

Moving data with Globus at Fermilab HPPC Facilities

Yujun Wu
Fermilab

Globus Online Usage at HPPC

□ Three Globus Online (GO) endpoints

- `fulla#cc3`: Cosmology data transfer in/out of Lustre file system;
1Gb network connection, Globus GridFTP server
 - `lqcd#fnal`: USQCD data transfer in/out of Lustre file system
10Gb network connection, Globus GridFTP server
 - `usqcd#fermi`: Direct transfer with Fermilab mass storage tape system (dCache/Enstore);
10Gb network connection, dCache GridFTP
- ◆ The transfer accounting info are reported to OSG Gratia service (grid accounting);

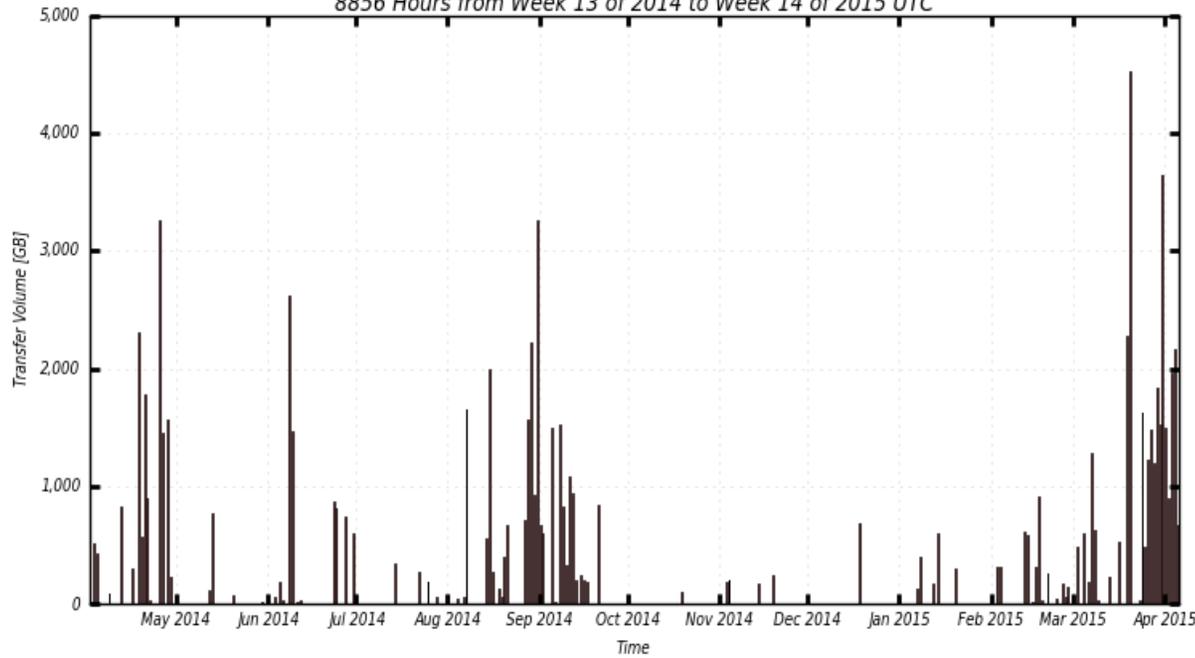
Volume of Gigabytes Transferred By Facility

365 Days from Week 14 of 2014 to Week 14 of 2015

1,200,000

Volume of Gigabytes Transferred By Facility

8856 Hours from Week 13 of 2014 to Week 14 of 2015 UTC



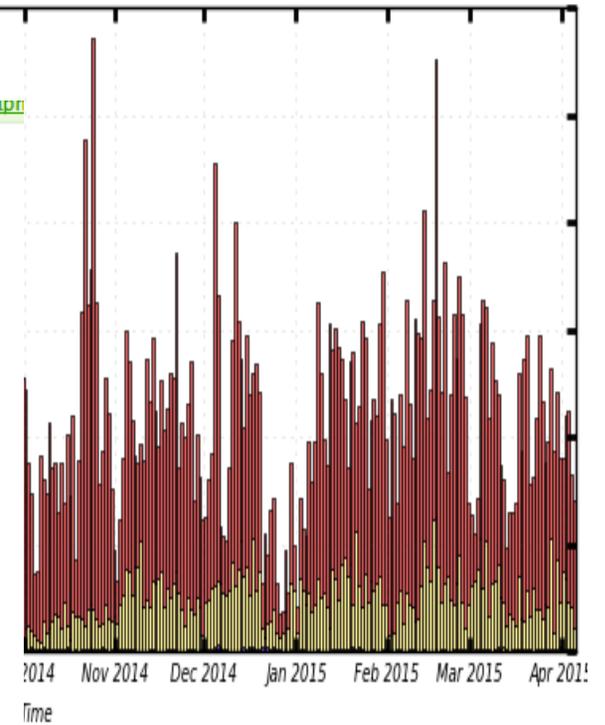
■ FNAL_LQCD

Maximum: 4,536 GB, Minimum: 0.00 GB, Average: 115.26 GB, Current: 672.05 GB

■ UCD

■ Other

Gratia



- USCMS-FNAL-WC1-SE
- FNAL.GOV
- FermiGrid-Bestman
- FNAL_LQCD
- U.S.GOV
- FG-IF-GRIDFTP-MINERVA
- FG-IF-GRIDFTP-UBOONE
- UMissHEP

Maximum: 1,143,472 GB, Minimum: 72,217 GB, Average: 414,425 GB, Current: 281,978 GB

Gratia transfer plots between April 2014 and April 2015. Most transfers in FNAL_LQCD were through GO. FNAL_LQCD is one of the facilities monitored by OSG Gratia service (some transfers from other facilities may not through GO).

Experience with Globus

❑ The good part

- Many users like to use it. Many good features: easy to set up, fire and forget transfer; much faster transfers than scp;
- Work with existing Grid authentication/authorization infrastructure;
- Great help from Globus Support (**Thank you!**);

❑ Still need

- Need make sure users can access and are authorized to access the right data (of course!);
- Wish to have the user job event log as an admin when needed. endpoint admin interface?
- Sometime time consuming to debug issues and need involvement from Globus Support. One example:

A cosmology user tried to transfer data from Argonne ALCF to Fermilab Cosmology at HPPC, but only got ~KB/s transfer rate until help from Globus (Karl/Dan) to trace the transfers. Temp fix: adjust --perf-cc and --perf-p parameters to 1. Working with ALCF on tracing down the root cause;

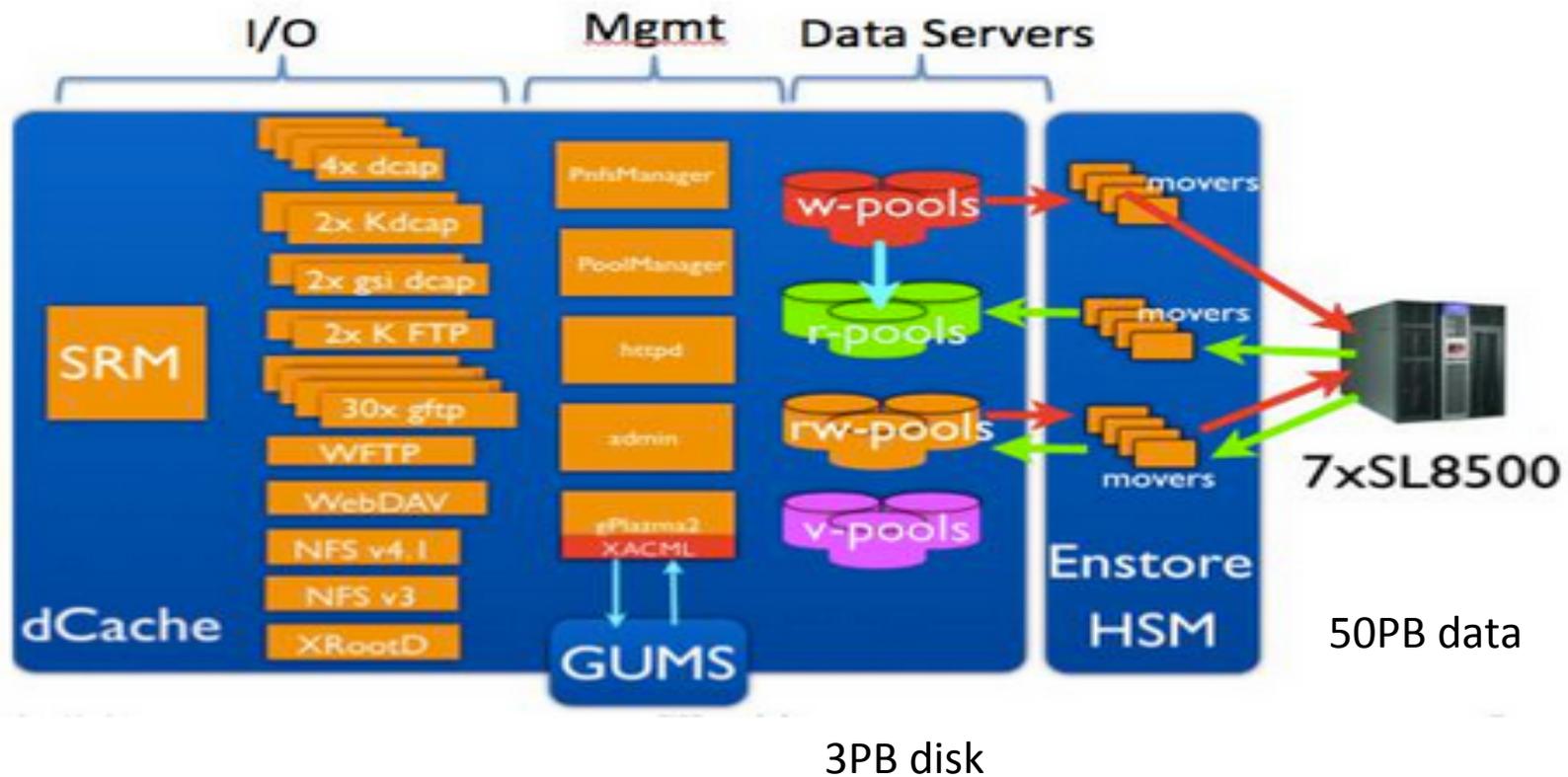
Experience with Globus Online (2)

- Continue to have compatibility issues between GO and dCache GridFTP server:
 - Experience problems with bulk recursive transfers when using GO against dCache GFTP servers;
 - The option "verify file integrity after transfer" does not work because of differences in checksum algorithms. GO supports only MD5 whereas dCache+Enstore currently does not support MD5, only Adler32;

- ◆ Right now, USQCD users are advised: **“There continue to be compatibility issues between GO and “door” nodes; globus-url-copy or gridftp may be a better choice for some endpoints. “**
 - USQCD All Hands' Meeting 2014

Big storage/big opportunities

- Fermilab is a pioneer in large data handling and storage management



Big storage/big opportunities(2)

- 1PB Lustre disk storage for LQCD and Cosmology;
- 3 PB dCache disk storage (2 PB more coming soon);
- 50PB in public tape storages (many more can be added)
- Great network connections to other institutes

Esnet 5 (100GE-based), Esnet 4 (10GE-based)

□ With the help from Globus, there could be many opportunities for research computing providers to help researchers to store, distribute and fully utilize the large amount of data and to collaborate between each other:

- Short-term opportunity storage for researcher's intermediate data;
- Long-term archive storage for experimental/research data;
- Fast data distribution among collaborators and compatibilities between different implementations;
- Research on efficiently utilizing the network bandwidth with end-to-end transfers;
- Large-scale storage technology research;

◆ One small example: debugging the data movement for Fermilab researchers (INCITE grant) has already involved help from both ALCF(Mira) and Fermilab resource providers and Globus Support;

Summary

- Globus Online is a very useful tool for our users and resource providers. More work is needed as the resource provider and some new features can be helpful from admin point of view;
- With the help from Globus, there could be many opportunities for the research computing providers to help researchers and collaborate between each other in (large) research data storing, distributing and utilization (publishing);
- Thank you for the good work from Globus. Hope we will continue to get excellent support from you!