ESnet
ENERGY SCIENCES NETWORK

FasterData Mobility Testing Framework

Internet2 TechEX23

Location: Minneapolis, MN

Date: September 19, 2023

Ken Miller, Science Engagement
Energy Sciences Network (ESnet)
Lawrence Berkeley National Laboratory

Motivation/Background

- Networks are an essential part of data-intensive science
  - Connect data sources to data analysis
  - Connect collaborators to each other
- Performance is critical, but **often** overlooked
  - Exponential data growth and not always aware of new data sources
  - Constant human factors and
  - Data movement and data analysis must keep up
- Effective use of wide area (long-haul) networks by scientists has historically been difficult
- Different IT groups manage various components of research workflow

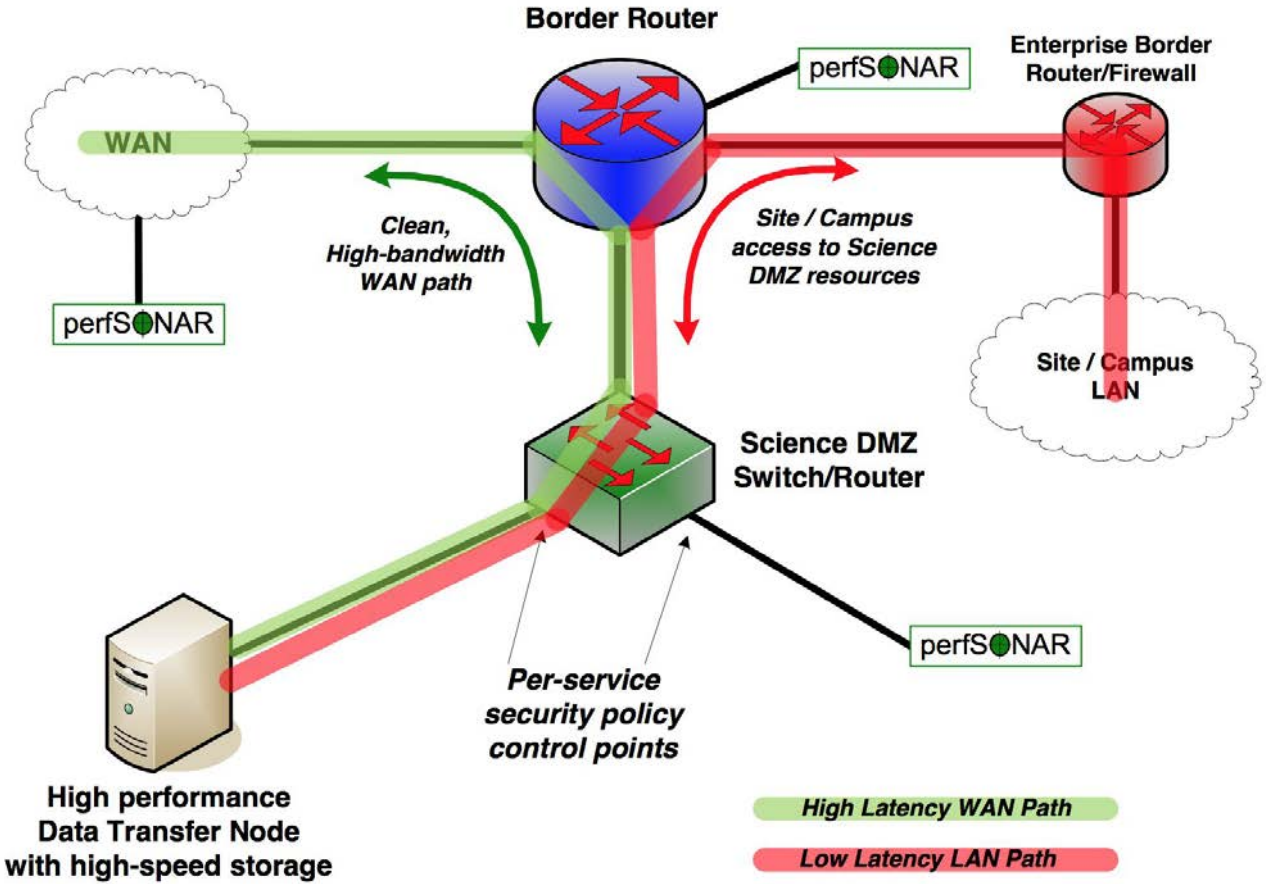
ESnet

Measure Twice, Copy Once

- With CC* and campus upgrades, how many sites are
  - measuring network speeds with perfSONAR?
  - measuring data transfer speeds to campus resources?
  - following up and engaging researchers to test/set data transfer expectations?
  - reviewing weekly Top 10 source/dest Netflow/IPFIX/sFlow or firewall logs for total volume transferred
- To baseline, how long does it take your site to transfer 1TB of data?
- **Our goal is a 10G connected DTN capable of 1 TB/hour (2.22 Gb/s) disk-to-disk, as a minimum, but ideally, 3TB/hour (6.67Gb/s)**
- **Our goal for clustered 10G, 25G, 40G, 100G DTNs is 6 TB/hr (13.23 Gb/s)**

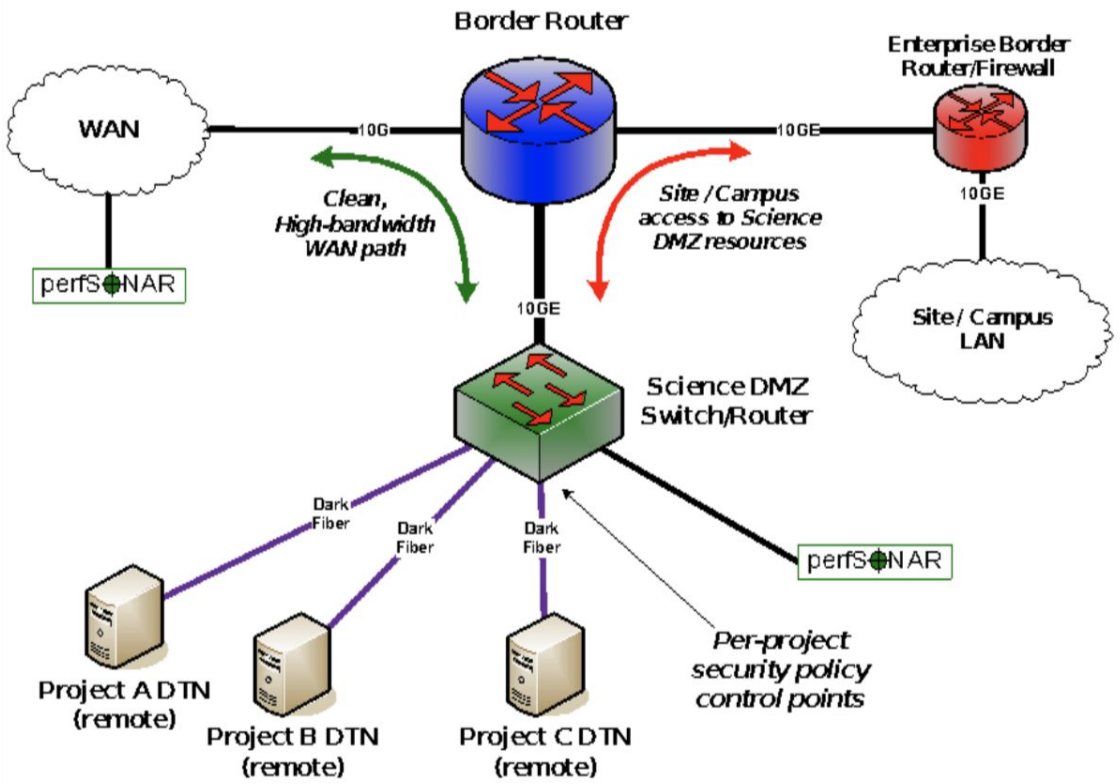
ESnet

Foundations: Science DMZ Design Pattern



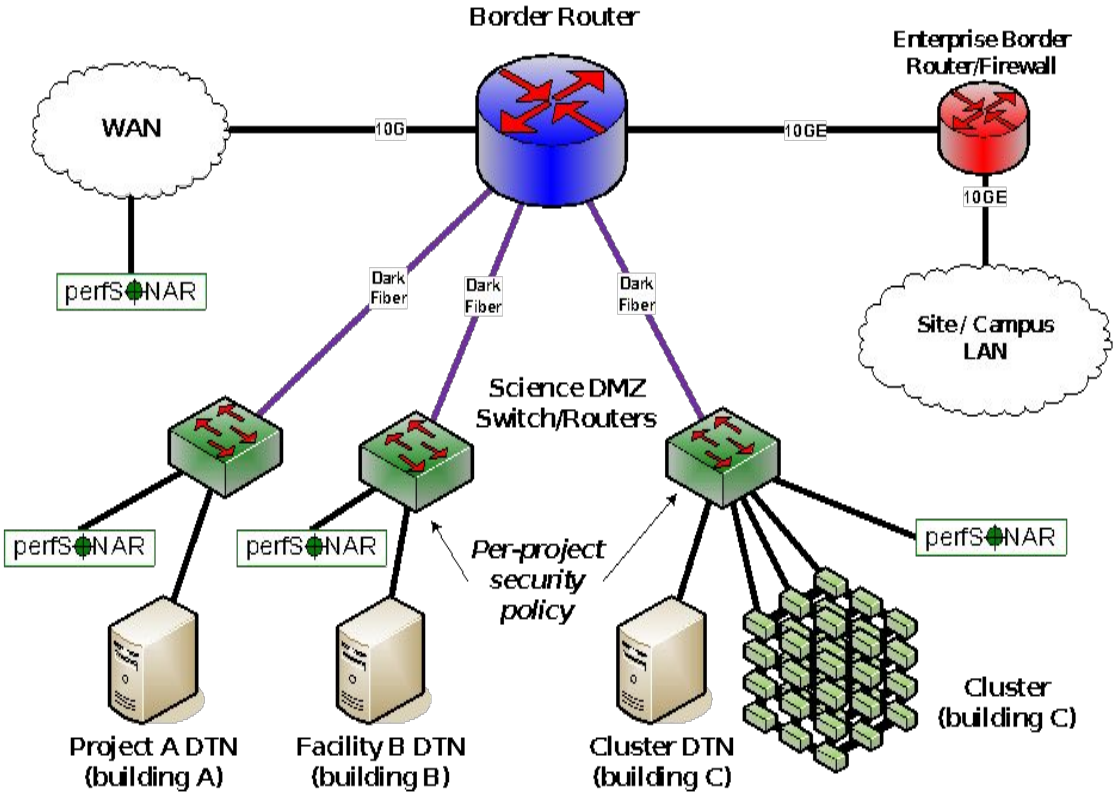
ESnet

Distributed Science DMZ – Dark Fiber to Instrument



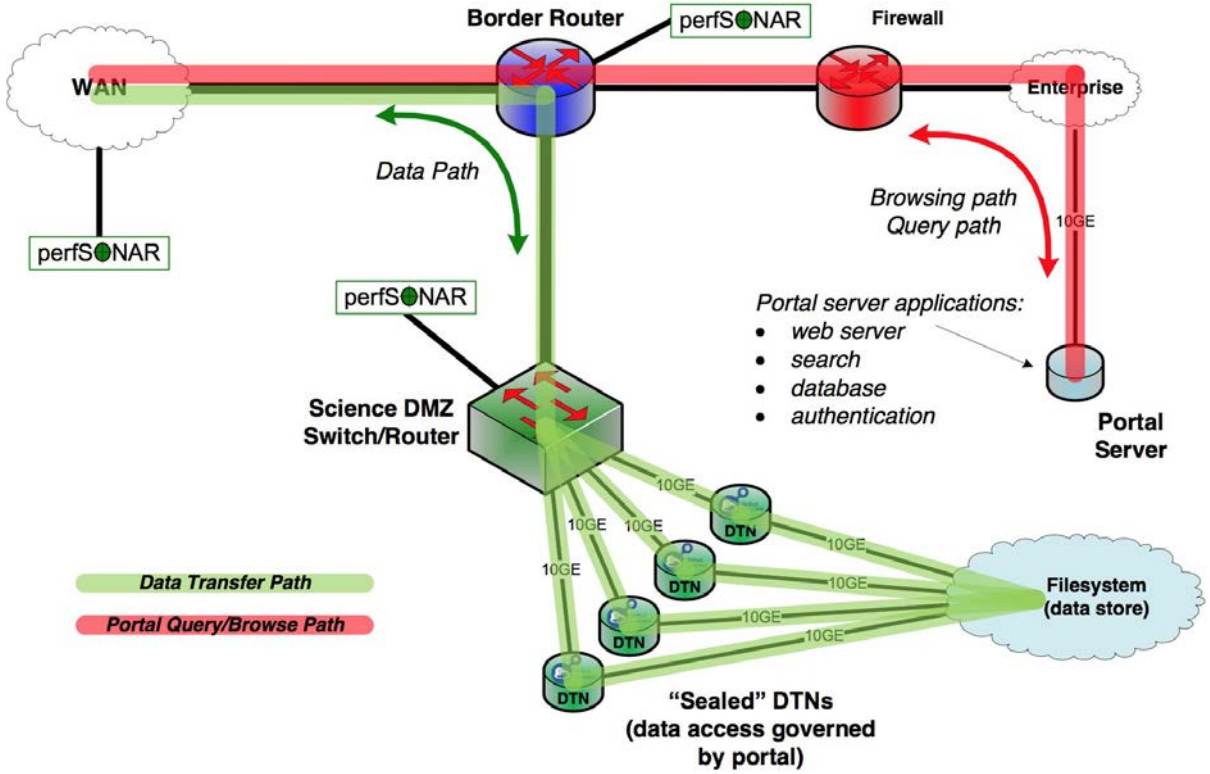
ESnet

Multiple Science DMZs – Dark Fiber to Dedicated Switches



ESnet

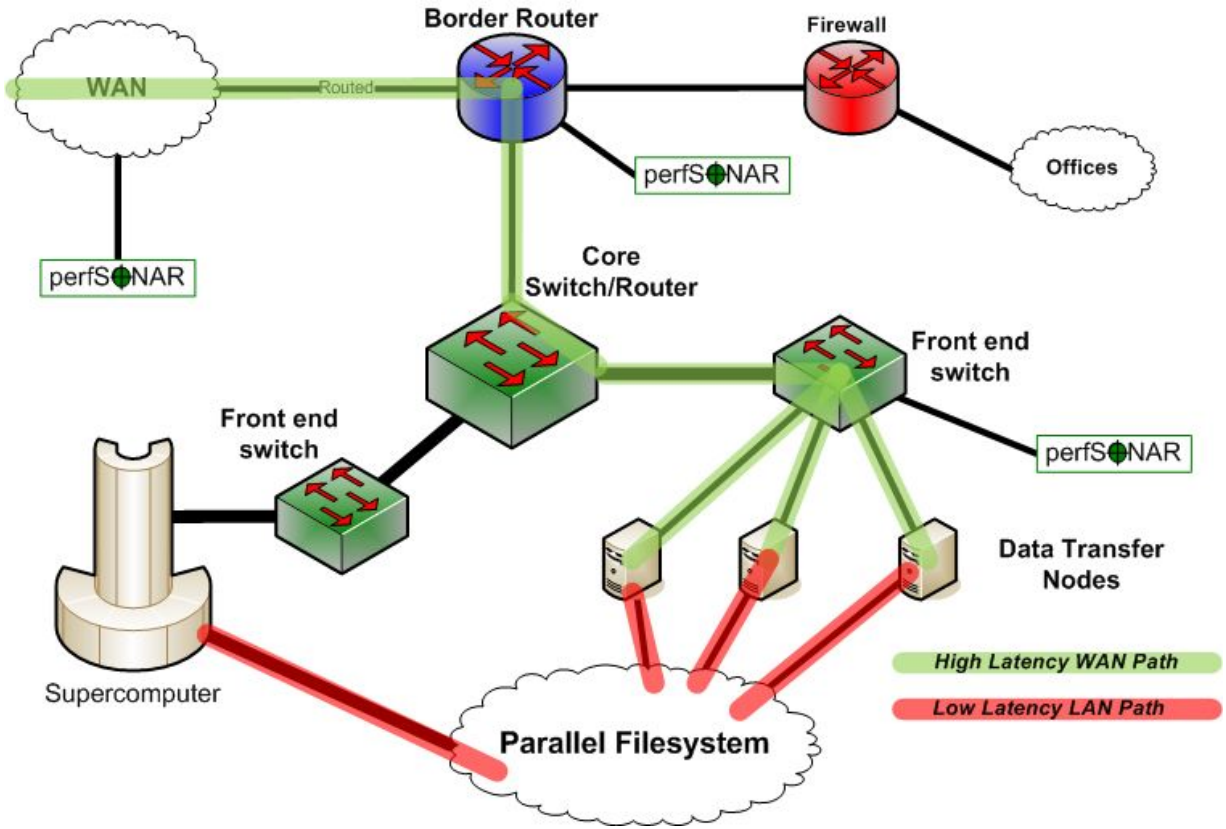
Modern Research Data Portal Leverages Science DMZ



https://peerj.com/articles/cs-144/

ESnet

HPC Center DTN Cluster Uses Science DMZ



ESnet

Data Mobility Improvement steps

- Poor performance Use Cases were baselined and selected for improvement
- Designed a Science DMZ and purchase equipment
- Network tuning with Installed perfSONAR at border
  - Tested/improved network performance
  - perfSONAR MadDash is green = Network is fixed
- You are done? Nope! **Has the Science improved?** Maybe?
- What speed did the researcher get with their data transfers? Now? What should they get?
- How do we test DTNs performance like we testing the network with perfSONAR?

ESnet

Next Steps – Building On The Science DMZ

- Enhanced cyberinfrastructure substrate exists and it works
  - Wide area networks (ESnet, Internet2, Regionals)
  - Science DMZs connected to those networks
  - DTNs in the Science DMZs
- What does the scientist see?
  - Scientist sees a science application
    - Data transfer
    - Data portal
    - Data analysis
  - Science applications are the user interface to networks and DMZs
- Large-scale data-intensive science requires that we build larger structures on top of those components

ESnet

Goal: High Performance DTNs Everywhere

- Every major site, cluster, and storage system should be able to do this
- Many DTNs are not configured correctly
- How can we help?
  - Standard benchmark data sets
  - High performance sources
  - Multiple community locations
- PetaScale DTN provided a baseline and improvement for national HPC centers
- Data Mobility Exhibition (DME) provided testing sequence and an understanding at the campus level
  - **Faster Data Mobility Framework provides an upgraded platform to baseline campus DTNs with Globus and expectations to reach**

ESnet

Performance At Different Data Scales

Data set size		1 Minute	5 Minutes	20 Minutes	1 Hour
		Time to transfer			
10PB		1,333.33 Tbps	266.67 Tbps	66.67 Tbps	22.22 Tbps
1PB		133.33 Tbps	26.67 Tbps	6.67 Tbps	2.22 Tbps
100TB	> 100Gbps	13.33 Tbps	2.67 Tbps		
10TB		1.33 Tbps	266.67 Gbps		
1TB		133.33 Gbps	26.67 Gbps	6.67 Gbps	2.22 Gbps
100GB	100Gbps	13.33 Gbps	2.67 Gbps	666.67 Mbps	222.22 Mbps
10GB	< 10Gbps	1.33 Gbps	266.67 Mbps	66.67 Mbps	22.22 Mbps
1GB		133.33 Mbps	26.67 Mbps	6.67 Mbps	2.22 Mbps
100MB	< 100Mbps	13.33 Mbps	2.67 Mbps	0.67 Mbps	0.22 Mbps

10G DTN

10G DTN min

- This table available at: http://fasterdata.es.net/fasterdata-home/requirements-and-expectations/

ESnet

Throughput required to move Y bytes in X time

Bits per second throughput

Data set size	1 Minute	5 Minutes	20 Minutes	1 Hour
10PB	1,333.33 Tbps	266.67 Tbps	66.67 Tbps	22.22 Tbps
1PB	133.33 Tbps	26.67 Tbps	6.67 Tbps	2.22 Tbps
100TB	13.33 Tbps	2.67 Tbps	666.67 Gbps	222.22 Gbps
10TB	1.33 Tbps	266.67 Gbps	66.67 Gbps	22.22 Gbps
1TB	133.33 Gbps	26.67 Gbps	6.67 Gbps	2.22 Gbps
100GB	13.33 Gbps	2.67 Gbps	666.67 Mbps	222.22 Mbps
10GB	1.33 Gbps	266.67 Mbps	66.67 Mbps	22.22 Mbps
1GB	133.33 Mbps	26.67 Mbps	6.67 Mbps	2.22 Mbps
100MB	13.33 Mbps	2.67 Mbps	0.67 Mbps	0.22 Mbps

Time to transfer

Data set size	8 Hours	24 Hours	7 Days	30 Days
1XB	277.78 Tbps	92.59 Tbps	13.23 Tbps	3.09 Tbps
100PB	27.78 Tbps	9.26 Tbps	1.32 Tbps	308.64 Gbps
10PB	2.78 Tbps	925.93 Gbps	132.28 Gbps	30.86 Gbps
1PB	277.78 Gbps	92.59 Gbps	13.23 Gbps	3.09 Gbps
100TB	27.78 Gbps	9.26 Gbps	1.32 Gbps	308.64 Mbps
10TB	2.78 Gbps	925.93 Mbps	132.28 Mbps	30.86 Mbps
1TB	277.78 Mbps	92.59 Mbps	13.23 Mbps	3.09 Mbps
100GB	27.78 Mbps	9.26 Mbps	1.32 Mbps	0.31 Mbps
10GB	2.78 Mbps	0.93 Mbps	0.13 Mbps	0.03 Mbps

Time to transfer

ESnet

4 x 250GB files in single directory

move Y bytes in X time

Bits per second throughput

Data set size	1 Minute	5 Minutes	20 Minutes	1 Hour
10PB	1,333.33 Tbps	266.67 Tbps	66.67 Tbps	22.22 Tbps
1PB	133.33 Tbps	26.67 Tbps	6.67 Tbps	2.22 Tbps
100TB	13.33 Tbps	2.67 Tbps	666.67 Gbps	222.22 Gbps
[illegible]	1.33 Tbps	266.67 Gbps	66.67 Gbps	[illegible]
1TB	133.33 Gbps	26.67 Gbps	6.67 Gbps	2.22 Gbps
[illegible]	13.33 Gbps	2.67 Gbps	666.67 Mbps	[illegible]
10GB	1.33 Gbps	266.67 Mbps	66.67 Mbps	22.22 Mbps
1GB	133.33 Mbps	26.67 Mbps	6.67 Mbps	2.22 Mbps
100MB	13.33 Mbps	2.67 Mbps	0.67 Mbps	[illegible]

Time to transfer

Data set size	8 Hours	24 Hours	7 Days	30 Days
1XB	277.78 Tbps	92.59 Tbps	13.23 Tbps	3.09 Tbps
100PB	27.78 Tbps	9.26 Tbps	1.32 Tbps	308.64 Gbps
[illegible]	2.78 Tbps	925.93 Gbps	[illegible]	30.86 Gbps
1PB	[illegible]7.78 Gbps	92.59 Gbps	13.23 Gbps	3.09 Gbps
[illegible]	27.78 Gbps	9.26 Gbps	[illegible]	308.64 Mbps
10TB	2.78 Gbps	925.93 Mbps	132.28 Mbps	30.86 Mbps
1TB	277.78 Mbps	92.59 Mbps	13.23 Mbps	3.09 Mbps
100GB	27.78 Mbps	9.26 Mbps	1.32 Mbps	0.31 Mbps
10GB	2.78 Mbps	0.93 Mbps	[illegible]	0.03 Mbps

Time to transfer

ESnet

FasterData Mobility Framework

- The scientific community cannot address universal problems in data movement for themselves
- Individual researchers do not control the resources
  - Computing centers
  - Data repositories
  - Science networks
  - Our community owns these – we have to do the work
- Science Engagement to teach scientists how to use the better platforms
- This is the path forward and this effort is about visibility and fixing problems

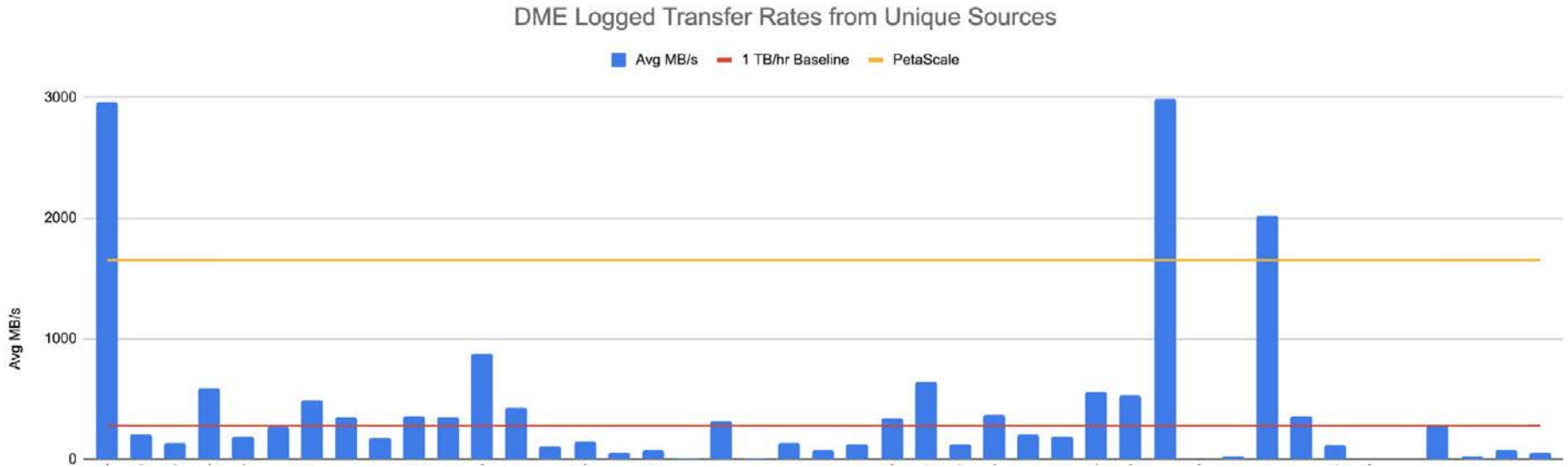
ESnet

FasterData Mobility Framework

- Built from the PetaScale DTN project, 1PB transfer in one week at HPC and Data Mobility Exhibition (DME) 1TB/hr at campuses
- Current and previous NSF CC* Awardees, along with the greater R&E community and DOE sites, are encouraged to participate
- Using reference data sets, and existing campus CI components, participants will work on scientific data movement capabilities:
  - Download/Upload data sets
  - Measure and baseline against 1TB/hr transfer rate or specific requirement, like PetaScale
  - Potentially improve transfer results locally from an end-to-end test

ESnet

Past DME Transfers from Unique Sources average by university



Each bar is the average data transfer rate from a unique site/campus DTN across any and all data sets

ESnet

Past DME Transfers to Unique Destinations

Each bar is the average data transfer rate to a unique site/campus DTN across any and all data sets

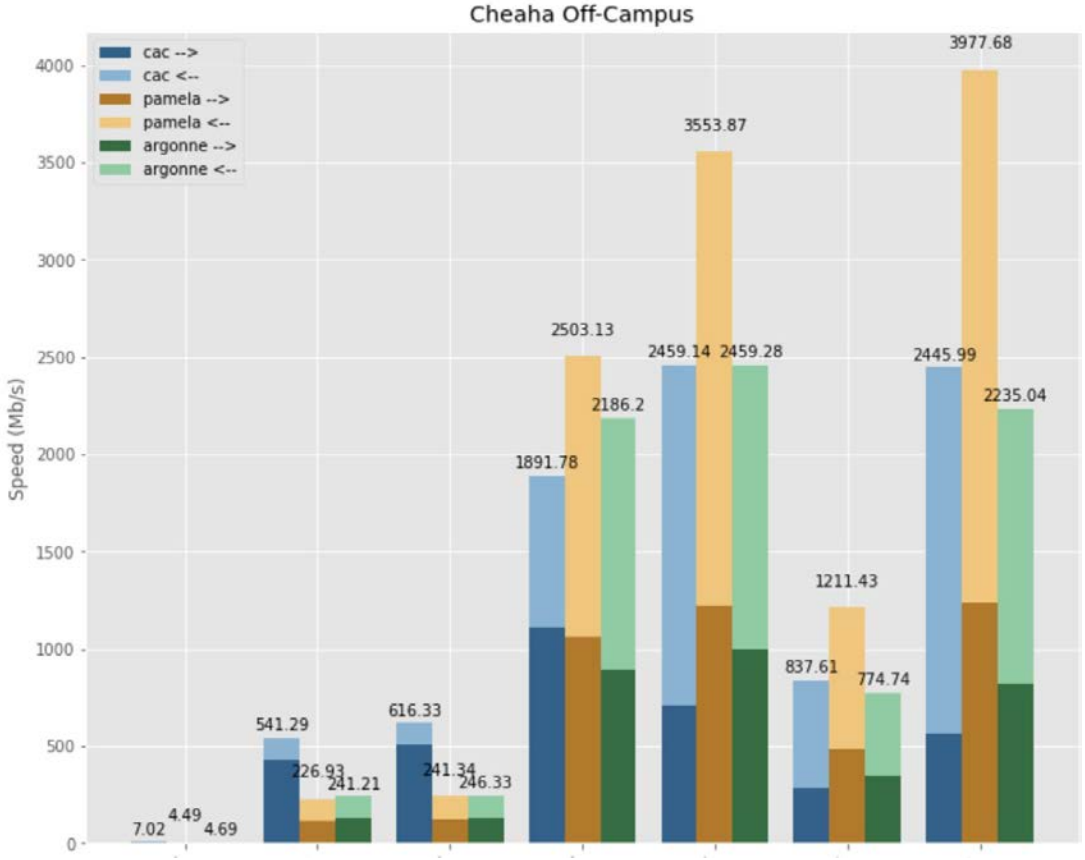
ESnet

DME Results

- Over 11,000 tests
- 2.45 PB transferred
- 52 unique sources uploaded to DME
- 118 Unique destinations downloaded from DME
- 4 site replied with all testing
- 1 site automated testing between outside/Science DMZ and inside the firewall

ESnet

Automated DME tests - Off Campus



ESnet

DME Lessons Learned

- Some sites only test perfSONAR and not data mobility
- Sites still think they need 100G perfSONAR and 100G DTNS
- Testing perfSONAR consistently will save future headaches
- Packet Pacing is good
- Clusters of DTNs scale and perform better

ESnet

DME Data Sets

ds01	100MB	10,000 x 10KB files in single directory
ds04	10GB	10,000 x 1MB files in 100 non-nested directories, 100 files/directory
ds06	100GB	100,000 x 1MB files in single directory
ds08	1TB	50 x 10GB; 350 x 1GB; 1,000 x 100MB; 5,500 x 10MB; 23,176 x 1MB files in single directory
ds10	1TB	100 x 10GB files in single directory
ds12	100GB	1 x 100GB file in single directory
ds14	5TB	50 x 100GB files in single directory
ds16	1TB	4 x 250GB files in single directory

ESnet

DME Data Sets

~~ds01~~	~~100MB~~	~~10,000 x 10KB files in single directory~~
~~ds04~~	~~10GB~~	~~10,000 x 1MB files in 100 non-nested directories, 100 files/directory~~
~~ds06~~	~~100GB~~	~~100,000 x 1MB files in single directory~~
ds08	1TB	50 x 10GB; 350 x 1GB; 1,000 x 100MB; 5,500 x 10MB; 23,176 x 1MB files in single directory
ds10	1TB	100 x 10GB files in single directory
~~ds12~~	~~100GB~~	~~1 x 100GB file in single directory~~
~~ds14~~	~~5TB~~	~~50 x 100GB files in single directory~~
ds16	1TB	4 x 250GB files in single directory

ESnet

Data Mobility Benchmark

- Try to benchmark your DTNs and Data Architectures monthly or after any changes.
- Download ESnet data Climate Data Sets from Wash-DTN1.es.net or another ESnet server to test your write speeds
  - https://app.globus.org/file-manager?origin_id=2a6a759c-5cfe-4402-ac5e-a06d9d7f7c37&origin_path=%2F
  - Climate-Small, ~245GB, 1496 files, 305 folders
  - Climate-Medium, ~245GB, 117 files, 1 folder
  - Climate-Large, ~245GB, 11 files, 1 folder
  - Climate-Huge, ~245GB, 2 files, 1 folder
- For larger systems, try the DME datasets:
  - https://app.globus.org/file-manager?origin_id=5837354e-7087-4d0d-b7bc-e3655f883899&origin_path=%2F
  - ds08, ~1TB, 30076 files, 1 folder
  - ds10, ~1TB, 100 files, 1 folder
  - ds16, ~1TB, 4 files, 1 folder
- Once downloaded, you can re-upload to test your read speeds.

ESnet

Data Transfer Rates by Audience

Host Transfer Rates	⅛ PetaScale (Minimum)	⅓ PetaScale	½ PetaScale		PetaScale: 1 PB/wk	PetaScale: 1 PB/day
	10G Capable DTN				10xG, 25G, 40G, 100G DTNs	
Data Transfer Rate/Volume (Researcher)	1 TB/hr	2 TB/hr	3 TB/hr		5.95 TB/hr	41.67 TB/hr
Network Transfer Rate (Network Admin)	2.22 Gb/s	4.44 Gb/s	6.67 Gb/s		13.23 Gb/s	92.59 Gb/s
Storage Transfer Rate (Sys/Storage Admin)	277.78 MB/s	555.54 MB/s	833.33 MB/s		1.65 GB/s	11.57 GB/s

A benchmark table is provided to gauge data architecture performance, which can vary depending on number of files, folders, size of files, distance between sites, CI performance (network, server, disk/filesystem), as well as data transfer tool.

ESnet

ESnet DTN testing of 1TB

(4 x 250GB files in single directory)

				DEST					
SOURCE	cern	denv	hous	star	sunn	wash		NCAR-GL	NERSC-Perlmutter
cern	X	18.29	18.45	27.10	31.23	27.69		34.48	20.75
denv	35.65	X	37.31	30.28	18.84	28.46		37.51	21.00
hous	35.24	32.93	X	30.47	38.27	28.40		36.47	21.63
star	33.91	32.74	33.92	X	33.99	34.16		27.42	21.80
sunn	35.36	34.72	36.69	30.58	X	29.25		29.21	21.85
wash	17.07	14.08	15.00	19.04	16.72	X		13.87	14.40
NCAR	28.25	28.33	37.86	26.29	36.85	29.25		X	17.48
NERSC-Perlmuter	29.75	25.99	25.73	27.00	29.23	25.08		25.99	X

ESnet

Other ESnet Data transfer developments:

- EScp
  - A Transfer Tool For Collaborative Science

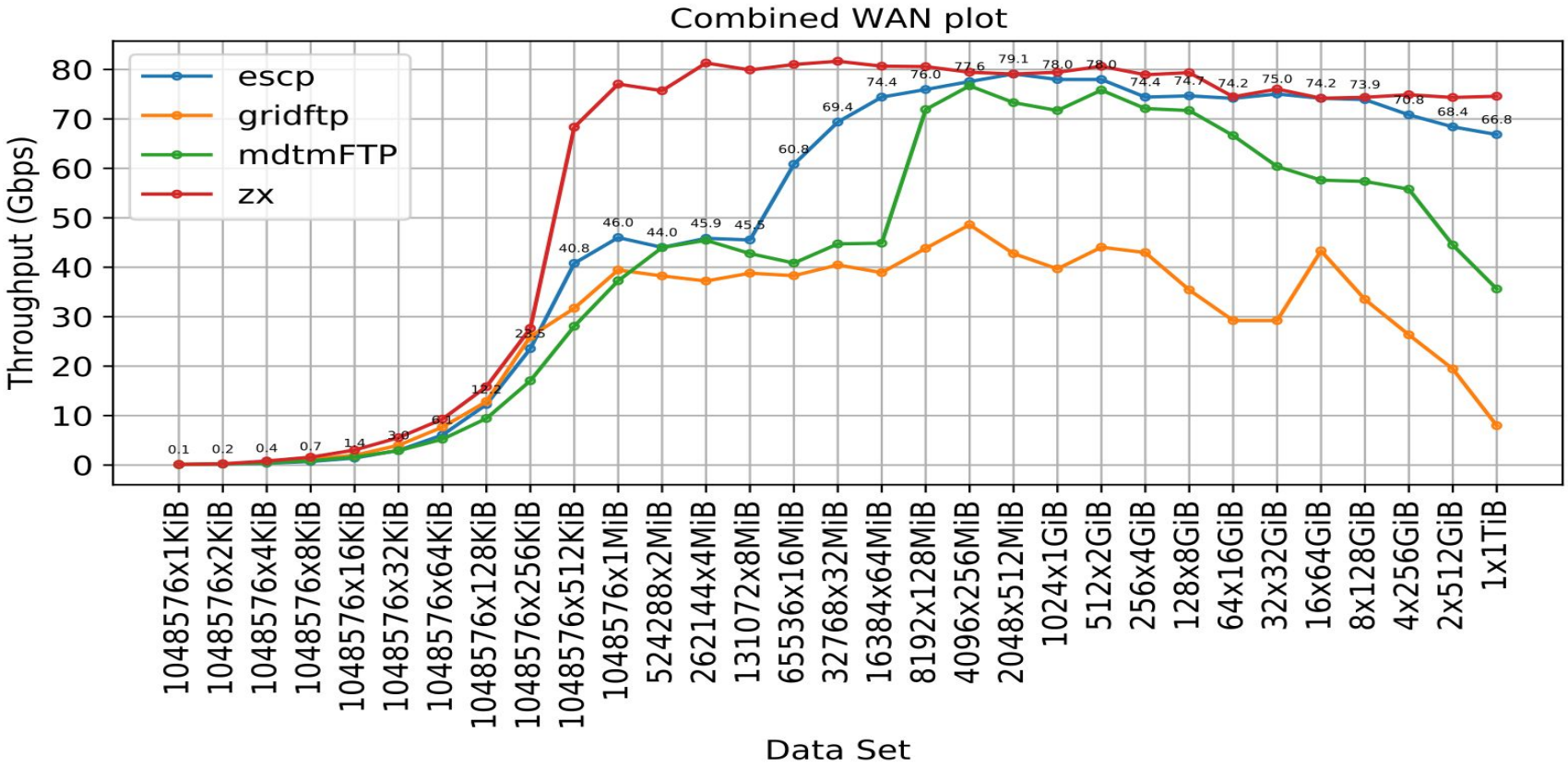
Background on EScp

- First used as a tool to help performance testing on our DTNs, called DTN tool;
  - Consistency of software allowed us to measure tuning parameters, including IPv6
- Added Python wrapper around DTN tool to allow it to emulate SCP functionality
- Released as an open source tool in August 2021

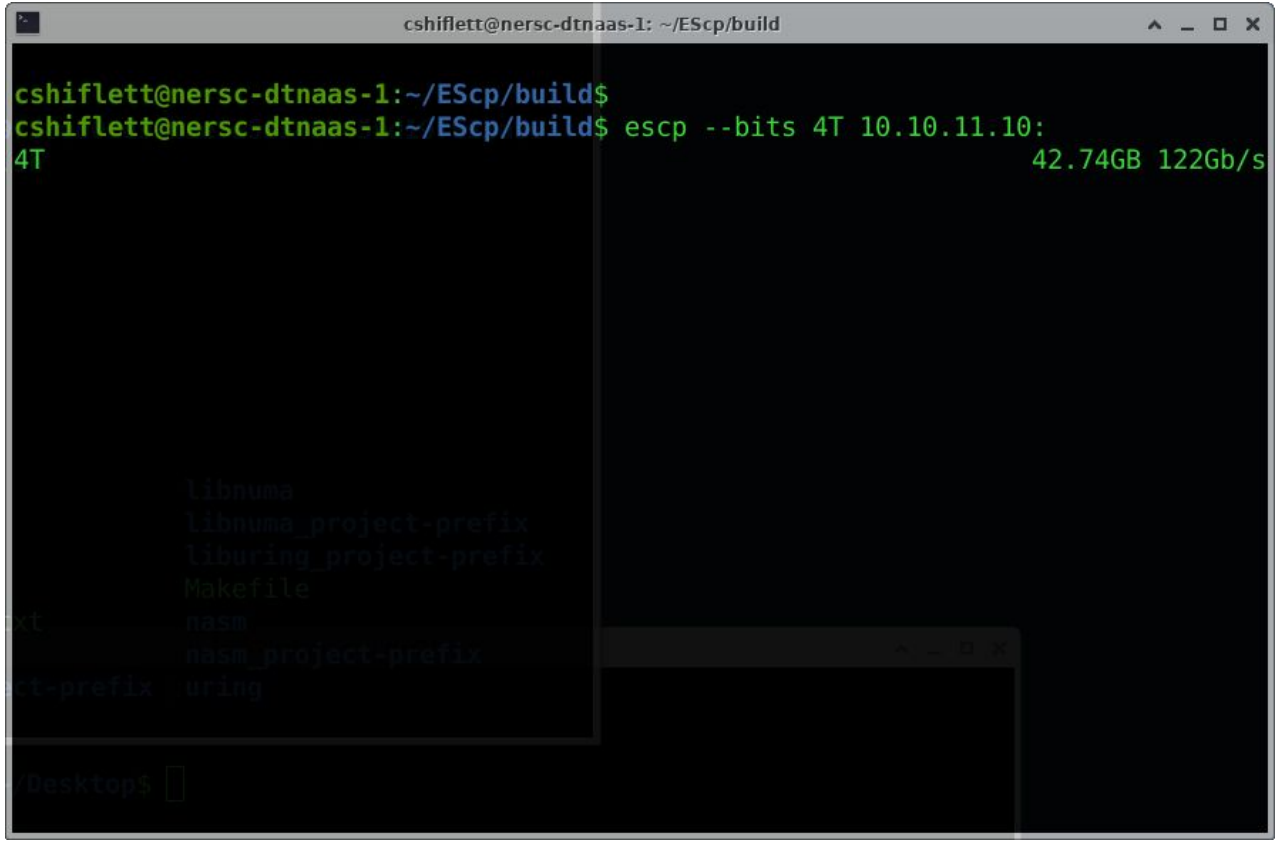
ESnet

Background on EScp (Continued)

- EScp wrapper combined with DTN tool creates a modern transfer tool
  - Encrypted
  - Checksummed
  - Block / Cache Aligned
  - Zero Copy
  - Versioned
  - Disk-to-Disk performance > 100gbit/s


Combined WAN plot
escp
gridftp
mdtmFTP
zx
Throughput (Gbps)
Data Set
0
10
20
30
40
50
60
70
80
0.1
0.2
0.4
0.7
1.4
3.0
6.1
12.2
28.5
40.8
46.0
44.0
45.9
45.5
60.8
69.4
74.4
76.0
77.6
79.1
78.0
78.0
74.4
74.7
74.2
75.0
74.2
73.9
70.8
68.4
66.8
1048576x1KiB
1048576x2KiB
1048576x4KiB
1048576x8KiB
1048576x16KiB
1048576x32KiB
1048576x64KiB
1048576x128KiB
1048576x256KiB
1048576x512KiB
1048576x1MiB
524288x2MiB
262144x4MiB
131072x8MiB
65536x16MiB
32768x32MiB
16384x64MiB
8192x128MiB
4096x256MiB
2048x512MiB
1024x1GiB
512x2GiB
256x4GiB
128x8GiB
64x16GiB
32x32GiB
16x64GiB
8x128GiB
4x256GiB
2x512GiB
1x1TiB

ESnet



- Screenshot of a transfer in progress

ESnet

Why EScp?

- Natural extension of ESnet efforts to increase science engagement.
  - Software and API's typically largest impediment to performance
- No high performance replacement currently available to replace SCP
- Science community needs a flexible tool to perform high speed file/block based transfers

ESnet

Current state of EScp

- 0.7 Branch should be released on github soon
- Upcoming demos at SC23
- Replaces Python Wrapper w/ Rust
  - Single statically linked executable
  - Debug using standard tools (gdb)
  - Rust calls C function directly which eliminates much of the Python Weirdness
  - Should be significantly faster w/ small files

Current state of EScp (continued)

- Fully lockless and optimized for high concurrency
- Ad-hoc release schedule
  - Developed in spare time and significantly affected by inf team workload
    - I also see it as a way of engaging with the scientific community and continued learning (i.e. Rust was new to me for this project).
- Many additional features planned; Mostly suggestions from ESnet or ESnet community.

ESnet

EScp Links:

- https://github.com/esnet/EScp
- mailto: cshiflett@es.net

ESnet

Other ESnet Data transfer developments:

- ~~EScp~~
  - ~~A Transfer Tool For Collaborative Science~~
- Wait 1TB/hr or 1 TB/minute?

ESnet

TestLab Data Transfer performance 1TB/min?

[0.0-2.0 sec]	29.92 GB	119.67 Gb/s
[2.0-4.0 sec]	32.12 GB	128.49 Gb/s
[4.0-6.0 sec]	35.36 GB	141.42 Gb/s
[6.0-8.0 sec]	43.22 GB	172.86 Gb/s
[8.0-10.0 sec]	44.83 GB	179.32 Gb/s
[10.0-12.0 sec]	43.38 GB	173.50 Gb/s
[12.0-14.0 sec]	44.83 GB	179.32 Gb/s
[14.0-16.0 sec]	45.34 GB	181.34 Gb/s
[16.0-18.0 sec]	45.70 GB	182.78 Gb/s
[18.0-20.0 sec]	46.62 GB	186.49 Gb/s
[20.0-22.0 sec]	47.00 GB	188.00 Gb/s
[22.0-24.0 sec]	47.35 GB	189.41 Gb/s
[24.0-26.0 sec]	48.07 GB	192.26 Gb/s
[26.0-28.0 sec]	48.05 GB	192.17 Gb/s
[28.0-30.0 sec]	48.08 GB	192.29 Gb/s
…		
[0.0-60.6 sec]	1385.80 GB	182.89 Gb/s bytes: 1385802235904

ESnet

Other ESnet Data transfer developments:

- ~~EScp~~
  - ~~A Transfer Tool For Collaborative Science~~
- ~~Wait 1TB/hr or 1 TB/minute?~~
- BBRv3 - testing ongoing

ESnet

Other ESnet Data transfer developments:

- ~~EScp~~
  - ~~A Transfer Tool For Collaborative Science~~
- ~~Wait 1TB/hr or 1 TB/minute?~~
- ~~BBRv3 - testing ongoing~~
- iperf3-mt

ESnet

FasterData Mobility Framework Servers

- ESnet locations
  - Sunnyvale, Ca
  - Chicago, IL
  - Washington, DC
  - Denver, CO
  - Houston, TX
  - CERN
- Previous DME test sites at Argonne and NCAR
- Collaborating with Globus and other groups like
  - Internet2 (Kansas)
  - TACC
  - GEANT

ESnet

Efforts like FasterData Mobility Framework… A rising tide lifts all boats

- Petascale DTN, Data Mobility Exhibition, and FasterDate Mobility Framework projects benefit all projects which use the HPC and campus DTNs
- Modern science data portal architecture
  - Data portals which use modern architecture benefit from DTN improvements
  - DTN scaling/improvements benefit all data portals which use the same pool
- Globus API supports this – see Globus World Tour
  - https://www.globusworld.org/tour/

- How would persistent testing benefit your campus?

ESnet

Possible Science Engagement Questions?

- What speeds are the Border Top 10 Source/Destination reporting?
  - https://tacc.netsage.io/grafana/d/xk26IFhmk/flow-data-for-circuits?orgId=1
- What is slow? Access, using, or gathering data from a resource?
- Are researchers mailing hard drives?
- What are researchers reporting as slow?
- What are researcher's expectations of a data transfer?
- What are IT's expectations of a data transfer?
- How do we test data transfer?

ESnet

FasterData Mobility Framework testing:

- https://fasterdata.es.net/DTN/
- https://fasterdata.es.net/DTN/data-transfer-scorecard/
- https://fasterdata.es.net/performance-testing/DTNs/

- For those that want to accelerate their campus results, 1:1 assistance with Engagement and Performance Operations Center (EPOC) is available: epoc@tacc.utexas.edu
-
- For any other questions: ken@es.net

ESnet