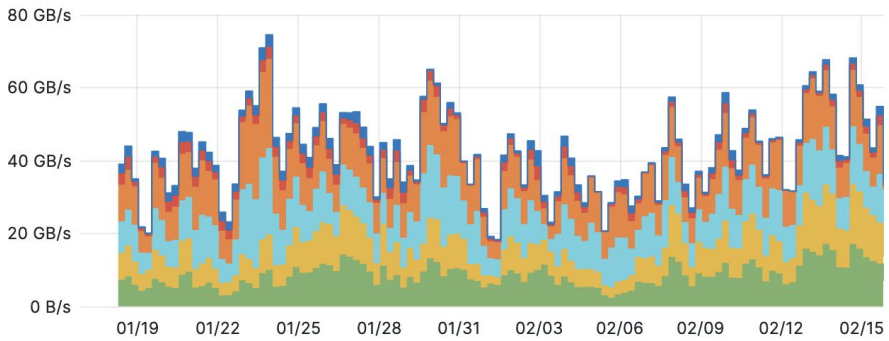
State of Storage

CdG 16 febbraio, 2024

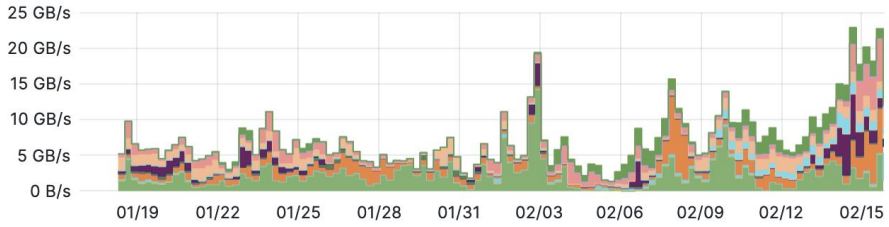
Business as usual

Last month

All servers network traffic out (reading)

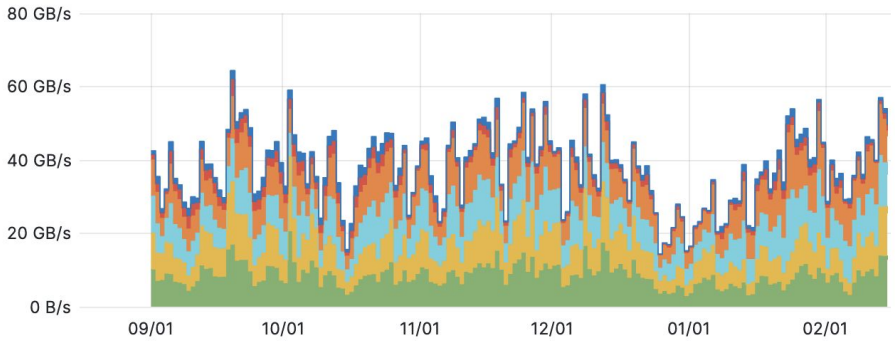


Gateway traffic out (non POSIX reading)

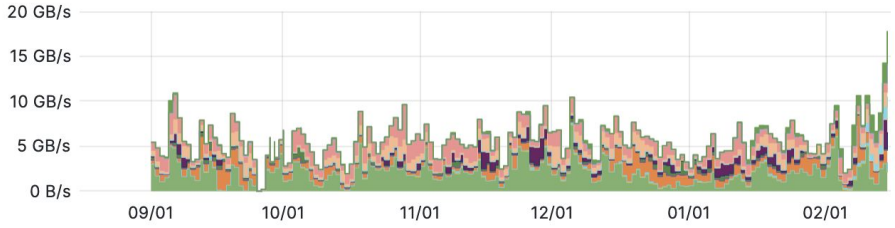


Last 6 months

All servers network traffic out (reading)



Gateway traffic out (non POSIX reading)



Disk storage in produzione

Installed: 53.64 PB, Pledge 2023: 69.6 PB, Used: 48.8 PB

Storage system	Model	Net capacity, TB	Experiment	End of support
ddn-10, ddn-11	DDN SFA12k	10120	ALICE, AMS	12/2022 (+10 spare hdd)
os6k8	Huawei OS6800v3	3400	GR2, Virgo	12/2023
md-1,md-2,md-3,md-4	Dell MD3860f	2308	DS, Virgo, Archive	05/2024
md-5, md-6 e md-7	Dell MD3820f	50	metadati, home, SW	11/2023 e 12/2024
os18k1, os18k2	Huawei OS18000v5	7800	LHCb	7/2024
os18k3, os18k5, os18k5	Huawei OS18000v5	11700	CMS	6/2024
ddn-12, ddn-13	DDN SFA 7990	5840	GR2,GR3	2025
ddn-14, ddn-15	DDN SFA 2000NV	24	metadati	2025
os5k8-1,os5k8-2	Huawei OS5800v5	8999	ATLAS	2027
Cluster CEPH	12xSupermicro SS6029	3400	ALICE, cloud, etc.	2027

Acquisti recenti e futuri

- Gara storage 2022 (14PB netti)
  - LENOVO DE6600: Collaudo non superato
    - Nuova proposta con apparati DDN SFA7990X
- AQ storage 2023-2024
  - Il vincitore è Huawei con sistemi OceanStore Micro 1500/1600
  - Richiesta fornitura di 64PB nel 2023
  - Installazione al Tecnopolo iniziata
- Gara Tape Library
  - Contratto fermo in AC, possibile ritardo di qualche mese
- Gare nastri
  - Ulteriori 30 PB di nastro per repack dei dati dalla libreria Oracle
    - Gara in preparazione

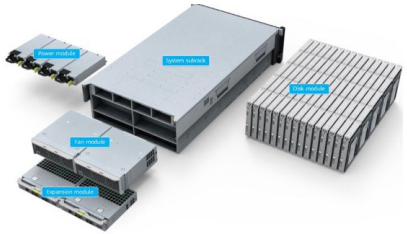


Figure 6 - Structure of the 4U high density disk enclosure

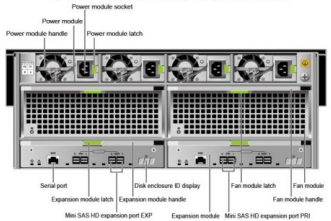


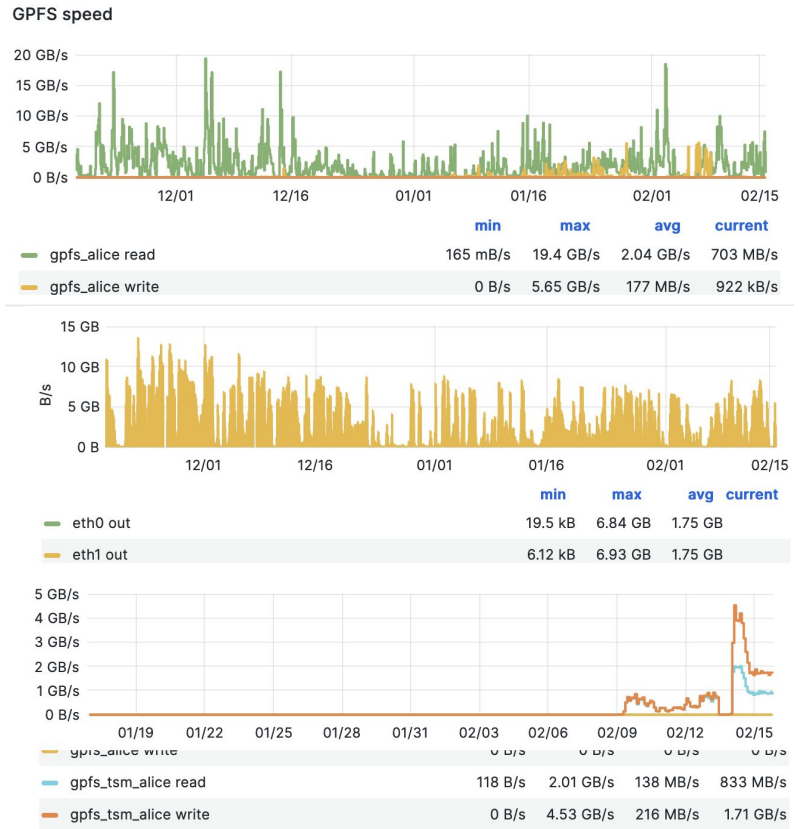
Figure 7 – Rear view of the 4U high density disk enclosure

ALICE

GPFS

CephFS

PROD -> tape

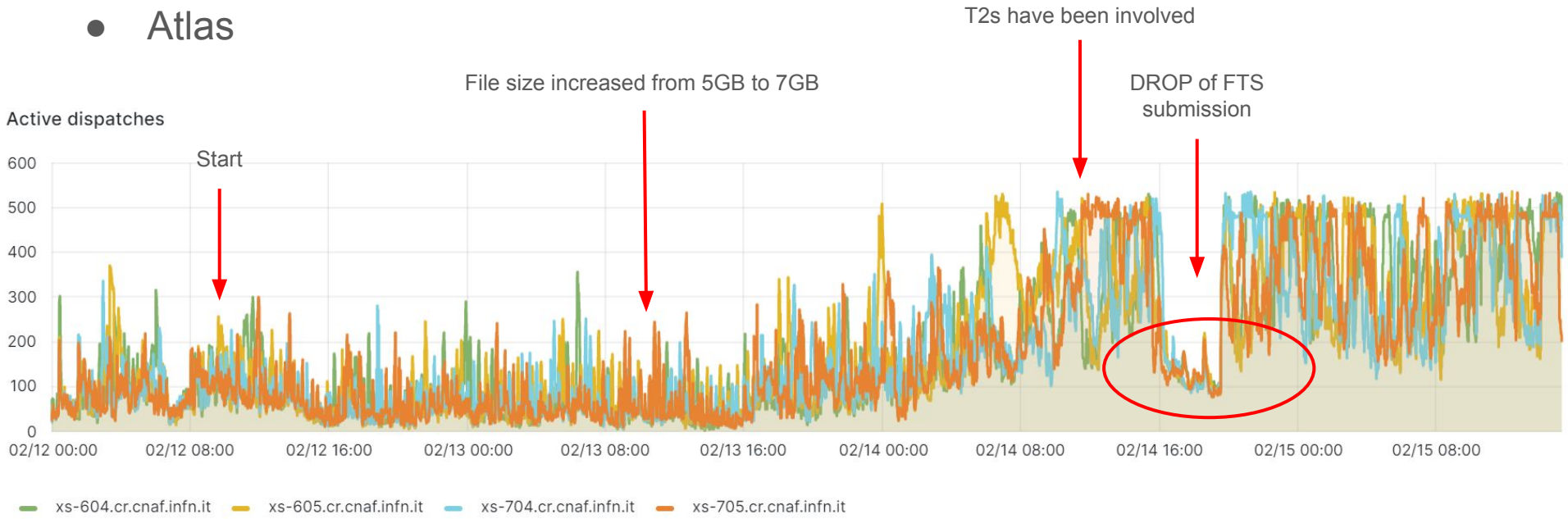


Current SW in PROD

- GPFS 5.1.2-11
- StoRM BackEnd 1.11.22 (latest)
- StoRM FrontEnd 1.8.15 (latest)
- StoRM WebDAV 1.4.2 (latest)
- StoRM globus gridftp 1.2.4
- XrootD 5.5.4-1
  - ALICE CEPH updated to 5.5.5-1.el8
- Ceph 16.2.6 (Pacific)

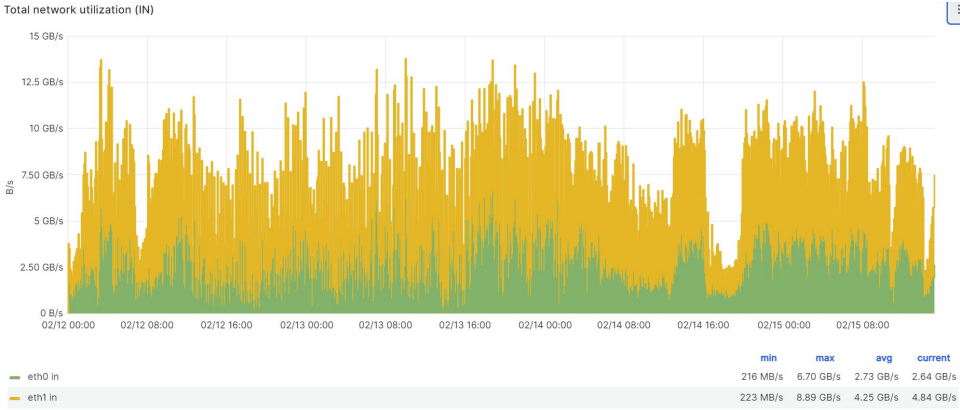
Data Challenge 2024

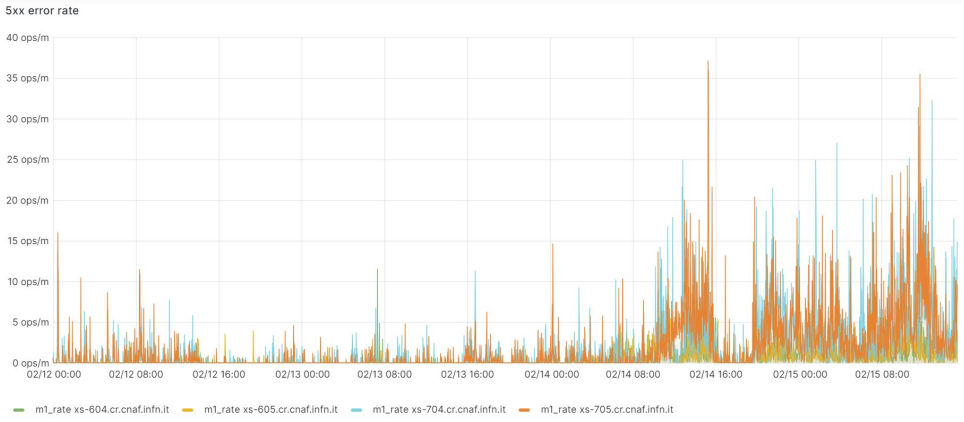
- Atlas



Data Challenge 2024

- Atlas
  - Threads saturation
  - 5** error rate increased
  - Total throughput not affected: between 5 and 13 GB/s (T0 → T1)
  - Archived week average (disk) **5.31GB/s**
  - Total (disk+tape) 7.21GB/s
  - Minimal target rate 3.03 GB/s
  - Higher target rate 3.46 GB/s



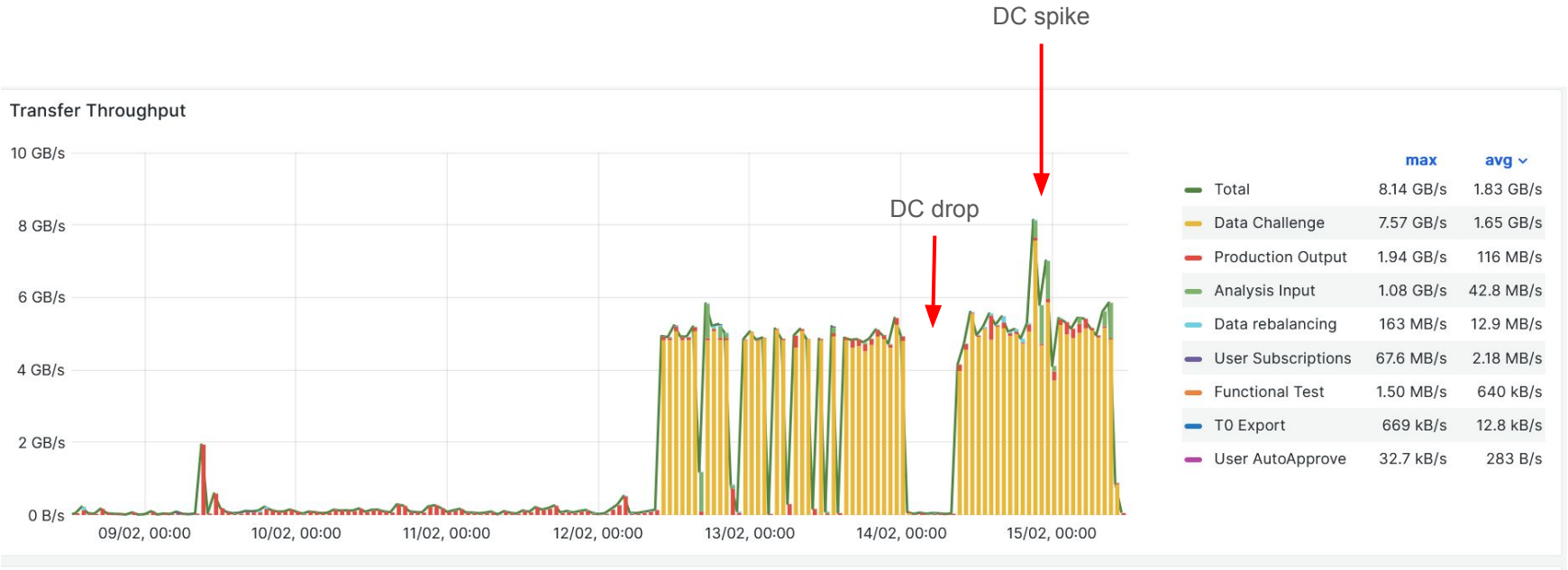


Data Challenge 2024

- Atlas
  - Issued also a lot of recalls from tape (not involved in the DC24)
  - `gpfs_tsm_atlas has 5246 files in 7 tapes being processed`
    `gpfs_tsm_atlas has 65378 files in 320 tapes waiting for recall`
  - It had a big impact on the tape infrastructure

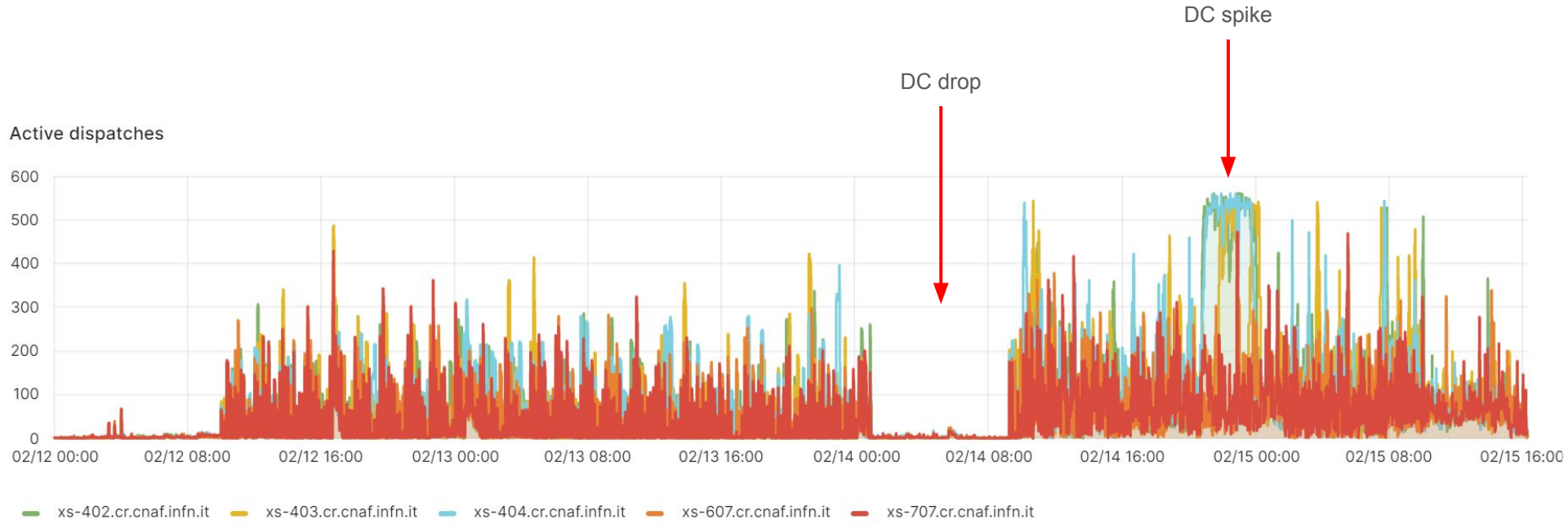
Data Challenge 2024

- CMS



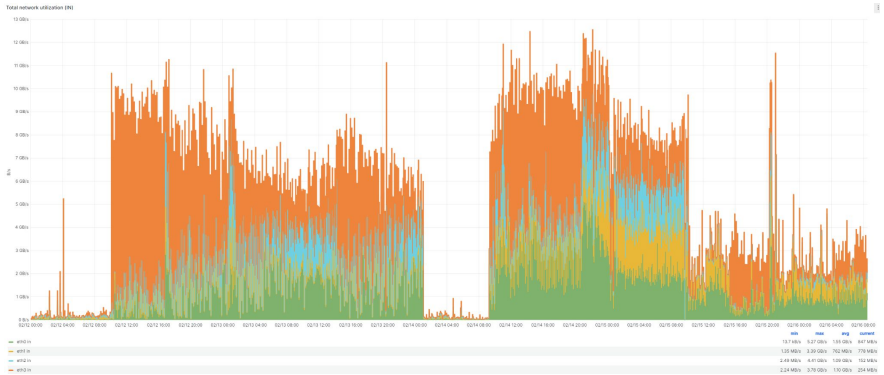
Data Challenge 2024

- CMS



Data Challenge 2024

- CMS
  - Thread saturation and high error rate only during the DC spike
    - Also SAM tests failed (see previous plot)
  - General throughput between 7 and 10 GB/s (T0 → T1)
    - Minimal target rate 4.25 GB/s
    - Higher target rate 7.13 GB/s
  - Single TPC transfer rates between 50 and 150 MB/s
  - Results in agreement with previous tests done by CMS

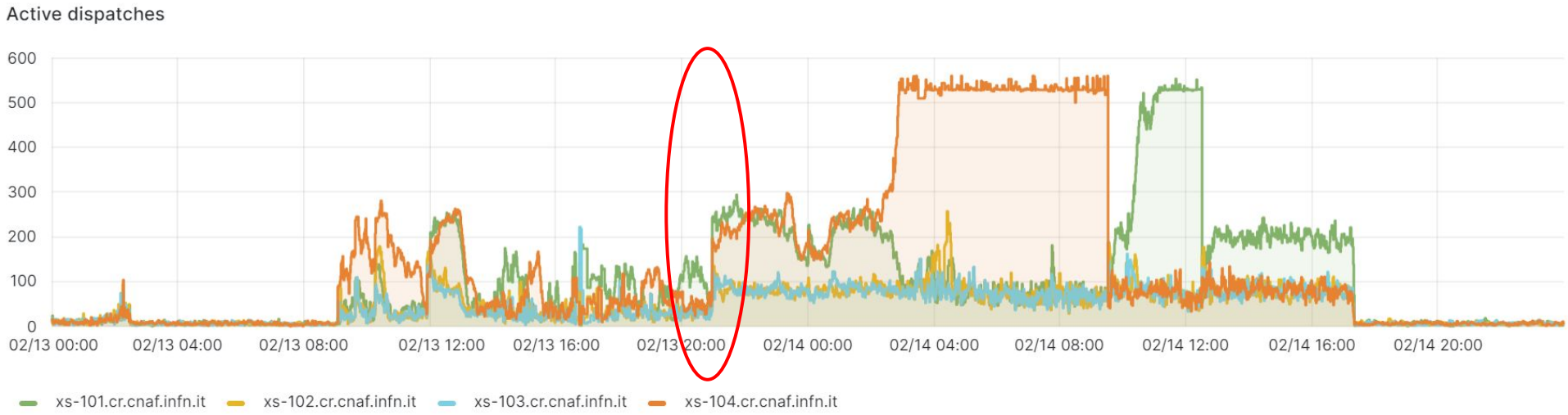


Data Challenge 2024

- LHCb
  - 309TB at 2.20GB/s (constant target rate)
  - CERN EOS → xfer-lhcb.cr.cnaf.infn.it → storm-fe-lhcb.cr.cnaf.infn.it

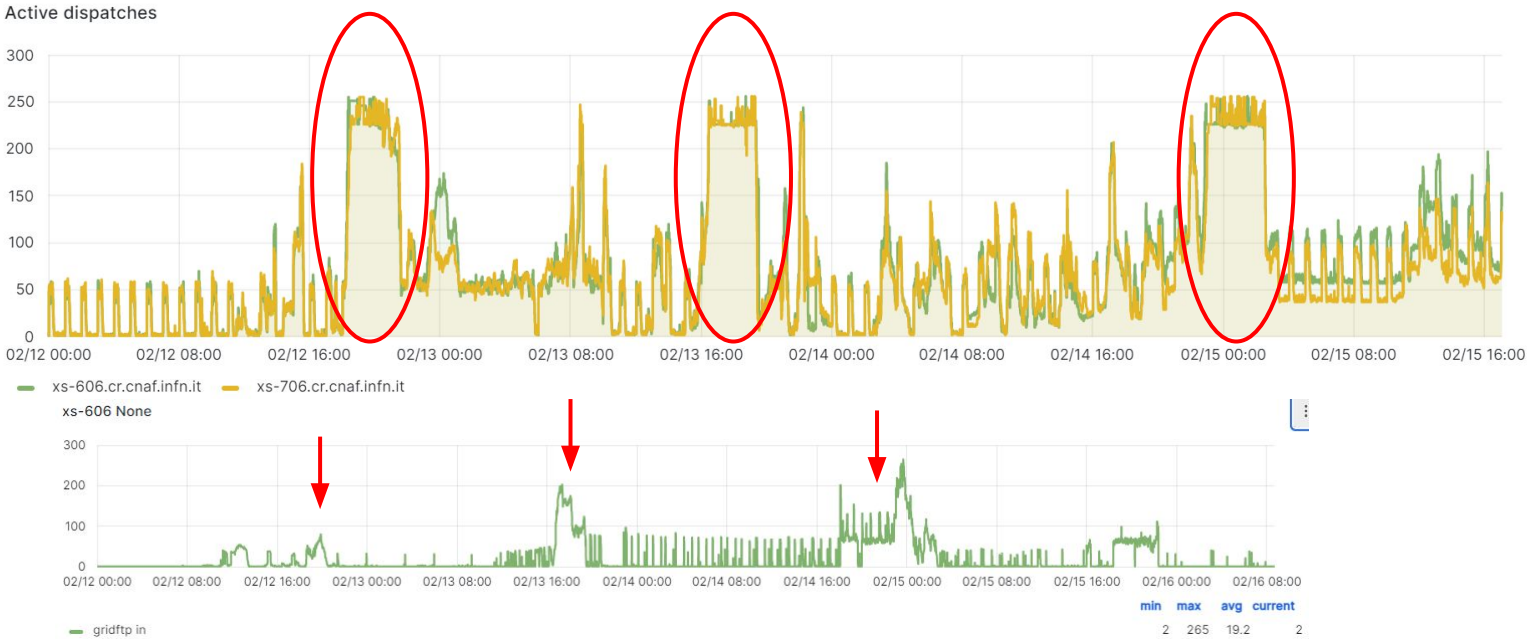
Data Challenge 2024

- LHCb
  - Bad threads management by StoRM WebDAV
  - Despite the target rate is achieved; halved throughput and high 5** error rate
  - Ongoing discussion with Lucio, Matteo and Chris to improve the situation (reduce TPCs rate?)



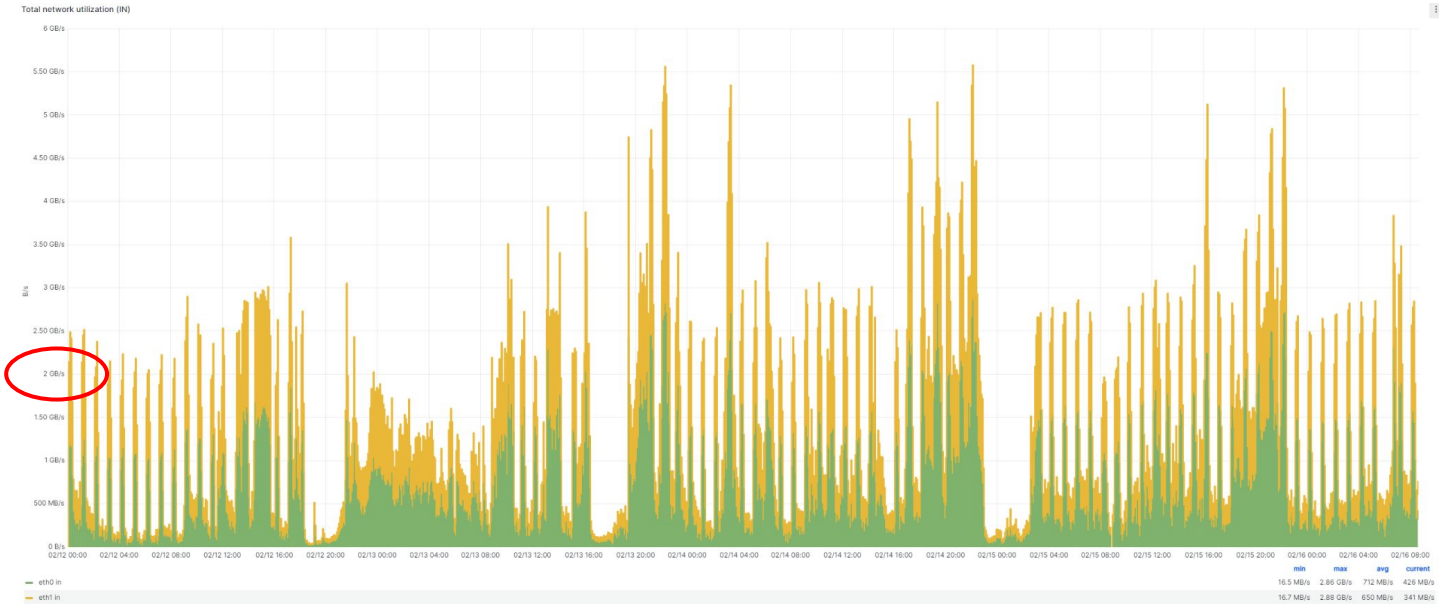
Data Challenge 2024

- Belle II
  - Errors and slowdowns in correspondence of threads saturation
    - Due to a big amount of concurrent gsiftp connections by Belle and Xenon



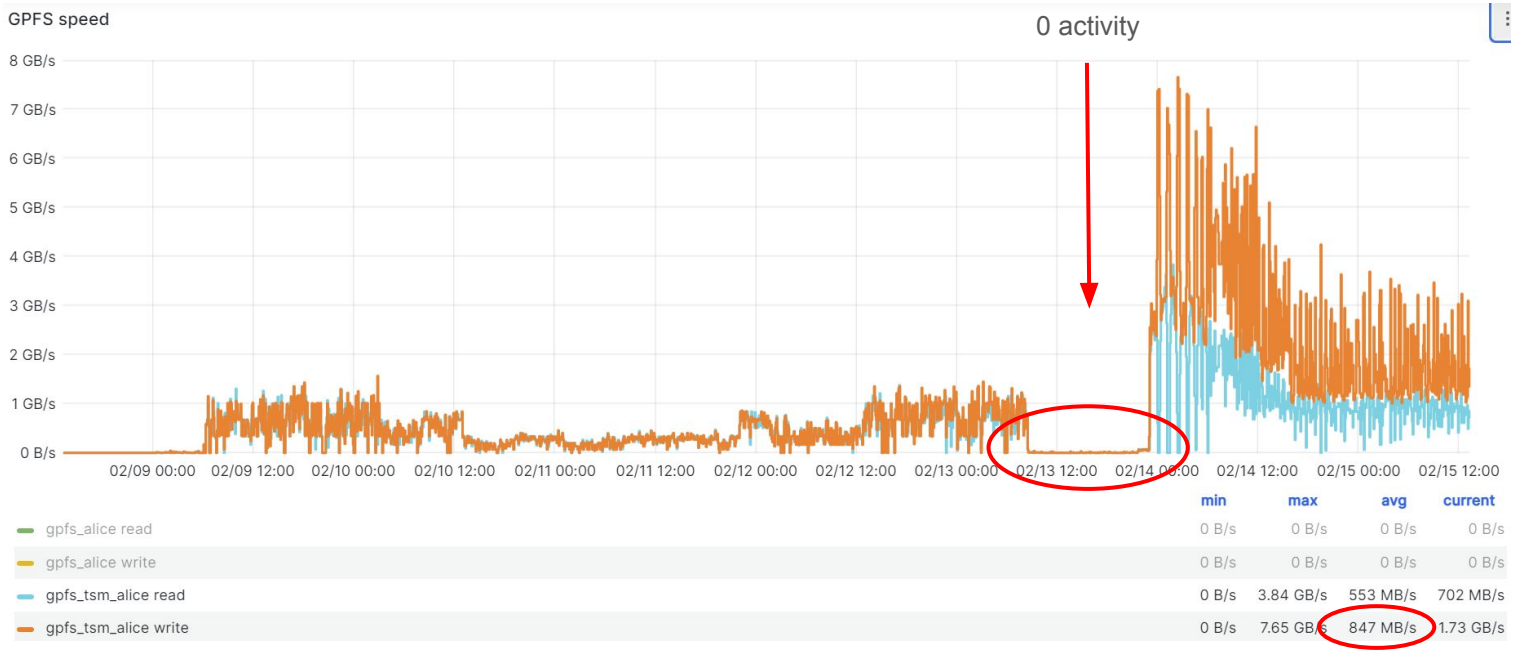
Data Challenge 2024

- Belle II
  - Minimal target rate 0.28 GB/s, higher target rate 0.46 GB/s
  - No LHC complessive throughput around 2 GB/s with spikes of 4-5 GB/s



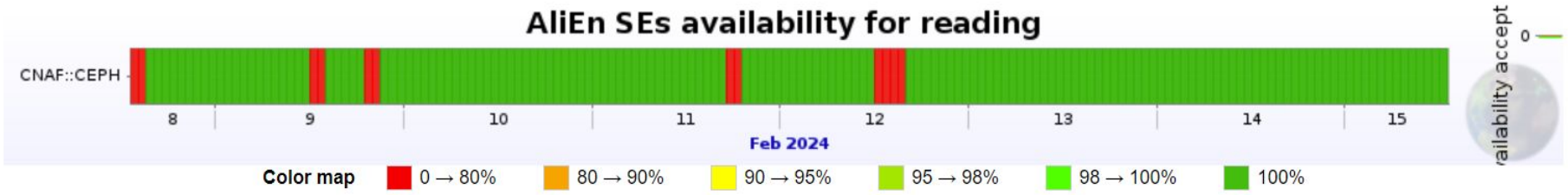
Tickets and more

- ALICE
  - Since February 9th Alice is transferring to CNAF (tape) a portion of the Pb-Pb data
    - 3.7PB at ~0.8GB/s



Tickets and more

- ALICE
  - Finishing XrootD configuration restyling of GPFS cluster:
    - Manage configuration files with Puppet
    - Revert to original configuration with all servers acting as managers as well
    - Upgrade to latest version in production (5.5.4-1.el7)
    - Check on the status of the service has been included within sensu framework of check and remediation
    - Waiting for the end of the Pb - Pb data transfer (approximately 2 months) to finalize the configuration with the tape cluster (xs-204, xs-304)
  - Misalignment between MonaLisa and our alerting system
    - We contacted Mario and Francesco to replicate MLsensor behaviour on our cluster



Tickets and more

- ATLAS
  - StoRM Tape REST installed and configured; no tests yet (ATLASGROUPTAPE)
- CMS
  - GridFTP still used, only for SAM tests
  - StoRM Tape REST installed and configured; no tests yet and it did not help in getting rid of GridFTP
  - Removed PhEDEx_* disco and tape data
  - GGUS 164856: among 5, 2 StoRM WebDAV endpoints had threads stuck in retrieving chains from the token issuer (IAM) to validate tokens
    - Restart of the service solved the problem
    - StoRM developers are investigating the issue
    - In the meantime, set an alerting check with remediator up for this issue

Tickets and more

- CMS
  - GGUS 164989: “connection to the xfer-cms.cr.cnaf.infn.it:8443 storage endpoint drops every 15 minutes” (?)
    - Closed with reason “We modified our script to avoid caching”
  - GGUS 165183: stage requests stuck from 6 days
    - Misconfigured parameter of GEMSS
    - Once fixed it, recalls started again but…
  - GGUS 165276: pending stage requests (6.5k submitted transfers still waiting to be completed)
    - They were at the “in progress” status in StoRM backend DB
    - We fixed and cleaned the database

Tickets and more

- LHCB
  - GGUS 164032 (in progress): User fts transfers from INFN are failing
    - Still in progress, but FTS and StoRM WebDAV developers have been involved
    - StoRM WebDAV provides storage tokens (macaroons) via the oauth/token endpoint, but FTS retrieves macaroons from the resource path. This should not be allowed by StoRM WebDAV (STOR-1602), no matter permissions of the storage area.
  - GGUS 164634 (on hold): Permission denied due to trailing slash
    - New versions of FTS and StoRM tape REST API will accept ending slash in the path
  - GGUS 164834 (closed): StoRM tape REST API was in debug mode and filled the log directory
  - GGUS 164961 (closed): StoRM Tape REST API could not handle tens of thousand requests simultaneously
    - Implementing index in the database solved the issue

Tickets and more

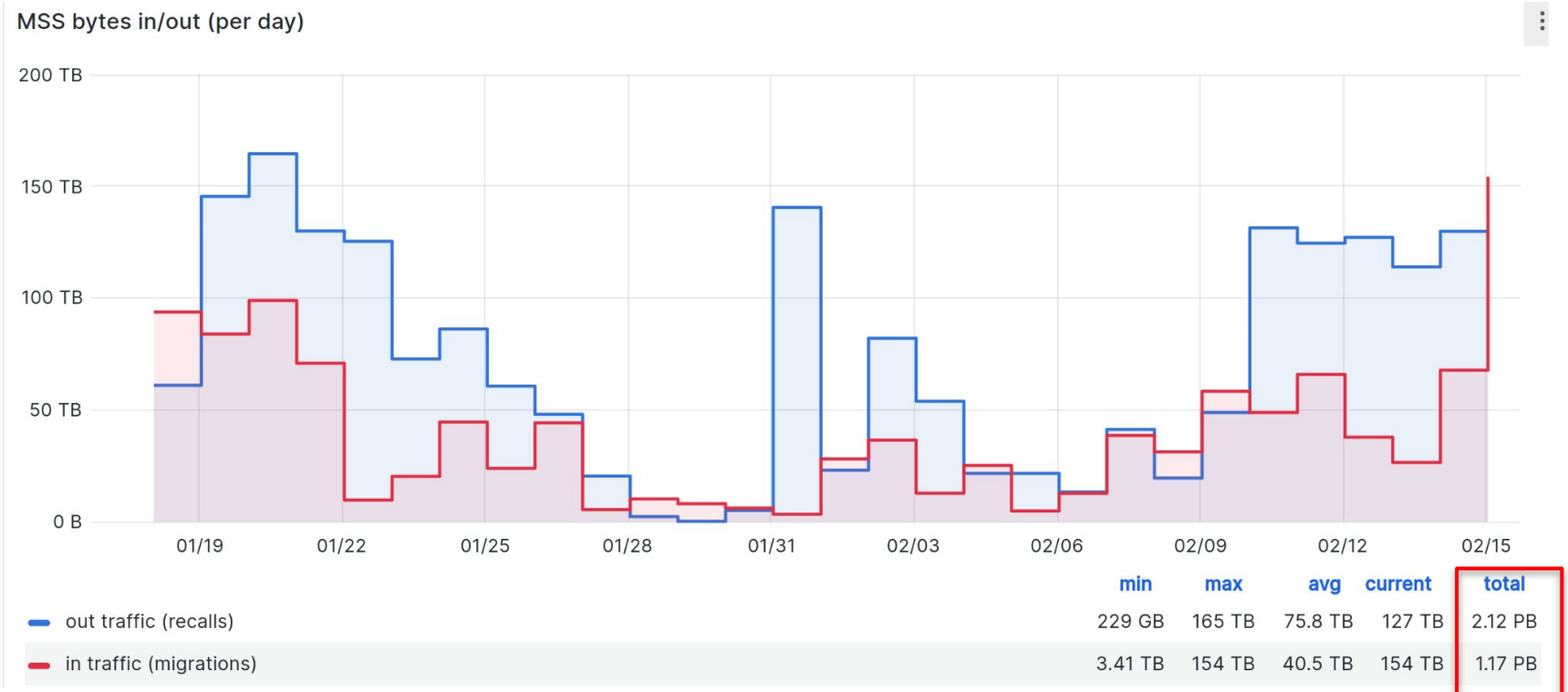
- LHCB
  - GGUS 165048: LHCb token authentication for disk storage
    - WLCG-scope-based token AuthZ implemented for disk storage area
    - StoRM WebDAV does not support full path scope
      - Access point/root path cannot be part of the scope path
    - Involved StoRM developers
      - Discussion ongoing
  - GGUS 165225: “DC24: LHCb activity”

Tickets and more

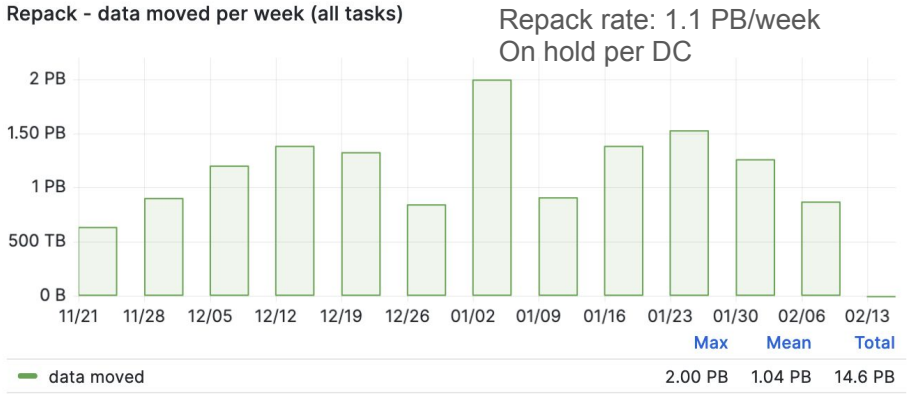
- Gsiftp protocol via StoRM backend is still available for a few experiments
  - Tests srm+https and feedback very welcome (Belle, Xenon)
  - Goal: remove gsiftp protocol and switch off GridFTP
- CTA-LST
  - GridFTP switched on for CTA-LST storage areas to allow Third-party copies from PIC (gsiftp)
- Dampe
  - GridFTP “plain” still used
    - Testing XrootD server at IHEP to perform the transfers to CNAF (WP6-Datacloud)
- Virgo
  - Intensive usage of stashcache (many access from WNs)
  - Hit the 2x10Gb=2.5GB/s limit traffic OUT

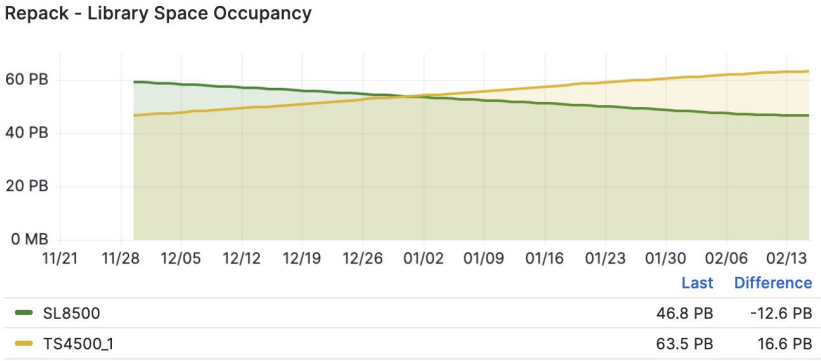
Stato tape

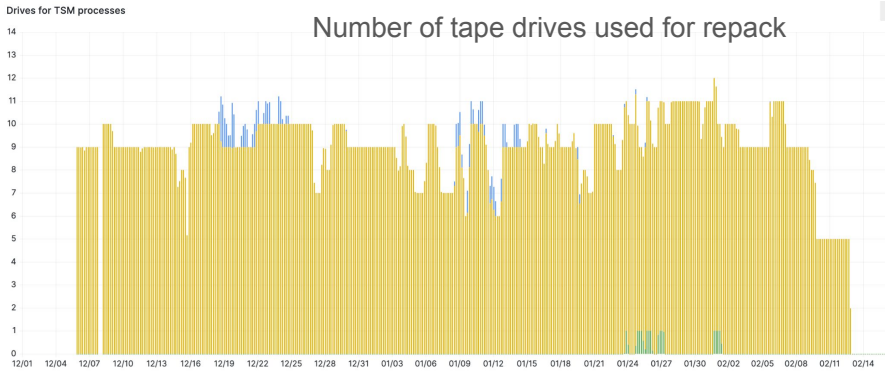
Last month

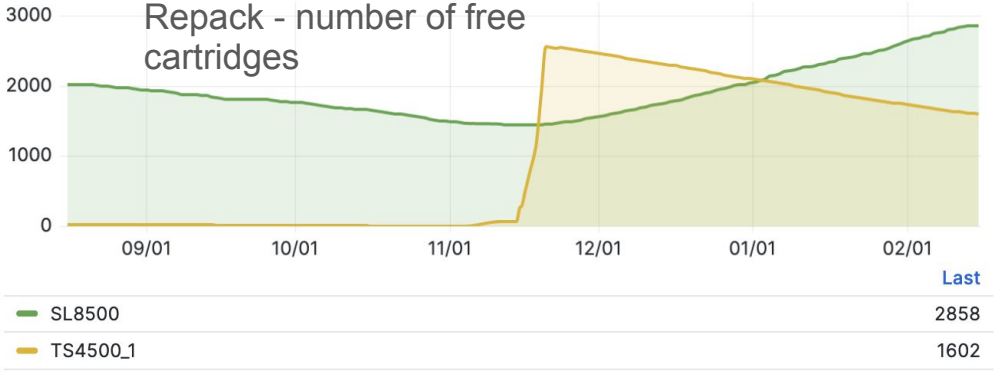


Tapes: Migration from Oracle to IBM library









Stato tape

- Liberi ~32 PB (Scratch tape sulla libreria IBM).
- Usati ~111 PB.
  - In preparazione gara per altri 30 PB

Library	Tape drives	Max data rate/drive, MB/s	Max slots	Max tape capacity, TB	Installed cartridges	Used space, PB	Free space, PB
SL8500 (Oracle)	16*T10KD	250	10000	8.4	~10000	**47**	-
TS4500 (IBM)	19*TS1160	400	6198	20	5104	**64**	**32**

BACKUP slides

ATLAS DC24

Target: (min) 24.2 Gbit/s = 3.03GB/s

(max) 27.7 Gbit/s = 3.46GB/s

Archived week average (disk)= 5.31 GB/s

Total (disk+tape) = 7.21GB/s