Globus-url-copy client

Introduction

globus-url-copy a command line tool that can do multi-protocol data movement supporting GridFTP. It is mainly for Linux/Unix users. It is possible to use globus-url-copy in these cases:

- User Local PC <==> CINECA HPC Cluster
- User Local PC <==> iRODS repository
- CINECA HPC Cluster A <==> CINECA HPC Cluster B
- CINECA HPC Cluster <==> iRODS repository

The following steps help you to easily transfer your data from/to CINECA cluster using globus-url-copy.

Client installation on your local workstation

Since 2018, it is possible to install the client using those provided by the Grid Community Forum (GridCF), a global community that provides support for core grid software (https://gridcf.org/).

The GridCF is attempting to support a software stack christened the Grid Community Toolkit (GCT). The GCT is an open-source fork of the venerable Globus Toolkit created by the Globus Alliance. The GCT is *derived* from the Globus Toolkit, but is not the Globus Toolkit. Further, the GridCF is not a part of the Globus Alliance.

Please refer to the Grid Community Toolkit official documentation https://gridcf.org/gct-docs/ for installation notes.

Get X.509 personal certificate

To use globus-url-copy tool, **you must have a valid x509 personal certificate.**
Please refer to the X.509 certificate page in case of troubles to obtain the certificate from a recognized CA.

Create your proxy credential

To use globus-url-copy, **you must have a valid proxy certificate on the machine on wich you have your source data.**
To reach this goal, refer to

- **"CASE 1)"** if you have requested a certificate to your certification authority
- or **"CASE 2)"** (down in this page) if you have requested a certificate to CA-CINECA .

CASE 1) if you have requested a certificate to your certification authority

Convert your certificate in a "pem" certificate. If it is a ".p12" or a ".pfx", please convert it by typing

```
bash$ openssl pkcs12 -clcerts -nokeys -in <name_certificate.{p12|pfx}> -out usercert.pem
Enter Import Password: <password used for backup of your .p12 certificate>
MAC verified OK

bash$ openssl pkcs12 -nocerts -in <name_certificate.{p12|pfx}> -out userkey.pem
Enter Import Password: <password used for backup of your .p12 certificate>
MAC verified OK
Enter PEM pass phrase: <password to encrypt your private key>
Verifying - Enter PEM pass phrase <password to encrypt your private key>
```

Set the right permission to the file just created:

```
bash$ chmod 644 usercert.pem
bash$ chmod 400 userkey.pem
```

Extract your own user DN (Distinguished Name) from the certificate, for example typing

```
bash$ openssl x509 -in usercert.pem -noout -subject | sed 's/subject= //'
```

To use globus-url-copy on MARCONI, GALILEO100 and MARCONI100 clusters, the extracted DN has to be added to our userdb profile (https://userdb.hpc.cineca.it/user) under the "personal data" section in the "X.509 certificate" field and following the specified syntax. To use globus-url-copy with the iRODS repository, you have been add as PI or collaborator to a DRES of type REPO.

Create the directory ~/.globus and copy here the usercert.pem ed userkey.pem.

```
bash$ mkdir ~/.globus

bash$ cp <some location>/usercert.pem ~/.globus

bash$ cp <some location>/userkey.pem ~/.globus
```

Then, download and install the certificates of the Certification Authorities by the command:

```
bash$ mkdir ~/.globus/certificates && cd ~/.globus/certificates
bash$ wget http://dist.eugridpma.info/distribution/igtf/current/accredited/igtf-preinstalled-bundle-classic.tar.
gz && tar -zxvf igtf-preinstalled-bundle-classic.tar.gz
bash$ wget http://dist.eugridpma.info/distribution/igtf/current/accredited/igtf-preinstalled-bundle-mics.tar.gz
&& tar -zxvf igtf-preinstalled-bundle-mics.tar.gz
bash$ wget http://dist.eugridpma.info/distribution/igtf/current/accredited/igtf-preinstalled-bundle-slcs.tar.gz
&& tar -zxvf igtf-preinstalled-bundle-slcs.tar.gz
bash$ wget http://dist.eugridpma.info/distribution/igtf/current/accredited/igtf-preinstalled-bundle-iota.tar.gz
&& tar -zxvf igtf-preinstalled-bundle-iota.tar.gz
```

Then, follow one of these instructions to create your proxy credential:

case 1.1) if both your X.509 certificate and the source data are in the same machine, create your proxy certificate (starting from your X.509 certificate) by using the command:

```
grid-proxy-init
```

case 1.2) if the X.509 certificate and the source data are on two different machines, login on the machine where the X.509 certificate is located and create the proxy by executing the command:

```
grid-proxy-init
```

Then store the proxy on the **grid.hpc.cineca.it** myproxy-server by typing the command

```
myproxy-init -l <username> -s grid.hpc.cineca.it
```

Finally, login into the machine where the source data are located and retrieve the proxy certificate by the command

```
myproxy-logon -l <username> -s grid.hpc.cineca.it
```

When you finish to use your proxy credential, destroy it by typing:

```
myproxy-destroy -s grid.hpc.cineca.it -l <username>
```

```
grid-proxy-destroy
```

NB The proxy will destroy itself 12 hours running from its "init". So after this time you have to create again the proxy for a new transfer. If you want to increase the proxy lifetime, use parameter "-t <hours>" in the myproxy-init command.

CASE 2) if you have requested a certificate to CA-CINECA

Login into the machine where your source data are located.

case 2.1) If the login machine is a CINECA HPC cluster, retrieve your proxy credential by the command

```
myproxy-logon -s grid.hpc.cineca.it -l <username>
```

where <username> and <password> are the same that you have on HPC CINECA machines.

case 2.2) If the login machine is your local workstation, download and install the certificates of the Certification Authorities by the command:

```
bash$ mkdir ~/.globus/certificates && cd ~/.globus/certificates
bash$ wget http://dist.eugridpma.info/distribution/igtf/current/accredited/igtf-preinstalled-bundle-classic.tar.
gz && tar -zxvf igtf-preinstalled-bundle-classic.tar.gz
bash$ wget http://dist.eugridpma.info/distribution/igtf/current/accredited/igtf-preinstalled-bundle-mics.tar.gz
&& tar -zxvf igtf-preinstalled-bundle-mics.tar.gz
bash$ wget http://dist.eugridpma.info/distribution/igtf/current/accredited/igtf-preinstalled-bundle-slcs.tar.gz
&& tar -zxvf igtf-preinstalled-bundle-slcs.tar.gz
bash$ wget http://dist.eugridpma.info/distribution/igtf/current/accredited/igtf-preinstalled-bundle-iota.tar.gz
&& tar -zxvf igtf-preinstalled-bundle-iota.tar.gz
```

Download the file 5be94fc8.0 and 5be94fc8.signing_policy and digit the command

```
bash$ openssl x509 -in 5be94fc8.0 -hash -noout
```

to obtain the hash of the certificate. Now, rename the two file downloaded as <hash>.0 and <hash>. signing_policy and cp these two file in ~/.globus /certificates.

Finally, retrieve your proxy credential by the command

```
myproxy-logon -s grid.hpc.cineca.it -l <username>
```

where <username> and password are the same that you have on HPC CINECA cluster.

Both for case 2.1) and case 2.2), wen you finish to use your proxy credential, destroy it by typing:

```
myproxy-destroy -s grid.hpc.cineca.it -l <username>
```

NB. The proxy will destroy itself 7 days from its "init". So after this time you have to create again the proxy for a new transfer.

Use the standard client

Now that you have a valid proxy on the machine with the source data, you can start to transfer your data bby using the standard client globus-url-copy. Please note that the client is already available on CINECA HPC clusters.

To transfer file from CINECA HPC Custer to yout Local PC:

```
$ globus-url-copy gsiftp://[username@]<ENDPOINT-CINECA HPC Cluster>/<remote_path/to/yourfile>   file:///home
/user/<local_path/to/yourfile>
```

To transfer file from your local PC to CINECA HPC Custer

```
$ globus-url-copy /path/to/your/local/file gsiftp://[username@]<ENDPOINT-CINECA HPC Cluster>/remote/path/
```

where the ENDPOINT-CINECA HPC Clusters are:

```
iRODS repository --> gftp.repo.cineca.it:2811
MARCONI machine --> gftp.marconi.cineca.it:2811
MARCONI machine for PRACE users --> gftp-prace.marconi.cineca.it:2811
GALILEO-100 machine --> gftp.g100.cineca.it:2811
MARCONI-100 machine --> gftp.m100.cineca.it:2811
```

Examples

1- for synching recursively a directory and its subdirectories to MARCONI (like with rsync)

```
$ globus-url-copy -cd -r -sync /path/to/your/dir/ gsiftp://[username@]gftp.marconi.cineca.it:2811/~/remote/dir/
```

2- for moving a big chunk of data from MARCONI to PICO, the parallel option can be used

```
$ globus-url-copy -p 4 gsiftp://[username@]gftp.marconi.cineca.it:2811/~/path/to/file gsiftp://gftp.pico.cineca.
it:2811/~/path/
```

3- for listing the file in your directory on iRODS repository.

```
$ globus-url-copy -list gsiftp://gftp.repo.cineca.it:2811/CINECA01/home/your-remote-dir/
```

Gridftp transfer via batch script

To perform a gridftp transfer longer than 10 cpu minutes, it is suggested to submit a job batch on the serial queue. In what follows the list of needed steps

1. Create a proxy certificate in a location available from all the nodes of the cluster (e.g. your $HOME directory)

```
$ grid-proxy-init -out $HOME/proxy.cert
```

2. Below an example of job script:

```
#!/bin/bash
#SBATCH --err=slurm_%J.err

#SBATCH --out=slurm_%J.out
#SBATCH --time=04:00:00 #max time 4h
#SBATCH --nodes=1 --ntasks-per-node=1 --cpus-per-task=1
#SBATCH --partition=bdw_all_serial #on GALILEO the partition is gll_all_serial, on DAVIDE the partition
is dvd_all_serial

globus-url-copy -cred $HOME/proxy.cert <file in local position> gsiftp://<gridftp endpoint>/remote/path/
```

GridFTP TCP Port Range configuration

Please note that GridFTP servers on our clusters are configured to use the port range 20000 - 25000 for the incoming and outgoing connections.