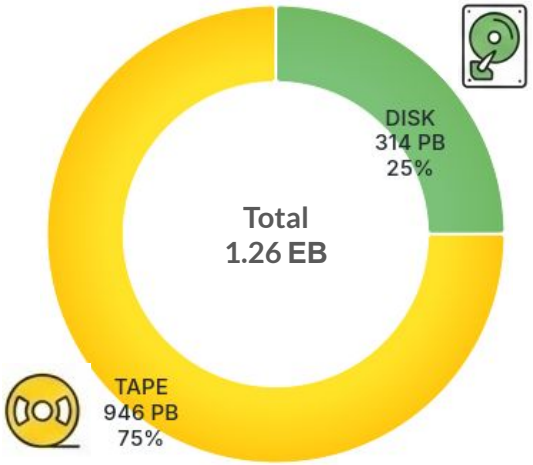
CMS Community Talk

Christos Emmanouil, on behalf of CMS Offline & Computing

XRootD & FTS Workshop September 14 – 18, 2026
CNRS Rhône Auvergne Délégation Lyon, France

Space Usage



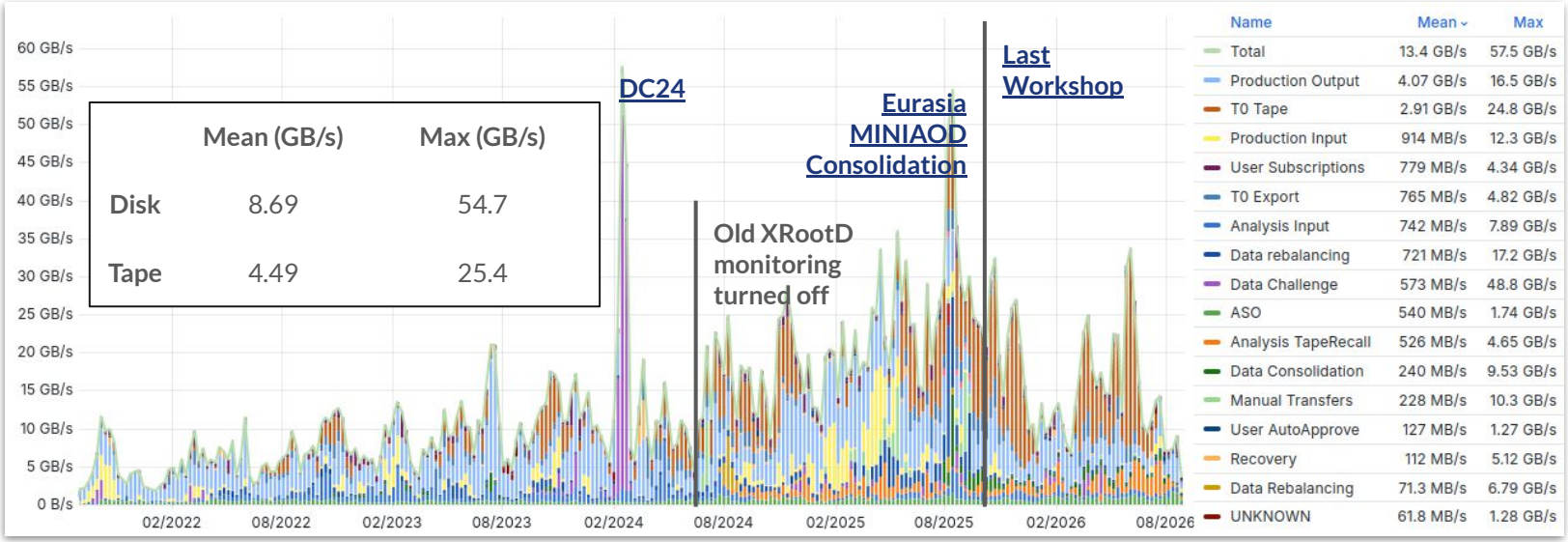
CMS holds ~314 PB on disk and ~946 PB on tape

- Approx. increase of ~47 PB on disk and ~144 PB on tape since the last workshop!
- Run3 data-taking and continuous MC (simulated data) production in parallel.

Run3 FTS Throughput

CMS currently relies on 2 FTS instances:

- **Americas** use FNAL, fallback CERN
- **Eurasia** use CERN, fallback FNAL



Last 5 years, 1 week binning, by activity

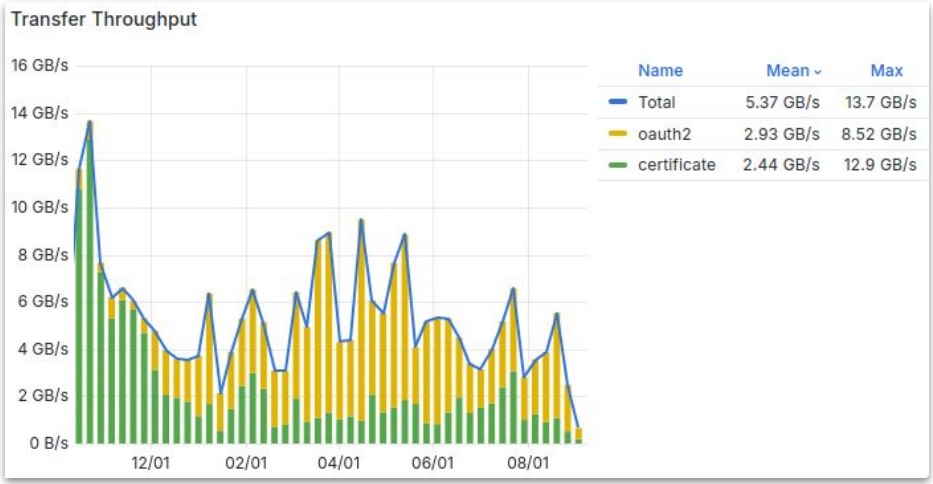
Moving Toward HL-LHC

- Overall, FTS3 has worked well for CMS.
- **DC24 (25% HL-LHC scale):** Pushed FTS3 to its limits, providing useful insights but only just reached CMS targets.
- **DC27 (50% HL-LHC scale):** Doubles the scale requirements.
- Further work is required to address future needs.

Tokens

DC27 is scoped to evaluate the system by explicitly using tokens.

- Since last workshop, FNAL FTS added token support and token transfers were enabled across most US sites.
- Last month, on average 75% of disk-to-disk transfers were token-based.
- Testing is still needed for 8 T2s and JINR.
  - Couple sites still need to update their EOS/XRootD versions to support tokens.
- CMS looks forward to tokens for tapes.
  - Most technologies in the stuck are ready but testing is required.



Since last workshop, 1 week binning, by authentication method

Larger File Size Requirements

CMS aims to test these features for CSA-28, targeting these requirements for the 2027 development pipeline.

Increased File Sizes
Driven by the data growth and in order to sustain efficiency, CMS plans to transition to 25 GB (mean) and up to 50 GB (max) files in Run 4, with a desire to increase further for Run 5.

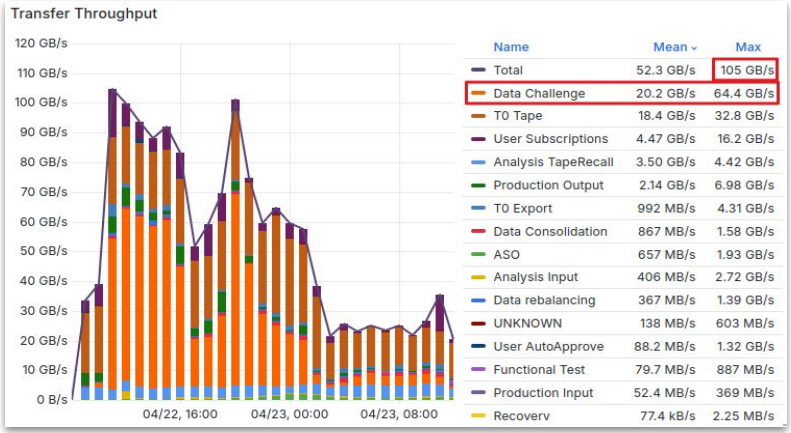
System Adaptations
Supporting these larger files requires key updates potentially across the entire transfer stack – in Rucio, in FTS and the underlying protocols (e.g. WebDAV):

- Implementation of file-size-dependent timeouts.
- Support for multi-stream transfers.
- The ability to resume incomplete transfers.

Experience with FTS4

Data Challenge is the activity used for testing.

- CMS started testing FTS4 at the end of March.
- Verified functionality for a single link (T2_CH_CERN → T1_DE_KIT_Disk ~350 MB/s).
- Achieved a maximum rate of 64 GB/s across 90 links and verified functionality with different disk technologies while pushing FTS to ~8K concurrent transfers and ~400K queued requests.



1 hour binning, by activity

Experience with FTS4

Priorities

- Priorities can be useful in critical situations, when a specific request should be accelerated.
- Confirmed that a Rucio rule with a higher priority number will place files in front of a single FTS queue.

In FTS, transfer requests are handled by queues. A separate queue is defined for every unique combination of VO, source, destination and activity.

Lower Priority Rule

created_at	Fri, 24 Apr 2026 07:04:19 UTC
locks_ok_cnt	1800
priority	3
rse_expression	T1_DE_KIT_Disk_Test
source_replica_expression	T1_IT_CNAF_Disk
state	OK
updated_at	Fri, 24 Apr 2026 07:26:33 UTC

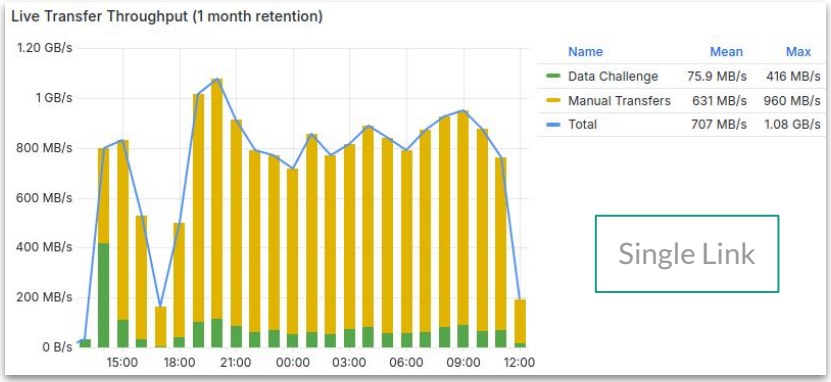
Higher Priority Rule

created_at	Fri, 24 Apr 2026 07:09:10 UTC
locks_ok_cnt	2995
priority	5
rse_expression	T1_DE_KIT_Disk_Test
source_replica_expression	T1_IT_CNAF_Disk
state	OK
updated_at	Fri, 24 Apr 2026 07:26:03 UTC

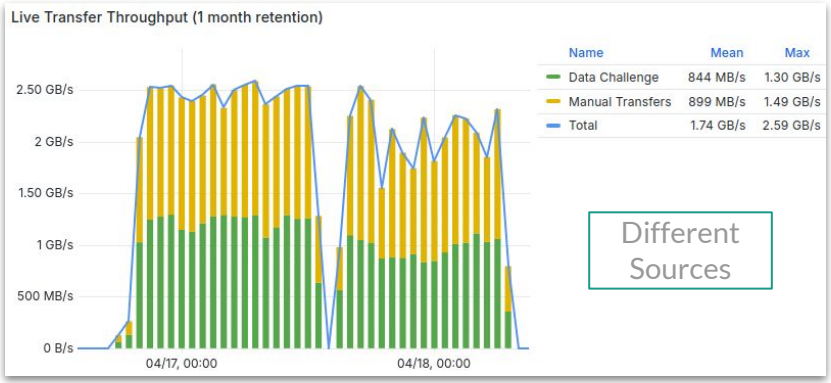
Submitted in the same link with the same activity, 5 min later with a higher number of files and priority, and completed half a minute earlier.

Experience with FTS4

Data Challenge share set to **0.1**
Manual Transfers Share set to **0.9**



Single Link



Different Sources

Activity Shares

- Shares are relative weights that ensure that important activities are not suppressed when a link is saturated.
- Each VO should define a default share which will be split among unregistered activities.
- Activity shares are not respected when the FTS scheduler selects which links will process; under high load, higher-share activities can still be suppressed by competing lower-share activities from different sources.
  - Is this the desired behaviour?

Testing Roadmap

- As of September, CMS activated FTS4 in production for 14 sites.
  - FTS4 was disabled for staging through a patch provided by ATLAS.
  - Usage of FTS4 was reverted due to a critical bug reported by the FTS team that could cause permanent data loss – CMS survived without issues and waits for updates.
    - Would have been possible to detect and prevent the incident?
- While 64 GB/s was a successful benchmark, further scale testing is required before DC27.
- Further testing of FTS4 features.
  - **Data Challenge Mode:** the scheduler favors faster links.
  - **Tape Factor:** the scheduler favors tape transfer; prevents staging issues (e.g. garbage collection).
  - Changing the activity of an already queued transfer to expedite urgent data movements (FTS-2425).

Tracked FTS Developments

These are items CMS knows the FTS team is actively tracking.

- **The New Optimizer (FTS-2426)**
  Replacing the FTS3 optimizer with a dynamic control loop that actively auto-tunes concurrent transfers while respecting endpoint limits to eliminate manual configuration.

- **Integrity-Based Overwrites (FTS-2120)**
  A refinement of the "overwrite when only on disk" feature to perform an integrity check, ensuring only truly corrupted files are overwritten.

- **Throughput-Based Shares (FTS-2427)**
  Transitions activity sharing from static transfer slots to real-time throughput to simplify production verification and ensure fair link utilization across varying file sizes.

Past CMS Requests to FTS

These are items CMS mentioned in the past but the status is unclear.

Priority Requests

- **Detailed WebDAV Errors**
  Restore granular error diagnostics degraded since WebDAV migration.
- **Extended Log Retention**
  Retain logs longer and share combined buffer ↔ tape logs.
- **Multi-Hop File Cleanup**
  Ensure deletion of intermediate files across multi-hop transfers.
- **Tape Archive Limit**
  Mechanism to manage tape pressure based on buffer occupancy.
- **Large-File Support**
  Timeout, multi-stream, and resume support for the 25–50 GB files planned for Run 4 (see slide 7).

Additional Ideas

- **Configuration Management**
  Track config changes and manage multiple FTS instances.
- **File-Based Search**
  Enable direct job searching in the Web UI using filenames.
- **Activity-Aware Link Selection**
  A mechanism that respects activity shares in link selection (see slide 9).
- **Federated Insights**
  Share operational insights across different VO FTS instances.
- **AI Integration**
  Explore AI-driven abilities like anomaly detection and debugging assistance.

Note: Items sourced from baseline input presentations CMS input for FTS4 @ WLCG Open Technical Forum (OTF) #5 and CMS community talk @ XRootD & FTS Workshop 2025

GFAL Decommission

- We need to ensure reproducibility of the actions FTS performs.
  - We often have to ask today: "What are the exact gfal commands FTS is executing?"
- Any GFAL replacement should take this into account.
  - Reproducibility of FTS transfers is crucial for debugging, enabling site admins and operations teams to investigate issues without relying on bringing FTS developers into the loop.
- This need highlights why a dedicated transfers debugging tool is needed following the GFAL decommission.

Summary

Actions are needed to reach HL-LHC targets and FTS4 will help achieve these goals by resolving key limitations of its predecessor.

- FTS3 has worked well for CMS.
- Significant progress on token usage has been made since the last workshop and we look forward to token support for tape transfers.
- Growing file sizes and rising demand toward DC27 and HL-LHC highlight the need for further development across the transfer stack.
- Further testing is needed for FTS4 to push system limits and validate new features. FTS feedback on our outstanding requests is very welcome.
- CMS looks forward to future improvements and remains committed to providing feedback.

Thank you!