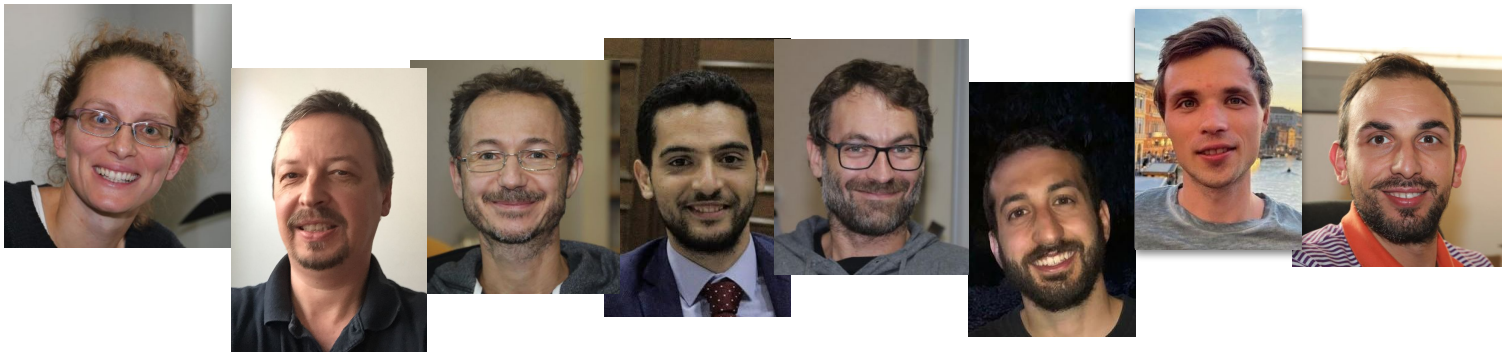
State of Storage

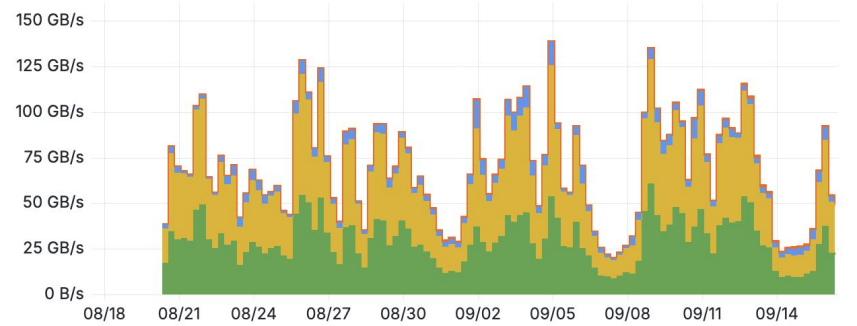
CdG 19 Settembre, 2025



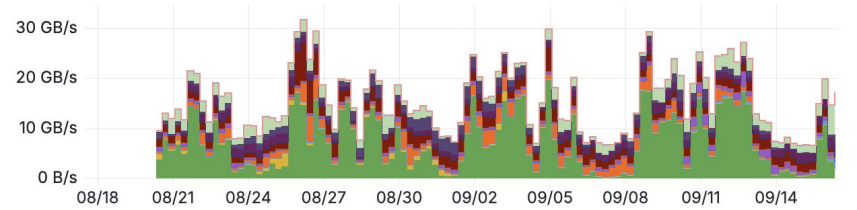
Back to Business as usual

Last month

All servers network traffic out (reading)

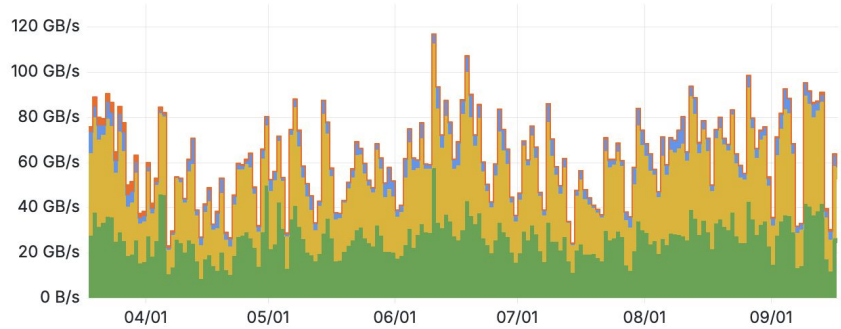


Gateway traffic out (non POSIX reading)

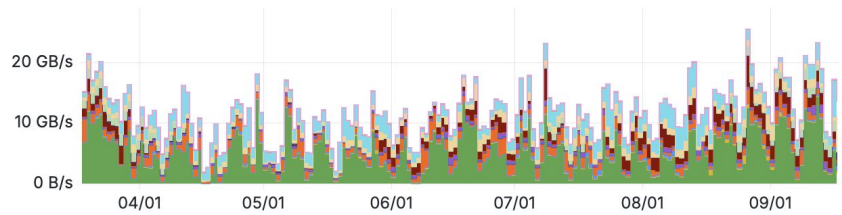


Last 6 months

All servers network traffic out (reading)



Gateway traffic out (non POSIX reading)



Disk storage in production **96.3**PB

Installed: **100.1**PB; Pledge 2025: **101**PB; Used: **77**PB

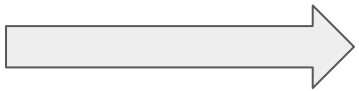
Storage system	Model	Net capacity, TB	Experiment	End of support
os5k8-1,os5k8-2	Huawei OS5800v5	8999	GR2	2027
ddn-12	DDN SFA 7990	4550	LHCb	2025
ddn-14, ddn-15	DDN SFA 2000NV (NVMe)	24	LHCb	2025
ddn-16	DDN SFA 2000NVX2 (NVMe)	96	LHCb metadata, hotdata	2031
ddn-17,ddn-18,ddn-19	DDN SFA 7990X	14000	LHCb	2031
od1k6-1,2,3,4,5,6	Huawei OD1600	60000(-10%!)	ALICE,ATLAS,CMS, GR2	2031
od1k6-7,8	Huawei OD1600	18000	CMS, GR2	2031
od1k5-1,2	Huawei OD1500 (NVMe)	400	Metadati, varie buffer	2031

Use of NVMe disks to improve stage-out

- Traditional (rotating) disks are not no longer suitable for some workflows
  - IO to tapes (new tape drives are capable of doing 400MB/s)
- Allocated 80TB of NVMe storage for each LHC exp (on a shared storage system) as disk cache (staging area for data going to tape)
- Defined “placement policy” to write data going to tape first to NVMe pool
- Defined migration policy to NOT keep data on buffer after stage-out to tape
  - “Migrate” instead of “premigrate” in terms of HSM
  - This is not the case for ALICE, as they re-read data from the buffer to calculate checksums.
- Data staged-in from tape are placed on HDD storage pool (as before)
- Significant increase in stage-out rate
  - We are able to get almost 400MB/s (the max rate for these tape drives)
- **To achieve this, we need experiments to control the data flow based on the available space in the NVMe buffer.**

Performance considerations

- Numbers requested at the tenders:
  - Streaming Reads and Writes (unidirectional) **3.5 MB/s/TB**
  - “Random” reads and writes 1MB blocks (simultaneous 50%+50%) **1.75 MB/s/TB** in ea direction.
- What we can expect for a FS of 18PB (LHCb case)
  - 63 GB/s only read or only write
  - 31.5 GB/s read and 31.5GB/s write.
- Optimal performance with up to 90% of FS occupation.
- Tape drive performance (IBM TS1177): 400 MB/s max
- To make 1.7 GB/s to tape we need
  - **5 tape drives** and **1000 TB** of cache on HDD (1.7 GB/s / 1.7MB/s/TB)
- But we can do it with much smaller buffer (NVMe)
  - 200 TB would be optimal even considering higher costs of NVMe disk (x3 of HDD)
  - 80 TB on Nvme (as we have now) filled up in 13 hours → 2 re-writes per day.
  - 13 hours of autonomy.

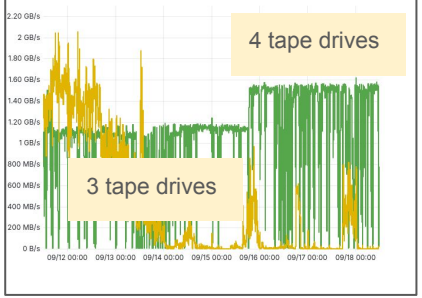


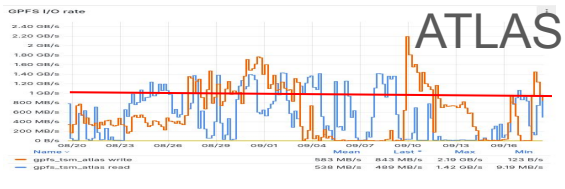
NVMe buffer and data stage-out to tape

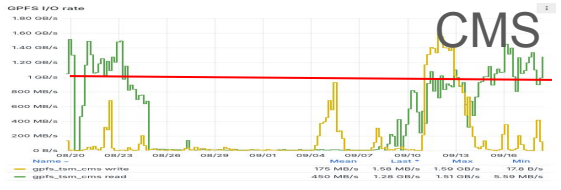
Rate of stage-in (recall, tape->disk cache) and stage-out (migrations, disk cache->tape) and usage of NVMe buffer.

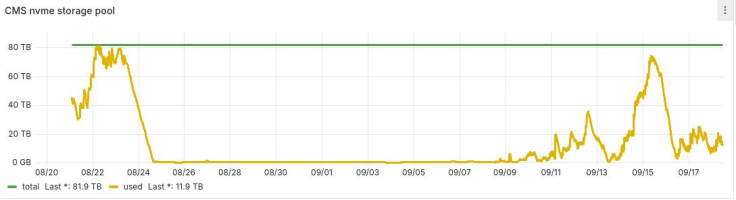
(Red line indicates rate of 1GB/s)

Rate of migrations with 3 and 4 tape drives (CMS)



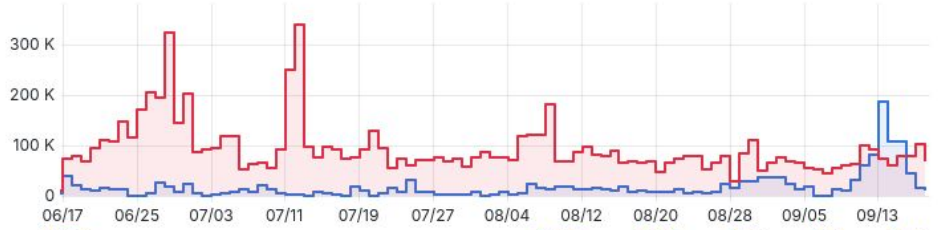






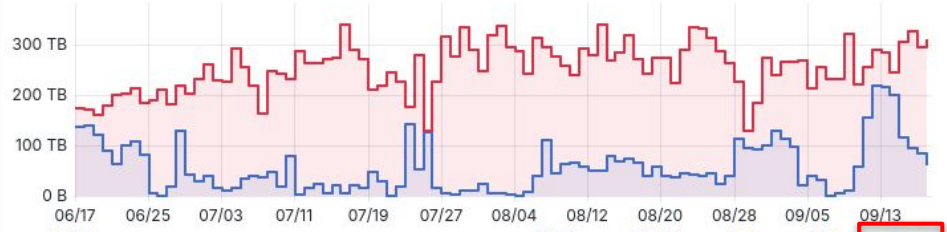
State of tape (last 3 months)

MSS files in/out (per day)



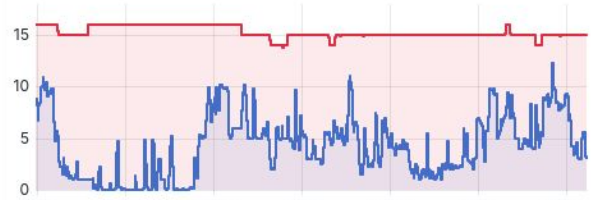
Name	Mean	Last *	Max	Min	Total
out files (recalls)	17.3 K	10.1 K	186 K	14	1.65 Mil
in files (migrations)	91.9 K	66.9 K	339 K	9.18 K	8.73 Mil

MSS bytes in/out (per day)



Name	Mean	Last *	Max	Min	Total
out traffic (recalls)	55.6 TB	60.8 TB	217 TB	44.2 GB	5.22 PB
in traffic (migrations)	253 TB	310 TB	339 TB	129 TB	23.8 PB

T10kD drive in use vs. active



Name	Mean	Last *	Max	Min	Total
drives in use	4.24	3	12.3	0	2383
drives active	15.3	15	16	13.8	8582

TS1160 drive in use vs. active

Name	Mean	Last *	Max	Min	Total
drives in use	6.70	5.94	17.1	0	3768
drives active	18.2	19	19	16.6	10205

TS1170 drive in use vs. active

Name	Mean	Last *	Max	Min	Total
drives in use	13.5	13.0	17.0	7	7609
drives active	17.0	18	18	14.5	9570

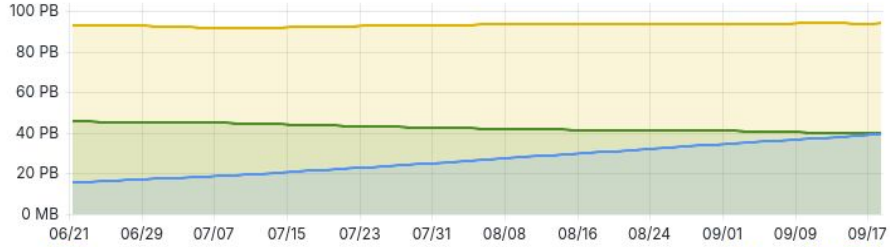
State of tape

- Installed 96 PB (TS4500-2) in June
- SL8500 repack in corso: → TS4500-2
  - 39.6 PB: still to migrate
- ~60 PB total available on IBM libraries
  - or only ~20 PB free considering space needed for repack

Library	Tape drives	Max data rate/drive, MB/s	Max slots	Max tape capacity, TB	Installed cartridges	Used Tape space, PB	Scratch Tape space, PB
SL8500 (Oracle)	16*T10KD	250	10000	8.4	9387	**51**	**29**
TS4500 (IBM)	19*TS1160	400	6198	20	5472	**106**	**3**
TS4500-2(IBM)	18*TS1170	400	7844	50	2082	**41.1**	**59**

Repack Oracle → IBM

Repack - Library Space Occupancy

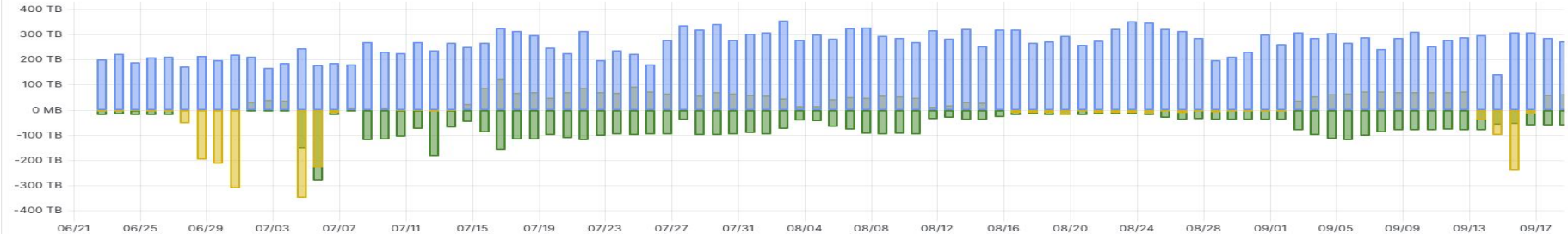


Name	Last	Difference
SL8500	39.6 PB	-5.87 PB
TS4500	93.9 PB	780 TB
TS4500_2	39.2 PB	23.7 PB

Repack - Library Scratch Tape

Name	Last	Difference
SL8500	3310	567
TS4500	165	12
TS4500_2	1191	196

Library Space Occupancy Difference



Name	Last	Mean	Total
SL8500	-61.0 TB	-66.0 TB	-5.87 PB
TS4500-1	63.3 TB	8.77 TB	780 TB
TS4500-2	273 TB	266 TB	23.7 PB

Current SW in PROD

- GPFS 5.1.9-11
- StoRM BackEnd 1.11.22 and StoRM FrontEnd 1.8.15 (latest, testing upgrade)
- StoRM WebDAV upgraded to 1.11.0 (latest, in beta)
  - Spring Boot 3, scitags, new monitoring metrics and log format, tape storage areas, nginx support, virtual threads
- StoRM Tape 0.9.0 (latest) and a new StoRM Tape Authz 1.1.0 (latest)
  - Telemetry via OpenTelemetry, new authorization policies (similar to StoRM WebDAV fine-grained authorization)
  - CMS has a different deployment with dedicated endpoint for polling requests
- XrootD 5.6.9 for ALICE (disk and tape), XrootD 5.7.3 for CMS, ATLAS, and no LHC, XrootD 5.5.5 for LHCb
  - One dedicated xrootd server for ATLAS (needed?)

Tickets and more

- ALICE
  - Nothing to report. Anything to be done to enable Scitags?
- ATLAS
  - GGUS-Ticket-ID: #1000341 (solved) “INFN-T1: T0 disk export to grid”
    - Network problems over the night of Aug 14th
  - GGUS-Ticket-ID: #1000271 (solved) “INFN T1 deletions errors”
    - Low deletion efficiency, ultimately caused by one endpoint banned by GARR
  - GGUS-Ticket-ID: #683593 (solved) “INFN-T1_DATADISK 2 possibly corrupted files”
    - To be invalidated, as already reported in the previous GGUS #683273
  - GGUS-Ticket-ID: #1000134 (on hold) “Transfer and staging errors”:
    - Several different issues and solutions in one ticket, but the debug of the low efficiency of transfers from CNAF tape to CNAF disk is still ongoing together with StoRM devels
    - Kind of duplicates of this one: GGUS-Ticket-ID: #1000072, GGUS-Ticket-ID: #683563
  - GGUS #683487 (solved): “File transfers timed out”
    - Switching to storm-webdav behind nginx, which improved performances.

Tickets and more

- CMS
  - Xrootd proxy for Leonardo installed
    - Current waiting for a robot certificate for A. Pascolini
  - GGUS-Ticket-ID: #1000548 (in progress) “Invalidating corrupted files to avoid FTS failures (making T1 CNAF unusable)”
    - We did provide a list of files without checksum in GGUS-Ticket-ID: #682009 ; these were not invalidated but they are being discussed in a jira issue (?)
  - GGUS-Ticket-ID: #1000366, #1000344 (solved) “Storage SAM test failures”
    - Problems with the filesystem

Tickets and more

- CMS (long-standing issues with tape archival)
  - GGUS-Ticket-ID: #682009 (in progress) “CNAF archived transfers and failure requests”
  - Several different issues with data archival: polling, when to declare successful/failed transfer to tape, missing 'transfer-done' messages in Rucio, and ultimately disagreement in monitored space and rates between site and CMS
  - Following two productive meetings in July and August with CMS Ops:
    - dedicated endpoint for /archiveinfo requests within the StoRM Tape REST API lead to strong decrease of 502 errors when polling files (solved)
    - generation of lists available in /info storage area containing all the files in tape storage area (CMS consistency check to be started soon)
    - generation of list with information on missing/present checksum for such files
      - changed a configuration parameter in GEMSS so that only files with the checksum extended attribute are staged out to tape asap, the other files being staged out to tape anyway after 5 days

Tickets and more: CMS mini data challenge

1st July: old deployment w/o nginx (wrt DC24, new homogeneous hw and 5 endpoints)

CMS was able to write at 10 GB/s while reading at 6 GB/s for a few hours. No crashes of StoRM WebDAV, no thread saturation (different from DC24), load average ~4 on each server but drop to 75% efficiency

DC 24 - CMS results

		CNAF					
		Expected	Observed	Ratio	Expected	Observed	Ratio
Day	Scenario	as DEST			as SRC		
1	T0 export	4.23	4.94	1.17	0	0	N/A
2	T0 export	4.23	4.93	1.17	0	0	N/A
3	T0 export, T1 export	4.23	5.4	1.28	4.25	3.94	0.93
4	T1 export	0	0	N/A	4.37	4.39	1.00
5	T1 export, prod-out	1.71	2	1.17	6.36	4.92	0.77
6	T1 export, prod-out	1.71	2.01	1.18	6.36	6.03	0.95
7	T1 export, prod-out	1.71	1.98	1.16	6.36	6.51	1.02
8	AAA	1.17	1.54	1.32	0	0	N/A
9	All	9.22	5.98	0.65	5.68	1.43	0.25
10	All	10.39	6.5	0.63	5.68	1.57	0.28
11	All	10.39	7.16	0.69	5.68	1.39	0.24
12	All	10.39	8.14	0.78	5.68	1.64	0.29

Tickets and more: CMS mini data challenge

1st July: old deployment w/o nginx (wrt DC24, new homogeneous hw and 5 endpoints)

CMS was able to write at 10 GB/s and reading at 6 GB/s for a few hours. No crashes of StoRM WebDAV, no thread saturation (different from DC24), load average ~4 on each server but drop to 75% efficiency

2nd July: new deployment with nginx

CMS was able to write at 10 GB/s and reading at 6 GB/s for a few hours. Load average ~1 on each server, many 502 errors due to some idle timeouts fixed in a new StoRM WebDAV release.

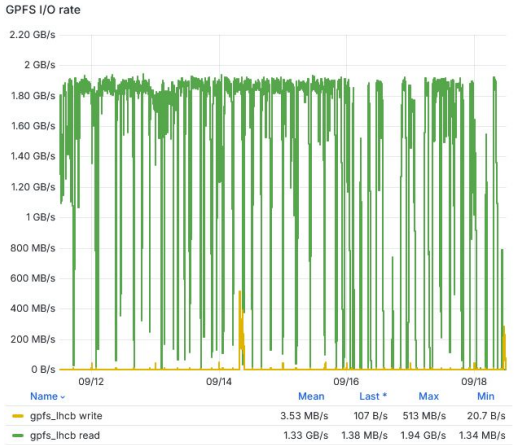
We would like to repeat the challenge with improved StoRM WebDAV and EL9 NSD servers

Tickets and more

- LHCb
  - GGUS-Ticket-ID: #1000539 (solved) “Files fail to be staged from CNAF-ARCHIVE”
    - Damaged tape; files removed from catalogue and from filesystem
  - GGUS-Ticket-ID: #1000287 (solved) “Many transfer failures to/from CNAF”
    - A problematic GARR router
  - GGUS-Ticket-ID: #1000269 (solved) “Corrupted files”
    - All corrupted files were written on July 27th via a server which was experiencing heavy load due to a monitoring check left there by mistake (GGUS-Ticket-ID: #1000204); thousands of files checked on disk and on tape, for all servers. Few hundreds corrupted files found and re-copied by LHCb
  - GGUS-Ticket-ID: #1000212 (solved)
    - Virtual threads introduced in StoRM WebDAV, which before used one system thread for each request, resulting in high load but idle CPU probably due to context switching overhead.
  - GGUS-Ticket-ID: #683620 (solved) “Failed transfers at INFN-T1”
    - New StoRM WebDAV deployment with nginx, aligned timeouts, fix of memory leak

Tickets and more

- LHCb
  - GGUS-Ticket-ID: #1000315 (solved) “CNAF SRR reports LHCb-Tape allocation is almost fully used”
    - Disk buffer filled up as stage-out was too slow. We asked LHCb to stop writing so to flush the buffer, but due to backlog LHCb kept writing at a faster rate than data was staged-out to the tapes (1.2GB/s in presence of writing activity, 1.6GB/s without).
    - We suggested to wait for buffer to empty completely before starting to write again, and we limited buffer quota (in SRR space report) to NVMe space
    - A tuning phase started for FTS parameters to find the sweet spot for the tape buffer in which rate_in = rate_out
    - NVMe space was also used for *.*dst* files to ease data access from sprucing jobs; we removed this placement policy so that now NVMe space is used for tape buffer only.



Tickets and more

- All no-LHC experiments can move away from *srm* and use *https* instead, given the StoRM Tape REST API is ready: tape-archive.cr.cnaf.infn.it
  - Already srm-less: Agata/Gamma, AMS, CTA-LST, Cygno, Dampe, DarkSide, GLAST, HyperK, JUNO, LHCf, MIBLAT/QCDLAT, Muoncoll, Newchim, PADME, Pamela, Theophys, Virgo
- Belle
  - GGUS-Ticket-ID: #1000052 (solved): permission error when accessing data
    - Again KEK LSC file (https://github.com/italiangrid/voms/issues/141), this time with nginx in front of StoRM WebDAV
- Cygno
  - Tape was being used as backup for extremely small files, so we blocked stage-out activity ‘til this was fixed by Cygno
- JUNO
  - GGUS-Ticket-ID: #683660 (solved): Buffer filled up by a different experiment
- KLOE
  - 355k files of KLOE1 (210TB) recalled from tape, currently checking checksum and then planning to rewrite them
- VIRGO
  - Currently configuring Pelican OSDF cache (OSG support ticket)

Ceph

- Object Storage Service available for several projects and offering different abstractions:
  - POSIX-Compliant FileSystem.
  - Block Devices.
  - S3 Object Storage.
- Cluster TeRABIT
  - FS for the HPC cluster.
  - Capacity of 8 PiB*.
  - TeRABIT, ICSC and INFN Cloud projects.
- Cluster Cloud@CNAF
  - Block Devices for OpenStack Volumes and Virtual Machines and FS for Kubernetes Clusters.
  - Capacity of 3.8 PiB*.
  - Cloud@CNAF users, CNAF and Virgo Kubernetes Clusters.

Ceph

- Cluster S3:
  - S3 compatible API for Object Storage.
  - Capacity of 4.6 PiB*
  - Cygno project and Cloud@CNAF users.
- Challenges :
  - Metadata Servers(MDS) that are responsible of POSIX Compliance can get overloaded leading to high memory consumption.
  - Clusters with high number of devices can lead to monitoring component failure.
  - Linux Device naming is causing problems for Ceph installation.
- Approaches:
  - Extensive monitoring and efficient distribution of MDS and their FS metadata.
  - Distributed metrics collection at the level of a single node.
  - Persistent Device naming using device level information.