LHCb DC24

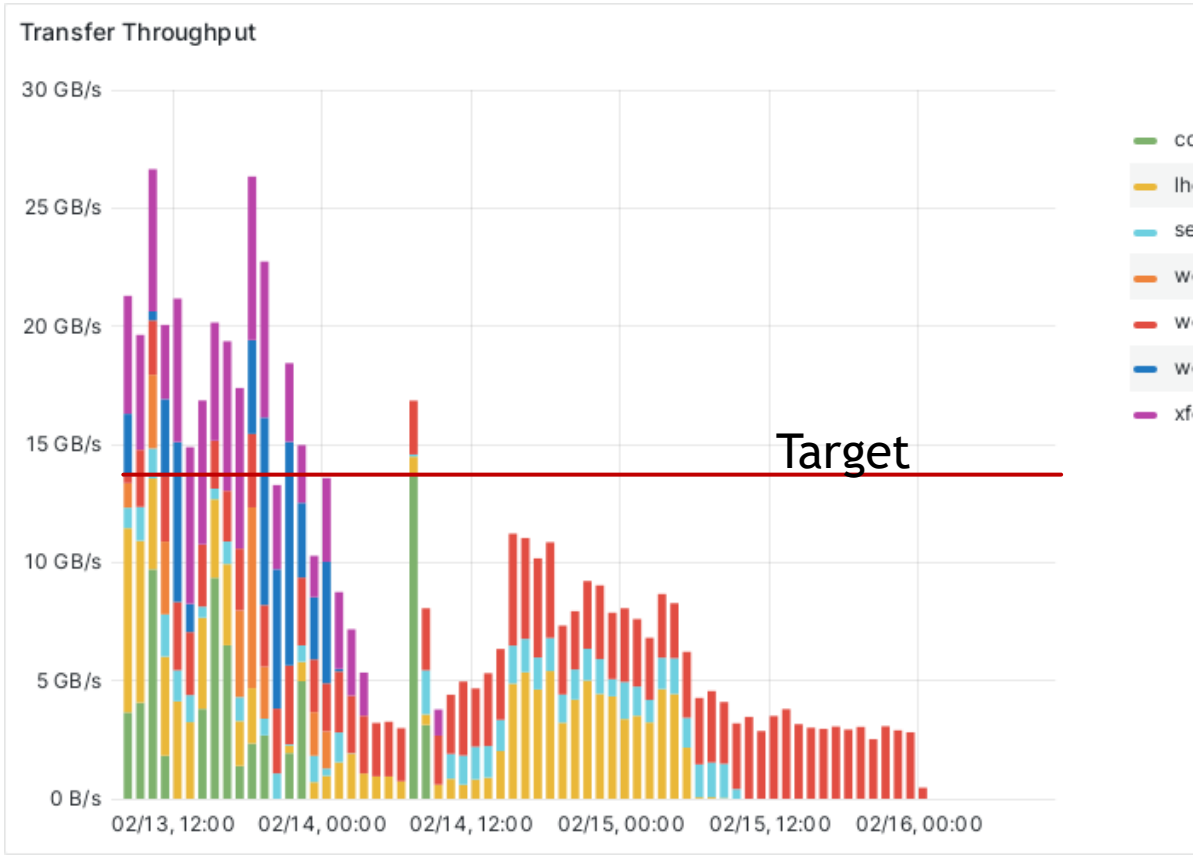
Christophe Haen, Alexander Rogovskiy

06.03.2024

Writing part

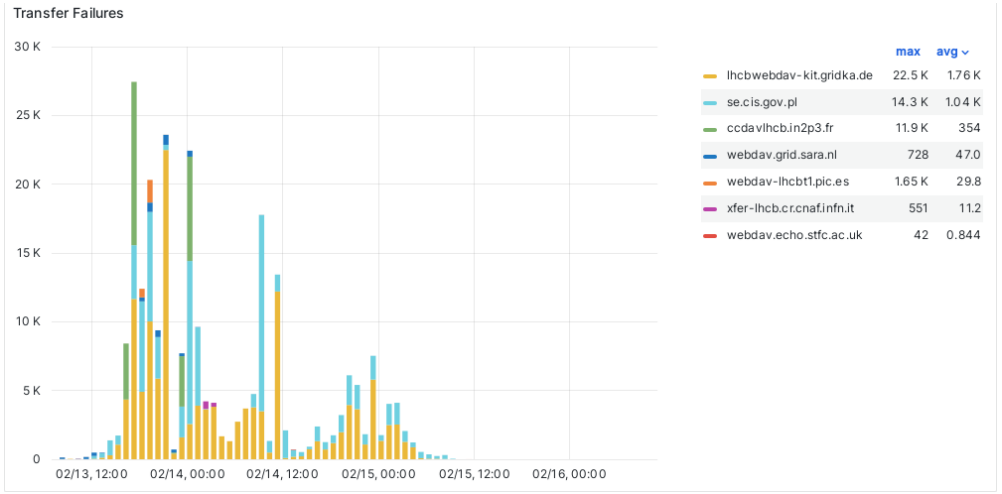
- Writing part emulates distributing data from T0 to T1 sites
- Distribution is done in 2 hops
  - First from CERN EOS to T1 Disk SE
  - Then from T1 Disk SE to T1 Tape SE (at the same T1, so it's a local copy)
  - After that files are deleted from Disk SE

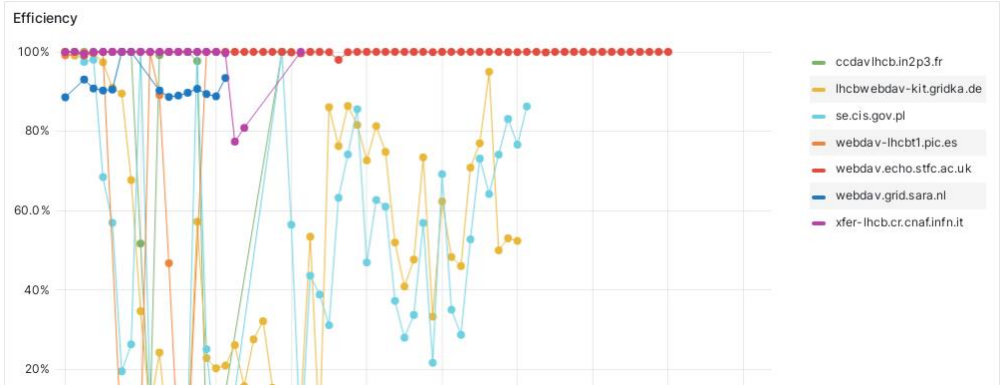
EOS -> Disk link



- Target throughput (14GiB/s) was achieved during the first day
- Lower throughput later
  - Some sites finished transferring their part during the first day so were no longer contributing to overall throughput
  - Submissions were slow and not optimal
  - Submission agent got stuck a few times, that was also a contributing factor

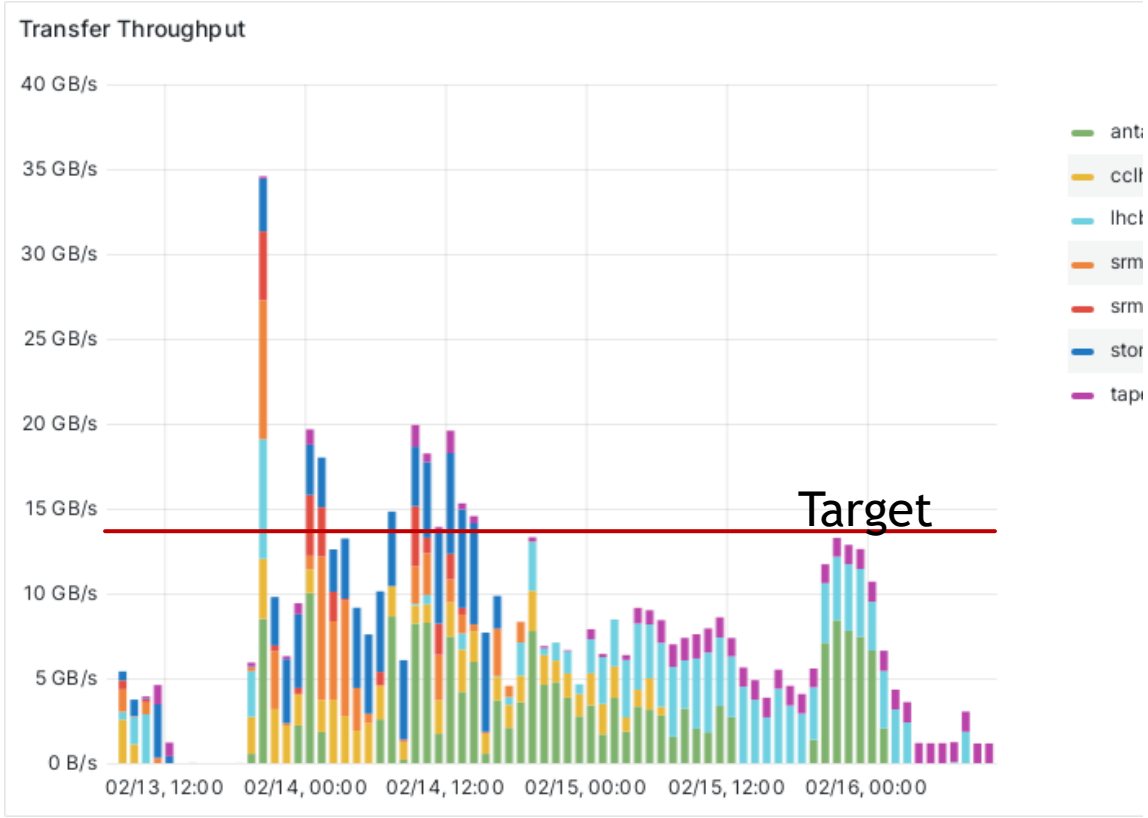
EOS -> Disk link





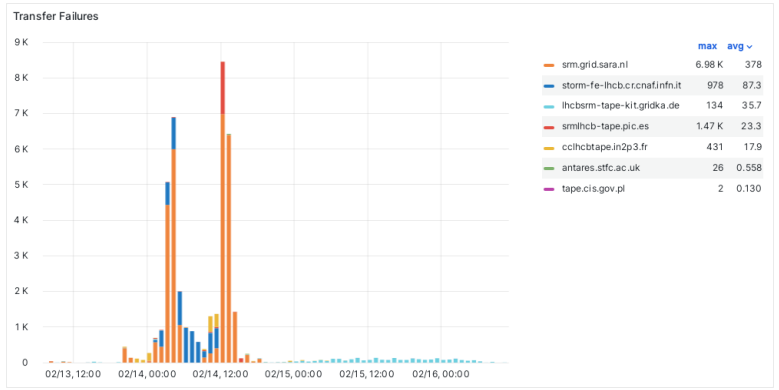
- Low efficiency for token-based sites, especially NCBJ and GRIDKA
  - Mostly because of the IAM server overload

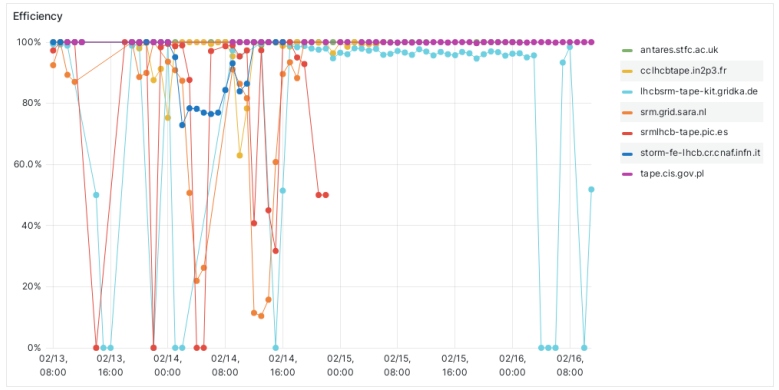
Disk -> Tape link



- Target threshold (14GiB/s) crossed several times
  - Max around 35GiB/s
  - Spikier throughput because of the nature of the link and submission agent problems

Disk -> Tape link



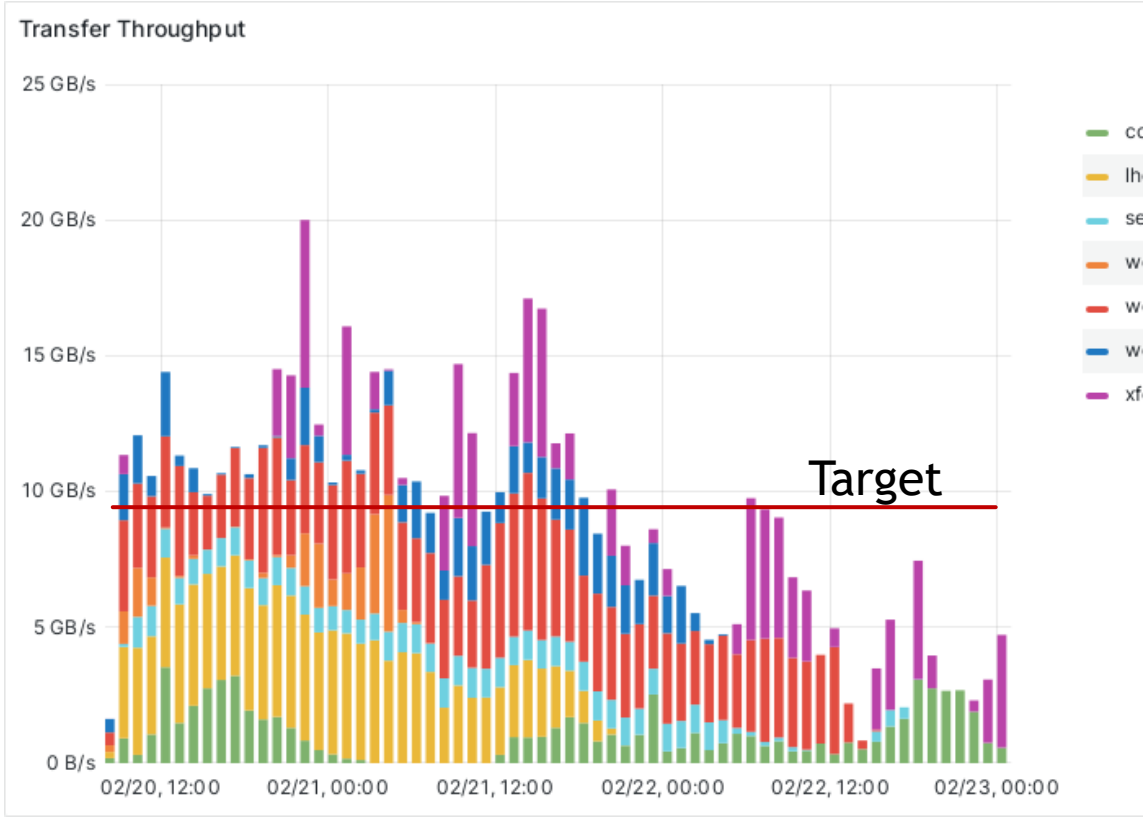


- Problems affecting the following sites:
  - SARA – disk buffer overflow
  - CNAF – low performance when reading and writing to the same storage simultaneously, due to STORM peculiarities
  - Others – FTS transfers got stuck
    - 4 files lost at RAL Tape SE, 3 at IN2P3 Tape; FTS is a suspect

Staging part

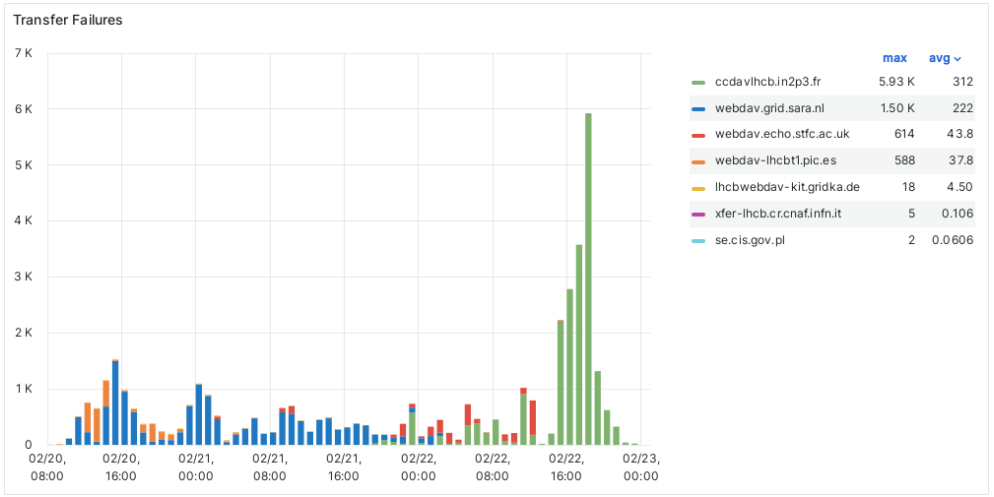
- Staging part emulates data processing after the data taking period
- Basically, just copying of files from local tape storage to local disk storage
  - This means no external traffic
  - Sites were asked to flush disk buffers of their tape SEs to allow for proper tape performance testing

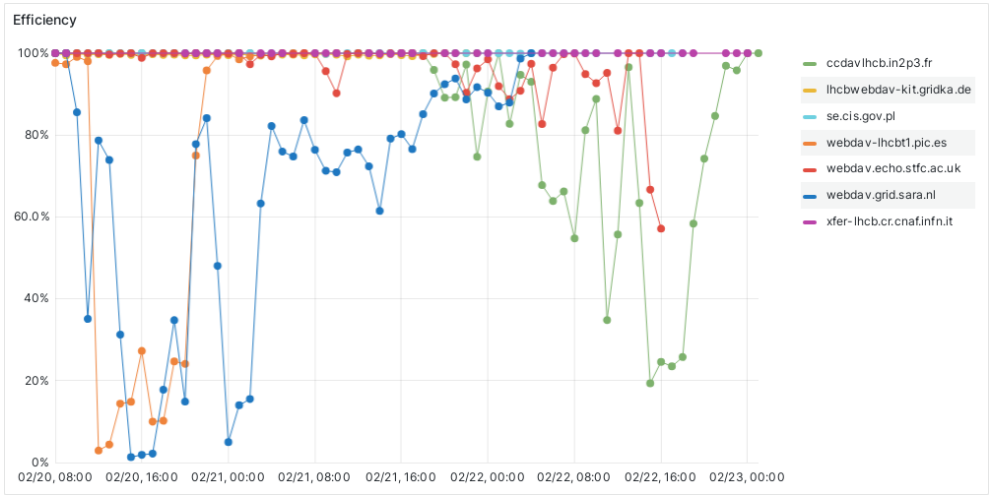
Staging



- Target throughput (9.58 GiB/s) was achieved during the first two days of the test
- Lower throughput later
  - Some sites finished transferring their part and were no longer contributing

Staging





- Problems at IN2P3 due to tape SE buffer overflow
  - Due SE limit on simultaneous requests some release requests were not processed
  - Site increased the limit
- Problems at SARA due to the limit on simultaneous internal transfer
  - Limit increased
- Problems at RAL due to gateway overloads
- Problems at PIC due to dCache's way of handling requests

Token setup

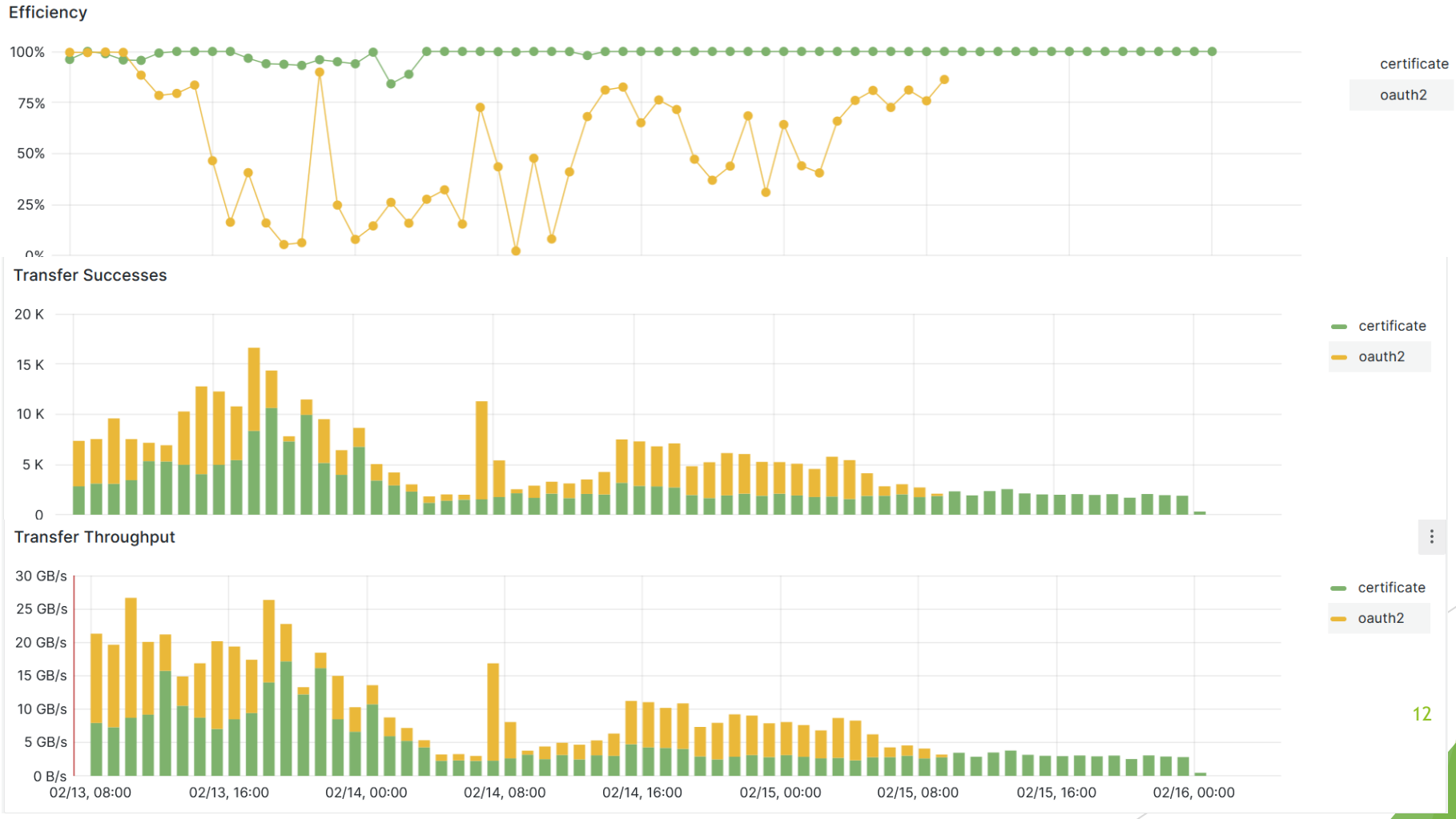
- All participating sites were requested to set-up tokens for DC24
- Unfortunately, only dCache sites (some) managed to set up tokens
- The problem is the following
  - we use full file path in storage scopes
    - So every token has `storage.modify:<full_path>` + `storage.read:<full_path>` scopes
  - FTS tries to make sure that all necessary directories exist before copying
    - Meaningless for some storages, e.g. RAL ECHO
  - To do so, it issues `PROPFIND $(basename <full_path>)` request
  - This request fails for xrootd and STORM since scope includes full path, not directory path
  - It looks like according to WLCG token spec (which is not very clear in this aspect) such requests should be allowed
  - It may be possible to restrict FTS to only copying on submission
    - Too many moving parts, so we did not use it

LHCb tokens during DC24

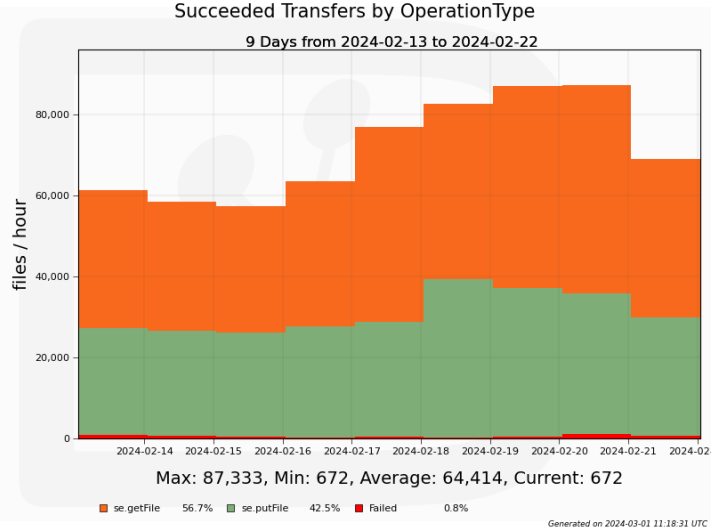
- GRIDKA, NCBJ, IN2P3 and PIC used tokens during writing part of the DC24
  - Tokens were used only on CERN->Disk link
- There were a lot of problems, namely:
  - Single point of failure
    - Poor performance of the IAM server affected all sites using tokens
  - Slow transfer submission
    - Since every transfer require at least 2 tokens, submission rate dropped (~0.5Hz on average)
      - Some links were starving as a result
  - Lack of proper monitoring
    - We were not able to see what's going on with the IAM server
  - Token refreshment problems
    - FTS is supposed to renew storage tokens before transfer starts if the lifetime left is short
    - Because of the number of requests LHCb IAM server was overloaded and responded very slowly
    - That resulted in many failed refreshments, and, eventually, failed transfers
      - The most affected sites were NCBJ and GRIDKA
  - Patched FTS Agent got stuck several times
    - Most probably because of the token related changes
- Reminder about pre-signed URL option that was agreed to be investigated after DC24
- Current token profile incompatible with LHCb's need

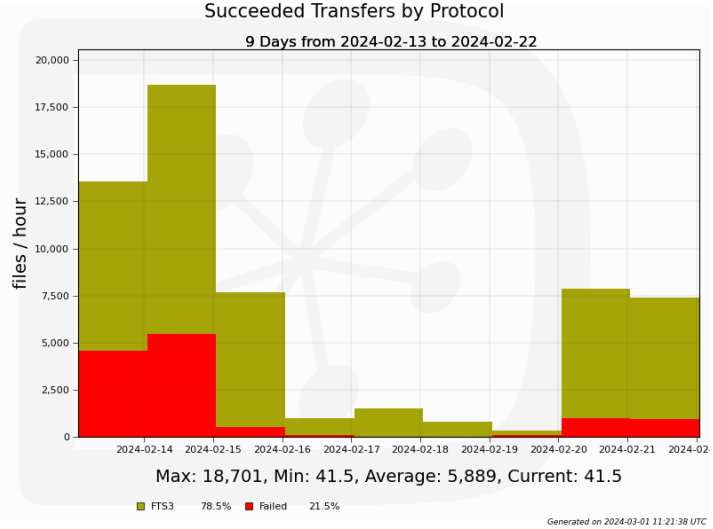
LHCb Tokens during DC24

- Efficiency of token-based transfers are much lower, compared to certificate-based



LHCb tokens during DC24





- Given the fact that most of the LHCb transfers comes NOT from FTS (see plots on the left), in “production mode” IAM server is going to get significantly more requests

Per-site results

Targets, GB/s			Achieved, GB/s			Ratio (achieved/target)		
Site	Write	Stage	EOS-Disk	Tape-Disk	Disk-Tape	EOS-Disk	Disk-Tape	Tape-Disk
CNAF	2.05	1.60	3.45	2.74	1.41	1.68	1.34	0.88
GRIDKA	2.74	1.66	2.50	1.65	3.35	0.91	0.60	2.01
IN2P3	1.53	1.20	2.56	1.42	1.05	1.67	0.93	0.88
NCBJ	1.02	0.89	0.953	0.602	0.798	0.93	0.59	0.90
PIC	0.51	0.40	1.21	0.553	1.05	2.37	1.08	2.63
RAL	3.96	2.40	2.68	2.64	3.28	0.68	0.67	1.37
SARA	1.15	0.80	2.77	1.39	1.17	2.40	1.20	1.46