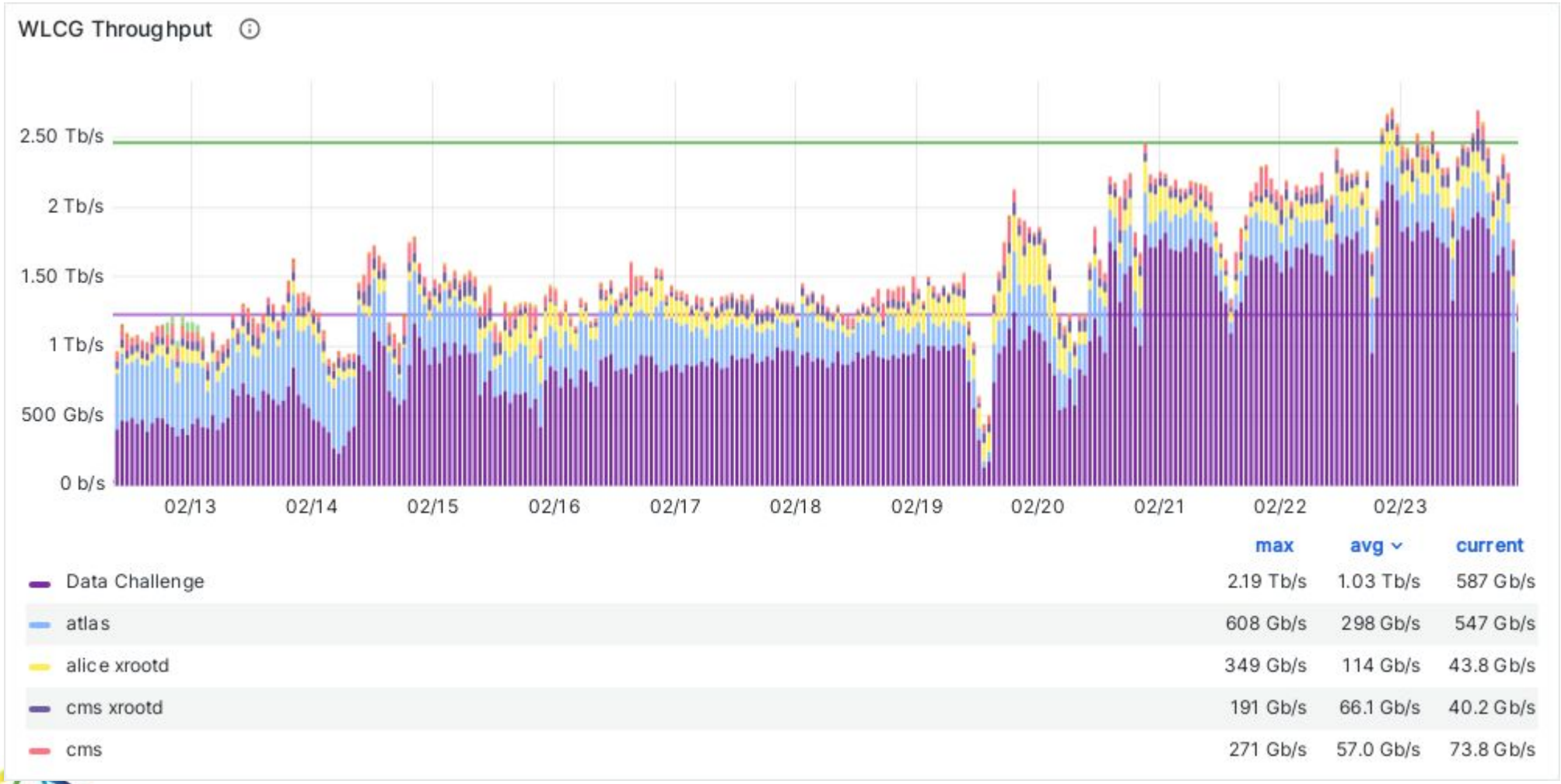
DC24 ATLAS retrospect

A. Forti, P. Vokac, M. Lassnig, S. McKee and the DDM team

DOMA general

6 March 2024

General assessment



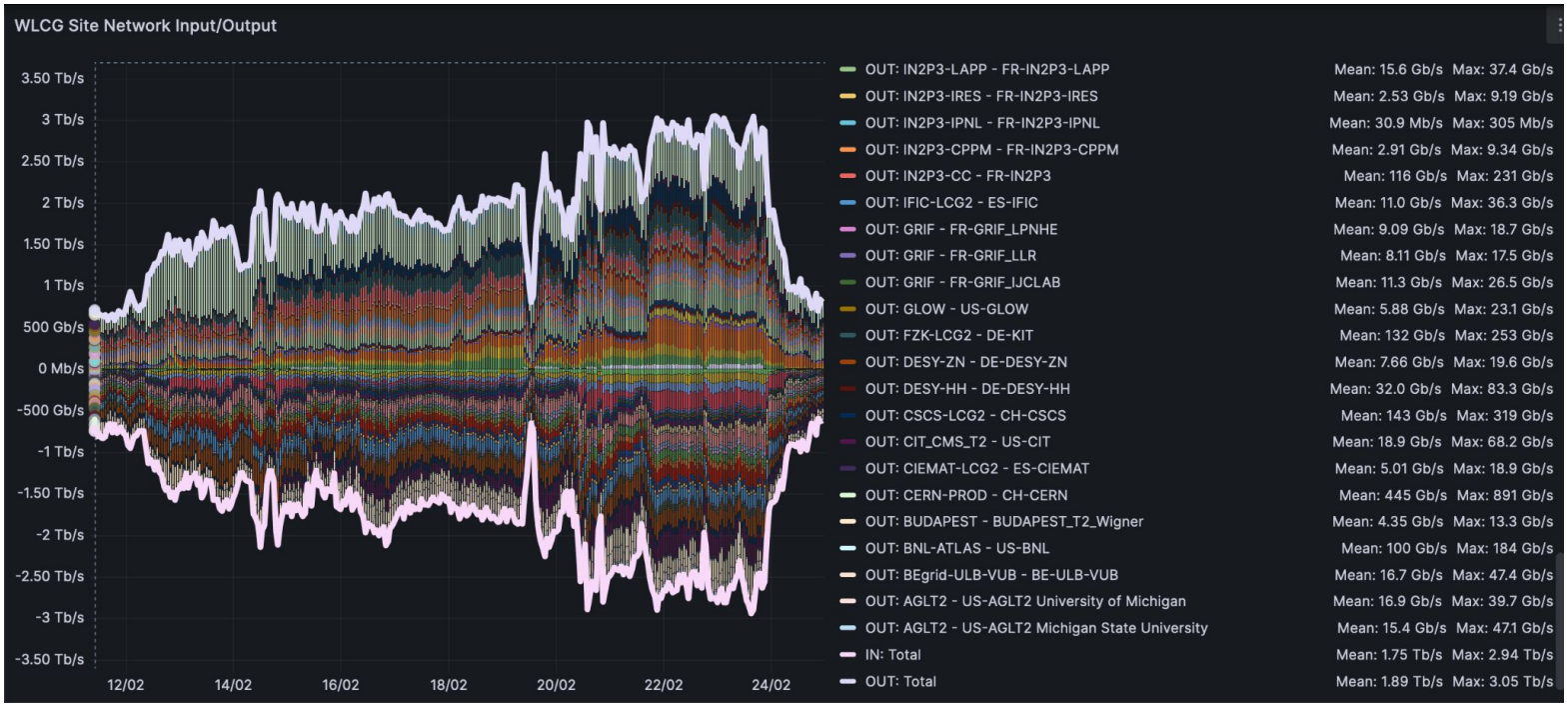
- Overall success! Generating that solid purple wall of increasing rates without affecting production much was a success!
- Getting to 2.5 Tb/s for ~9 hours was a success!

General ATLAS assessment

- Overall the challenge was a useful exercise which helped identify bottlenecks.
- Many problems, but not all of them, were the byproduct of how the challenge was run.
  - Injections on >1200 links every 15 minutes
    - ~2000 links if we include production
  - Short data sets lifetime 1h -> 2h -> 3h (with 3h space was running out in some places)
- But this still helped highlighting problems that wouldn't have been seen otherwise in the infrastructure

Network

- None of the bottlenecks were due to the network specifically
  - WLCG+non-WLCG traffic peaked at 3 Tb/s



- Some sites had the LHCOPN link down but had alternative paths in place. The backbone network wasn't the problem.
  - Bottlenecks were mostly due to storage configurations or storage hardware limitations

Sites

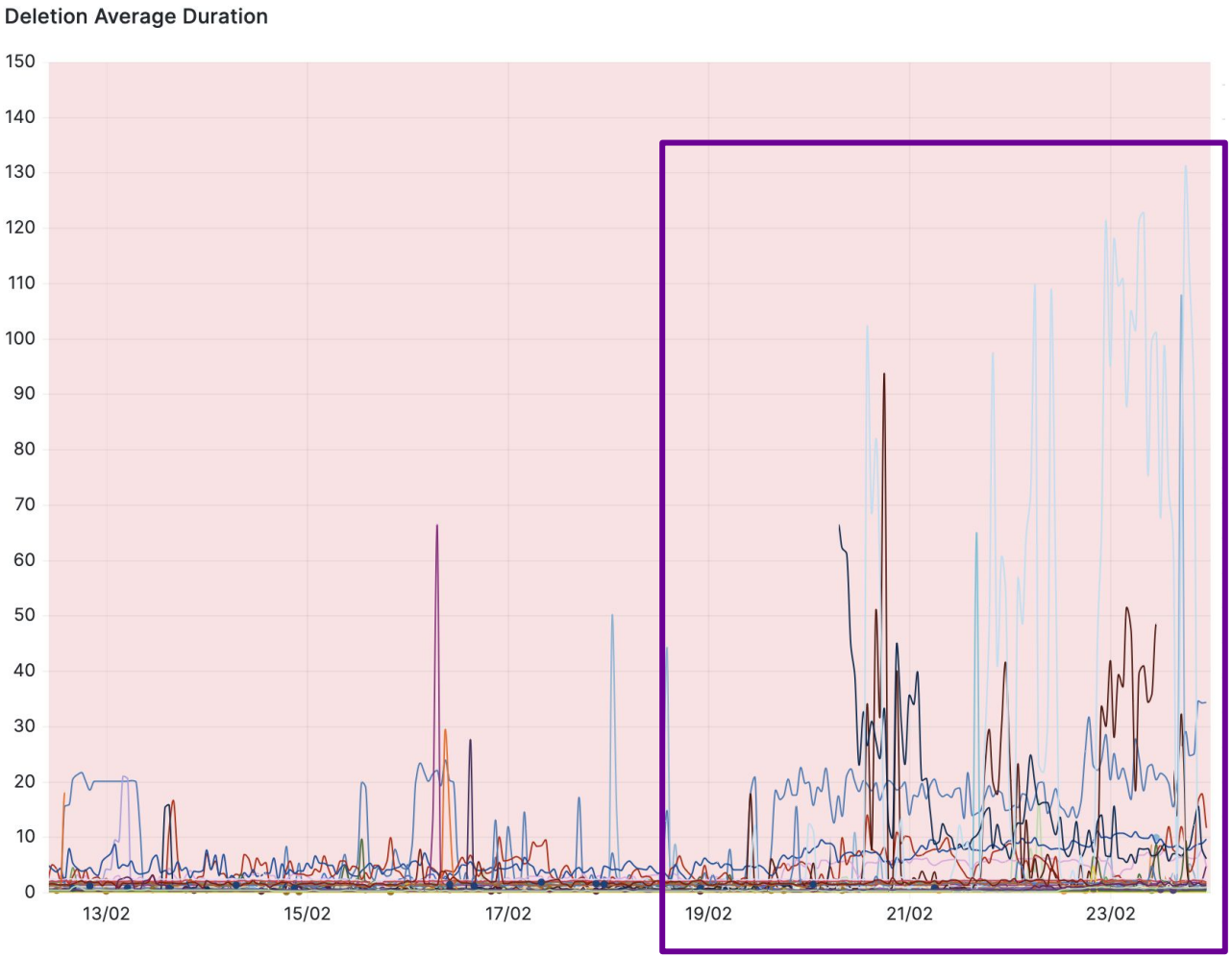
- Some sites struggled mostly due to storage limitations.
  - Either it wasn't possible to open enough parallel connections (IN2P3-CC)
  - or they had a problematic bug (NDGF),
  - or a bottleneck on the gateways due to hardware limitations (RAL).
  - Rates exceeding the expected values and storage not coping (INFN-CNAF)
- Some Tier2s also reported having problems
  - Lancaster had to double the number of gateways from 4 to 8,
  - SWT2 and other sites had a long wave of jobs in transferring state
  - MILANO and other sites saw a large amount of timeouts.
- Overall the number of problems reported, considering the amount of data pushed through, is reasonable
  - 17 problems were reported or GGUS tickets open (list in backup slides)

T0 - T1

- T0 export rates were not achieved
  - Particularly in the second week with increased number of transfers, compounded by the lack of prioritization in FTS, and slightly higher rates
- T0 exports test will need to be rerun before the WLCG/HSF workshop at DESY

T1 Site	Minimal (T0→T1)			Flexible (T0→T1)			Flexible (T0+T1→T1)		
	model	reality	[%]	model	reality	[%]	model	reality	[%]
BNL-ATLAS	60.0	25.9	43	68.4	21.2	31	82.1	57.1	70
FZK-LCG2	32.0	34.1	107	39.0	13.2	34	59.4	43.2	73
IN2P3-CC	38.0	36.4	96	44.2	1.4	3	59.1	21.4	36
INFN-T1	23.0	22.0	96	28.3	8.9	31	39.4	47.6	121
~~NDGF-T1~~	~~15.0~~	~~0.7~~	~~5~~	~~24.4~~	~~0.0~~	~~0~~	~~52.2~~	~~0.0~~	~~0~~
SARA-MATRIX	15.0	17.9	119	19.3	32.8	170	36.2	84.6	234
pic	11.0	13.8	126	13.3	4.2	32	18.1	35.7	198
RAL-LCG2	38.0	12.5	33	44.4	29.7	67	56.9	48.4	85
TRIUMF-LCG2	25.0	26.0	104	29.3	12.5	43	38.6	54.0	140
∑ (no NDGF)	242.0	188.6	78	286.3	123.9	43	389.8	392.0	101

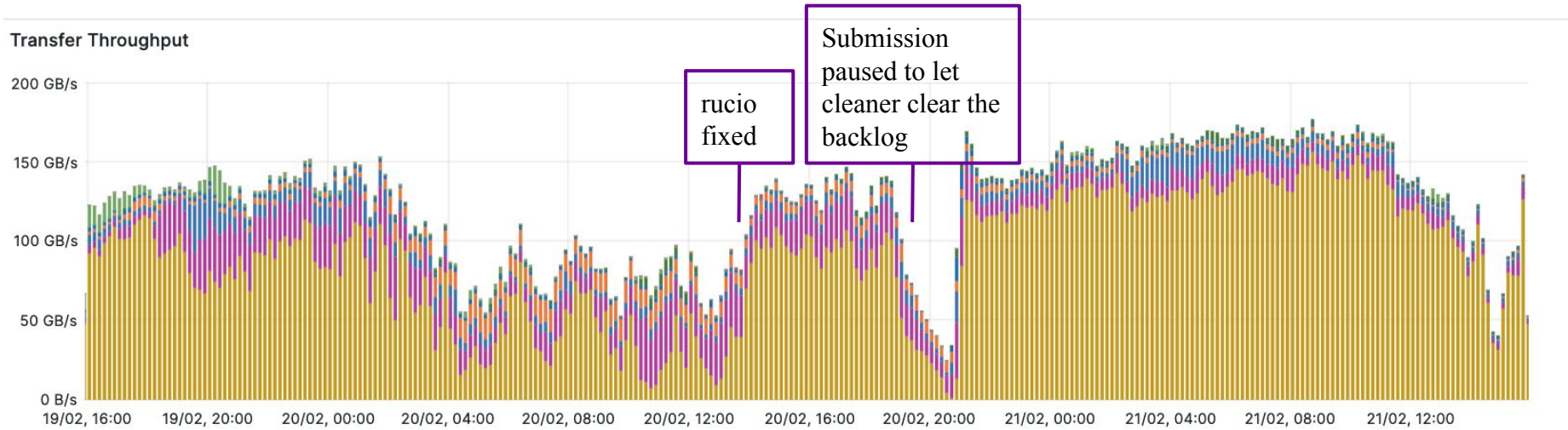
Deletions



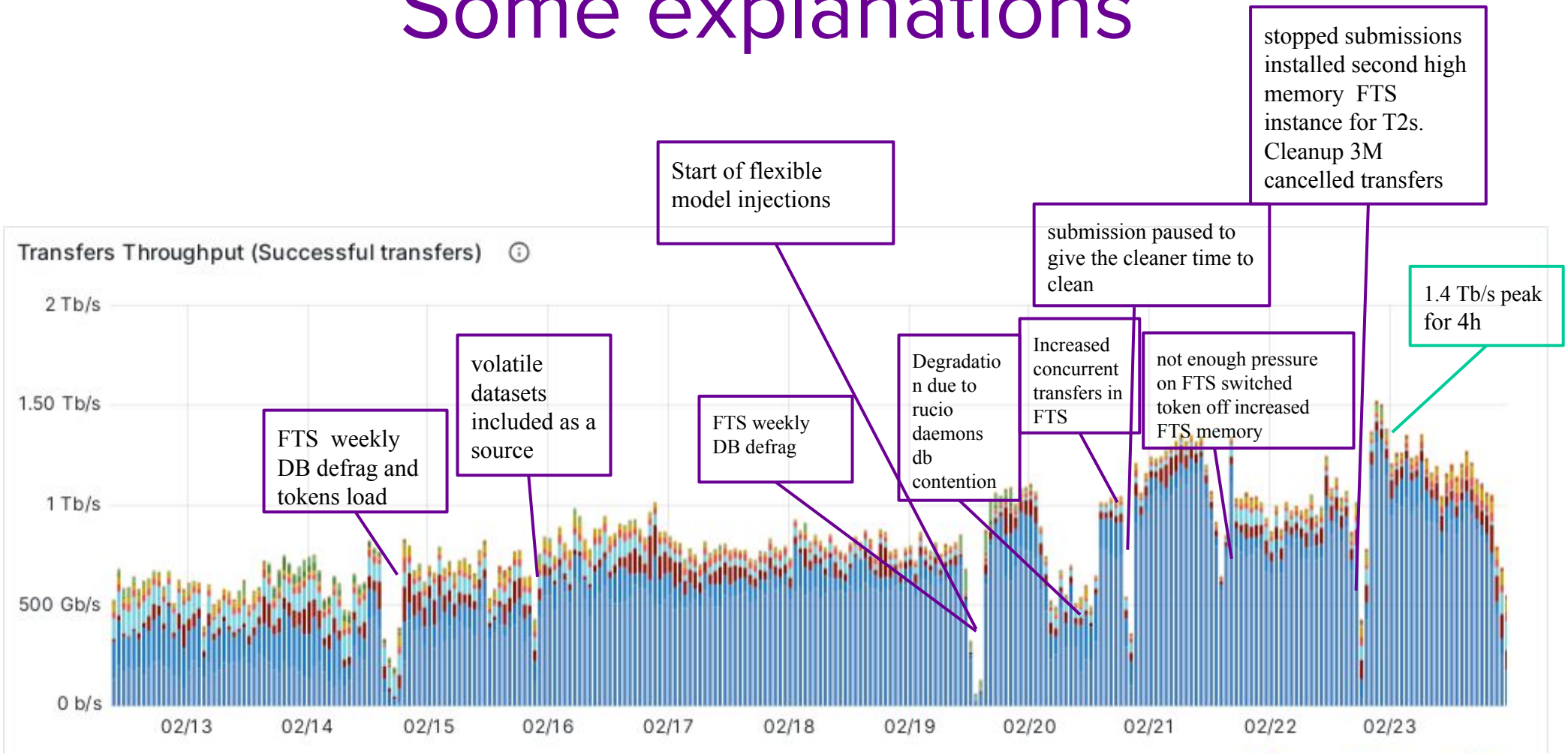
- Increase in deletion time particularly during the second week.
  - Problem was general but some sites had much higher times than others
    - It will need further investigation to see how it maps with storage types

rucio

- Behaved generally well
- Hot patched to avoid a database contention between the submitter and the cleaner daemons.
  - Patch will be added to production release
- Also increased the number of submitters and cleaners to exclude them from possible bottlenecks.



Some explanations



- Fair to say FTS wasn't expecting this high load
  - Next DC will need better common preparation
  - It will need a dedicated development roadmap

Conclusions

- Positive: system was definitely stressed and it cracked in places
  - Aim of a challenge is finding bottlenecks not only achieve rates
- Limitations of certain setups were highlighted and, where possible, corrected on the fly.
- In other places it will require more thinking
  - FTS may need some development identified during the data challenge
    - A process like the TAPE REST API went through with all stakeholders contributing to the requirements?
  - Some sites storage bottlenecks need to be corrected
- Request to sites to help with the more detailed retrospective by providing a report for their sites
  - T1s but also (larger) T2s

Backup

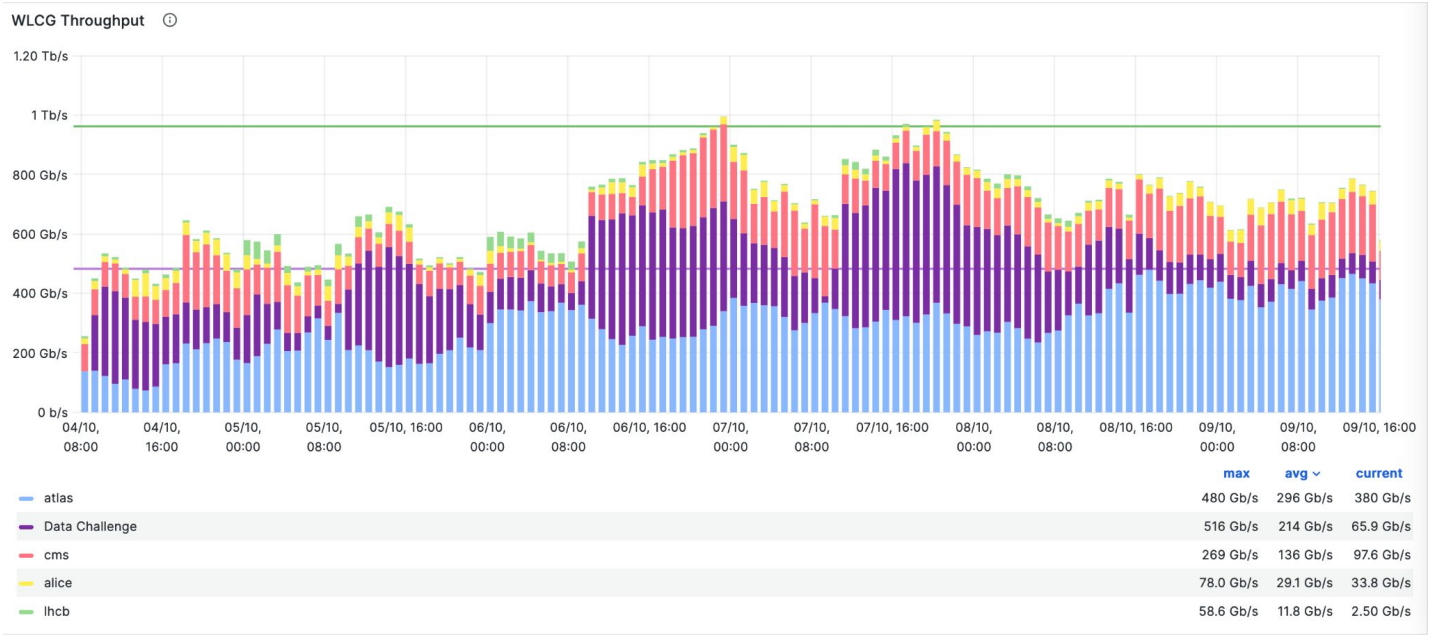
Sites reported problems

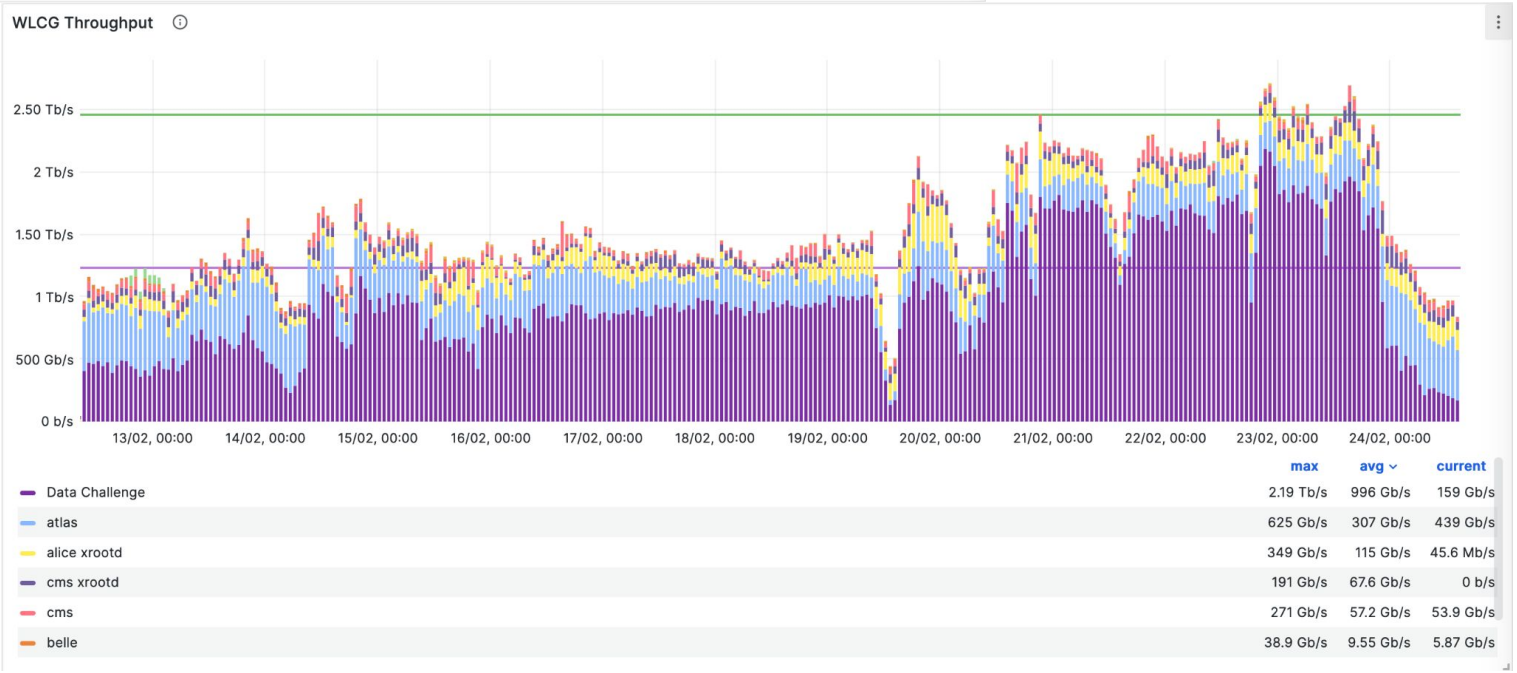
- Slow deletion at RAL - https://ggus.eu/index.php?mode=ticket_info&ticket_id=165358
- Worsened NDGF-T1 - https://ggus.eu/index.php?mode=ticket_info&ticket_id=164846
- Timeouts to Milano - https://ggus.eu/index.php?mode=ticket_info&ticket_id=165356
- SSL errors to CNAF - https://ggus.eu/index.php?mode=ticket_info&ticket_id=165355
- Timeouts to FZK - https://ggus.eu/index.php?mode=ticket_info&ticket_id=165393
- Timeouts at OU_OSCER_ATLAS - https://ggus.eu/index.php?mode=ticket_info&ticket_id=165362 , https://ggus.eu/index.php?mode=ticket_info&ticket_id=165379
- Expired tokens in the FTS causing problems at DESY-HH - https://ggus.eu/index.php?mode=ticket_info&ticket_id=165397
- Timeouts at UKI-SCOTGRID-GLASGOW - https://ggus.eu/index.php?mode=ticket_info&ticket_id=163552
- "Unexpected server error" to NIKHEF during pre-DC24 test - https://ggus.eu/index.php?mode=ticket_info&ticket_id=165263
- "Unexpected server error" to UKI-NORTHGRID-LANCS-HEP - https://ggus.eu/index.php?mode=ticket_info&ticket_id=165394
- Dark data caused by heavy load at TRIUMF-LCG2 - https://ggus.eu/index.php?mode=ticket_info&ticket_id=165343
- IFIC tickets is not blaming DC24 but errors stopping with the end of it - https://ggus.eu/index.php?mode=ticket_info&ticket_id=165395
- IN2P3-CC being overloaded and HC putting the site into test for lack of storage free connections
  - Cured by reducing the number of connections in FTS but this meant reduced rates
- SWT2 large wave of jobs in transferring state (concurrent with a wave of evgen jobs)
- FZK - QMUL slowed down transfers
- INFN-T1 - (one) stuck doors while retrieving JWK from IAM (GGUS:165355, STOR-1603)
- Some sites applied storage limits tuning, e.g. FZK (GGUS:165393), TRIUMF (GGUS:165364)…, there were few more and not all communicated with GGUS => for final report we should also ask sites what they observed (e.g. SARA internal? throughput saturated their links (GGUS:165359), INFN also observed huge traffic (GGUS:165355), we don't fully understand much higher throughput on some links, …).

FTS problems

- FTS couldn't cope with the amount of transfers we were putting through some of the reasons below
- Changes in transfer protocols: http doesn’t have threads like gridftp used to have
  - Large increase in concurrent transfers ➜ max increased from hundreds to several thousands per link/storage
- The weekly defragmentation of the database, i.e. a standard maintenance operation, blocked transfers twice
- Cancelled jobs were accumulating in the DB making it unresponsive
  - Should be removed automatically ➜ might be a problem of communication with rucio (?)
- Memory had to be increased on fts3-atlas
  - It was recognised the only way to scale right now is to add more memory - that is a valid choice too of course
- Had to install a second high memory instance on fts3-pilot and move all the T2s on the second instance to achieve the necessary rates
  - Before deciding to end DC24 few sites had been moved to FTS BNL too to spread the load further,
- While tokens have been a success story they were a secondary goal for ATLAS and had to be switched off to achieve throughput
  - They created a drop on the 14th,
  - Tokens refresh was switched off to ease the load on the 20th
  - Tokens were eventually switched off completely on the 21st because without refresh there were failures despite the token 6h lifetime
- The optimizer needs to be reviewed
  - Cycle eventually was taking 3 hours and couldn't be restored to a working state easily.
  - It wasn’t possible to switch it off
  - It could benefit with scaling with the number of active transfers
  - 2/4 of the optimizer settings are not useful without gridftp threads and could be eliminated
- Extensive manual tuning of links and storage to try to optimize the throughput
- FTS doesn't have any concept of priority of transfers other than per activity
  - Could have created two activities but within an activity it should be possible to prioritize links according to some weight in the configuration and faster links should be prioritized automatically

DC21 → DC24


WLCG Throughput
1.20 Tb/s
1 Tb/s
800 Gb/s
600 Gb/s
400 Gb/s
200 Gb/s
0 b/s
04/10, 08:00
04/10, 16:00
05/10, 00:00
05/10, 08:00
05/10, 16:00
06/10, 00:00
06/10, 08:00
06/10, 16:00
07/10, 00:00
07/10, 08:00
07/10, 16:00
08/10, 00:00
08/10, 08:00
08/10, 16:00
09/10, 00:00
09/10, 08:00
09/10, 16:00
max avg current
atlas 480 Gb/s 296 Gb/s 380 Gb/s
Data Challenge 516 Gb/s 214 Gb/s 65.9 Gb/s
cms 269 Gb/s 136 Gb/s 97.6 Gb/s
alice 78.0 Gb/s 29.1 Gb/s 33.8 Gb/s
lhcb 58.6 Gb/s 11.8 Gb/s 2.50 Gb/s


WLCG Throughput
2.50 Tb/s
2 Tb/s
1.50 Tb/s
1 Tb/s
500 Gb/s
0 b/s
13/02, 00:00
14/02, 00:00
15/02, 00:00
16/02, 00:00
17/02, 00:00
18/02, 00:00
19/02, 00:00
20/02, 00:00
21/02, 00:00
22/02, 00:00
23/02, 00:00
24/02, 00:00
max avg current
Data Challenge 2.19 Tb/s 996 Gb/s 159 Gb/s
atlas 625 Gb/s 307 Gb/s 439 Gb/s
alice xrootd 349 Gb/s 115 Gb/s 45.6 Mb/s
cms xrootd 191 Gb/s 67.6 Gb/s 0 b/s
cms 271 Gb/s 57.2 Gb/s 53.9 Gb/s
belle 38.9 Gb/s 9.55 Gb/s 5.87 Gb/s