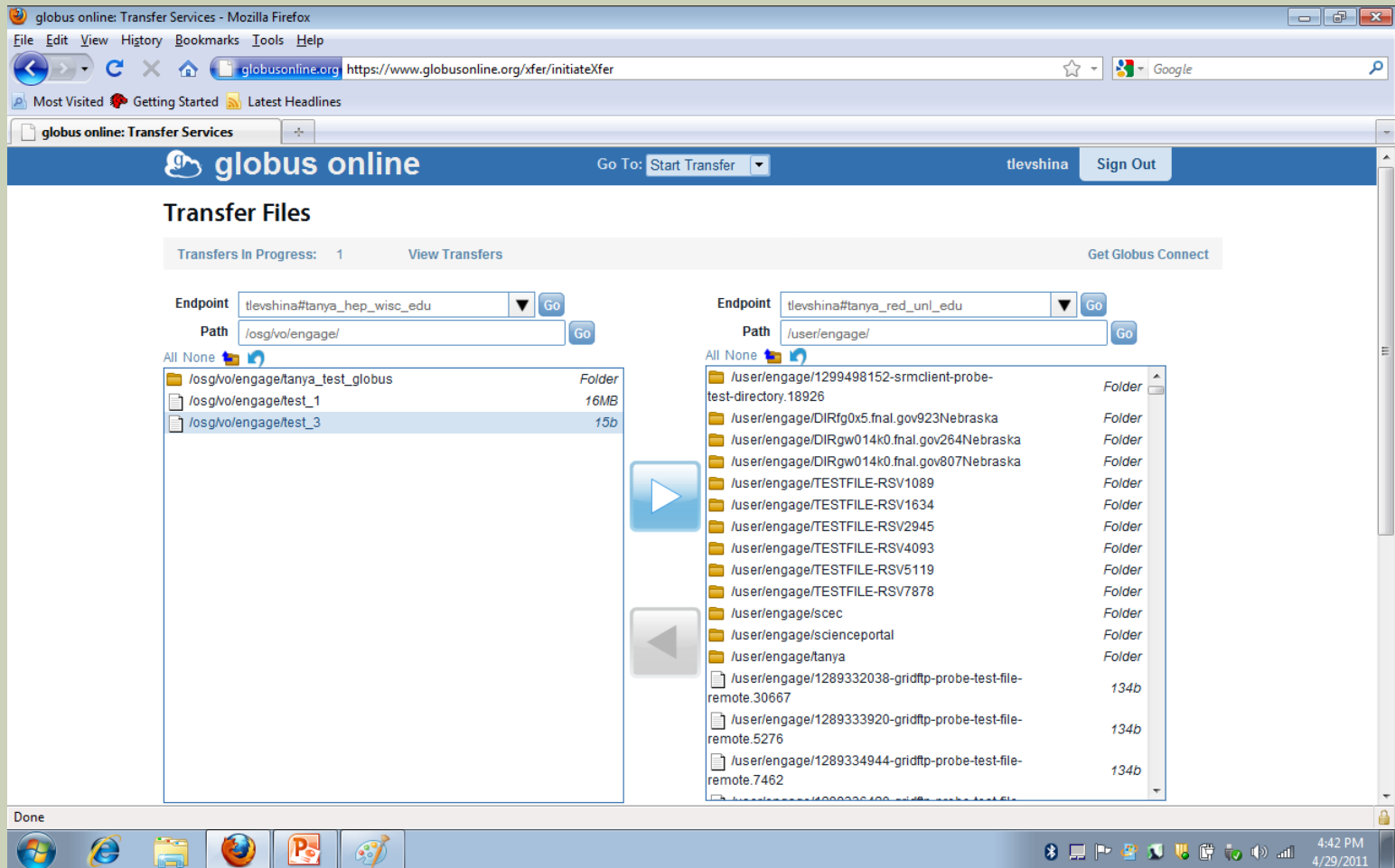
- Storage Admin:

- Creates public endpoints that work for your VO


- User

- Prepares the list of files you want to transfer from A to B
 - Gets proxy certificate and activate endpoint with your proxy or use myProxy server
 - Initializes the transfer using UI, CLI or Rest API
 - Monitors the progress of transfer
 - Gets notification when job succeeded

TRANSFER WEB UI



WEB MONITORING (I)

 globus online

Go To: View Transfers

tlevshina Sign Out

Transfer Activity

Cancel 1 of 8 View 25 Records

	Status	ID	Task Progress	Username	Completion Time	Request Time
		e72a0...	<div><div></div>574 / 500</div>	<div><div>Transfer Details</div><div>Task ID: e72a05ec-729d-11e0-8f02-123139152864</div><div>Status ACTIVE</div><div>Origin tlevshina#tanya_red_unl_edu</div><div>Destination tlevshina#gw_hdfs</div><div>User tlevshina</div><div>Directories 0</div><div>Request Time 04/29/2011 08:19 PM</div><div>Deadline 04/30/2011 08:18 PM</div><div>Completion Time</div><div>Bytes Transferred 30651973632</div><div>Files 5000</div><div>Task Statistics</div><div>Pending 3174</div><div>Cancelled 0</div><div>Failed 0</div><div>Succeeded 1827</div><div>Expired 0</div><div>Retrying 0</div><div>View Event Log</div></div>		
		ebe54...	<div><div></div>1 / 1</div>			PM
		e5f18...	<div><div></div>5000 / 1 / 50</div>			PM
		90bd2...	<div><div></div>1 / 1</div>			AM
		80629...	<div><div></div>5001 / 500</div>			AM
		9fd20...	<div><div></div>1 / 1</div>			PM
		6e857...	<div><div></div>5001 / 500</div>			PM
		45bf5...	<div><div></div>1 / 1</div>			PM
		95f16...	<div><div></div>5001 / 500</div>			PM
		fa1eb...	<div><div></div>1 / 5000 / 50</div>			PM
		22ead...	<div><div></div>4999 / 2 / 50</div>			PM
		5c63e...	<div><div></div>1 / 1</div>			AM
		adffe...	<div><div></div>5001 / 500</div>			AM
		9c319...	<div><div></div>5000 / 1 / 50</div>			PM
		02291...	<div><div></div>1 / 1</div>			PM
		a283c...	<div><div></div>5001 / 5001</div>	tlevshina	04/25/2011 11:23 PM	04/25/2011 10:45 PM
		f7e4c...	<div><div></div>1 / 1</div>	tlevshina	04/25/2011 08:20 PM	04/25/2011 08:17 PM
		53f35...	<div><div></div>0 / 1 / 1</div>	tlevshina	04/25/2011 08:16 PM	04/25/2011 08:05 PM
		3d444...	<div><div></div>1 / 1</div>	tlevshina	04/25/2011 05:53 PM	04/25/2011 05:48 PM
		749db...	<div><div></div>2 / 2</div>	tlevshina	04/25/2011 05:48 PM	04/25/2011 05:43 PM
		946dc...	<div><div></div>1 / 1</div>	tlevshina	04/25/2011 05:00 PM	04/25/2011 05:00 PM
		6d626...	<div><div></div>0 / 1 / 1</div>	tlevshina	04/25/2011 05:00 PM	04/25/2011 04:59 PM

WEB MONITORING (II)

globus online

Go To: View Transfers

tlevshina

Sign Out

Event Log (e72a05ec-729d-11e0-8f02-123139152864)

<< View Transfer Activity

<<< 1 of 36 >>>

View 100 Records

Date	Time	Code	Description	Details
04/29/2011	08:19:08 PM	STARTED	The operation was started or restarted	Starting sync scan.
04/29/2011	08:21:11 PM	TIMEOUT	A network operation timed out	Timeout occurred while running globus-url-copy
04/29/2011	08:21:14 PM	STARTED	The operation was started or restarted	Starting sync scan.
04/29/2011	08:21:16 PM	STARTED	The operation was started or restarted	Starting at offset 0
04/29/2011	08:21:18 PM	STARTED	The operation was started or restarted	Starting at offset 0
04/29/2011	08:21:19 PM	SUCCEEDED	The operation succeeded	bytes=16777216 mbps=41.943
04/29/2011	08:21:19 PM	STARTED	The operation was started or restarted	Starting at offset 0
04/29/2011	08:21:22 PM	SUCCEEDED	The operation succeeded	bytes=16777216 mbps=37.283
04/29/2011	08:21:22 PM	STARTED	The operation was started or restarted	Starting at offset 0
04/29/2011	08:21:22 PM	SUCCEEDED	The operation succeeded	bytes=16777216 mbps=53.687
04/29/2011	08:21:22 PM	STARTED	The operation was started or restarted	Starting at offset 0
04/29/2011	08:21:25 PM	SUCCEEDED	The operation succeeded	bytes=16777216 mbps=41.943
04/29/2011	08:21:25 PM	STARTED	The operation was started or restarted	Starting at offset 0
04/29/2011	08:21:25 PM	SUCCEEDED	The operation succeeded	bytes=16777216 mbps=37.283
04/29/2011	08:21:25 PM	STARTED	The operation was started or restarted	Starting at offset 0
04/29/2011	08:21:28 PM	SUCCEEDED	The operation succeeded	bytes=16777216 mbps=49.710
04/29/2011	08:21:28 PM	STARTED	The operation was started or restarted	Starting at offset 0
04/29/2011	08:21:28 PM	SUCCEEDED	The operation succeeded	bytes=16777216 mbps=40.672
04/29/2011	08:21:28 PM	STARTED	The operation was started or restarted	Starting at offset 0
04/29/2011	08:21:31 PM	SUCCEEDED	The operation succeeded	bytes=16777216 mbps=40.672
04/29/2011	08:21:31 PM	STARTED	The operation was started or restarted	Starting at offset 0
04/29/2011	08:21:32 PM	SUCCEEDED	The operation succeeded	bytes=16777216 mbps=33.554
04/29/2011	08:21:32 PM	STARTED	The operation was started or restarted	Starting at offset 0
04/29/2011	08:21:35 PM	SUCCEEDED	The operation succeeded	bytes=16777216 mbps=37.283

GO CLI

- Requirements (available in OSG Client):

- grid-proxy-init
- gssish

- Create/modify/activate/list endpoint

```
$ gsissh -t tlevshina@cli.globusonline.org endpoint-list
go#ep1                20:40:53
gw_hdfs                162:35:31
hep_wisc_edu_1         89:53:38
tanya_hep_wisc_edu     89:53:38
tanya_red_unl_edu      190:11:12
```

```
$ gsissh -t tlevshina@cli.globusonline.org endpoint-list -v tanya_red_unl_edu
Name                : tanya_red_unl_edu
Host(s)             : gsiftp://red-gridftp1.unl.edu:2811
Subject(s)          :
MyProxy Server      : n/a
Credential Status   : ACTIVE
Credential Expires: 2011-05-07 20:17:26Z Credential Subject: /DC=org/DC=doegrids/OU=People/CN=Tanya Levshina
508821/CN=614089312/CN=383514127
```

- List directories and files

```
gsissh -t tlevshina@cli.globusonline.org ls tanya_red_unl_edu:/user/engage/tanya_test_ftp_1
```

- File copy (scp)

```
$ gsissh -t tlevshina@cli.globusonline.org scp tanya_red_unl_edu:/user/engage/tanya_test_ftp_1 tanya_hep_wisc_edu:/osg/vo/engage/test_1
Task ID: c8c776ba-72ad-11e0-aa60-1231380e306c
[XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX] 1/1 0.00 mbps
```

- Get status and transfer details

```
$ gsissh -t tlevshina@cli.globusonline.org status c8c776ba-72ad-11e0-aa60-1231380e306c
Task ID      : c8c776ba-72ad-11e0-aa60-1231380e306c
Request Time: 2011-04-29 22:12:34Z
Command     : scp tanya_red_unl_edu:/user/engage/tanya_test_ftp_1 tanya_hep_wisc_edu:/osg/vo/engage/test_1
Status      : SUCCEEDED
```

- Access to event logs

- File Transfer

- Read EOF or CNTRL-D terminated list (can specify sync level)
- Can set expiration time

REST API

- Python and Java API is available, not too difficult use Rest API with PERL
- Good documentation
- Easy to integrate with existing code
 - Integrated and tested with DES workflow during two days of integration-fest at Fermilab
- Access to more detailed information about transfer status and event
- Adequate support
- Requirements:
 - Python 2.6 or higher

TESTS RESULTS

■ DES

- 31,000 files
- 184 GB
- 27% improvement in transfer time
- DES workflow requires simple transfer verification (file length comparison after transfer). So, total transfer and verification time was similar to the time of original workflow.

■ LIGO emulation

- 5,000 files
- 16MB each (78GB)
- GridFTP servers on top of HDFS
- Average transfer time between Nebraska and Wisconsin is slightly better than transfer time shown by LIGO tests with multiple srm-copy commands with concurrency 5

ISSUES

- Number of retries per file is not supported. Only expiration time can be specified.
- SRM is not supported, have to explicitly list all `grdftp` servers you want to use. GO `gidftp`-servers load balancing could interfere with SRM load balancing.
- Do not distinguish between “FATAL” and intermittent errors during file transfer: continue retries in case of FATAL error.
 - Several OSG SEs uses HDFS that doesn't allow “seek” operation, so any attempt to retransmit data in case of failure with offset >0 fails.
- GO does not provide interfaces to do verification (get stats) of a list of files.
- Some features are missing from API (for example: activation with certificate).
- Doesn't provide space accounting.

HOW CAN IT BE USED IN THE OSG ENVIROMENT?

- In order to use OSG public storage we need to:
 - Discover what OSG SEs support a specific VO (discovery tools can do it).
 - Use pigeon tools to verify that SE is working for this VO.
 - Use GO CLI or Rest API to dynamically create endpoints.
 - Prepare list of files or directories for transfer.
 - Use GO CLI or Rest API to initiate transfer.
 - Use GO CLI, Rest API or WEB UI to monitor and check status.
- Still missing:
 - Accurate account of space utilization and availability.

IF YOU WANT TO TRY IT NOW...

- Register with GO
- Generate proxy certificate
voms-proxy-init
- Create your own endpoints for the sites that are working for you (e.g. Nebraska and Wisconsin)
endpoint-create
- Configure them with gridftp servers.
endpoint-add, endpoint-activate
 - **Tip:** Use discovery tools to find gridftp servers for the site:
\$ *get_gridftp_storage_element_id --vo Engage --show_site_name | grep -i Nebraska*
Nebraska red-srm1.unl.edu
\$ *get_gridftp_url --vo Engage --storage_element_id red-srm1.unl.edu*
gsiftp://red-gridftp1.unl.edu:2811/mnt/hadoop/user/engage/TESTFILE
- Test:
ls
scp
- Need help?
 - support@globusonline.org
 - tlevshin@fnal.gov