U.S. DEPARTMENT of ENERGY

BERKELEY LAB

ESnet
ENERGY SCIENCES NETWORK

Host Tuning Campaign For HL-LHC

Eli Dart
Network Engineer, Science Engagement
dart@es.net

LHCONE 55
Karlsruhe, Germany
8 October 2025

Framing

- We need significant performance increases for HL-LHC
  - But also for DC27
  - How much equipment do sites need to buy?
- Network pipes are being deployed
  - How do we use the network well?
  - Some of this is network config - a lot is host system tuning
- Several different reasons to do host tuning
  - Overall system functionality/performance
  - Get the performance we pay for
    - Default config isn't necessarily the best config
    - We could increase performance without buying hardware
  - Potentially reduce scaling pressure on other parts of the WLCG
- This talk is intended to start a conversation
  - I don't have the answers, but the community does

Area of focus: per-file data transfer rate

https://monit-grafana.cern.ch/d/dd6c5e41-6946-47b9-bb1d-50ac89bc7499/user-tunikode-fts-queue-metrics?orgId=11&var-source=davs:%2F%2Fatlasdtn01.usatlas.bnl.gov&var-destination=davs:%2F%2Fantares.stfc.ac.uk&var-source_se=davs:%2F%2Fcmsdcadisk.fnal.gov&var-dest_se=davs:%2F%2Frdr.echo.stfc.ac.uk&tab=query&from=now-30d&to=now&var-bin=1d

Per-file transfer speed

Per-file = throughput / active_count

Timestamp	Decision	Running	Queue	Success rate (last 1min)	Throughput	EMA	Diff	Explanation
2025-03-18T00:26:18Z	1000	278	10305	100.00%	1.88 GiB/s	1.88 GiB/s	0	Good link efficiency. Increment. Hit upper range limit
2025-03-18T00:25:39Z	1000	268	10352	100.00%	1.77 GiB/s	1.77 GiB/s	0	Good link efficiency. Increment. Hit upper range limit
2025-03-18T00:25:01Z	1000	277	10359	100.00%	1.77 GiB/s	1.77 GiB/s	0	Good link efficiency. Increment. Hit upper range limit

Time	data.source_se	data.dest_se	data.active_count	data.throughput
Mar 4, 2025 @ 15:29:30.359	davs://eoscms.cern.ch	davs://rdr.echo.stfc.ac.uk	766	4,736,329,630.439
Mar 4, 2025 @ 15:28:54.315	davs://eoscms.cern.ch	davs://rdr.echo.stfc.ac.uk	805	4,768,777,080.548
Mar 4, 2025 @ 15:28:17.267	davs://eoscms.cern.ch	davs://rdr.echo.stfc.ac.uk	841	4,741,144,937.96
Mar 4, 2025 @ 15:27:39.228	davs://eoscms.cern.ch	davs://rdr.echo.stfc.ac.uk	8[illegible]	4,752,035,[illegible]

Source	Avg per-file rate
CNAF T1	4 MB/s
IN2P3 T1	8 MB/s
FNAL T1	6 MB/s
Caltech	3 MB/s
Nebraska	3 MB/s

24

Key question: what rate should we see?

Key question: what are the implications for changing this?

Slide credit: Katy Ellis, RAL, LHCONE54

ESnet

Host interface to the network

- Fasterdata has TCP configuration
  - fasterdata.es.net
  - This is useful, but only if the host networking is the bottleneck
- DTN TCP stack is only part of the picture
  - TCP config doesn't affect disk controllers
  - If packet loss is limiting performance, bigger buffers won't help
- TCP config is important, but only part of the picture
  - Valuable to have a recommendation for WLCG sites
  - But: tuning TCP may not increase performance on its own

Storage architecture

- Multiple different storage systems in WLCG
  - Some use Ceph, some GPFS, there are others as well
  - Useful to identify best practice for commonly deployed designs
- Interaction with hard drives is important
  - A single disk spindle can do about 100MB/sec to 500MB/sec
  - Single SSD/NVMe device is much faster
  - Some storage designs assign a file to one disk
  - Some storage designs stripe files across multiple disks
- It is likely that host internals are much faster than hard drive spindles - internal performance mismatch
- Key need: understand stored files per spindle vs. concurrent files in flight

Host interface to storage

- Do the DTNs have their own internal disks?
- Is the storage SSD/NVMe?
- Does the data transfer host mount a network filesystem?
  - NFS
  - Lustre
  - Something else
  - Data transfer hosts and storage hosts may need separate tuning
- Probably many variations in WLCG
  - What are the common designs?
  - What are future plans?

Zoom out

- How does all this affect the larger WLCG?
- If we increase per-file data rate, does that take pressure off of FTS?
  - If files move more quickly, can we have fewer files in flight?
  - Does FTS make assumptions?
  - Does Rucio make assumptions?
- Lots of work is being done on FTS
  - Informed by previous data challenges
  - What will a future FTS look like?
  - How does that affect what data transfer hosts need to do?

How to move forward?

- Is a host tuning campaign worthwhile?
  - We need a lot of performance increase
  - How much can we get without buying hardware?
- Can changes be implemented at scale?
  - How close to defaults is the current config?
  - How willing are sysadmin staff to change defaults?
- What is the right path forward?
  - Start with information gathering?
  - Start with a small number of interested sites?
- What can we do before DC27?

U.S. DEPARTMENT of ENERGY

BERKELEY LAB

ESnet
ENERGY SCIENCES NETWORK

Discussion

Eli Dart

dart@es.net

https://my.es.net/

https://www.es.net/

https://fasterdata.es.net/