Data transfer service for HBP

1st HPAC Platform Training 11-12 December 2018

Cristiano Padrin  Michele Carpené
Introduction
HPAC Platform

• A federated infrastructure for data and scalable compute services consisting of:
  – the central HBP supercomputer;
  – satellite HPC and data facilities;
  – cloud storage;
  – high-fidelity visualization capabilities.
HPAC Platform

Co-funded by the European Union
Description
Data transfer service

• The data transfer service has been deployed and configured on the HPAC sites:
  – to allow users to move datasets between all sites;
  – based on UNICORE File Transfer Protocol (UFTP).

• UFTP is a high-performance data transfer service based on FTP:
  – fully integrated with the UNICORE authentication mechanisms;
  – accessible via UNICORE Rest API and different UNICORE clients.
How to
Use UFTP

UFTP is available via two different interfaces:
• the UNICORE command line client (ucc):
  – a full-featured client for the UNICORE middleware able to start both direct data transfer (end-user machine to server) and third party transfer (server to server);
• Jupyter notebooks available in the Collaboratory web portal:
  – via UNICORE Rest API;
  – enables users to integrate data transfer steps inside more complex workflows.
Which network

• All HPAC sites are integrated in the PRACE Network.

• If both the UFTP endpoints have access to the PRACE network, UFTP is configured to use it for the data transfer.

• Otherwise the public network is used.
Data transfer service

Server to server data transfer on PRACE network

End-user client to server data transfer on Public network

User request for server to server data transfer
Examples
In [2]:
```python
!pip install --upgrade "hbpp-service-client==1.1.0"

hdr = get_bbpp_client().task.oauth_client.get_auth_header()
print 'Authorization' + hdr
```

Requirement already up-to-date: hbpp-service-client==1.1.0 in /opt/conda/envs/python2/lib/python2.7/site-packages
Requirement already up-to-date: validators in /opt/conda/envs/python2/lib/python2.7/site-packages (from hbpp-service-client==1.1.0)
Requirement already up-to-date: requests<3.0,>=2.18 in /opt/conda/envs/python2/lib/python2.7/site-packages (from hbpp-service-client==1.1.0)
Requirement already up-to-date: six>=1.4.0 in /opt/conda/envs/python2/lib/python2.7/site-packages (from validators->hbpp-service-client==1.1.0)
Requirement already up-to-date: decorator>=3.4.0 in /opt/conda/envs/python2/lib/python2.7/site-packages (from validators->hbpp-service-client==1.1.0)
Requirement already up-to-date: urllib3<1.25,>=1.21.1 in /opt/conda/envs/python2/lib/python2.7/site-packages (from requests<3.0,>=2.18->hbpp-service-client==1.1.0)
Requirement already up-to-date: charset<3.1.0,>=3.0.2 in /opt/conda/envs/python2/lib/python2.7/site-packages (from requests<3.0,>=2.18->hbpp-service-client==1.1.0)
Requirement already up-to-date: idna<2.5,>=2.4 in /opt/conda/envs/python2/lib/python2.7/site-packages (from requests<3.0,>=2.18->hbpp-service-client==1.1.0)
Requirement already up-to-date: certifi>=2017.4.17 in /opt/conda/envs/python2/lib/python2.7/site-packages (from requests<3.0,>=2.18->hbpp-service-client==1.1.0)
You are using pip version 9.0.1, however version 18.1 is available.
You should consider upgrading via the 'pip install --upgrade pip' command.
```

In [ ]:
UFTP client configuration

```bash
vim /home/username/.ucc/preferences

authenticationMethod=UNITY
unity.address=https://unity-jsc.fz-juelich.de/hbp-unicore-soapidp-oidc/saml2unicoreidp- soap/AuthenticationService
oauth2.bearerToken=eyJhbGciOiJSUzI1….ldVQpWYKQ

truststore.type=directory
truststore.directoryLocations.1=/certificate_path/*.pem
truststore.directoryEncoding=PEM

registry=https://hbp-unic.fz-juelich.de:7112/HBP/services/Registry?res=default_registry
output=.
uftp.streams=4
uftp.client.host=login02.marconi.cineca.it
uftp.encryption=false
uftp.compression=false
uftp.server.host=login02.marconi.cineca.it
protocols=UFTP BFT
```
UFTP client configuration


[ucc ls] UCC 1.7.12, http://www.unicore.eu
[ucc ls] Reading properties file </home/mcarpene/.ucc/preferences>
[ucc ls] Current directory is </home/mcarpene/Scrivania/UNICORE/ucc/unicore-ucc-7.11.0/bin>
[ucc ls] Output goes to <.
[ucc ls] Checking registry connection.
[ucc ls] Registry connection status: OK
[ucc ls] Current directory is </home/mcarpene/Scrivania/UNICORE/ucc/unicore-ucc-7.11.0/bin>
[ucc ls] Output goes to <.
[ucc ls] Detailed listing.
-r-- 1073741824 2018-01-29 10:00 /HBP_UFTP/Cineca_1GB
-r-- 497270784 2018-10-23 14:32 /HBP_UFTP/500M.file
...

Data transfer client - server

```
```

...  

[ucc cp] Using preferred protocols: [UFTP, BFT]
[ucc cp] Have 5 extra parameters for protocol UFTP
[ucc cp] UFTP: parameter <uftp.client.host> determined as <login02.marconi.cineca.it>
[ucc cp] File transfer URL : https://grid.hpc.cineca.it:9111/CINECA-MARCONI/services/FileTransferUFTP?res=a7d2c8ea-fada-4b79-bf5b-8273ebabca84 provacarpene.txt 100% 6 1,9B/s

**Speed rate: 90/130 MB/sec via Public Network with 1GB file**
Data transfer server - client


[ucc cp] Using preferred protocols: [UFTP, BFT]

[ucc cp] Have 5 extra parameters for protocol UFTP
[ucc cp] UFTP: parameter <uftp.client.host> determined as <login02.marconi.cineca.it>

[ucc cp] File transfer URL : https://grid.hpc.cineca.it:9111/CINECA-MARCONI/services/FileTransferUFTP?res=d0b3bdf-2215-4b5d-bdb0-d096e76b9d46
provacarpene.txt                                        100%  6 1,9B/s

Speed rate: 90/130 MB/sec via Public Network with 1GB file
Actually CINECA, CSCS, BSC, JSC, CEA supports UFTP and have been federated.

UNICORE services are currently published in the HBP UNICORE Registry.

Also KIT is going to provide a new UFTP server installation.
Data transfer server - server


... [ucc cp] Using preferred protocols: [UFTP, BFT]
[ucc cp] Synchronous transfer = true
[ucc cp] Using preferred protocol: UFTP
[ucc cp] Have 5 extra parameters for protocol UFTP
[ucc cp] Have filetransfer instance: https://unicoregw.cscs.ch:8080/DAINT-CSCS/services/FileTransfer?res=ad73fab0-29f7-4930-808b-2750d173b8bd /homeb/zam/hbptests/uftptest/1GB 100% 1,0G
real 0m43.931s
user 0m21.928s
sys 0m0.560s
In [5]:

```python
import requests
from bs4 import BeautifulSoup

# Pull the HTML from the URL
html = requests.get('https://example.com').text

# Parse the HTML
soup = BeautifulSoup(html, 'html.parser')

# Extract the data you need from the HTML
```

In [1]:

```python
# Your code here
```
UFTP setup on HBP sites
CINECA UFTP installation

CINECA UNICORE setup

- FEC - frontend
  - Gateway
  - Registry
  - XUUDB

- Cineca LDAP
- Unity HBP
- OIDC Server

- HBP LDAP

- MARCONI
  - UFTPD
  - TSI

- GPFS - $WORK

- Other Sites
  - Public Network
  - PRACE Network
JSC UFTP installation

- UNICORE Gateway
  https://hbp-unic.fz-juelich.de:9112
- JSC UNITY
  https://jsc-unity.fz-juelich.de
- HBP Registry
- JURICA UNICORE/X
- XUUDB
- JUDAC UNICORE/X

PRACE

UFTP D

TSI

PLANNED!

judac.fz-juelich.de

PLANNED!

Shared file systems
$WORK, $DATA, $HOME, ...

GPFS

PLANNED!

HBP OIDC
Token validation
User authentication

Automated from JSC user management incl. HBP users

jureca*.fz-juelich.de
CSCS UFTP installation

UNICORE setup at CSCS

Legend
- Data feed
- Workflow
- Future Work

CSCS Kerberos

HBP LDAP

CS4S LDAP

Piz Daint

Lustre scratch

UFTP gnitetti.cscs.ch

TSI

GPFS

PRACE Network

Other sites

INTERNET

CSCS

Not yet automated

Elogin daint101.cscs.ch

TSI

Rest API

UNICORE/X ronco.cscs.ch

XUUDB

UNICORE GW brissago.cscs.ch

UNITY

OIDC

user mapping
BSC UFTP installation

Unicore setup at BSC

Marenosrum

Prace Network
Public Network

dt02.bsc.es
UFTP D
TSI

GPFS Storage

Unicore Gateway
Unicore X
xuubdb
unicore-hbp.bsc.es

Unity HBP
OIDC
HPB LDAP
Server-server via PRACE NW

Performance tests are currently on going between CINECA and CSCS. CSCS will deploy a test version of the UFTPD server on a dedicated virtual node, hooked up to the PRACE NW interface. Server-server data transfer will be performed between CINECA, CSCS and also JSC.
THANKS!