

WLCG DOMA TPC Updates

Brian Bockelman
HOW2019

Third-Party-Copy (TPC)

- Third party copy - moving data from site A to site B - is one of the core, essential activities on the grid.
 - Why is it suddenly interesting?
- Actually, many individuals have been working on improving approaches and techniques continuously!
 - It is suddenly *more* interesting because, as a community, we have realized a need to replace the functionality found in the Globus Toolkit (particularly, GridFTP & GSI).
 - It is not the sole reason but rather the catalyst.

Working Group

<https://twiki.cern.ch/twiki/bin/view/LCG/ThirdPartyCopy>

- We formed the working group shortly after CHEP.
 - Aim is to **put in place viable alternate protocol(s) for GridFTP.**
 - Participants are currently working on both XRootD and HTTP/WebDAV.
- We break down the work to three phases:
 1. **Prototype / implementation:** Demonstrate viability of protocols.
Ensure all storage implementations have a valid alternate in production.
 2. **Early deployment:** Ensure rollout of alternates at all sites with >3PB storage.
 3. **Widespread deployment:** Rollout to remaining WLCG sites.

You are here

Site Participants

Production sites, testbeds, developer instances

- RAL
- Prague
- Glasgow
- Brunel
- Manchester
- Lancaster
- CERN DPM trunk
- CERN DPM release candidate
- DESY developer testbed
- DESY dedicated testbed
- AGLT2
- BNL
- Imperial College
- PIC
- CERN EOS pre-production
- INFN-T1
- Queen Mary University
- NERSC
- SLAC
- OU
- Nebraska
- Bonn
- FNAL
- Beijing
- DynaCloud CERN
- DynaCloud CERN (Grid instance)
- IN2P3
- Brussels
- Florida
- SURFSara (production & test)
- NDGF
- University of Victoria
- TRIUMF

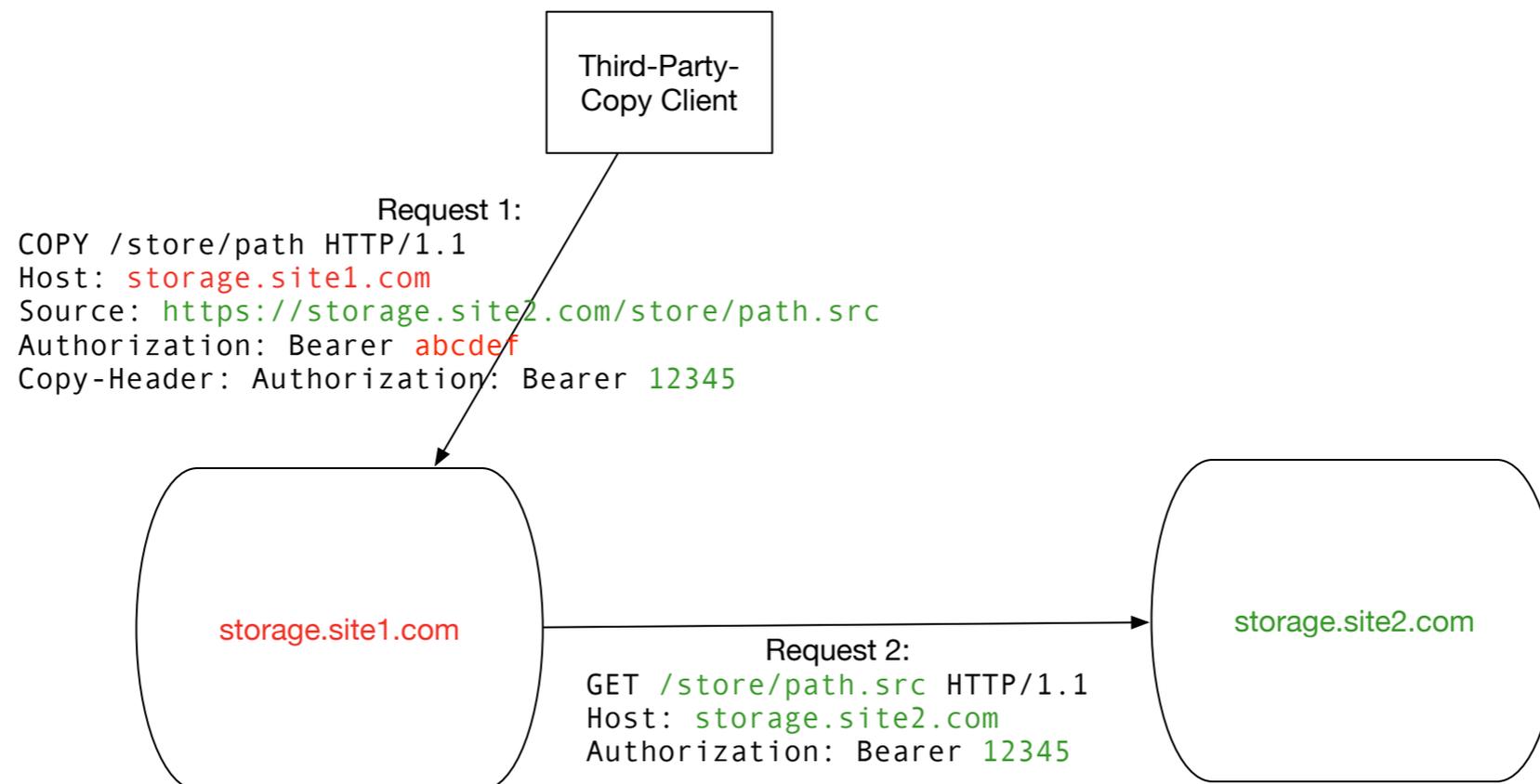
Storage Participants

More on storage status later!

- We have active participation from both those implementing the protocols (such as the XRootD, StoRM, and dCache teams) ...
 - XRootD
 - dCache
- ... And from those who use the implementations as part of a larger storage element.
 - DPM
 - EOS
- Example: XRootD implements a server exporting both HTTP and xrootd protocols; Echo utilizes this as a core piece of their storage.
 - StoRM
 - Echo

Basic Idea

- Both protocols convert the third-party-copy to a normal download that is performed by an “active” side.
- HTTP can do either push or pull: as only one side needs to understand COPY, the other side can be a “pure” HTTP server - nginx, Apache, etc.



Note: at least theoretically, the two requests could be done with different protocols!

Trials & Tribulations

- **HTTP/WebDAV:**
 - Currently have 5 independent implementations of TPC in WebDAV (StoRM, XRootD, dCache, DPM, Dynafed).
 - HTTP is not as simple as it looks! We have to ensure everyone utilizes the same critical headers, agree on header values, etc. Requires some amount of bug-hunting and communication between developers.
 - Took a more forward-looking approach to authentication (token-based) for the data movement — avoids need for X509.
- **XRootD:**
 - XRootD team updated the security protocol to delegate X509 credentials. Transfers are now done with the credentials of the user, not the server.
 - Significant improvements to protocol documentation.
 - Now have two independent implementations (SLAC & dCache),.

Tribulations

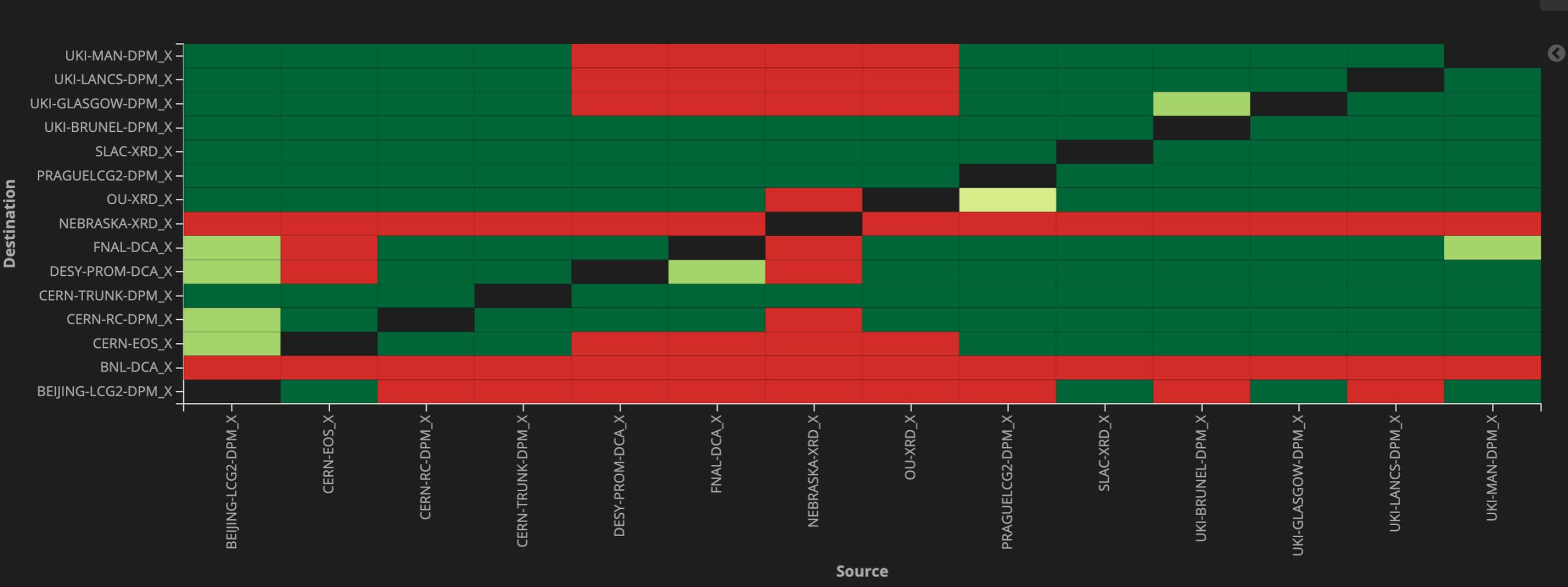
- In late 2018, we setup a dedicated Rucio instance to drive transfers between sites.
- Around January 2019, started scale testing HTTP transfers.
- At this point, regularly moving >500TB / week across participating scale test sites.
- So many sites participating in the test transfers that we are breaking the Kibana plots. Split out over the next few slides!

Transfer Connectivity

Do not pay attention to any given site - who is “red” and who is “green” when the snapshot was taken. This evolves daily.

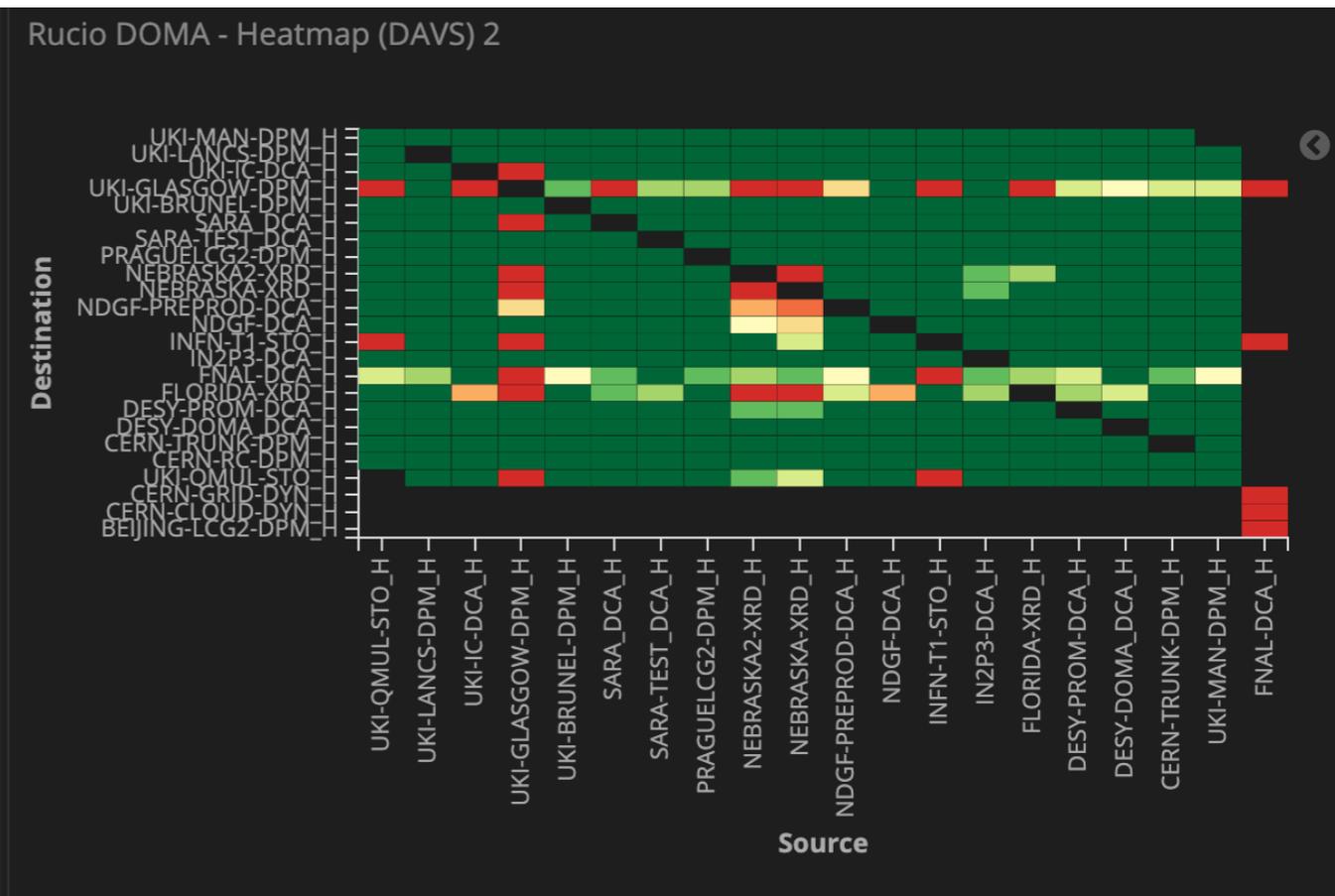
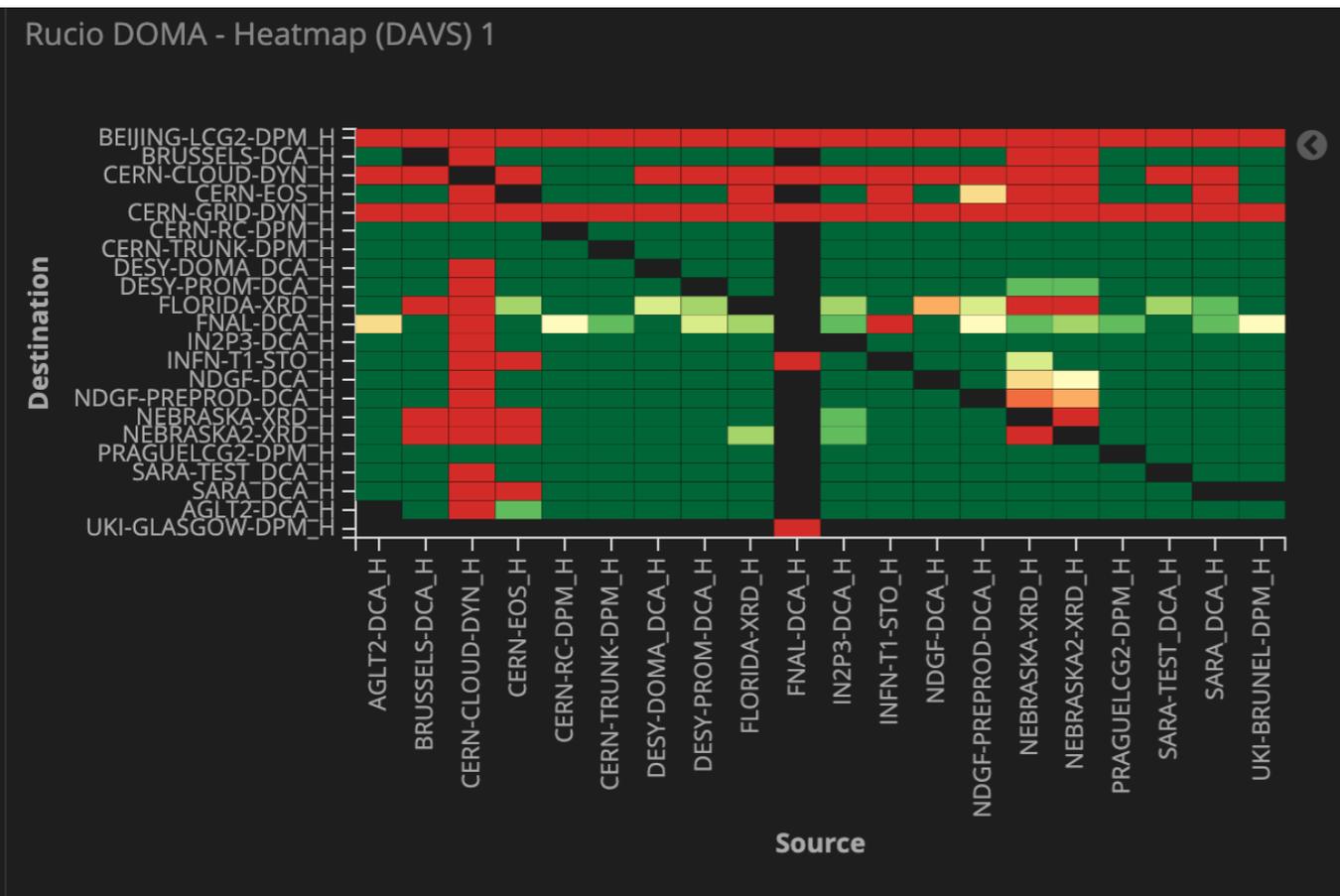
Focus on the breadth of the testing!

Rucio DOMA - Heatmap (Root)

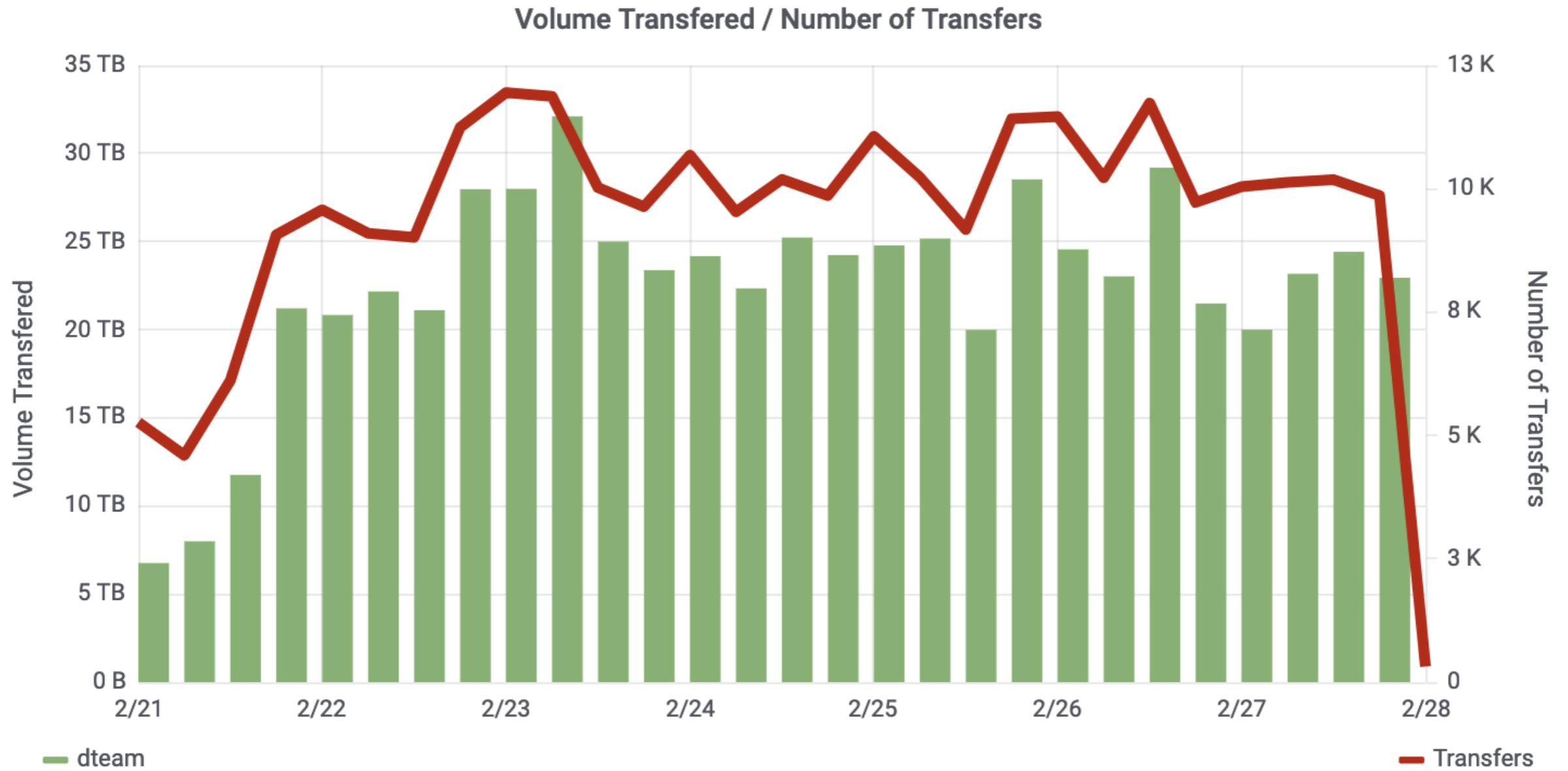


Transfer Connectivity

Do not pay attention to any given site - who is “red” and who is “green” when the snapshot was taken. This evolves daily.
Focus on the breadth of the testing!



Data Rates



Each bar is a 6-hour time window.

Data Transfers

- Both XRootD and HTTPS implementations support load-balancing transfers over multiple TCP streams.
 - Honestly, this appears to provide minimal benefit these days. TCP is much better than it used to be — and, as a community, we tend to have multiple files in flight as opposed to a single 100Gbps transfer.
- We had some internal questions about the cost of encryption for HTTPS. Luckily, almost all the heavy lifting is done by hardware, no longer software.
 - Quick tests show that a modern server can perform TLS encryption at ~385Gbps; serving over HTTPS, one can mostly fill a 40Gbps connection (tests limited by disk I/O).
 - A 7-year-old server can do ~20Gbps of encryption and serve at 10Gbps over HTTPS (tests limited by network card).

Nightly Smoke Tests

- It's difficult to digest a simple site status from transfer matrices: Is it my site that is broke? Is it the other end?
- Paul Millar has implemented some simple tests for HTTP TPC functionality against a "known good" endpoint.
 - These get sent out nightly. It's driven by a "bash script invoking curl"; any admin should be able to reproduce failures at their site.

Paul Millar 

Smoke test report 2019-03-14T14:34+01:00

To: wlcg-doma-tpc (WLCG DOMA Third Party Copy Deployment)

DOMA-TPC smoke test 2019-03-14T14:34+01:00

SOUND ENDPOINTS

```
AGLT2      dCache
CERN-TRUNK DPM
DESY-PROM  dCache
DESY-DOMA  dCache
FNAL       dCache
IN2P3      dCache
INFN-T1    StoRM
LRZ-LMU    dCache
NDGF       dCache
NDGF-PREPROD dCache
NEBRASKA2  xrootd[*]
NEBRASKA   xrootd[*]
PRAGUELCG2 DPM
SARA       dCache
SARA-test  dCache
UKI-BRUNEL DPM
UKI-IC     dCache
UKI-LANCS  DPM
UKI-QMUL   StoRM
```

PROBLEMATIC ENDPOINTS

```
UKI-MAN    DPM      Of 23 tests: 22 successful (95%), 1 failed (4%)
CERN-RC    DPM      Of 23 tests: 21 successful (91%), 2 failed (8%)
FLORIDA    xrootd[*] Of 23 tests: 21 successful (91%), 2 failed (8%)
BRUSSELS   dCache    Of 23 tests: 10 skipped (43%), 13 attempted (56%):
UKI-GLASGOW DPM      Of 23 tests: 10 skipped (43%), 13 attempted (56%):
CERN       EOS[*]    Of 23 tests: 13 skipped (56%), 10 attempted (43%): 6 s
TRIUMF     DynaFed[*] Of 23 tests: 12 skipped (52%), 11 attempted (47%): 6
UKI-RAL    DynaFed[*]/ECHO Of 23 tests: 13 skipped (56%), 10 attempted (43%):
BEIJING-LCG2 DPM      Of 23 tests: 18 skipped (78%), 5 attempted (21%):
BEIJING-TEST DPM      Of 23 tests: 18 skipped (78%), 5 attempted (21%):
```

Storage Status

Can you do transfers with FTS?

- **XRootD (SLAC):** Implementations of both protocols in release (4.9.0). Due to some bug fixes, OSG is holding off until the 4.9.1 release.
 - Non-POSIX sites (such as HDFS) need to write special plugins to use the xrootd protocol.
- **Echo:** Working on integrating the xrootd protocol. Hit some technical issues with their object ID formats.
- **DPM:** With Xrootd 4.9.0, everything works *as long as* the site is using the new DOME infrastructure.

Storage Participation

- **dCache:** HTTP is fully supported for all supported release series of dCache (3.2 and beyond).
 - XRootD TPC is in 4.2; the delegation portions should be present in 5.0 (transfers are working in the developer instance).
- **EOS:** XRootD TPC (no delegation) works in production. HTTP TPC works in a branch.
 - Timeline for XRootD TPC with delegation is unclear.
- **StoRM:** HTTP TPC works in latest release.

Next Steps

- We need a few brave sites to be “first across the breach” to switch over production transfers.
 - Particularly, this is well-tested for Rucio: need ATLAS site volunteers! I understand that there were two volunteers at the ATLAS Site Jamboree.
 - University of Bonn has been experimenting with changing to WebDAV & XRootD for production TPC downloads. Happily and helpfully have been filing tickets against source that break.
- Need to close out “Phase I” and work on “Phase II” - all SEs on WLCG greater than 3PB include a non-GridFTP method.
 - Will soon be mining REBUS and reaching out to sites.

We Want You!

<https://twiki.cern.ch/twiki/bin/view/LCG/ThirdPartyCopy>

- If you have >3PB of data, I'd encourage you to participate in the transfers! *Get your name on the list!*
- Peruse the above documentation and get HTTP and/or XRootD working.
- Add your site info to the table so we can test the endpoint.
- Join the mailing list, introduce yourself. Ask to be added to the functionality test matrix.

Questions?

<https://twiki.cern.ch/twiki/bin/view/LCG/ThirdPartyCopy>

