GridPP

UK Computing for Particle Physics

Data Transfer Approaches

Brian Davies

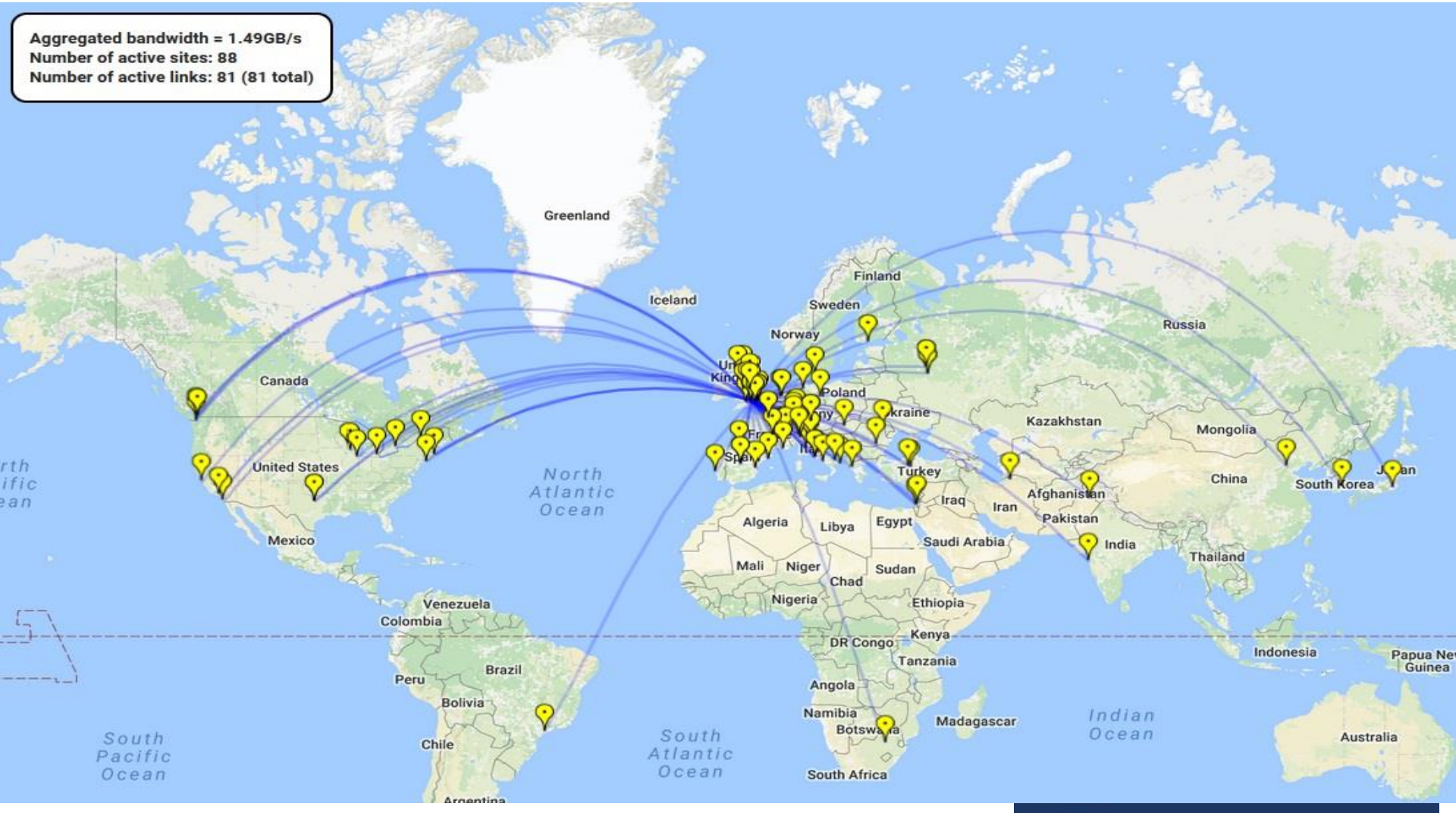
GRIDPP/SKA

Manchester, November 2016

Science & Technology Facilities Council

GridPP
UK Computing for Particle Physic

Transfers to a single site/1day/1VO

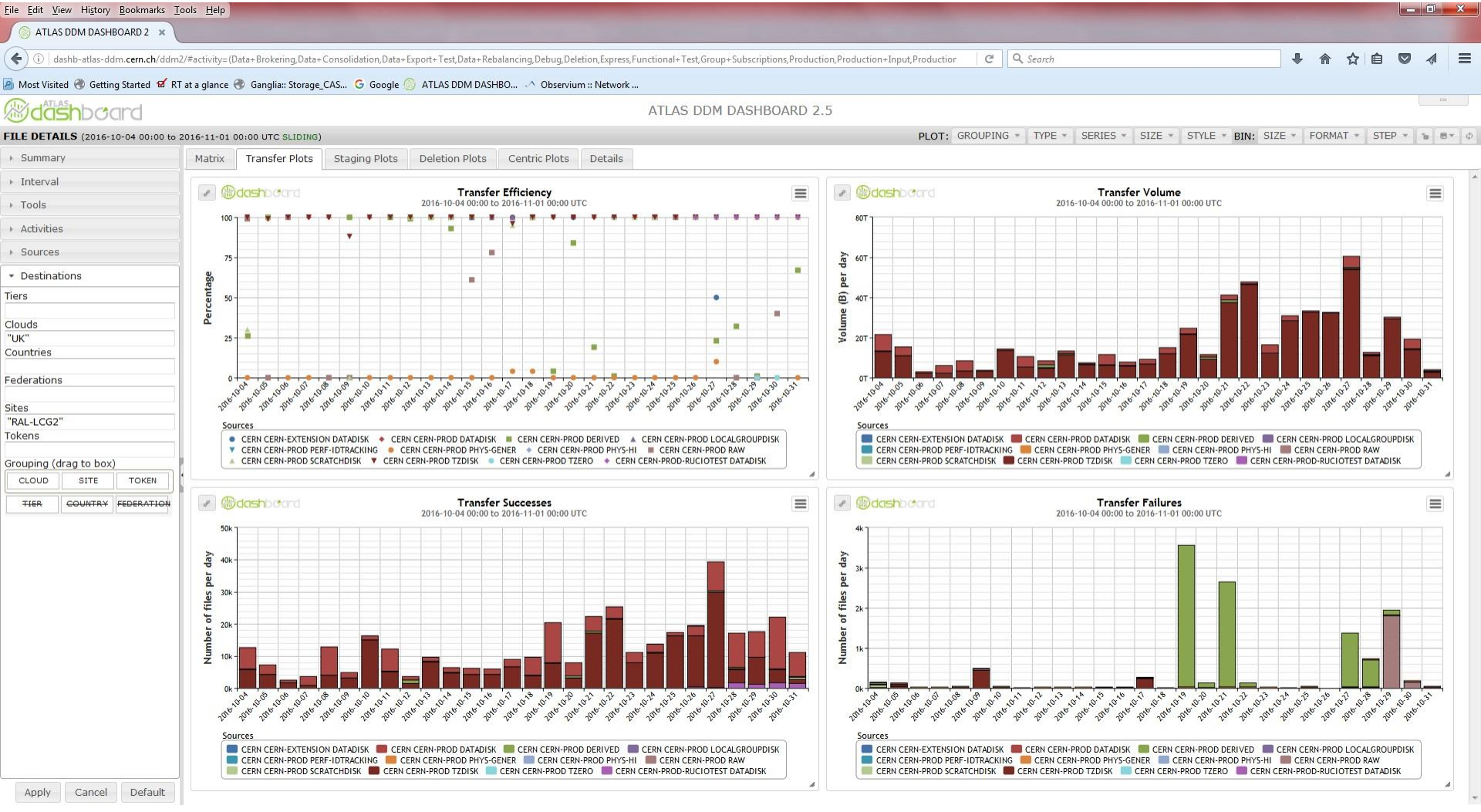


02 November 2016

Science & Technology Facilities Council

GridPP TB/day between sites now normal

UK Computing for Particle Physic

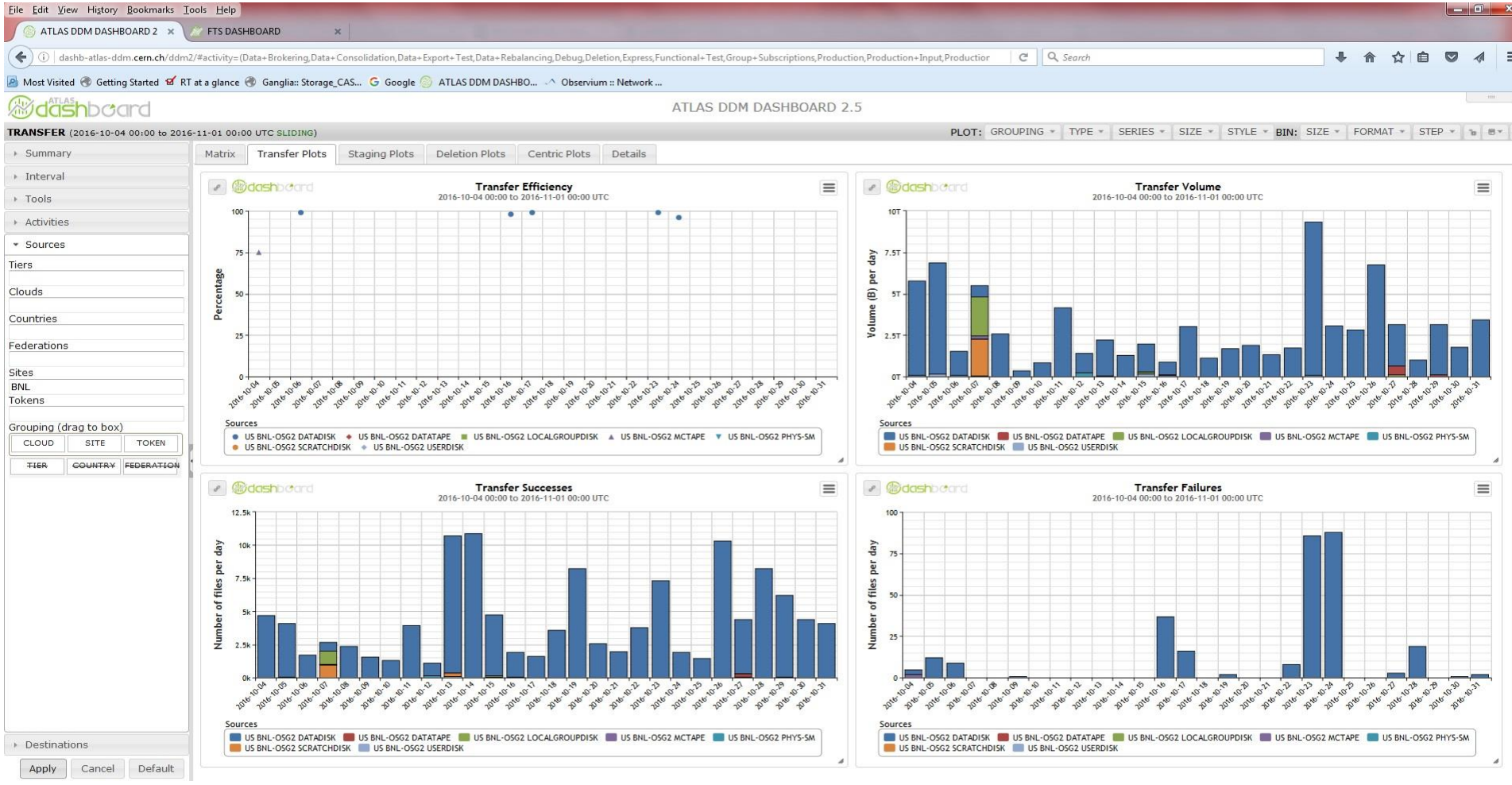


02 November 2016

Science & Technology Facilities Council

GridPP
UK Computing for Particle Physic

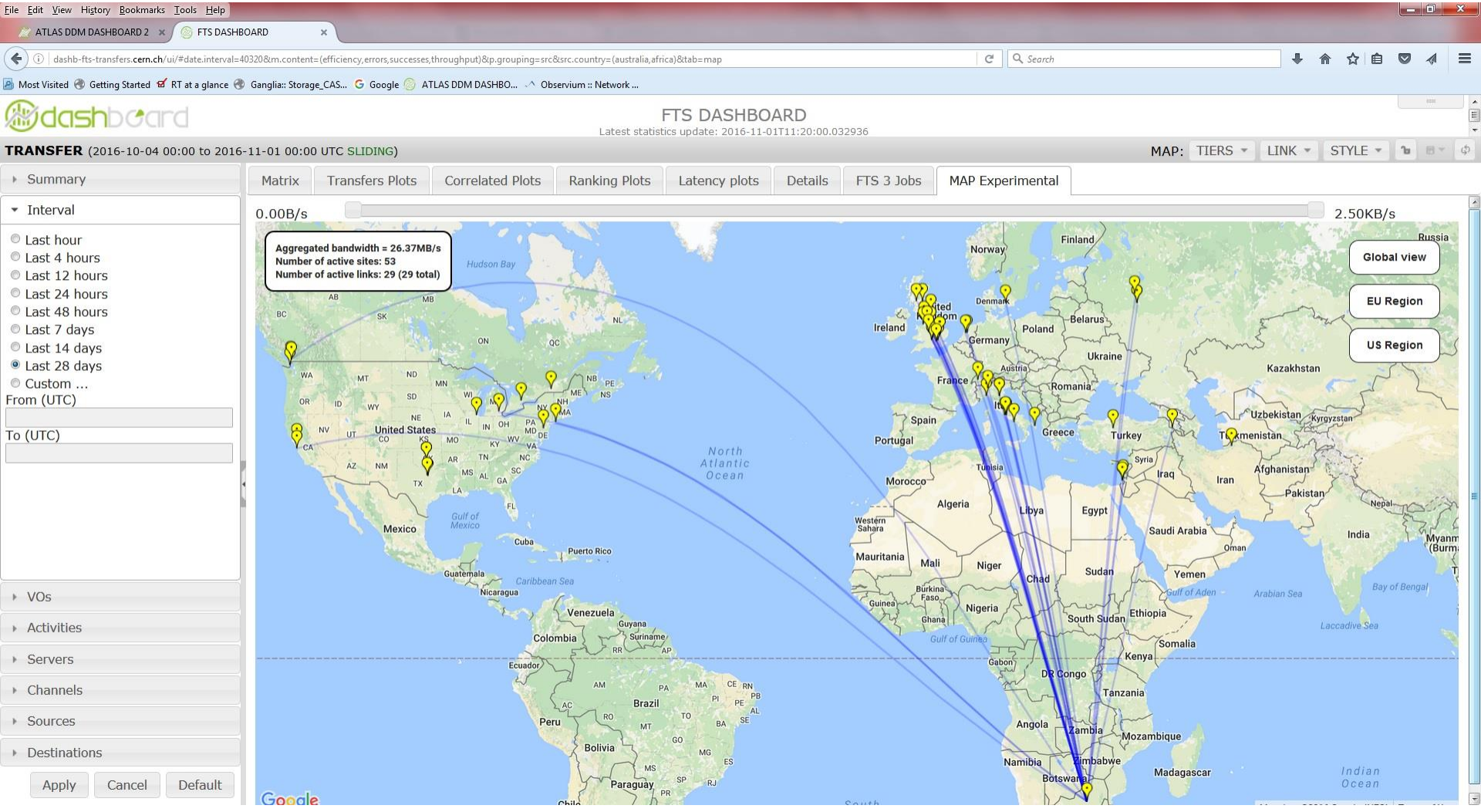
With Decent rates for Intercontinental transfer



02 November 2016

Science & Technology Facilities Council

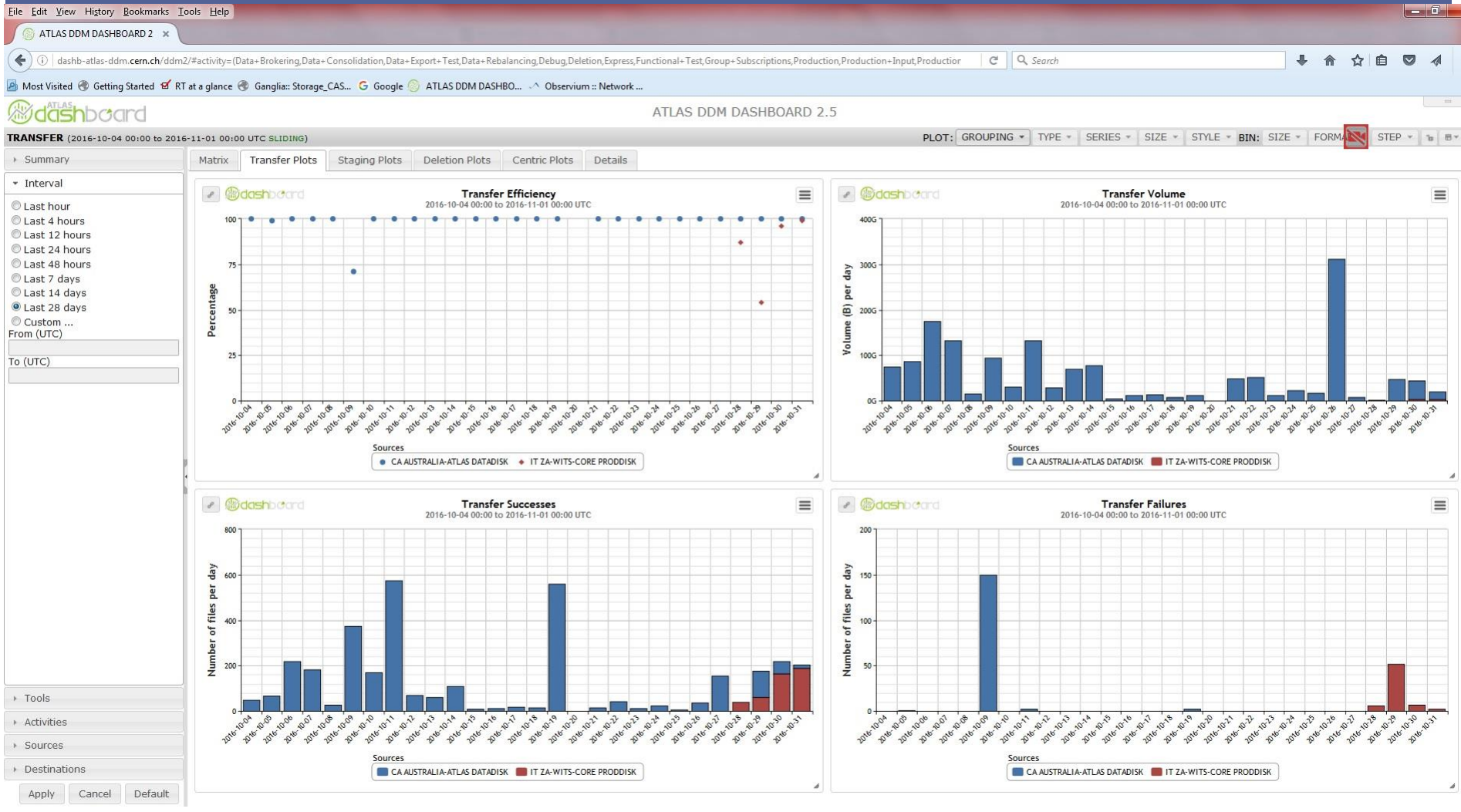
GridPP UK Computing for Particle Physics — Already Transferring from AU/SA



02 November 2016

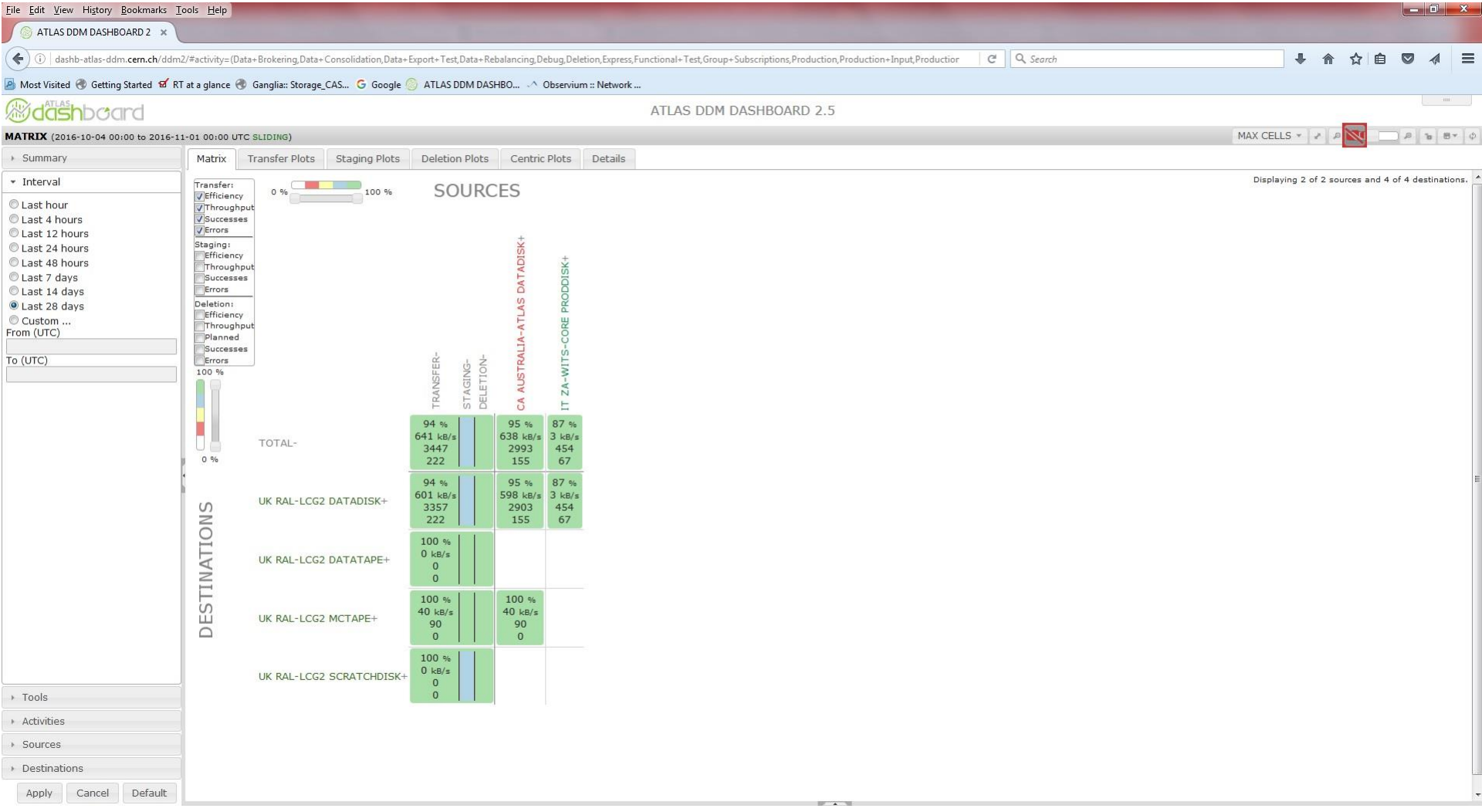
Science & Technology Facilities Council

Rates Lower from smaller sites



02 November 2016

Transfers to Disk and Tape Seamless*



02 November 2016

WLCG Has a lot of Data transfers to monitor

- 167 Sites in 43 Countries on six Continents
- Storage endpoints containing 250PB (disk) 300PB (tape)
  - Organised and chaotic access
  - Supporting Single/Multiple endpoints for Single/Multiple Virtual Organisations
  - Vary in size and scope
    - 10TB-10s of PB of Total Storage (Disk and Tape)
    - 1/10 GE NICs, 1/10/100 Gbps, R&E networks and private OPN
    - 10TB-1PB filesystems/object stores, 1-300 diskservers per site
    - Multiple filesystems (XFS,HDFS,CEPH,GPFS,Lustre)
- Central Production and User initiated
- Last two years WLCG has moved 0.5EB of data
  - Over 1billion files.
- WN jobs produce a lot of data which also has to be stored/moved
  - One VO runs 200k concurrent jobs which last 10mins to 72 Hours.
  - 0-100s of Input files, 2-3 Output files
    - Individual file open times 1-10000s

Data movements vary greatly

- File size from ~10B to ~10GB
- Latency between hosts from 0.1ms to 350ms (just for the UK)
- Different workflows require different data movement
  - WAN SE<->SE, SE->WN, WN->SE
  - LAN WN<->SE, SE<->SE
    - Different Tools to monitor different workflows
- Different storage middleware
  - Native gridFTP, BeSTMan, DPM, dCache, SToRM
- Different transfer protocols
  - gsiFTP, http/WebDaV, xrootd, NFSv4.1, S3

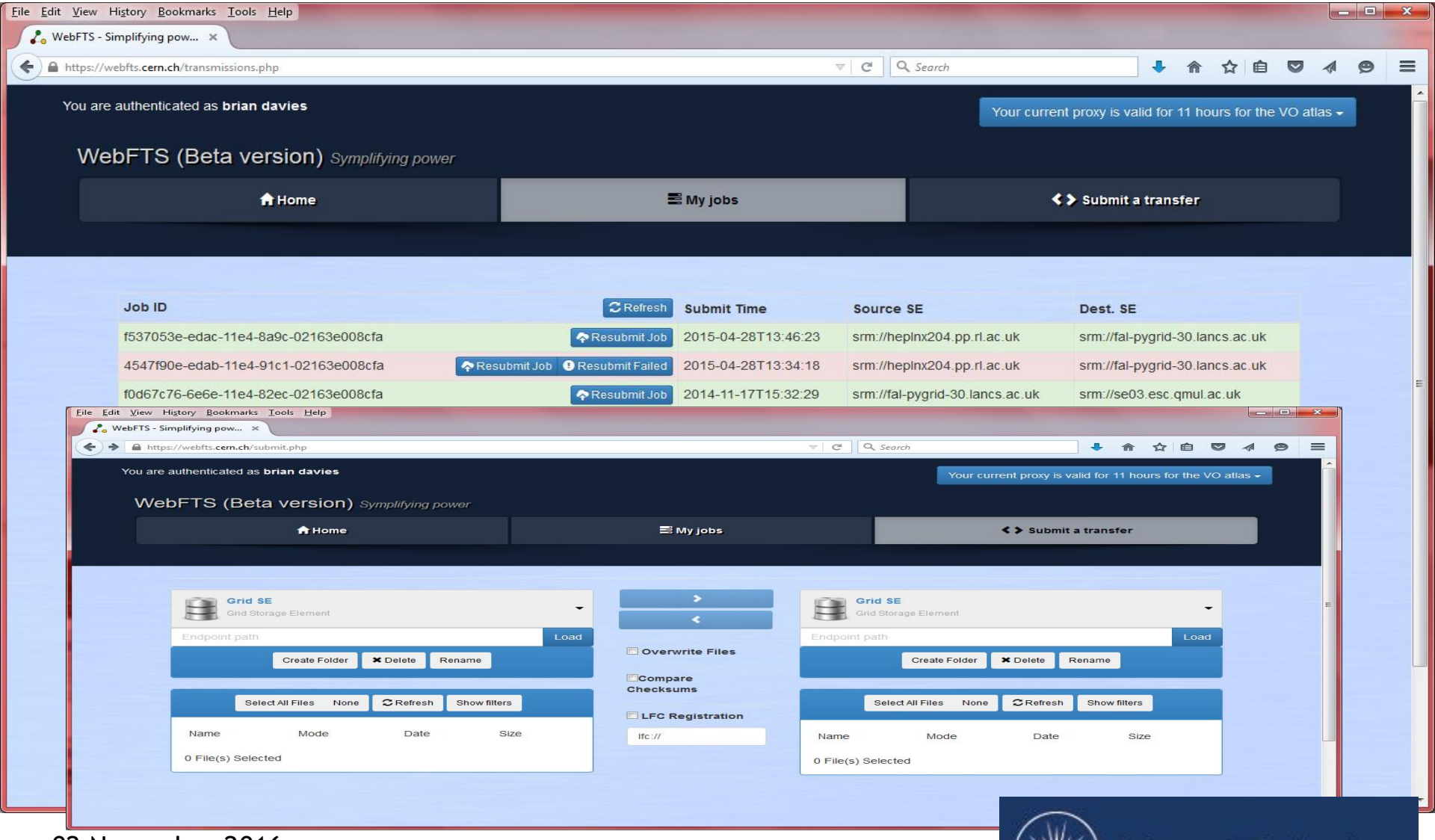
File Transfer Service (FTS) Moves data!

- EGI Middleware Stack
- Can handle many VOs
  - 22 (HEP and non-HEP)
- Checksum validation of files
- Retry of failed transfers
- Auto-optimisation of transfer parameters to maximise throughput
- Ability to set limits suitable for varied storage setups
- Web friendly GUI also available!! Federated Failover
  - Mainly use Command line tools or higher level control systems.
- Handle many file transfers (~1.5M a day)
  - Single to thousands of files per single submission

File Transfer Service (FTS) Moves data!

- High Available service
- 1/4 of main FTS services used by WLCG run by RAL
- IpV6
- Work with Developers

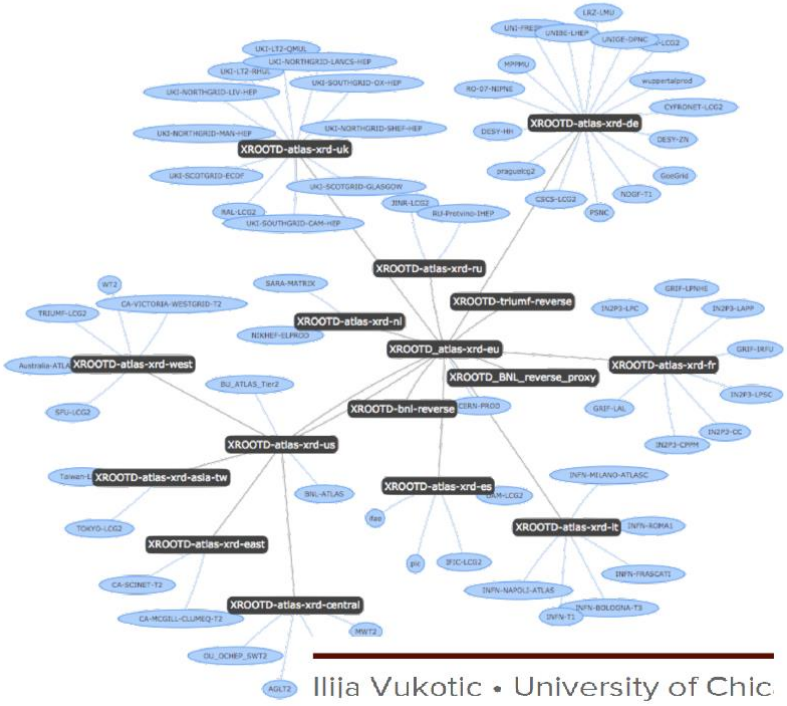
Web GUI


File Edit View History Bookmarks Tools Help
WebFTS - Simplifying pow...
https://webfts.cern.ch/transmissions.php
Search
You are authenticated as brian davies
Your current proxy is valid for 11 hours for the VO atlas
WebFTS (Beta version) Symplifying power
Home
My jobs
Submit a transfer
Job ID
Refresh
Submit Time
Source SE
Dest. SE
f537053e-edac-11e4-8a9c-02163e008cfa
Resubmit Job
2015-04-28T13:46:23
srm://heplnx204.pp.rl.ac.uk
srm://fal-pygrid-30.lancs.ac.uk
4547f90e-edab-11e4-91c1-02163e008cfa
Resubmit Job
Resubmit Failed
2015-04-28T13:34:18
srm://heplnx204.pp.rl.ac.uk
srm://fal-pygrid-30.lancs.ac.uk
f0d67c76-6e6e-11e4-82ec-02163e008cfa
Resubmit Job
2014-11-17T15:32:29
srm://fal-pygrid-30.lancs.ac.uk
srm://se03.esc.qmul.ac.uk
https://webfts.cern.ch/submit.php
Grid SE
Grid Storage Element
Endpoint path
Load
Create Folder
Delete
Rename
Select All Files
None
Refresh
Show filters
Name
Mode
Date
Size
0 File(s) Selected
Overwrite Files
Compare Checksums
LFC Registration
lfc://

History/Future of Data movement

- Scp/rsync
- Globus-url-copy
  - Single machines
  - gridftp
- Multiple Machines with Storage Resource Management Frontends controlling Hierachal Storage Systems
  - Disk/Tape
  - Multiple transfer protocol
- Distributed filesystems with gateway head nodes
- Lustre/GPFS
- Object Store /Cloud infrastructures with new/old protocols
  - CEPH/Amazon S3/Swift/gsiFTP/http/NFS4.1/xrootd
- Globus Online

Example of global network for Federated Data Access



- Job Access data from any site in world
- Automated re-direction
- Xrootd transfer protocol

FAX backup transfer mechanism also monitored

02 November 2016

DiRAC Data Movement Case

- DiRAC an HPC collaboration Astro/Cosmo/PP theoretical modellers.
- Needed a method to archive PB of data from various sites to Tape Store (at Rutherford Appleton Laboratory.)
- Use knowledge base from GridPP
- Bespoke solution for specific needs.
  - No Data Management
- Took FTS experience and simplified the middleware stack to bare bones. (Simplified Authen'|Author')

- Four sites
  - Durham progressing (L.Heck)
  - Leicester setup and ready for transfers (J.Wakelin)
    - Site had no previous GrridPP Resources
  - Currently Setting up Cambridge (No previous GridPP experience) and ECDF (Colleagues have GridPP Experience.)
- Document available to setup further sites
- Looking to compliment with Globus Online/Connect

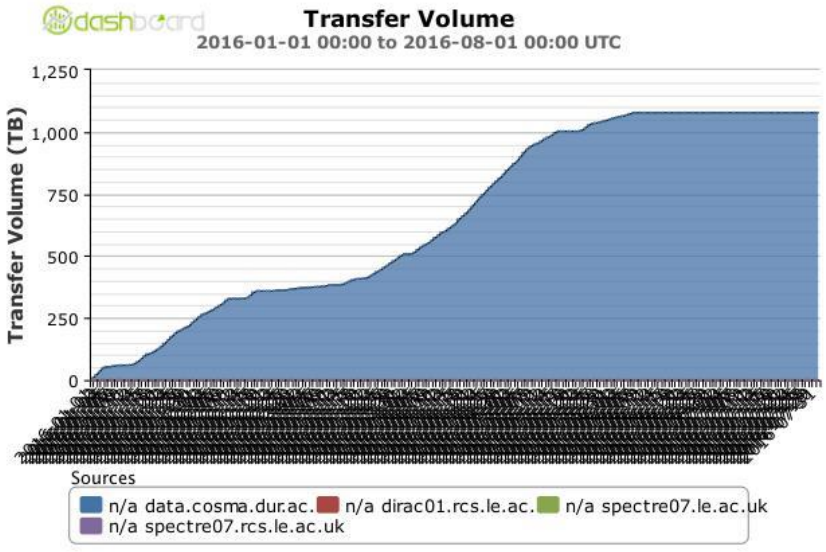
Vo.dirac.ac.uk

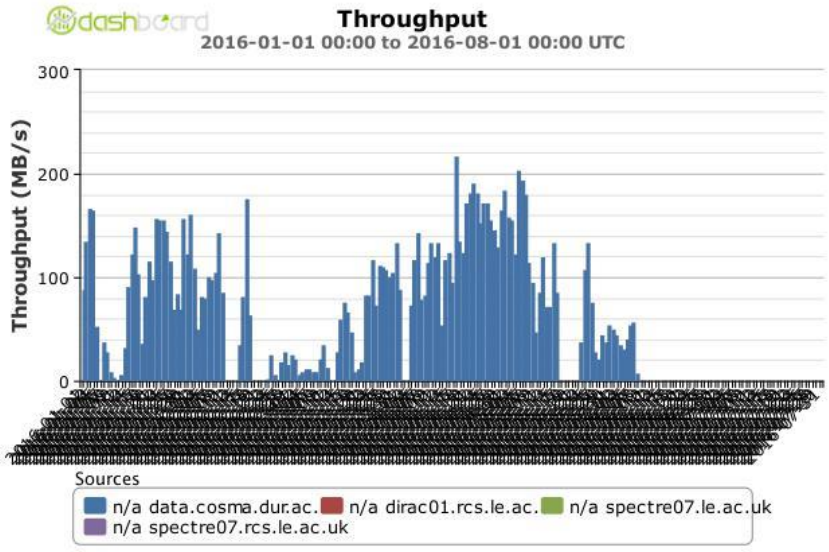
1.2PB copied in 5 Months

(Periods of reduced activity as site admin calculates what to move

Peak Values 350MB/s

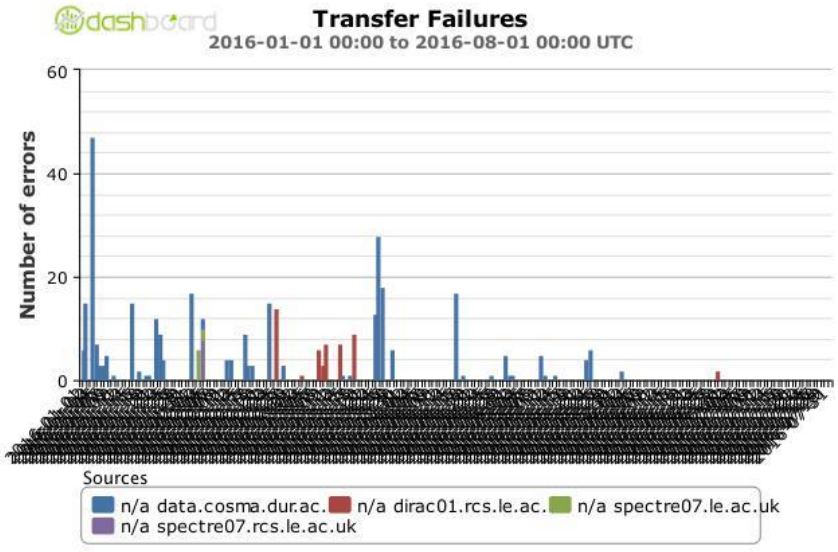
Daily Averages can reach over 200MB/s



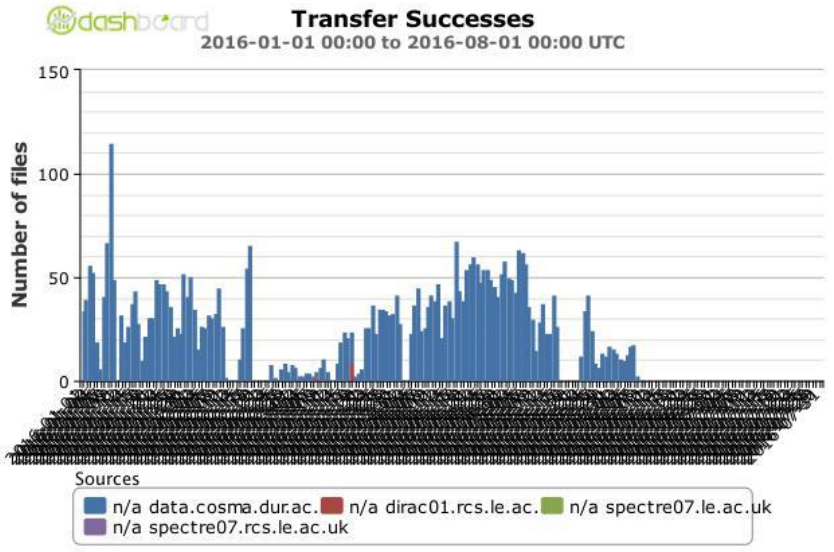


Vo.dirac.ac.uk

Low failure rate



Low number of transfers since change to large file size



Durham

126k files in total

Mean file size 9.45GB (96% files are smaller than this.)

Largest file 1.69TB

Median 6.5KB

3966 files larger than 100GB

Newer 250GB files leave unused space on tapes.

22TB over 126 Tapes (2.2% inefficiency (0.75/0.5 % for ATLAS/CMS.)

File/Object size effects transfers!

(Rates and Mechanisms.)

Moving data v' simple compared to managing it

(Over to Sam…)