WLCG/DOMA Data Challenge 2024

Final Report

Authors: Arora Aashay, Agostini Federica, Arsuaga Rios Maria, Balcas Justas, Balci Berk, Barisits Martin, Benjamin Doug, Betev Latchezar, Carder Dale, Chauhan Rahul, Christidis Dimitrios, Chudoba Jiri, Dart Eli, Davila Diego, Dewhurst Alastair, Ellis Katy, Fernández Casaní Álvaro, Flix Molina Josep, Forti Alessandra, Gardner Robert William, Garonne Vincent, Garrido Borja, Giacomini Francesco, Glushkov Ivan, Haen Christophe, Hoeft Bruno, Lassnig Mario, Lehman Tom, Litmaath Maarten, Luehring Frederick, Lukasczyk Mark, Manrique Andres, Mascetti Luca, McKee Shawn, Miccoli Roberta, Morganti Lucia, Murray Steven, Musheghyan Haykuhi, Nappi Antonio, Ozturk Hasan, Pacheco Pages Andres, Paparrigopoulos Panos, Pardi Silvio, Paspalaki Garyfallia, Patrascoiu Mihai, Perez-Calero Yzquierdo Antonio, Robinson Kate, Rogovskiy Alexander, Sapunenko Vladimir, Shah Asif, Timm Steven, Vianello Enrico, Vokac Petr, Wissing Christoph, Yang Xi, Zani Stefano

Editors: Lassnig Mario, Wissing Christoph

Executive Summary

The WLCG/DOMA Data Challenge 2024 (DC24) was executed to rigorously test the functionalities and capabilities of the Worldwide LHC Computing Grid (WLCG) in preparation for the High-Luminosity LHC. DC24 was the second in a series of increasing challenges and targeted 25% of the expected HL-LHC throughput. DC24 was structured to stress-test various data transfer tools and methodologies, optimise network configurations, and investigate potential limitations in our infrastructure. The challenge set forth aggressive targets: 1.2 Tbps for the minimal model focusing on Tier-0 Export to the Tier-1 centres, and 2.4 Tbps for the flexible model, including complex experiment data flows. The primary objectives of DC24 also included validating the scalability and performance of data management tools like FTS (File Transfer Service) and Rucio, and ensuring robust authentication mechanisms using tokens. During DC24, the Belle-2 and DUNE experiments executed network exercises as well. Although their throughput was orders of magnitude lower than the LHC experiments, many sites and network paths were shared, and no interference was observed.

The challenge yielded numerous significant achievements. The minimal model was easily achieved thanks to a multi-month preparatory effort involving various ramp-up challenges. The flexible model was reached during the second half of the challenge and sustained for multiple hours. The challenge also identified various performance bottlenecks, including issues with token refresh operations and database overloads. DC24 offered the first opportunity to gain operational experiences using token-based authentication for data transfers. About half of the transfers injected for the challenge used tokens already. Significant tuning and dynamic adjustments were essential to maintain high transfer rates during the challenge. While token-based transfers were successfully tested, significant issues related to token refresh operations led to timeouts and transfer failures, particularly at highly loaded sites. New network technologies for load balancing, guaranteed bandwidth, and congestion control were also successfully evaluated.

Most of the sites did not observe problems with their storage nor suffered from network saturations. A few sites identified bottlenecks in their local infrastructure and are now in a position to apply upgrades or tuning of parameters. Overall, the challenge was considered very useful.

Based on the experiences from the previous DC21, the monitoring capabilities were greatly improved before DC24. New capabilities to tag dataflows on the network were demonstrated and will allow enhanced monitoring in the future. Some small-scale differences between different monitoring systems were spotted during the challenge and are subject to further investigation.

DC24 was instrumental in identifying and mitigating various performance bottlenecks, configuration issues, and operational hurdles. The insights and improvements derived from this challenge are pivotal for future scalability and operational robustness of our infrastructure. The preliminary date of the next Data Challenge will be autumn 2026, targeting 50% of the HL-LHC traffic.

1 Table Of Contents

2 Introduction

With LHC Run-4, a significant increase in luminosity is anticipated for the ATLAS and CMS experiments. This increase is primarily attained through stronger focusing of the colliding proton beams, resulting in a higher number of proton-proton scatterings during each bunch crossing. Consequently, the experiments must manage more complex events, involving a greater number of particle tracks that require processing and storage. Additionally, in addition to the more complex events, the experiments also intend to increase the rate of events that will be permanently stored. These heightened demands notably impact the requirement for computing and storage resources, as well as network capacity for data distribution.

WLCG has been tasked with demonstrating the readiness of the globally distributed Grid infrastructure for the High-Luminosity LHC (HL-LHC) era, beginning with LHC Run-4, through a series of data challenges that progressively increase in data volume and complexity. As of the time of writing, the commencement of LHC Run-4, signifying the start of the HL-LHC era, is scheduled for 2029.

In 2021, a planning document for HL-LHC regarding network requirements was prepared based on the DAQ TDRs for HL-LHC by ATLAS and CMS, as well as observations from Run-3 for the ALICE and LHCb experiments. For the latter two, the network requirements for Run-4 are expected to remain at the level of Run-3. The planning introduces two main scenarios.

The **minimal model** encompasses solely the data export from CERN (Tier-0) to the Tier-1 centres, where the data are archived in addition to CERN. The modeling yields an aggregated bandwidth requirement of **4.8 Tb/s**, including typical overprovisioning to accommodate spikes and timely fluctuations in network demand.

The **flexible model**, while not detailed, accommodates data flux between all tiers, including reprocessing and production of MC events, in addition to the described data archiving. This results in an aggregated bandwidth requirement of **9.6 Tb/s**.

The data challenges progress with increasing rates, culminating in the final challenge demonstrating the required HL-LHC capabilities close to the start of Run-4. It's noteworthy that the start of Run-4 shifted from 2027, when the initial planning for a series of data challenges commenced, to 2029 by the time preparation for the 2nd data challenge began. Due to the adjusted timeline, the target rate of the 2nd challenge was slightly reduced to 25%, compared to the originally planned 30%.

3 Experiments

3.1 ALICE

ALICE successfully completed the DC24 exercise, transferring Compressed Time Frame (CTF) data from CERN's O2 buffer storage to T0 and T1 custodial storage elements. The transfers were done according to the already established data sharing rates for Run3 and Run4. Daily processing and analysis activities continued throughout the exercise, adding background load to storage elements and networks across all computing centre tiers providing resources for the experiment.

Crucially, data transfers ran uninterrupted throughout the DC24 period (February 12-23) and met or exceeded target average rates. No major errors occurred in any part of the transfer chain, and no interference was observed from other participating VOs.

3.1.1 Goals and metrics

The ALICE experiment's participation in DC24 coincided with the planned Compressed Time Frames (CTF) data transfer of the Pb-Pb data collected in 2023. The data was transferred to the custodial storage at CERN (CTA) and the 7 T1s (GridKA, CNAF, RAL, KISTI, CCIN2P3, NDGF, and SARA) supporting the experiment. The volume of data to be transferred was 33.7PB, shared as follows: two-thirds to CERN CTA and one-third to the T1s, as outlined in the computing model and corresponding to the amount of custodial storage provided by each computing centre. Table 3.1.1a illustrates the individual centres' shares of data and target transfer rates. The latter are determined to complete the data transfers in approximately the same time, depending on the data volume to be transferred to each custodial SE. The main metrics for success are the average transfer rate and its stability.

Computing centre	**Share of data volume in %**	**Data volume in PB**	**Target transfer rate Gbit/s**
CERN	66	22.2	80
GridKA	8	2.7	4.8
CNAF	11	3.7	6.4
RAL	1	0.3	2.4
KISTI	3	1.0	8
CCIN2P3	6	2.0	16
NDGF	4	1.3	10.4
SARA	1	0.3	2.4
Sum	**100**	**33.7**	**100**

Table 3.1.1a: *Participating computing centres, data volume, share and transfer rate for ALICE.*

3.1.2 Transfer tools, tuning of transfer streams and monitoring

The tools utilised for data movement include the JAliEn transfer scheduler, responsible for managing the transfer request queue, alongside the xrd3cp (Third Party Copy) XRootD transfer tools and the ALICE storage tokens for authentication and authorization. These tools, which are supported by all SE systems on the Grid, are currently in production use for ALICE data transfers and have not been modified for the DC24 exercise. Rate tuning was achieved through the adjustment of concurrent data streams, ensuring a consistent transfer rate to each of the target custodial SEs.

Monitoring is conducted through the MonALISA framework, which integrates essential storage and network monitoring tools, as well as data integration and visualization. The transfer data accumulated in MonALISA is also exported to the WLCG common transfer monitoring system.

3.1.3 DC24 operation, achieved rates

The average achieved rates during the period of PDC24 are illustrated in Table 3.1.3a. Target rates were not only achieved but also surpassed at all centres, with most of the target rates increased during the exercise at the request of the T1s to test the limits of the individual storage systems.

Centre	**Target rate GB/s**	**Average achieved GB/s**
CNAF	0.8	0.98 (+20%)
IN2P3	0.4	0.6 (+40%)
KISTI	0.2	0.25 (+22%)
GridKA	0.6	1.12 (+90%)
NDGF	0.3	0.35 (+15%)
NL-T1	0.1	0.25 (+150%)
RAL	0.1	0.58 (+500%)
CERN	*10*	*14.2 (+40%)*

Table 3.1.3a: Average achieved rates.

In total, the sum average target rate to the T1s was 3.75 GB/s, which is 40% higher than the target rate, and at T0 was 14.2 GB/s, also 40% higher than the target.

The plots below depict the time series of transfers for the different target sites over the entire period of data transfer, of which DC24 is a subset.

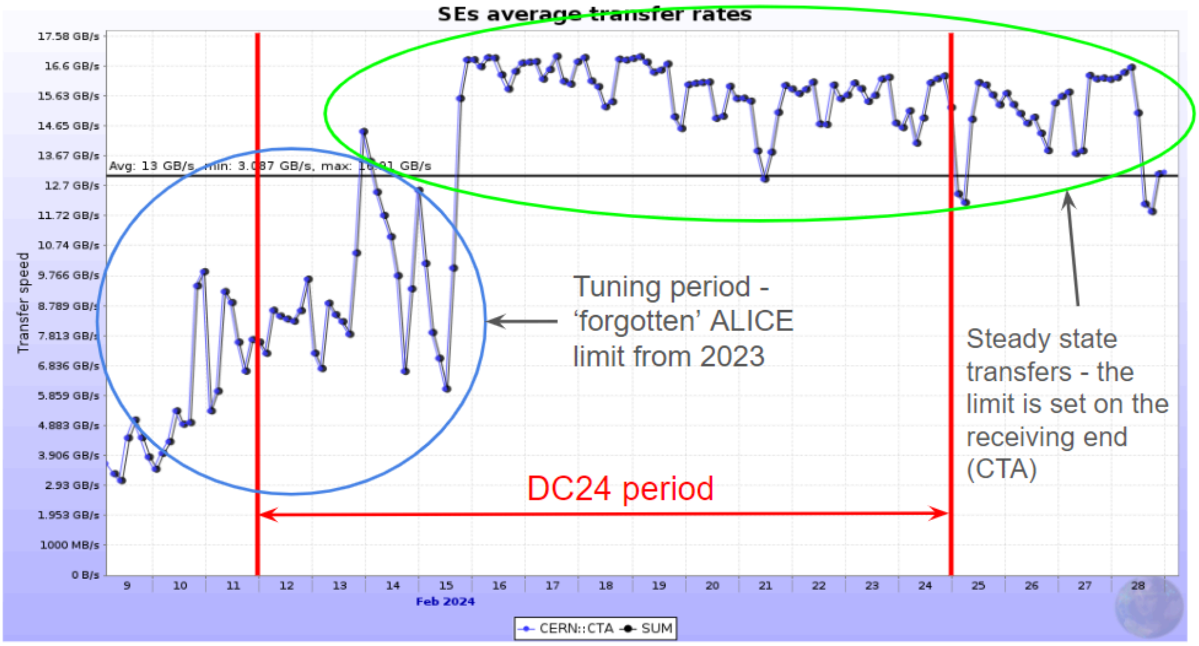


Figure 3.1.3b: Data transfer rate to T0 with DC24 period shown in red

The transfers to the T0 CTA required several days of tuning, primarily due to a transfer limit in the ALICE software that was not properly set. Once this issue was resolved, the transfers reached the target rate and remained stable throughout the DC24 period and beyond.

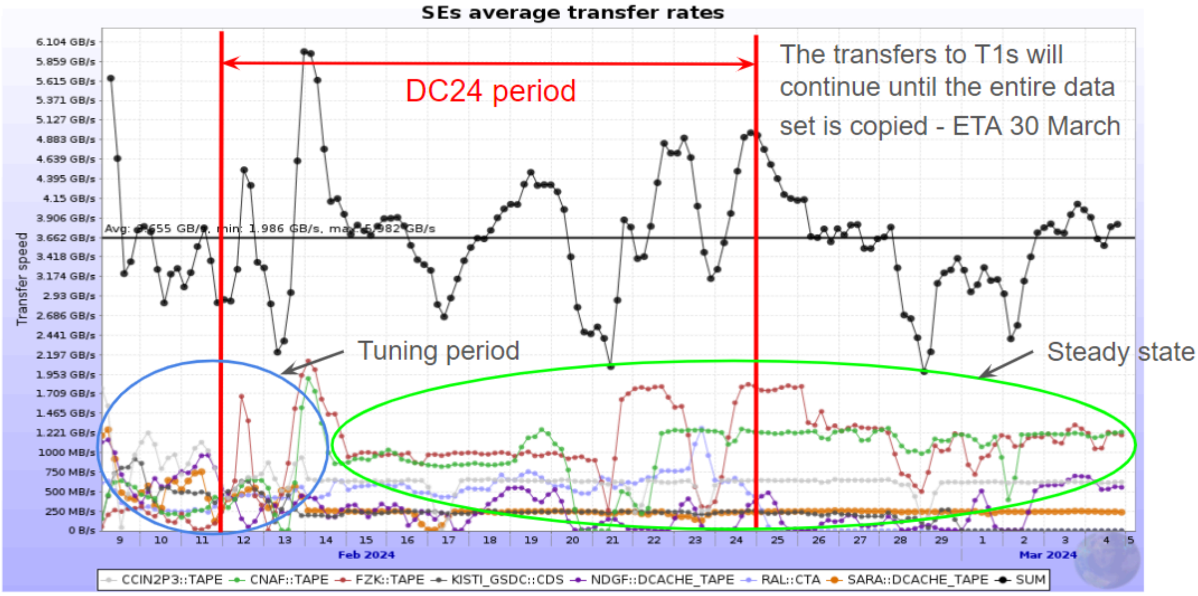


Figure 3.1.3c: Data transfer rate to T1s with DC24 period shown in red

After the initial tuning of rates to each T1, a steady state of transfers was achieved and maintained for the duration of DC24. It is worth noting that the rate was increased to CNAF and GridKA in the second part of the exercise to assess the performance of the storage under heavier load.

3.1.4 Incidents

The full list of incidents during the DC24 exercise is provided below:

1. The T0 rate could not achieve 10 GB/s in the first 3½ days due to a forgotten limit on the maximum active transfer threads in the ALICE transfer system.
2. A 3-hour interruption of transfers to GridKA occurred due to a dead XRootD service on a disk buffer, which was resolved by restarting the service.
3. A 24-hour interruption at CNAF was caused by a too high rate observed on the disk buffer, pinpointed to reads for md5sum calculation. This issue was resolved by adding an SSD buffer.

All of the above incidents were relatively trivial running issues, and no structural or software problems were identified.

3.1.5 Additional WAN traffic

As agreed for the DC24 exercise, in addition to the data transfers to the T0/T1s custodial storage elements, the usual wide area network traffic induced by other experimental workflows should continue and be present during DC24. The plot below illustrates the ALICE WAN traffic generated primarily by the analysis payload running on the Grid.

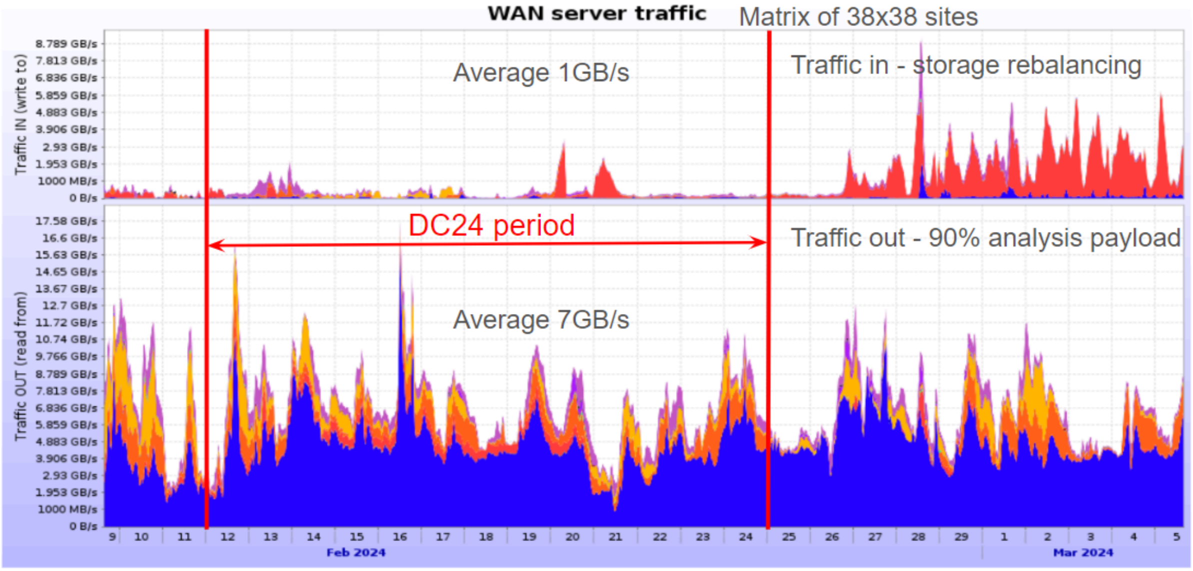


Figure 3.1.5a: Wide Area Network traffic induced by the ALICE payloads running on the Grid. This traffic is in addition to the data transfer shown in the previous plots. DC24 period is shown in red.

3.1.6 Conclusions

ALICE employed real data transfer in DC24 and exercised the entire custodial storage chain, including source storage, network, target storage, and transfer software. The target DC24 rates were exceeded by 40% at both T0 and T1s, with consistent transfers to all target storage elements and minimal intervention required at only two sites. No network issues were encountered, and no interference was observed from the activities of other VOs. Conversely, no adverse effects on other data management activities within ALICE were reported. All data was transferred without errors. We consider the DC24 exercise to be a success for ALICE.

3.2 ATLAS

Overall, the challenge proved to be a valuable exercise aimed at stressing all systems and identifying bottlenecks in the infrastructure, not only in the network but also in storage and services, which ultimately proved to be more problematic compared to the network. ATLAS achieved global rates it had never reached before, running at 1.4 Tb/s for several hours with a combination of injected and production traffic, similar to what was done during the 2021 Data Challenge (DC21). However, there was a difference from DC21 in the ratio of traffic types: while in DC21, production traffic was dominant and carried most of the workload, in DC24, injected rates were dominant. During the 12-day challenge, ATLAS transferred a total of 108 PB, of which 67.5 PB were injected.

3.2.1 Goals and metrics

ATLAS divided the challenge into three parts, progressively injecting more data into a more complex mesh of links as described by the agreed models in the introduction.

The rates were calculated for every link, taking into consideration the original HL-LHC Data Challenges planning document rates, the reported bandwidth of the sites, the bandwidth usable by ATLAS, and the active (injected rates) versus passive (only production traffic) participation in the challenge. ATLAS also calculated the number of TB necessary for 24 injections and the number of deletions, as well as the number of concurrent transfers necessary to sustain the calculated rates. Table 3.2.1a displays the numbers for the Tier1s.

Table: DC24 (src)	Site WAN (Gb/s)	Common to all scenarios	DC24 minimal scenario			DC24 flexible scenario			FTS active inbound / outbound
	Usable by ATLAS	T0 Export	Total Gb/s & bandwidth		Space [TB/24h] (deletions/hour)	Total Gb/s & bandwidth		Space [TB/24h] (deletions/hour)	
Site			∑ ingress	∑ egress		∑ ingress	∑ egress		
CERN-PROD	891	257.0	23.4	282.5	246 (3505)	88.9	392.8	937 (13330)	454 / 2037
BNL-ATLAS	400	60.0	84.5	67.1	892 (12681)	119.8	124.9	1263 (17964)	719 / 851
FZK-LCG2	144	32.0	55.9	35.5	590 (8386)	92.9	65.5	980 (13939)	473 / 410
IN2P3-CC	177	38.0	59.8	43.0	631 (8976)	93.5	77.7	987 (14032)	543 / 429
INFN-T1	62	23.0	36.3	26.0	383 (5447)	61.2	46.1	645 (9177)	230 / 209
NDGF-T1	149	15.0	44.6	23.3	471 (6692)	95.6	33.7	1009 (14345)	593 / 106
SARA-MATRIX	238	15.0	31.0	16.4	327 (4650)	60.1	30.2	634 (9020)	164 / 139
pic	85	11.0	17.1	12.5	181 (2570)	29.0	20.9	306 (4355)	141 / 150
RAL-LCG2	177	38.0	64.7	40.3	683 (9709)	92.8	81.0	978 (13915)	1595 / 663
TRIUMF-LCG2	100	25.0	38.2	27.8	402 (5723)	60.0	50.9	632 (8996)	322 / 434

Table 3.2.1a: Tier1s ATLAS numbers

A similar larger table was produced for the Tier2s.

These numbers were used to calculate the traffic per link, which in turn was used as input to generate the traffic.

3.2.2 Method to generate the traffic

The DC24 traffic comprised both production traffic and traffic generated using the same tools as production, using production datasets. Much of the original data in the selected threshold (average size of 5-7 GB, approximately 1000 files) originated from temporary datasets, which were automatically removed as computational tasks progressed. Eventually, the input datasets had to be refreshed every 12 hours to ensure a sufficient number of fresh datasets were available to sustain our injection rates.

A script was employed to generate Rucio rules and submit the transfers to FTS, which was updated several times, with the history saved in git. The tool proved to be highly stable, and for the next challenge, there is a proposal to integrate its functionality directly into Rucio. This would make it extremely straightforward for sites to adopt it for smaller tests between challenges.

One significant constraint in running a challenge of this size using real data is the limited amount of space available at sites. This necessitated DC24 to be executed with a rapid cycle of injections and deletions. Coupled with the large number of links to inject, this put considerable strain on the entire infrastructure. In numerical terms:

- Number of links to inject: 200 in the minimal model, >1200 in the flexible model
- Total links including production traffic: approximately 2000
- Injection cycle interval: 15 minutes
- Lifetime of datasets tested: 1 hour, 2 hours, and 3 hours
- When the lifetime of the datasets was set at 3 hours, some sites ran out of space, so eventually, 2 hours was used for most of the challenge.
- With 1-hour granularity in WLCG monitoring, it is necessary to wait a few hours to see the effect of configuration changes.

3.2.3 DC24 operation, achieved rates

The challenge, as agreed upon in WLCG/DOMA, spanned over 12 days, and ATLAS adhered to the following schedule:

	Monday	Tuesday		T0 export	
	12/02/2024	13/02/2024		blue: minimal scenario	
ATLAS	T0 → T1	T0 → T1		red: flexible scenario	
	Wednesday	Thursday	Friday	Saturday	Sunday
	14/02/2024	15/02/2024	16/02/2024	17/02/2024	18/02/2024
ATLAS	T0 → T1 ↔ T1 → T2	T0 → T1 ↔ T1 → T2	T0 → T1 ↔ T1 → T2	T0 → T1 ↔ T1 → T2	T0 → T1 ↔ T1 → T2
	Monday	Tuesday	Wednesday	Thursday	Friday
	19/02/2024	20/02/2024	21/02/2024	22/02/2024	23/02/2024
ATLAS	T0 ↔ T1 ↔ T1 ↔ T2 ↔ T2 ↔ T0	T0 ↔ T1 ↔ T1 ↔ T2 ↔ T2 ↔ T0	T0 ↔ T1 ↔ T1 ↔ T2 ↔ T2 ↔ T0	T0 ↔ T1 ↔ T1 ↔ T2 ↔ T2 ↔ T0	T0 ↔ T1 ↔ T1 ↔ T2 ↔ T2 ↔ T0

Table 3.2.3a: ATLAS schedule table

- 2 days exclusively for T0 export (9 links to inject)
- 5 days minimal (~200 links to inject)
- 5 days flexible (~1200 links to inject)

The results on a per day basis for each T0 and T1s, and for the sum of T2, are reported in Table 3.2.3b. Two observations can be made:

1. Some Tier1s were not performing particularly well in ingress.
2. The increase in the number of injected links at high rates in the second week degraded the rates of other Tier1s and Tier2s as well. This was due to a number of problems in the central services, which will be detailed in the incidents subsection.

T0 also appears to be predominantly red and orange, but this is a reflection of the first two points.

Day	Scenario	BNL-ATLAS		FZK-LCG2		IN2P3-CC		INFN-T1		NDGF-T1		pic	
		dst	src	dst	src	dst	src	dst	src	dst	src	dst	src
1	T0 → T1	25.68	N/A	29.76	N/A	35.6	N/A	21.84	N/A	12.56	N/A	10.48	N/A
2	T0 → T1	35.1	N/A	13	N/A	41	N/A	23.52	N/A	9.79	N/A	14.5	N/A
3	T0 → T1 ↔ T1 → T2	61.6	67.1	47.4	42.2	43.8	39.3	32.1	28	7.72	26.5	18.4	10.8
4	T0 → T1 ↔ T1 → T2	65.3	79.7	61.8	58.5	64.6	47.2	31.8	50.1	4.92	22.7	30.3	15.2
5	T0 → T1 ↔ T1 → T2	63	116	81.3	78.4	75.6	56.6	37.8	52.3	7.59	18.1	32.7	13.1
6	T0 → T1 ↔ T1 → T2	73.7	98.9	85	77.9	71.1	51	39.1	60	4.8	20.2	29.5	21.8
7	T0 → T1 ↔ T1 → T2	65.7	94	79.6	102	63.6	44.8	33.7	69.5	2.2	11.2	33.6	43.8
8	T0 ↔ T1 ↔ T1 ↔ T2 ↔ T2 ↔ T0	52.8	77.3	59.5	56.5	38.9	50.8	33.7	20	2.99	33.1	24.5	19.1
9	T0 ↔ T1 ↔ T1 ↔ T2 ↔ T2 ↔ T0	87.9	80.7	51.6	63.6	40.1	34.8	46.1	48.6	2.41	33	39.3	28.8
10	T0 ↔ T1 ↔ T1 ↔ T2 ↔ T2 ↔ T0	90	95.9	43.7	97.5	39.6	36.8	47.6	50.5	21.9	32.4	54	43.4
11	T0 ↔ T1 ↔ T1 ↔ T2 ↔ T2 ↔ T0	110	96.8	58.8	82.1	42.1	44.6	55.9	53.4	16.3	44.8	50.7	38.3
12	T0 ↔ T1 ↔ T1 ↔ T2 ↔ T2 ↔ T0	89.8	84.2	52.4	51.8	34	38.7	64.6	56.4	27.2	67.2	48	38.3

Day	Scenario	RAL-LCG2		SARA-MATRIX		TRIUMF-LCG2		T2 summary		T0 summary	
		dst	src	dst	src	dst	src	dst	src	dst	src
1	T0 → T1	12.16	N/A	12.64	N/A	19.92	N/A	N/A	N/A	N/A	188
2	T0 → T1	12.5	N/A	18.9	N/A	24.2	N/A	N/A	N/A	N/A	201
3	T0 → T1 ↔ T1 → T2	16.7	40.2	34.3	65.3	33.3	27.6	299	141	19.8	141
4	T0 → T1 ↔ T1 → T2	25.2	44.7	35.8	92.2	35.5	28.3	346	124	19.6	173
5	T0 → T1 ↔ T1 → T2	23.1	52.2	36.3	89.2	49.2	46.3	387	134	25.9	197
6	T0 → T1 ↔ T1 → T2	27.4	23.6	30.6	95.5	40.9	41.1	337	104	20.3	201
7	T0 → T1 ↔ T1 → T2	27.6	20.4	47.2	86.5	53.7	43.4	341	91.7	17.1	190
8	T0 ↔ T1 ↔ T1 ↔ T2 ↔ T2 ↔ T0	29.4	47.1	37.7	29.1	37.3	19.9	400	311	54	100
9	T0 ↔ T1 ↔ T1 ↔ T2 ↔ T2 ↔ T0	32.3	39.1	59.4	84	51.7	42.7	447	330	89.8	139
10	T0 ↔ T1 ↔ T1 ↔ T2 ↔ T2 ↔ T0	43.9	43	92.9	72.3	62.8	52.5	435	337	94.4	97
11	T0 ↔ T1 ↔ T1 ↔ T2 ↔ T2 ↔ T0	51.9	56	111	73.8	66.8	42.1	445	406	127	138
12	T0 ↔ T1 ↔ T1 ↔ T2 ↔ T2 ↔ T0	72.7	58.8	115	70.8	72.9	31.5	418	407	158	174

>90%
70-90%
50-70%
<50%

Table 3.2.3b: Daily results for each T1, T0 and the sum of T2

In Figure 3.2.3c, the achieved global rates for each of the intervals, including DC24 and production traffic, are depicted. The minimal model proved to be quite successful, achieving 805 Gb/s while the expected rates were 682 Gb/s, resulting in a 118% achievement. However, the flexible model experienced many drops and was not as successful, with an average of 1.0 Tb/s achieved compared to the expected 1.4 Tb/s, resulting in only 71% achieved over the five days. Additionally, 1.4 Tb/s was reached for only 4 hours on the last day. For consistency, the global rates for the first 2 days are also included, but these are not representative of the T0 export test, which ran throughout the entire 10 days, as shown in Figure 3.2.3d.

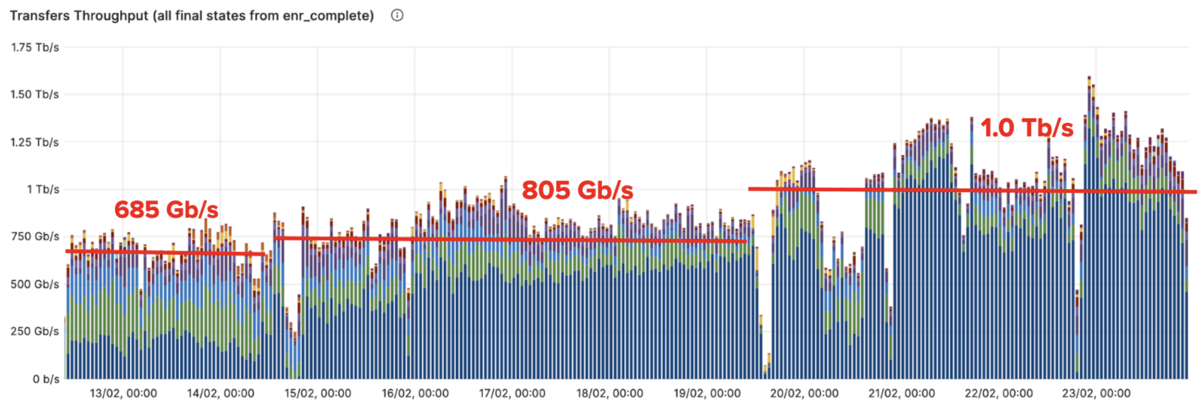


Figure 3.2.3c: Global rates for each of the DC24 intervals

The T0 Export activity never achieved the expected rates during the challenge and notably degraded during the last 5 days. The expected rate was 257 Gb/s, but we only achieved 192 Gb/s in the first 2 days. The average rate for the following 5 days was slightly lower but more consistent. However, T0 Export completely dropped during the last 5 days. This was mostly due to the inability of FTS to prioritise links within an activity. Since it operates on a FIFO basis, the Tier0 export to Tier1s within the DC24 activity was treated the same as Tier2 to Tier2 transfers. This is something to consider for the next Data Challenge, either by splitting activities even more, or by implementing smarter scheduling in FTS. Both approaches will need to be discussed.

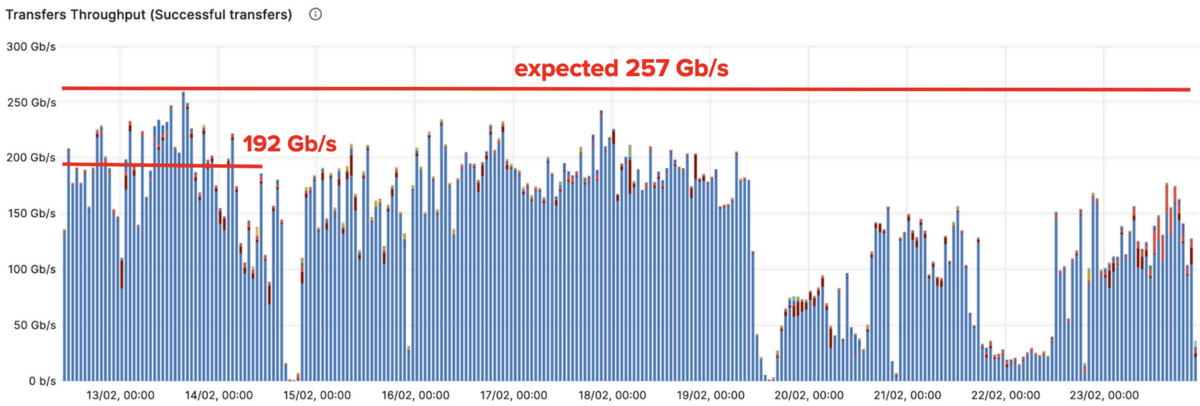


Figure 3.2.3d: T0 export during DC24

Due to this significant performance issue, we decided to re-run the T0 export test after the challenge, once at a time and then all together again over one day. In Table 3.2.3e, the results on a per T1 basis are reported. A general improvement can be observed. BNL and FTS encountered storage problems on the common day. RAL requested to test both on the single day and in the collective day, Antares, their tape system, which will be the real destination for T0 export without multi-hopping through disk in the next years. Antares had never been tested before with this kind of use case, and the test provided valuable information. SARA-MATRIX was testing an 800 Gb/s link directly from CERN and was injected with much larger rates than those requested by DC24 on the single-site test, while on the collective day, the rates were more in line with the other sites but still exceeded the rates. In any case, this demonstrates that T0 export was badly affected by not being prioritised rather than the sites having some problem. For future DCs, it is recommended to make the "T0 Export" part more realistic with respect to FTS transfer priorities and transfer paths closer to the TAPE system. Ideally, simulated "T0 Export" should reach the tape buffer, while additional reprocessing and production transfers should be directed to the disk with different DC activity and priority.

Site	T0 Export	DC24 best rates on day 1,2	% of expected rates	T0-T1 one T1at the time	% of expected rates	T0-T1 collective	%
BNL-ATLAS	60	31.5	53%	61.3	102%	5.36	9%
FZK-LCG2	32	26.4	83%	42.2	132%	7.84	25%
IN2P3-CC	38	43	113%	50.9	134%	57	150%
INFN-T1	23	19.3	84%	33.5	146%	36.08	157%
NDGF-T1	15	13.8	92%	28.2	188%	30.88	206%
SARA-MATRIX	15	12.2	81%	274.1	1827%	24.96	166%
pic	11	12.3	112%	18.1	165%	17.28	157%
RAL-LCG2	38	15	39%	27.2	72%	4.24	11%
TRIUMF-LCG2	25	23.9	96%	27.2	109%	35.2	141%

Table 3.2.3e: Post-DC24 rerun of the T0 Export tests

3.2.4 Incidents

There were several incidents during the challenge that resulted in quite a few drops in our rates. Figure 3.2.4a is associated with each drop or peak, most of which were caused by central services.

FTS was responsible for most of the drops in ATLAS. The FTS team did an admirable job of keeping it operational by tuning and expanding on the fly, reaching a number of transfers that they had never reached before. Detailed explanations will be provided in the FTS section, but here we want to list what happened according to our notes:

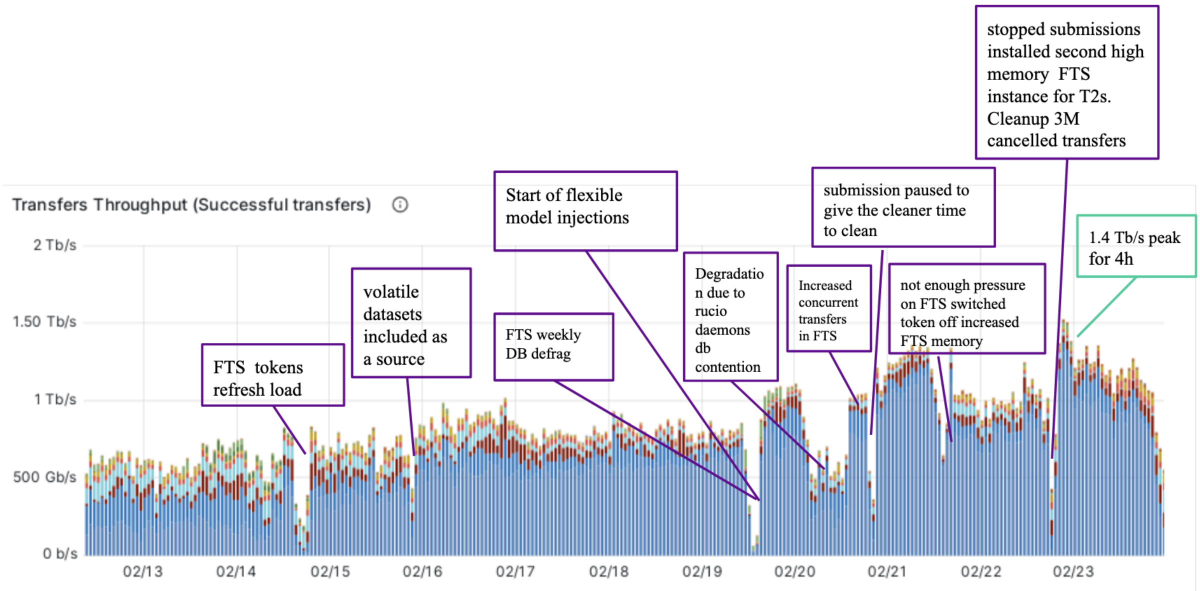


Figure 3.2.4a: ATLAS summary of rate drops motivations

- Unlike GridFTP, The HTTP TPC protocol doesn't support parallel streams for file transfers.
  - It is difficult to use increasing available and required network bandwidth.
  - A large increase in concurrent transfers caused the number of concurrent transfers to increase significantly, from hundreds to several thousands per link/storage, indicating a need for scalability in handling future increases.
  - This is an important aspect because in the future the number of transfers will be even larger. And the way we use FTS will not likely change.
- The weekly defragmentation of the database, i.e., a standard maintenance operation, blocked transfers on the 19/02/2024.
- Cancelled jobs were accumulating in the DB, making it unresponsive.
  - Eventually, 3 million transfers had to be manually deleted.
  - These should be removed automatically.
- Memory had to be increased on fts3-atlas.cern.ch
  - It was recognised that the only way to scale is to add more memory. There is an ongoing development to fix this.
- A second high memory instance was installed on fts3-pilot, and all the T2s were moved to the second instance to achieve necessary rates.
  - Before the end of DC24 a few sites had been moved to FTS BNL for load spreading.

- While tokens have been a success, they were a secondary goal for ATLAS and had to be switched off to achieve the target throughput:
    - They created a drop on 14/02/24
    - Tokens refresh was switched off to ease the load on 20/02/2024
    - Tokens were eventually switched off completely on 21/02/2024 because without refresh there were failures despite the 6h token lifetime and 2h Rucio rule lifetime.
- The optimiser needs to be reviewed:
    - The Cycle eventually was taking 3 hours and couldn't be restored easily.
    - It wasn't possible to switch it off.
    - It could benefit from scaling with the number of active transfers as well as a fixed number +-2. There is an existing FTS ticket for this.
    - It would benefit from automatically scaling down with the number of failures. This might already be implemented, but it wasn't clear if it was tied to the maximum number of transfers when we discussed it. If it is then it shouldn't depend on it.
    - 2/4 of the optimiser settings are not useful without GridFTP parallel streams.
- Extensive manual tuning of links and storage to try to optimise the throughput was necessary during the challenge and sometimes conflicting
    - We tried to tune the parameters at least 30 times, something could be automatized.
- FTS doesn't have a concept of transfers priorities other than per activity. Within an activity there is no prioritisation.
    - We could have created two activities for the challenge but within an activity it should be possible to prioritise links according to some weight in the configuration and faster links should be prioritised automatically.
    - It was pointed out that it is possible to pass a priority value when the transfer is created by Rucio, but this never came up before, during or after the challenge. It is something we should test, though this may be more difficult to change.

Rucio exhibited better performance until the second week, when a database contention between the submitter and the cleaner was observed, causing slow submission rates. A hot patch was applied to address this issue. Additionally, a further drop occurred due to pausing submission to allow the cleaner to catch up, after which submissions were resumed. To mitigate potential bottlenecks, the number of submitters was increased by a factor of three and the number of cleaners by a factor of four.

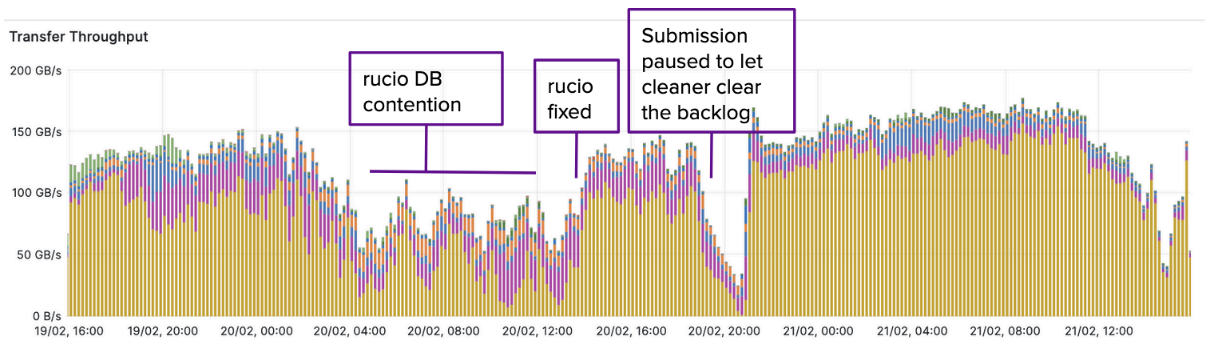


Figure 3.2.4b: Drops caused by Rucio internal DB contention.

Improving dc_inject logging to record the injected volume accurately is necessary.

3.2.5 Sites

Overall, 17 tickets were opened or reports were made about problems caused by the Data Challenge traffic. Considering the amount of traffic pushed through, this is reasonable.

At some Tier1s, internal staging and consolidation activities, which create traffic from disk to tape, were reducing the available bandwidth. This affected both the number of FTS available connections and the real bandwidth because the gateways used were the same. While internal site traffic is explicitly removed from the challenge plots, its effects are something that should be discussed, either to avoid them or to include them in future planning discussions.

RAL, NDGF, and IN2P3-CC experienced storage limitations that heavily impacted the rates at these Tier1s.

- RAL had to tune their gateways network cards. The tuning they applied is already affecting production traffic, which occasionally now sustains higher rates than theData Challenge.
- NDGF had a bug in dcache which heavily affected the incoming traffic. This issue has been resolved after the challenge, and the T0 export rates were twice what was expected.
- IN2P3-CC couldn't sustain more than 600 connections in ingress and 600 in egress. Particularly during the second week, this was not enough to sustain all the connections, and their rates dropped dramatically. They have already strengthened their storage to correct these problems.
- BNL performance was also lower than expected. This is still under investigation. They can reach much higher rates within the US, and in previous tests from CERN, there was a problem where the storage was outperforming the injections. However, during the challenge, this effect was less prominent. It could be that the higher latency from non-US sites requires a higher level of injections. It is certain that we need more tests to determine the exact cause.

3.2.6 Outlook

ATLAS found DC24 to be a positive experience that helped identify several bottlenecks. Most of the time, these were due to storage or central services rather than the network. This underscores the importance of testing the entire infrastructure, not just the network. Some of the problems have already been resolved by the sites or the services, while others will require further development. In preparation for the next challenge, it would be useful to run smaller tests more frequently. This can be achieved by streamlining dc_inject and integrating it into Rucio, allowing more operators to perform injections without needing superuser access to the database. Closer cooperation with other teams should also be part of the preparation. Another issue with the next challenge is that it will be at 50% of the HL-LHC rates, and it will be difficult to prioritise production. In particular, tape-to/from-disk transfers should be avoided during the challenge.

3.3 Belle II

3.3.1 Rationale

Belle II has joined the WLCG Data Challenge 2024 with the objective of exercising RAW Data replication in scenarios of maximum luminosity. According to current estimations, the experiment will collect 40TB/day, which will be stored at the KEK Data Centre. A secondary copy will be replicated to 6 RAW Data Centres as follows: 30% to BNL, 20% to CNAF, and 15% to IN2P3CC.

Considering that the average speed needed to transfer 40TB/day is 3.7 Gbit/s outbound at KEK to all RAW Data Centres, for the Data Challenge 2024, we have defined 2 targets:

- Minimum: The target speed to achieve is 3 times 3.7 Gbit/s = 11.1 Gbit/s.
- Maximum: The target speed to achieve is 5 times 3.7 Gbit/s = 18.5 Gbit/s.

In addition to the main test, some additional transfers were performed to check the routing among the RAW Data Centres.

3.3.2 Tools and data set

The Belle II Data Challenge was conducted using the production infrastructure, namely the Rucio instance and the FTS servers at KEK. To generate synthetic traffic on the network, a predefined dataset consisting of 8000 files, each 5GB in size, was stored at KEK and registered in Rucio.

For conducting the test, a Python script was developed to automate data replication management within the Rucio service. It operates on a cyclical basis to monitor the replication of data across various storage sites. During each cycle, the script checks for existing replication rules associated with specific datasets. If no rule is found for a particular site, it verifies the presence of data replicas at that site. If replicas are absent, a new replication rule is created to ensure data redundancy and availability. Furthermore, the script also manages deletion rules. When a replication rule exists but the replication is completed, the script triggers a deletion instruction.

As reported, the RAW Data Centres responsible to maintain the secondary copy of RAW Data are BNL, CNAF, DESY, KIT, IN2P3CC, and UVic. Table 3.3.2a shows the nominal value of

				Minimal x3		Maximal x5	
Storage Name	Site	Country	#5G Files	Ingress (Gbps)	Egress (Gbps)	Ingress (Gbps)	Egress (Gbps)
KEK-TMP-SE	KEK	JP	8000	0,0	11,1	0,0	18,5
BNL-TMP-SE	BNL	US	2400	3,3	0	5,6	0
CNAF-TMP-SE	CNAF	IT	1600	2,2	0	3,7	0
DESY-TMP-SE	DESY	DE	800	1,1	0	1,9	0
KIT-TMP-SE	KIT	DE	800	1,1	0	1,9	0

IN2P3CC-TMP-SE	IN2P3CC	FR	1200	1,7	0	2,8	0
UIVc-RAW-SE	UIVc	CA	1200	1,7	0	2,8	0

Table 3.3.2a: Nominal bandwidth to achieve per site.

During the test, UVic-RAW-SE was in the upgrading phase. However, we finally decided to maintain the same nominal goal by rerouting the relative part of traffic over the other Data Centres.

3.3.3 The Data Challenge

The tests were conducted between February 12th and 23rd, 2024, running transfers concurrently with those executed by LHC experiments. The Belle II Data Challenge was divided into 6 different phases, summarised in Table 3.3.3a. The first test (Test 1) was conducted under optimal conditions, utilizing the newest FTS server kek2-fts03.cc.kek.jp and achieving the highest peak throughput recorded during the DC (50 Gbps), with an average of 22 Gbps over more than 60 hours of continuous activity. Initially, we began with 500 concurrent transfers. However, in some cases, the faster storage systems were overloaded with all 500 active transfers. To better balance the load, we then limited the maximum number of concurrent transfers to 200 per destination, while still maintaining the 500-transfer limit at the source.

During tests 2 and 3, we switched to an older FTS server, kek2-fts01.na.infn.it, due to errors observed on kek2-fts03.na.infn.it. The FTS appeared less aggressive, never reaching the limit of 500 concurrent transfers. However, in both time windows, the minimum target was reached.

In tests 4 and 5, we tuned the script to run transfers among RAW Data Centres with the goal of creating a dataset for successive flow analysis, attempting to highlight which part of the traffic goes via LHCONE or via LHCOPN.

Finally, in test 6, we returned to the FTS server kek2-fts03.na.infn.it, reaching a maximum average bandwidth of up to 26 Gbps in the last 14 hours of activity during the peak of traffic generated by LHC experiments.

	DATE	Test	TOT	Peak (1h)	Average
1	12/02/2024 9:00 to 14/02/2004 23:00	KEK vs RAW DC (kek2-fts03 - v3.12.1)	606 TB/61h	50 Gbps	22,0 Gbps - Reached Max goal
2	15/02/2024 9:00 to 15/02/2024 16:00	KEK vs RAW DC (kek2-fts01 older)	39,9 TB/7h	25 Gbps	12,6 Gbps - Reached Min goal
3	16/02/2024 6:00 to 17/02/2024 19:00	KEK vs RAW DC (kek2-fts01)	194 TB/38h	24 Gbps	11,3 Gbps - Reached Min goal
4	19/02/2024 8:30 to 19/02/2024 21:30	KEK vs RAW DC + RAW DCs vs RAW DCs (kek2-fts01 and kek2-fts03)	80 TB/13h	27 Gbps	13,7 Gbps - Mixed traffic
5	21/02/2024 10:00 to 22/02/2024 9:00	RAW DCs vs RAW DCs (kek2-fts03 - v3.12.)	141 TB/23h	46 Gbps	13,6 Gbps - Mixed traffic
6	23/02/2024 0:00 to 23/02/2024 14:00	KEK vs RAW DCs (kek2-fts03 - v3.12.)	178 TB/15h	46 Gbps	26 Gbps - Reached Max goal

Table 3.3.3a: Summary of tests, grouped in 6 different tests.

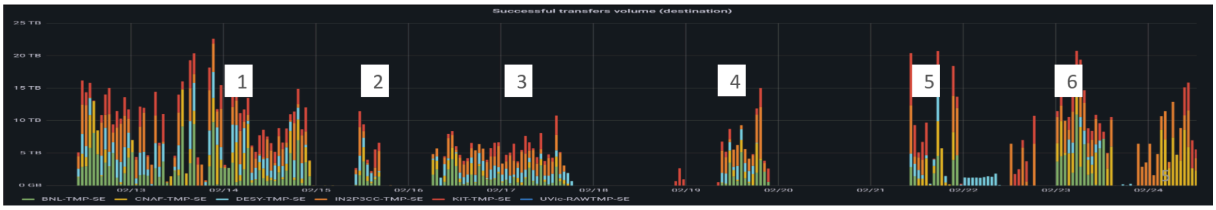


Figure 3.3.3b: Traffic in TB/h from the dashboard of Belle II Rucio monitoring system

3.3.4 Analysis of the peaks

The collected graphs in Figures 3.3.4a, 3.3.4b, and 3.3.4c demonstrate significant agreement among the different monitoring sources. For instance, multiple peaks exceeding 4 TB within a 10-minute interval recorded on the RUCIO monitoring system, equivalent to speeds surpassing 50 Gbps, correspond to peaks close to 60 Gbps observed on the KEK Network Monitoring Dashboard, encompassing all Data Challenge traffic and Belle II production activity. Similarly, on the GEANT dashboard, peaks exceeding 60 Gbps are evident, including non-Belle II traffic.

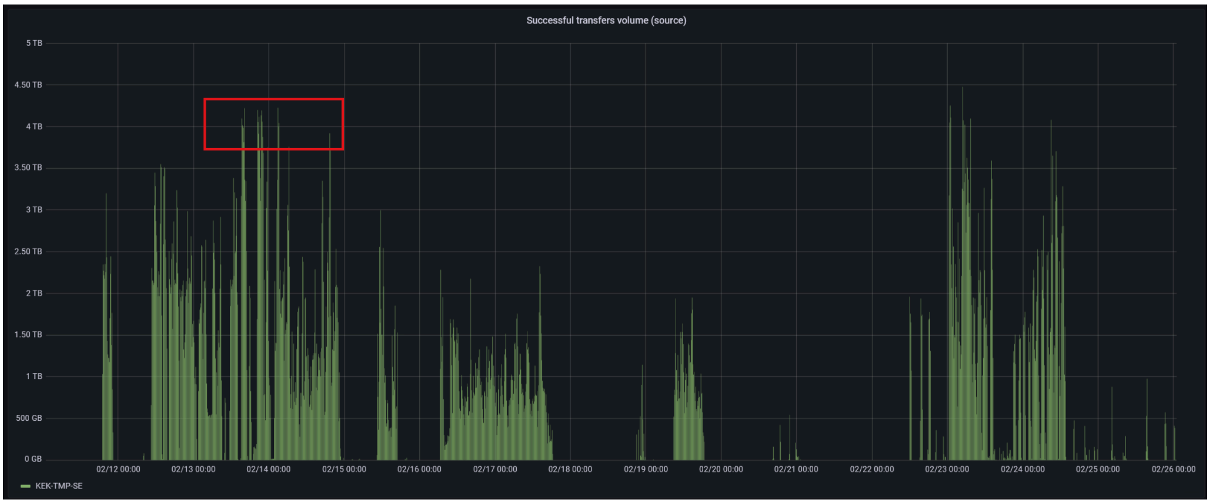


Figure 3.3.4a: Rucio Dashboard Isolating the functional test. Each bin represents 10 minutes.

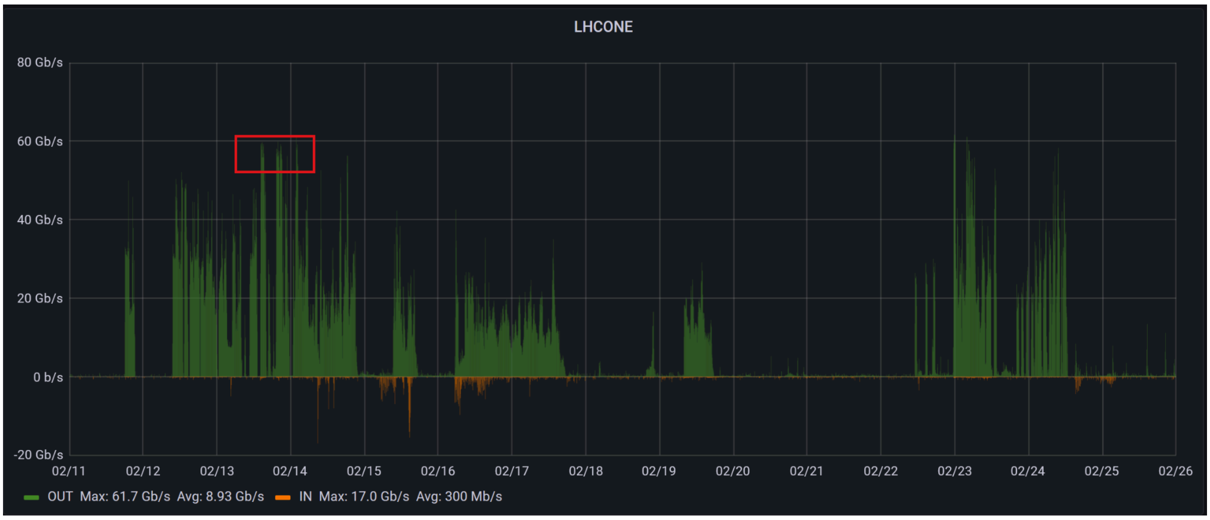


Figure 3.3.4b: KEK Network Monitoring Dashboard - LHCONE Link

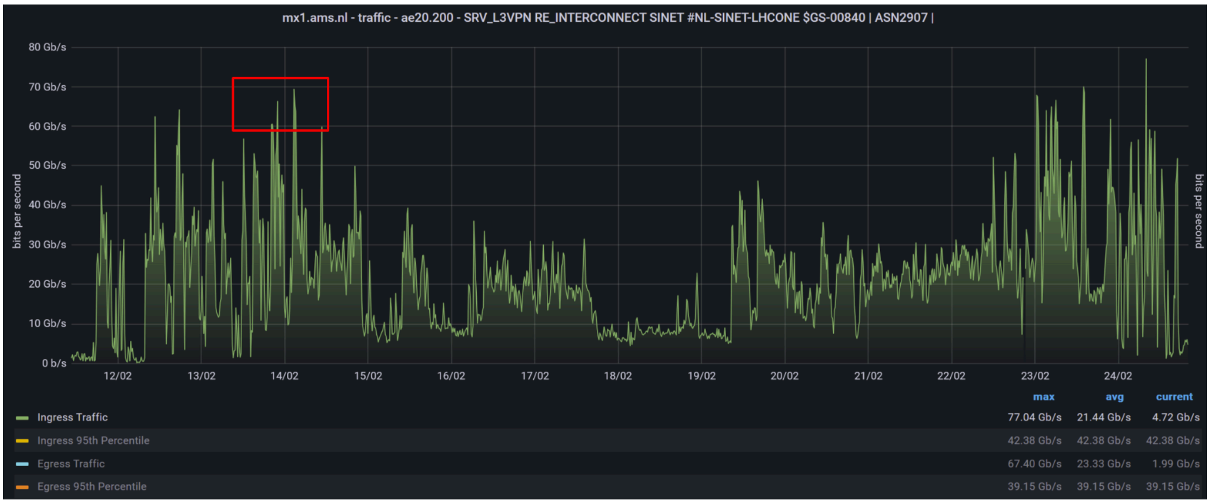


Figure 3.3.4c: SINET-LHCONE Peering from GEANT Dashboard

3.4 CMS

3.4.1 Summary

CMS executed a full programme of transfers, totalling approximately 60PB in volume over the twelve days of the challenge. Different scenarios were demonstrated during the first eight days, with the maximum target of 1Tbps set for the final four days. Production traffic continued throughout, seemingly unaffected by the challenge, and these transfers counted towards the desired rates. Almost all of the CMS T0, T1, and T2 sites were involved, totalling 53 possible disk sites, with a maximum of 200 links, plus production, tested at any one time. No tape endpoints were employed. Real data was used and was deleted and retransferred multiple times. DC24 injection rates hugely surpassed the usual FTS production transfer rates for CMS.

3.4.2 Transfer tools

The normal CMS setup of Rucio and FTS was used during the challenge, with the dc_inject tool employed to generate Rucio rules according to the target menu of the day. The menu was changed at approximately 10 am CET each day. For almost all transfers, the WebDAV protocol was used. JSON Web Tokens were enabled on around half of the sites as part of preparations for the challenge and were the authenticating method for approximately 50% of the total throughput. However, the granularity of token usage was low, so significant pressure was not put on the token issuer.

Note that in normal production, CMS streams a significant volume of data directly to running jobs via its AAA system, which is an XRootD federation. This was neither tested nor reported in DC24. The AAA scenario, mentioned in the next section, merely tried to account for the additional data transfer rate from particular sources. Nevertheless, it would be interesting to try to replicate the AAA scenario more faithfully in future challenges.

3.4.3 Scenarios tested

The following scenarios were tested for one or more days of the challenge:

- T0 export (CERN to T1s)
- T1 export (T1s to nearby, or non-nearby T2s)
- T1s<->T1s
- T2s->T1s
- “AAA” (CERN or FNAL to T2s)
- Special (sites that had requested additional rates, typically applied as T2s<->T2s

The requested rates are summarised as follows: T0 export - 250Gbps; T1 export - 250Gbps; “Production output” (T1s<->T1s and T2s->T1s) - 250Gbps; AAA - 250Gbps. Special rates were added as needed/when requested. The target on days 9-12 was 1000Gbps.

3.4.4 Schedule

The menu, i.e., list of source/destination pairs with requested rate, was typically changed each day at 10 am CET. There was an attempt to compensate for missing rates caused by sites that were unavailable. Towards the end of the challenge, additional rates were sometimes added at different times, as requested.

Date	12 Feb	13 Feb	14 Feb	15 Feb	16 Feb	17 Feb	18 Feb	19 Feb	20 Feb	21 Feb	22Feb	23 Feb
	T0 export	T0 export	T0 export	T1 export	T1 export	T1 export	T1 export	AAA	T0 export	T0 export	T0 export	T0 export
			T1 export		Prod. output	Prod. output	Prod. output		T1 export	T1 export	T1 export	T1 export
									Prod. output	Prod. output	Prod. output	Prod. output
									AAA	AAA	AAA	AAA
Scenario(s)	1	1	1,2	2	2,3	2,3	2,3	4	1,2,3,4	1,2,3,4	1,2,3,4	1,2,3,4
Rate (GB/s)	**31**	**31**	**62**	**31**	**62**	**62**	**62**	**31**	**125**	**125**	**125**	**125**
Rate (Gb/s)	250	250	500	250	500	500	500	250	1000	1000	1000	1000

Table 3.4.4a: CMS Run Schedule

The team found DC24 a challenging but rewarding experience. We learned a lot about running Rucio and FTS at a much higher scale, identifying a number of issues never seen before. Transfer monitoring was observed continuously to ensure issues were dealt with swiftly, particularly in the second week. However, even the first week was revealing as a learning experience, and also concerning token issues. Several sites used the continuously high data rate to make adjustments to their setup, and some made significant changes.

3.4.5 Results

Figure 3.4.5a shows the total data rates during the 12 days of the challenge, coloured by activity. Note that the total rates include even traffic not included in any scenario, such as data going to and from tape sites. The target rates for each day are indicated.

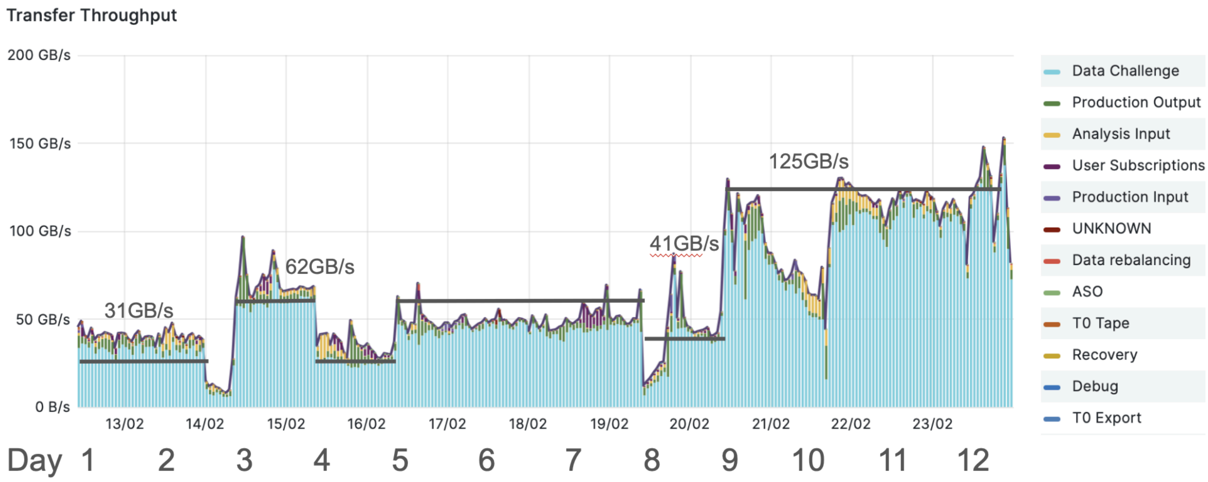


Table 3.4.5a: Total data rates

In the first two days, the 31GB/s rate for T0 export was consistently achieved. On day 3, double that rate was achieved with contributions from T0 export and T1 export, where transfers were made from T1s to their nearest T2s. On day 4, transfers were made only from T1s to non-nearest T2s. On days 5 to 7 (Friday-Sunday), the same menu was maintained for three days, continuing the T1 export scenario to different, non-nearest T2s, but also adding T1 to T1 and T2 to T2 transfers. On these days, the target rate was not achieved, and this was partly blamed on throttling of the number of parallel transfers by FTS. This FTS configuration was adjusted on the morning of day 8, before changing the menu.

On day 8, the target was 31GB/s of "AAA" transfers using CERN and FNAL as the only sources for injections, plus a specially requested full mesh of transfers between US T2 sites, summing to 10GB/s, giving a total target for the day of 41GB/s. For the remaining four days, the nominal target was 125GB/s, but in reality, some days towards the end of the challenge period did have additional rates at the request of particular sites wishing to test their network. This period was by far the most difficult to achieve and sustain the desired rates despite considerable effort to do so.

3.4.6 Issues

A number of issues were observed. Some were mistakes made by the CMS team, others were associated with FTS and Rucio. Figure 3.4.6a shows the main periods identified by CMS as unassociated with either sites or the network. These are the following:

1. FTS 'got stuck': This is related to the use of tokens in FTS.
2. Small files/blocks: The dc_inject tool was requested to start from small files rather than large files. This was an error made by the team and was corrected later in the day.
3. Insufficient Rucio deletions daemons running / Deletions too slow: If deletions cannot keep up with transfers then we cannot maintain a steady state. The number of Rucio deletion daemons was increased, and in some cases sites were assigned their own daemon. Some sites showed particularly slow deletions at times.

4,5. Brief periods without injections to allow for dataset refresh and deletions to catch up.

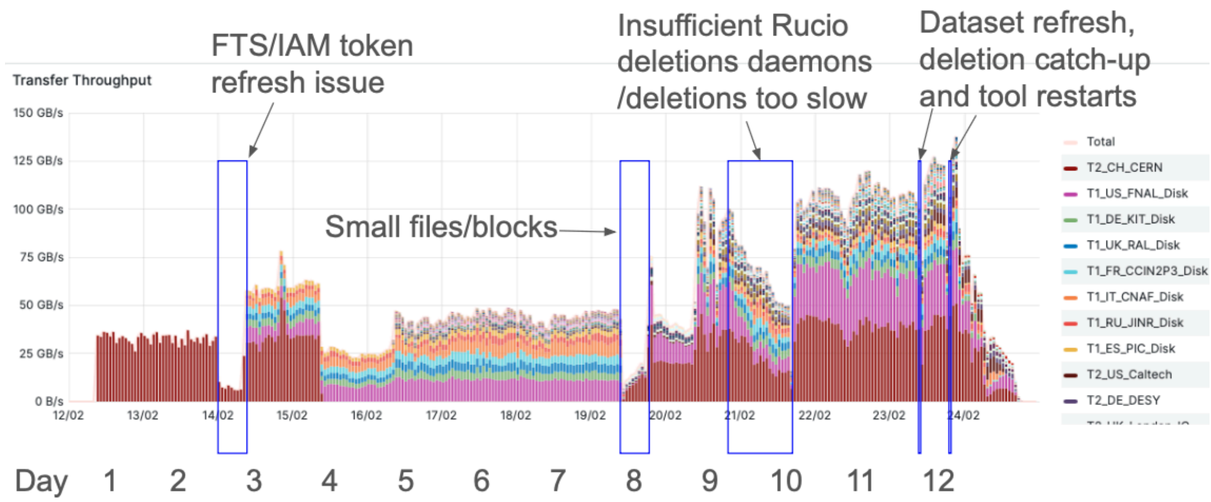


Figure 3.4.6a: Diagram of observed issues.

3.4.7 Analysis

Table 3.4.7a shows the results by day of the challenge, and split by transfer type where injections occurred on that day. It was not possible to split by the previously defined scenarios due to double-counting rates in multiple categories and being unable to split in the monitoring. For example, AAA transfers overlapped with T0 export (from CERN) and with T1 export (from FNAL). To allow for known problems, the blocked-out periods in Figure 3.4.6a are excluded. The values in Table 3.4.7a are ratios of the following:

- Expected rate: The desired rate for that transfer type injected using the dc_inject tool
- Observed rate: The average rate coming from FTS monitoring incl. injections and production
- $Ratio = \frac{Observed\ rate}{Expected\ rate}$

The colouring is as follows:

- Green - ratio is greater than or equal to 0.9
- Yellow - ratio is greater than or equal to 0.7
- Orange - ratio is greater than or equal to 0.5
- Red - ratio is less than 0.5

Day	T0 Export	T1 Export	T1s<->T1s	T2s<->T2s	AAA CERN to T2s	AAA FNAL to T2s	T2s->T1s	Special	Σ scenarios
1	1.11								1.11
2	1.05								1.05
3	1.11	0.99							1.05
4		0.83							0.83
5		0.79	1.09	0.59					0.79
6		0.86	1.10	0.56					0.81
7		0.83	1.11	0.59					0.81
8	1.29			0.92	1.18	0.98			1.08
9	0.61	0.54	0.77		0.96	0.74	0.73	0.90	0.70

10	0.83	0.62	0.67		1.05	0.67	0.70	0.83	0.75
11	0.71	0.64	0.80		0.92	0.60	0.85	0.84	0.73
12	0.82	0.70	0.92		0.86	0.67	0.89	0.22	0.71

Table 3.4.7a: Day results of the challenge

As previously noted, targets were easily reached in the first few days, involving transfers from CERN to the Tier 1s, and also from the Tier 1s to their nearest Tier 2s. During days 4-7, performance degraded slightly as T1 export transfers occurred between Tier 1s and non-nearest Tier 2s. This is not unexpected. One noticeable area of low performance in this period was in the transfers between Tier 2 sites. At least part of this effect was attributed to FTS throttling transfers between these sites, almost all of which use the 'default' settings, combined with a lack of attention over the weekend. Early on day 8, the number of concurrent transfers allowed by FTS for sites with the default settings was increased from 200 to 300, but unfortunately, the same test was not repeated again.

Day 8 focused on the AAA scenario with traffic from FNAL to American Tier 2s, CERN to the European Tier 1s, here as T0 export, and CERN to non-American Tier 2s. Due to a request from USCMS, a mesh of transfers between US Tier 2s was also added, and this result appears in the T2s to T2s column. All transfer types on this day reached their targets.

Throughout days 9-12, the overall goal was the same - to push hard on multiple scenarios simultaneously with a target of 1Tbps. From an operational point of view, there were two main concerns: not overloading the FTS, as was the case with ATLAS, and monitoring the rate of deletions. Results in Table 3.4.7a show clearly that many transfer scenarios were unable to be sustained as easily as in the days with fewer simultaneous scenarios. CERN remained a very reliable data source, as can be seen in the AAA CERN to T2s column. Collectively, the Tier 1s were under strain as they were pushed to the maximum as both source and destination, but improvements were made by the final day, as evidenced by the green entry for transfers between Tier 1s. Only the Special rates on the final day showed a red entry, indicating a very low observed measurement compared to the target. This can be attributed to a small number of sites.

A similar analysis was done per Tier 1 site as both a source and as a destination, with the results shown in Table 3.4.7b. Again, these results are the ratio of observed and expected data rates averaged per day of the challenge, with the five periods mentioned previously excluded.

		JINR		FNAL		IN2P3		RAL		PIC		KIT		CNAF	
Day	**Scenario**	DEST	SRC	DEST	SRC	DEST	SRC	DEST	SRC	DEST	SRC	DEST	SRC	DEST	SRC
1	T0 Export	1.42	N/A	1.13	N/A	1.09	N/A	0.76	N/A	1.18	N/A	1.16	N/A	1.17	N/A
2	T0 Export	1.46	N/A	1.12	N/A	1.10	N/A	0.50	N/A	1.17	N/A	0.94	N/A	1.17	N/A
3	T0Export, T1Export	1.31	0.62	1.08	0.88	1.33	1.03	0.72	0.99	1.18	1.06	1.10	1.06	1.28	0.93
4	T1 Export	N/A	0.37	N/A	0.91	N/A	1.12	N/A	0.76	N/A	1.05	N/A	0.95	N/A	1.00
5	T1-Export, Prod-out	1.18	1.72	1.15	0.87	1.25	0.89	0.98	1.01	1.21	1.09	1.23	0.77	1.17	0.77
6	T1-Export, Prod-out	1.14	2.42	1.18	0.88	1.47	0.88	0.72	0.81	1.17	1.03	1.19	0.76	1.18	0.95
7	T1-Export, Prod-out	1.19	2.19	1.15	0.87	1.22	0.87	0.81	1.04	1.20	0.98	1.21	0.73	1.16	1.02

8	AAA	1.30	N/A	N/A	1.10	1.39	N/A	1.31	N/A	1.31	N/A	1.70	N/A	1.32	N/A
9	All	0.38	0.34	0.87	0.84	0.57	0.57	0.95	1.02	1.25	0.86	0.86	0.56	0.65	0.25
10	All	0.70	0.34	0.98	0.74	0.58	0.65	0.56	0.99	0.70	0.66	1.03	0.98	0.63	0.28
11	All	0.63	0.33	0.91	0.73	0.43	0.76	0.77	1.05	1.09	0.84	0.91	1.09	0.69	0.24
12	All	0.40	0.54	0.92	0.86	0.89	1.00	0.85	1.15	1.21	0.87	1.13	0.89	0.78	0.29

Table 3.4.7b: Analysis of activities per Tier-1

It is evident that the first 8 days did not pose significant challenges for many of the Tier 1 sites. The OPN connection from CERN to RAL was disrupted during the initial 4 days, and JINR also encountered network issues intermittently. As depicted in Table 3.4.7a, it was the final 4 days that truly pushed the limits of the Tier 1s. Mitigating circumstances are outlined by the Tier 1 sites in Section 6. One notable oversight on the CMS side was the FTS configuration for IN2P3, which evidently failed to update on day 10. This was reattempted in time for day 12, and the subsequent change in performance is apparent.

3.4.8 Outlook

DC24 was a positive experience for CMS; we set ambitious targets and achieved them, although only for short periods at the highest rates. We learned a lot about scaling the system. The extensive pre-tests were useful for preparing the team and allowing the Tier 1s to adjust their setup in advance. There's much here that can prepare us for future, even more ambitious challenges. Next time, it would be interesting to focus more on the Tier 2s, use tokens more fully, and better simulate the “AAA” transfers. It would also help with the smooth running of the next challenge to force Rucio to delete some unlocked data before the start of the challenge to give deletions some additional headroom.

3.5 DUNE

3.5.1 Introduction

DUNE (Deep Underground Neutrino Experiment) is a neutrino experiment studying neutrino oscillation parameters (mass ordering, matter vs antimatter asymmetry, unitarity), proton decay, supernova neutrinos, and more. Ultimately, there will be four very large LAr TPC (17 kT) at 4850 ft underground in Lead, SD (Homestake Mine) with a near detector onsite at Fermilab being designed (3 sub-detectors, two that move). There are two prototype detectors in a test beam at CERN - (ProtoDUNE Horizontal Drift (HD) - ProtoDUNE Vertical Drift (VD)). Figure 3.5.1a shows how the neutrino beam will be produced at Fermilab (FNAL) and sent to the Sanford Underground Research Facility (SURF). The data will be sent from the detectors underground at SURF to the surface and then onto FNAL where the far detector raw data and the near detector (located on the FNAL site) raw data will be sent around the world for processing and data archiving (tapes). The SURF site is still under construction. Until it is available, the SURF site will be emulated by other sites in testing.

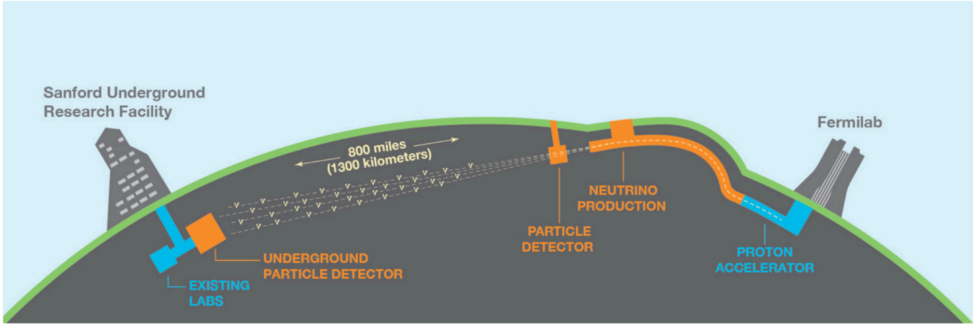


Figure 3.5.1a: DUNE Experiment Setup.

3.5.2 Initial Test plan

Initially, DUNE planned to perform three tests during DC24.

The first test - Far Detector (FD) raw data to archival storage - was to simulate the archival of 25% of the raw data rate from the SURF to FNAL, where one copy of the data would be archived to tape and a second copy to archival sites in Europe. This translated to a data rate of 1 GB/s from SURF (BNL was to be used as a stand-in source of FD raw data) to FNAL and then from FNAL to Europe. Prior to DC24, tests were performed between BNL and FNAL that achieved a sustained rate of 3 GB/s.

The second planned test was FD raw data keep-up processing. This is the most significant test for DUNE at this time because in June 2024, the protoDUNE HD/VD experiment will take beam data at CERN and DUNE plans to process data at the same rate as it is taken. During this test, the input data was to come from two sites, one in the US (FNAL) and one in Europe (PIC). It was planned to maintain continuous processing workload at distributed sites commensurate with 25% of the nominal "FD" raw data rate (1 GB/s). Compute resources across sites in Europe and North America were to be used for processing. The DUNE workflow system (justIN) was to be used to match the locality of jobs with the locality of data at nearby RSEs - i.e. US sites pull from FNAL and European sites pull from PIC. Prior to the DC24, the input dataset needed for the keep-up processing test was sent to PIC.

The third planned test was the most network-intensive (3.5 GB/s) for 4 hours and was to simulate the transfer of SuperNova Burst raw data from SURF (BNL was to be used as a stand-in) to FNAL and onto additional RSEs around the world.

3.5.3 Results

Test #1 - "FD" Raw Data to archival storage was not performed because of the backlog of output data being returned to FNAL as part of the output of the "FD" Raw Data keep-up processing.

Test #2 - "FD" Raw Data keep up processing. Since there was production work to do, it was decided to use the far detector horizontal drift and the far detector vertical drift Monte Carlo data reconstruction using the DUNE workflow system (justIN) to emulate the processing chain expected during protoDUNE HD/VD and ultimately DUNE data taking. This sample was chosen because this MC needed to be processed for analysis by DUNE collaborators and it somewhat emulated the type of processing expected during keep-up processing of ProtoDUNE-HD data. This is more I/O intensive than anticipated keep-up processing because the input file sizes were smaller, 2 GB per file vs 6-8 GB per file, and the increased number of output files (2 - HD MC and 4 - VD MC vs 1 for raw data reconstruction). The HD MC reconstruction processed produced one large file (roughly the size of the input, about 2 GB per file) and one small file (histograms - hundreds of MB per file). The VD MC reconstruction produced one large file (roughly the size of input, about 2 GB per file) and three small files (less than 300 MB per file). Each of the MC samples was divided into 6 sub-samples (each for a different initial neutrino type). During the processing, each MC sub-sample was further divided into justIN workflow tasks of 5000 input files each. Figure 3.5.3a shows the number of running DUNE jobs during most of the keep-up processing test. DUNE did not have the means to collect XRootD transfer information from the jobs since the XRootD monitoring is not enabled in RSEs use, and DUNE has not yet enabled the same detailed process monitoring from each job (i.e. prmon) used by ATLAS and CMS.

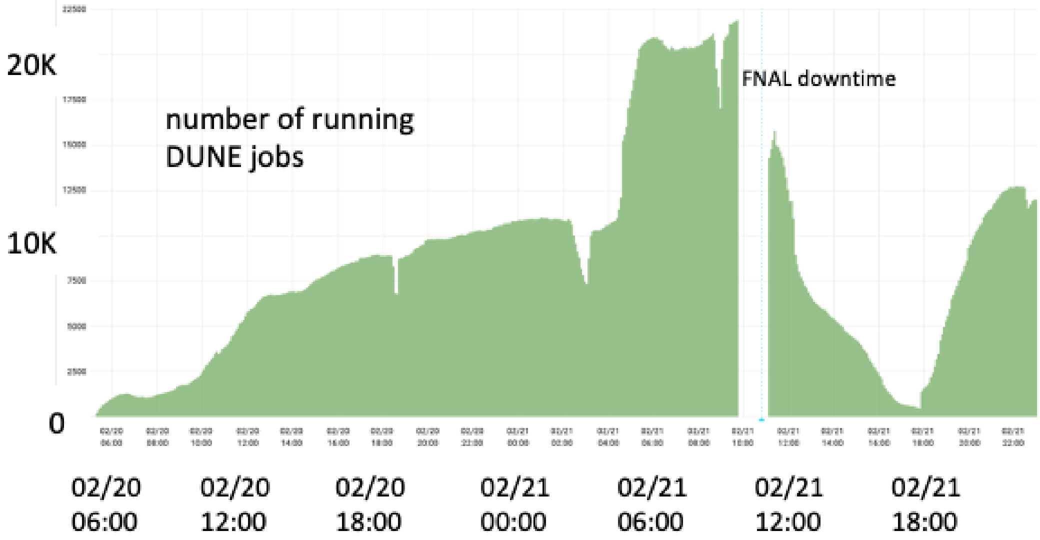


Figure 3.5.3a: Keep up processing (Test #2)

Figure 3.5.3b shows the distribution of processing sites for one of the justIN workflow tasks run during the DC24 keep-up processing test. Each colour represents a different site, and the vertical axis represents the number of jobs running at any given time.

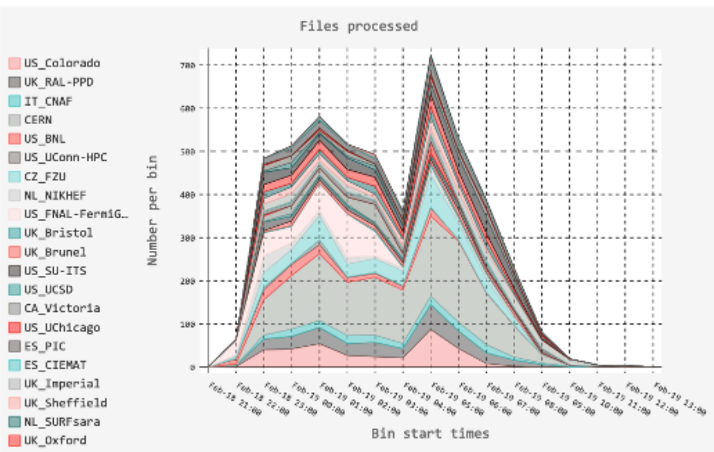


Figure 3.5.3b: File processing rate across different Compute sites in US, Canada and Europe

Since there was no XRootD client-side monitoring deployed, DUNE used internal site monitoring from PIC to show the data streaming from PIC to European compute centres. The actual data rates vs time can be seen in Figure 3.5.3c with the vertical axis being in Gbits/s.

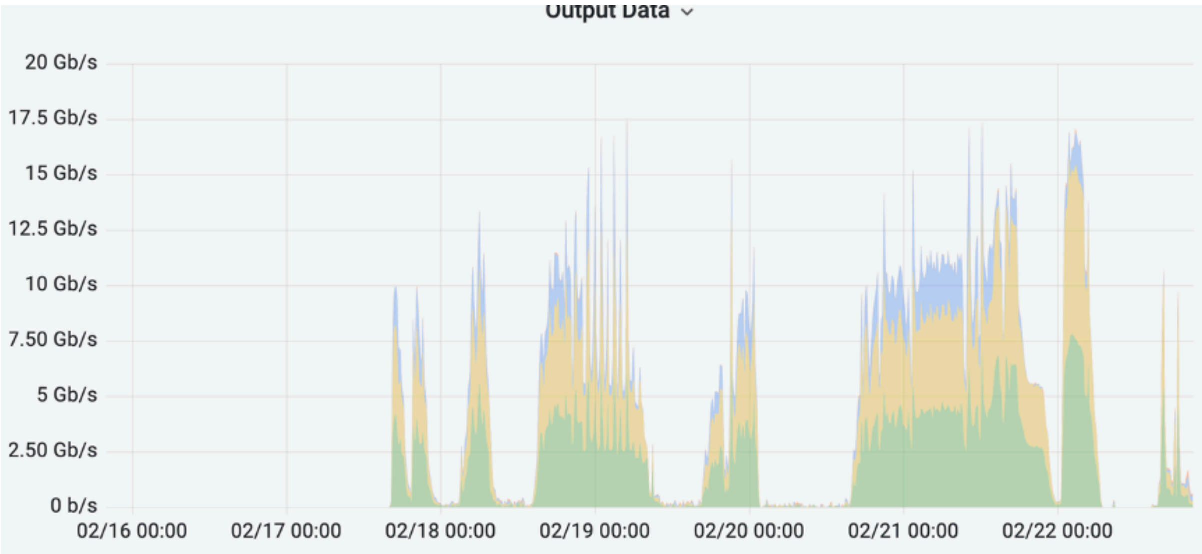


Figure 3.5.3c: Data Streaming from PIC to European sites processing data during the keep-up processing test.

As we scaled up, we then observed a large increase in the number of held jobs. This was tracked to the condor_schedd at RAL (to which justIN submits) being grossly overloaded to the point of it getting killed with the out-of-memory killer. The holds of jobs were occurring all at submit time. Over the course of the week, the memory on this condor_schedd was expanded from 16GB to 48GB, and the spool space was also increased significantly. We reached 22,000 simultaneous running jobs at one point, but the stable level is approximately 16,000 jobs. As a result of this testing, a second HTCondor condor_schedd was deployed at RAL for the justIN workflow system.

Due to the rest of the WLCG Data Challenge going on simultaneously, the FTS3 transfers back to Fermilab were much slower than usual. The result was a backlog of as many as 160,000 files waiting to be transferred back to Fermilab. In particular, those transfers which were trying to use XRootD across the long haul were timing out at all sites where they were attempted. We have now tuned our third-party copy transfer protocols across the board to use davs where possible. We also discovered in the process that it is the preferred third-party copy protocol of the destination that governs which protocol is used.

This run of testing has tested Rucio at data rates and number of clients that we had never reached before. The scalability of the new Rucio version very much paid off, and we are now running 9 Rucio server pods where we previously had just one. The key conveyor daemons are running more threads, and the total number of connections allowed to the database has been increased to 500. This run of testing also has found many of the limitations of the justIN and HTCondor system, several of which have already been addressed.

Test #3, the SuperNova Burst data transfer test, was attempted and aborted because of the backlog of Monte Carlo (MC) output created during the keep-up processing test. The faster transfers between BNL and FNAL were held up by the slower transfers from Europe, in particular RAL, to FNAL.

3.6 LHCb

3.6.1 Overview

For LHCb, DC24 consisted of two parts: the writing part and the staging part. The writing part took place during the first week of the challenge, and the staging part – during the second week. Only Tier-0 and Tier-1 sites participated in DC24 for LHCb. The writing part emulated data distribution from Tier-0 to Tier-1 sites. Files were copied from CERN EOS to Tier-1 tape storage. This copy process, however, was performed in multiple steps:

1. Copy a file from CERN EOS to Tier-1 Disk storage.
2. Copy the file from Tier-1 Disk storage (where it was uploaded before from CERN) to tape storage at the same Tier-1 site.
3. Remove the file from Tier-1 Disk storage.

That means that only the first link (CERN -> Tier-1-Disk) generated external traffic, while the second link (Tier-1 Disk -> Tier-1 Tape) generated local Tier-1 traffic. The above steps apply to individual files, not the whole datasets (that means that we do not wait for all files to be copied to Tier-1 Disk storage before starting transfers to Tier-1 Tape SE). The copying is done in such a way because production workflows distribute data similarly (since we need to process raw data from the detector ASAP, we copy it to disk first to allow jobs to access it).

Transfer rates are calculated on both links independently, and then compared with the target rate. Site's target rates for the writing part are the same for both links.

The staging part emulates data processing activities after the data-taking period. From the DC perspective, it only involves copying (and staging) data from Tier-1 Tape storage to Tier-1 Disk storage. Sites were asked to clean disk buffers of their tape storage prior to the start of this part. Note that this part does not generate external traffic at all since all transfers were local to Tier-1 Data Centres. The target rate for the staging part applies to throughput on the Tier-1 Tape -> Tier-1 Disk link.

Transfers were submitted at the beginning of every part, additional submissions only happened for failed FTS jobs to trigger (additional) retries.

Target rates for individual sites (for both parts) apply to the average throughput achieved on the link. Note that data volumes and transfer speeds may vary depending on the site, so the duration of writing and staging parts for individual sites may vary as well. For tape storage, a transfer is considered as finished as soon as the file is uploaded to SE's disk buffer (so FTS feature that allows checking whether a file has been copied to tape or not was not used).

The cumulative target rate for the writing part was 14GiB/s, while for the staging part it was 9.58GiB/s. Table 3.6.1a below summarises DC24 results for individual sites.

Target, GiB/s			Achieved, GiB/s			Ratio Achieved/Target		
			Writing part		Staging part	Writing part		Staging part
Site	Writing	Staging	CERN-Disk	Disk-Tape	Tape-Disk	CERN-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	0.6	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	0.9
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

Table 3.6.1a: Cumulative target rates

3.6.2 Writing part

The writing part took place during the first week of the Data Challenge. It started on the 13th of February at around 8 am UTC and ended on the 16th of February at around 1 am UTC. Initially, transfer submissions happened on the 13th of February; afterwards, only failed transfers (i.e. corresponding files) were resubmitted. That means there were no additional data injections.

The amount of data that each site received depended on the amount of resources it provides. Individual percentages of the whole DC dataset that sites were supposed to receive are given in Table 3.6.2a.

CNAF	GridKA	IN2P3	NCBJ	PIC	RAL	SARA
15.79	21.14	11.81	7.89	3.93	30.54	8.90

Table 3.6.2a: Distribution percentages per Tier-1.

The above percentages coincide with the site's shares of raw LHCb data. The cumulative size of the DC24 dataset that was distributed during the writing part is around 2PiB. Below is the plot showing the distribution of this dataset among the sites.

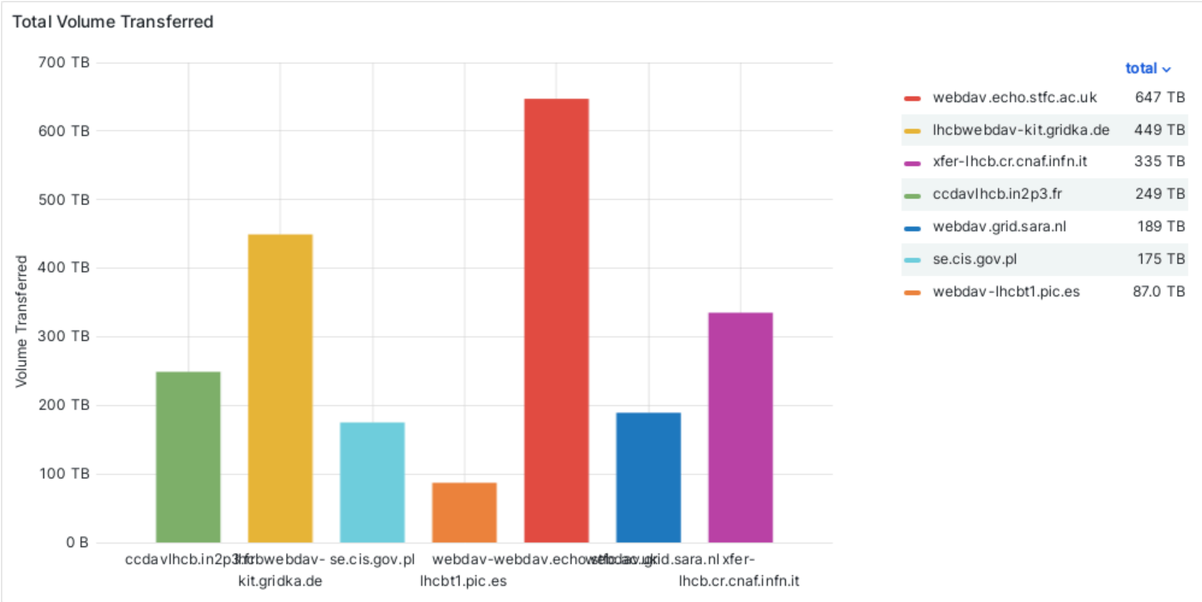


Figure 3.6.2b: Cumulative size of distributed data

3.6.3 CERN->Tier-1 Disk link

Below is the throughput plot for the CERN->Tier-1 Disk link:

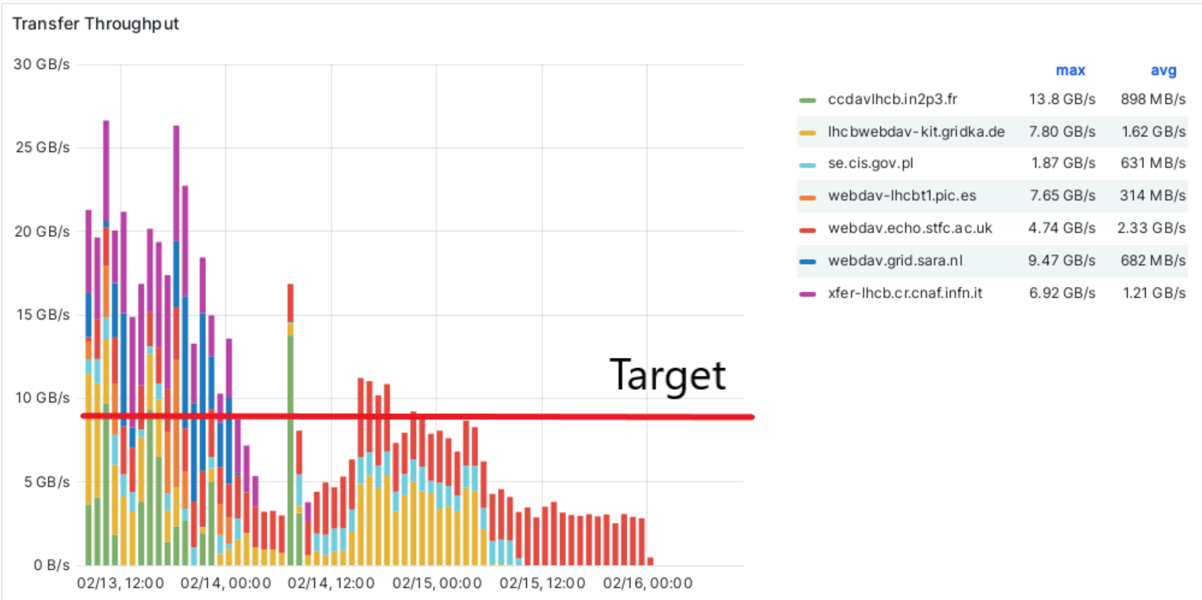


FIgure 3.6.3a: Throughput for CERN to Tier-1 disks.

As can be seen, target throughput was achieved during the first day when all sites were contributing. After that, some sites (e.g. CNAF and SARA) finished transferring their parts and were no longer contributing, that is why the cumulative throughput decreased. One can also see that for some sites (e.g. SARA and PIC), there is a big gap in the middle of the first day. This is caused by slow submissions from DIRAC FTS3 Agent. The slowness, in turn, is caused by slow token retrieval (see below).

Sites that have not achieved target performance on this link are GridKA, NCBJ, and RAL. For GridKA and NCBJ, it was caused by very low transfer efficiency, which, in turn, was caused by the IAM server problems (see below). For RAL, LHCOPN link was down during the writing part of the challenge, and ECHO gateways network tuning was not optimal. Below are the efficiency and error plots for the link in question.

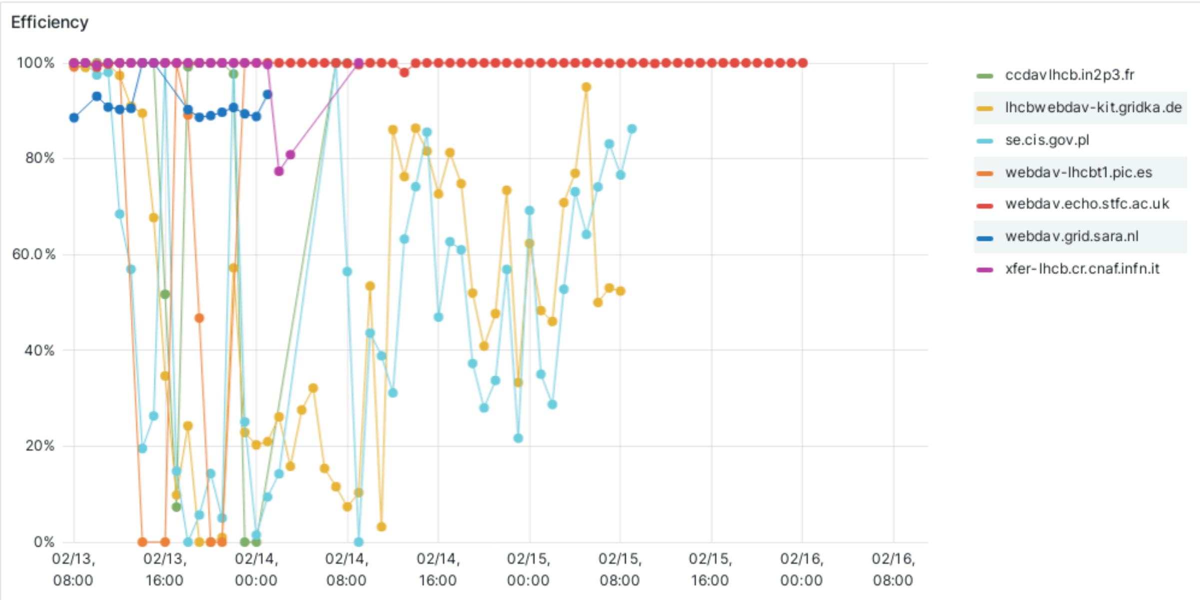


Figure 3.6.3b: Link efficiencies

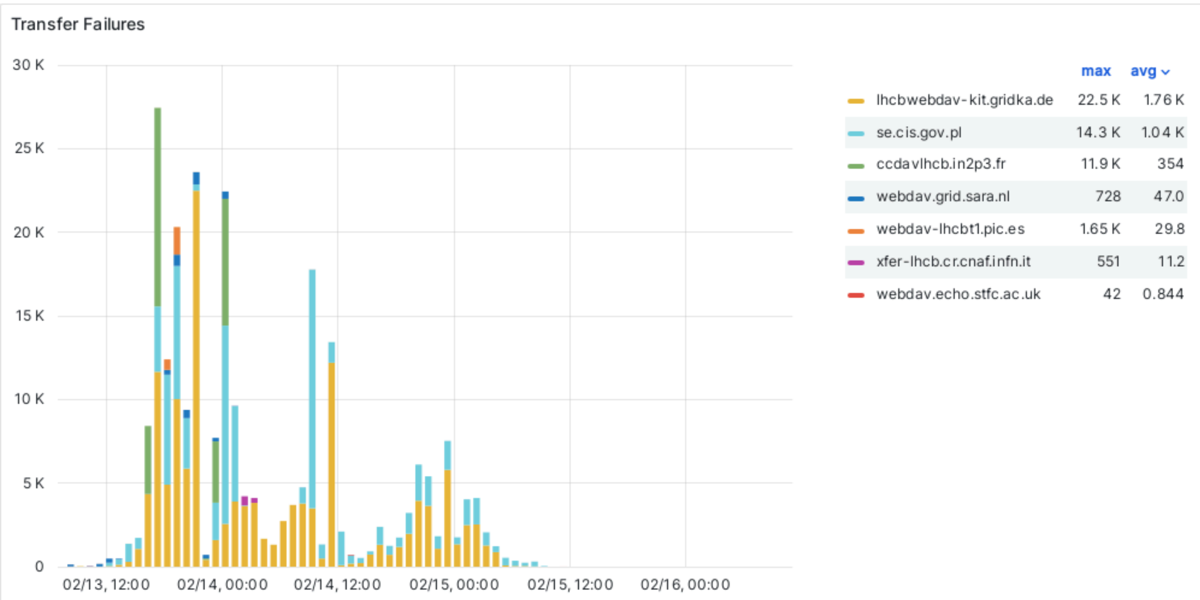


Figure 3.6.3c: Errors on the links

As can be seen from the above plots, sites using tokens for authentication (IN2P3, GridKA, NCBJ, PIC) are the major sources of errors. There were also some efficiency drops at CNAF (due to the nature of their storage, which does not perform very well when there are simultaneous reads and writes to/from it due to local FS overload) and SARA.

3.6.4 Tier-1 Disk -> Tier-1 Tape link

Below is the throughput plot for Tier-1 Disk->Tier-1 Tape link.

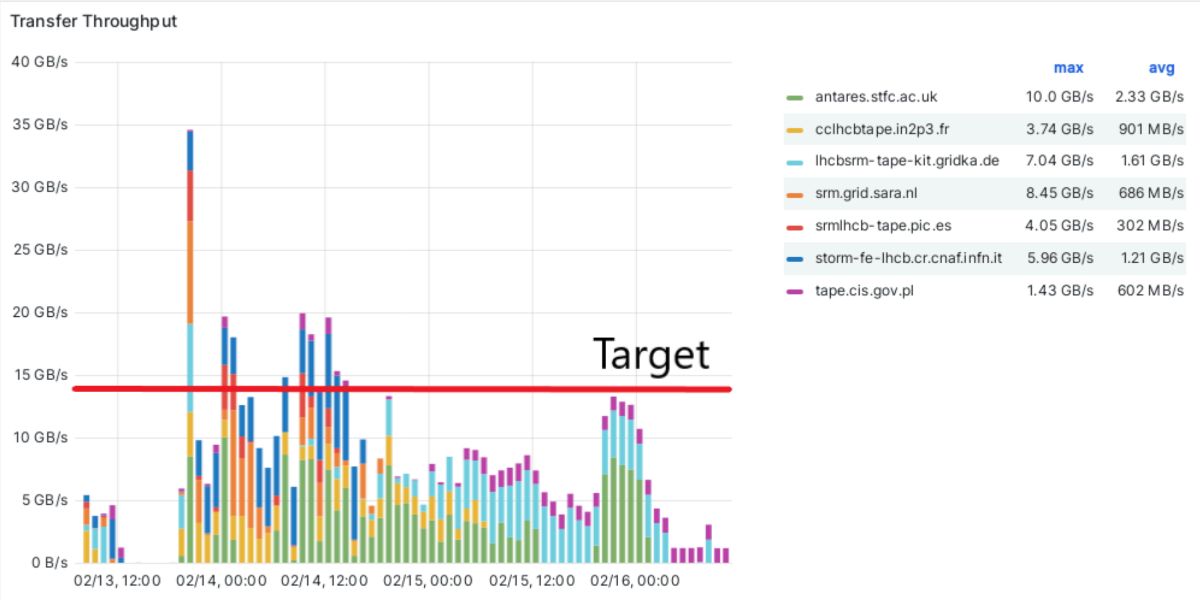


Figure 3.6.4a: Transfer throughput on the Tier-1 Disk to Tape links

As can be seen, the throughput is spikier due to the nature of the link. There was also a significant delay in tape transfer submissions during the first day, most probably because of the submission delays introduced by tokens (see below). The following sites have not achieved target throughput on this link: GridKA, IN2P3, NCBJ, and RAL. This is most probably caused by low performance on the CERN -> Tier-1 Disk link (probably with the exception of IN2P3, which had very good performance on the first link). Below are the efficiency and error plots for the link in question.

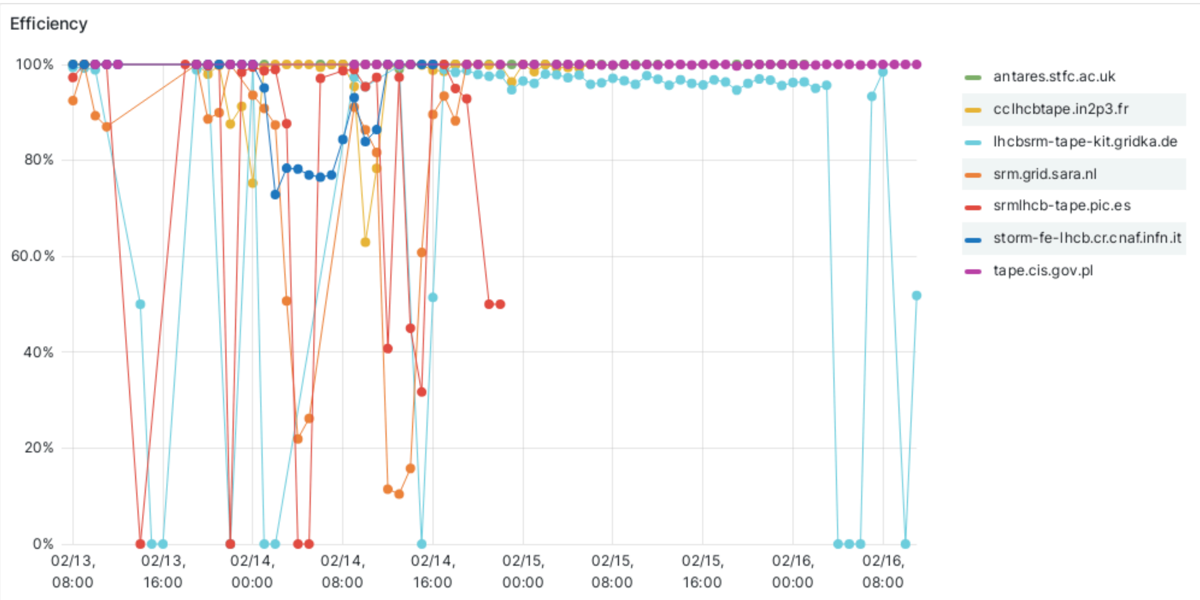


Figure 3.6.4b: Efficiency on the links

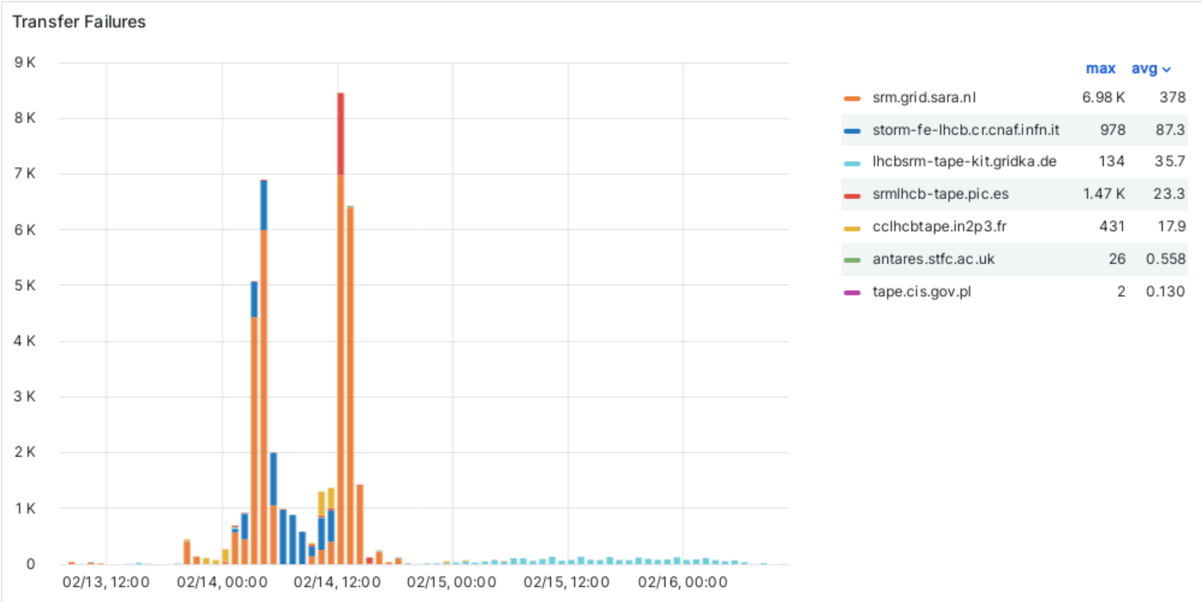


Figure 3.6.4c: Errors on the links

Here the major source of errors is SARA, where the disk buffer of the tape storage was overflown. Issues at CNAF (caused by the storage overload, see above), PIC, and GridKA can be seen as well. Failures at GridKA seem to be caused by transfers getting stuck (see e.g. FTS jobs "b8562f5e-cbde-11ee-a5a0-fa163e5a69c7" and “b7aed498-cbde-11ee-b32d-fa163e5a69c7”).

3.6.5 Summary

In general, the writing phase was successful as a relatively sustained target rate (cumulative one) was achieved on both links when all sites were contributing. However, it was not entirely seamless, and several issues were identified.

3.6.6 Staging part

The staging phase commenced on the 20th of February, at approximately 8 am UTC. During this phase, files that were previously copied to Tier-1 tape storages were recalled. Sites were instructed to clear disk buffers of their tape SEs before the start of the staging phase. Subsequently, the files were copied to local Tier-1 disk storages. As mentioned earlier, this phase did not generate any external traffic.

Below is the throughput plot for this part.

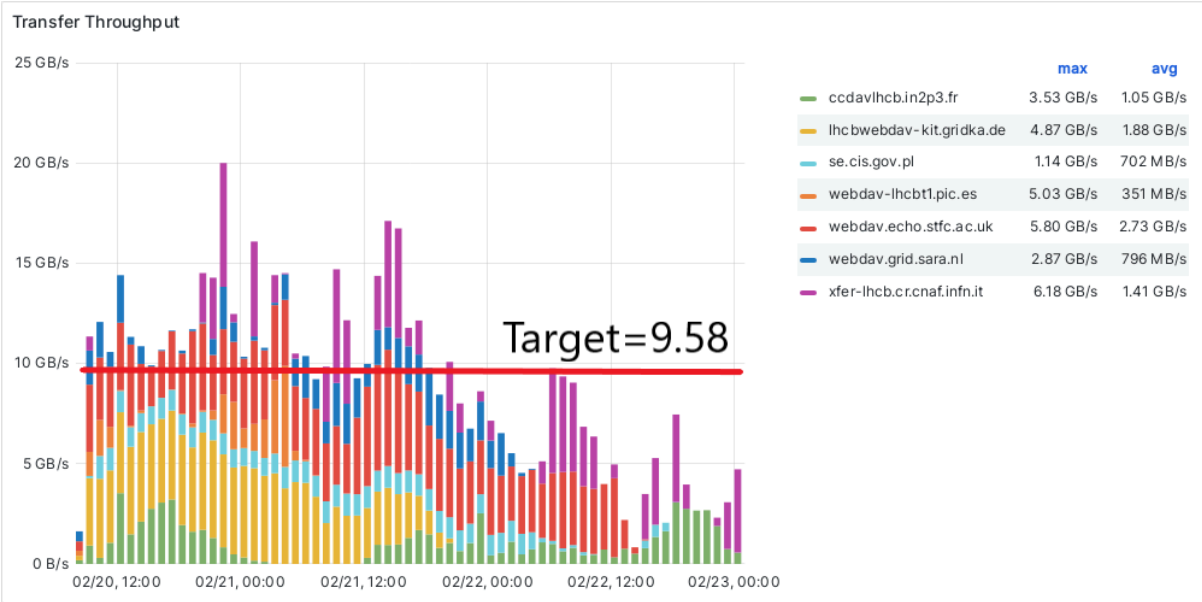


Figure 3.6.6a: Throughput for recall from tape

As can be observed, the target throughput was attained initially when all sites were contributing. Subsequently, as some sites completed transferring their portions, the cumulative rate decreased, as expected. The following sites did not achieve the target throughput on this link: CNAF, IN2P3, and NCBJ. For IN2P3, the likely cause is the low transfer efficiency (see below).

Below are the efficiency and error plots for the link in question.

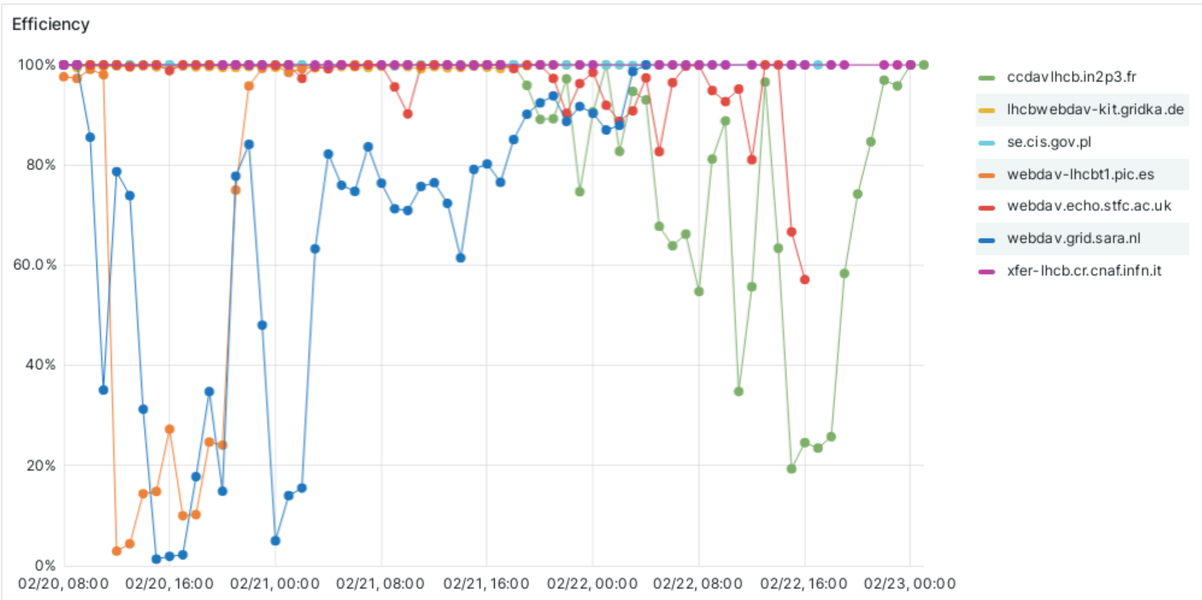


Figure 3.6.6b: Recall efficiency

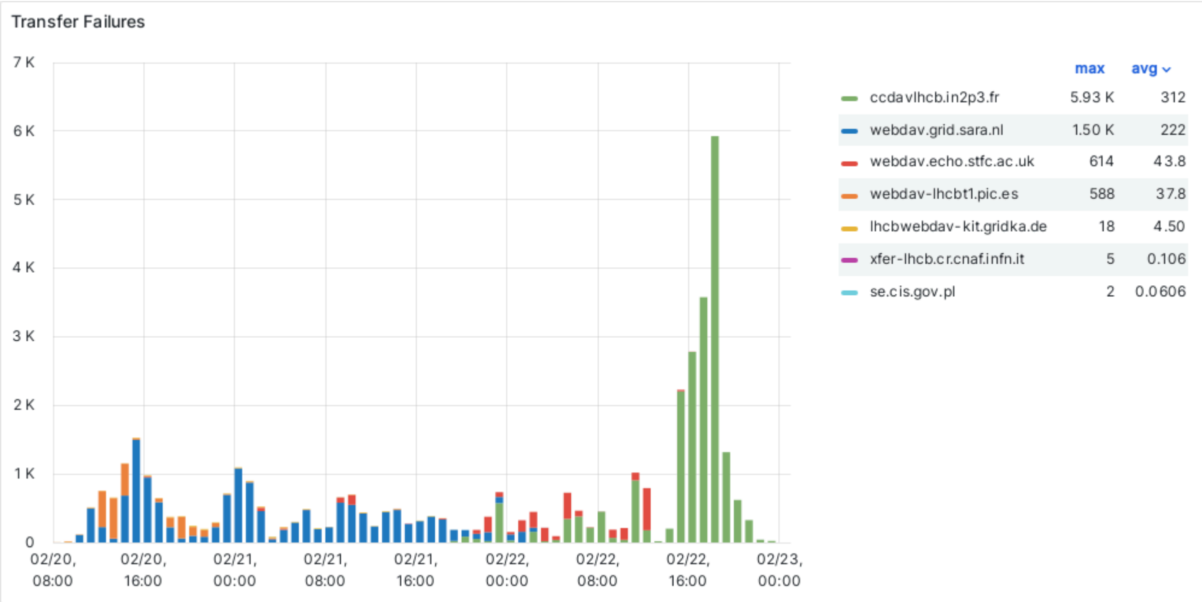


Figure 3.6.6c: Recall errors

As can be observed, the major sources of errors are SARA (where there was a configuration issue) and IN2P3 (where tape buffers were overflown during the staging activities). It was discovered that 4 files at RAL and 3 files at IN2P3 were lost after the writing part (i.e. files were successfully copied to tape SEs, but were not present when the staging part started). FTS was initially suspected, but after detailed investigation, it was determined that this occurred due to storage issues, and FTS behaviour was normal. In general, the staging phase was also successful as a sustained target rate was achieved, and some problems were discovered.

3.6.7 Token authentication

Prior to the start of the DC24, all participating sites were requested to implement token authentication for their disk storages. Only 4 dCache-based sites successfully implemented it: GridKA, IN2P3, NCBJ, and PIC. The issue with other storage software products (specifically StoRM and XRootD) was related to authentication problems with certain types of requests. Namely, LHCb uses tokens with the following scopes:

storage.read:<full_path> + storage.modify:<full_path>

where <full_path> is the file path with respect to the configured prefix. For every transfer FTS (when in default mode) tries to check (and create when necessary) all subdirectories prior to copying the file. To check directory existence, it issues PROPFIND request for the directory that is supposed to contain the file (i.e. `basename <full_path>`). It seems like this request should be allowed according to the WLCG token specification, though it failed to pass authorization for the storage implementations in question. Unfortunately, developers were not able to apply the necessary amendments before the DC24. Thus, token authentication was employed solely on the CERN -> Tier-1 Disk link (as other links involve tape storage) for the following sites: GridKA, IN2P3, NCBJ, and PIC. The DIRAC FTS3 agent was also patched to enable the submission of FTS jobs with tokens.

Overall, approximately 45% of all data on the CERN -> Tier-1 Disk was transferred via token-based transfers (see picture below).

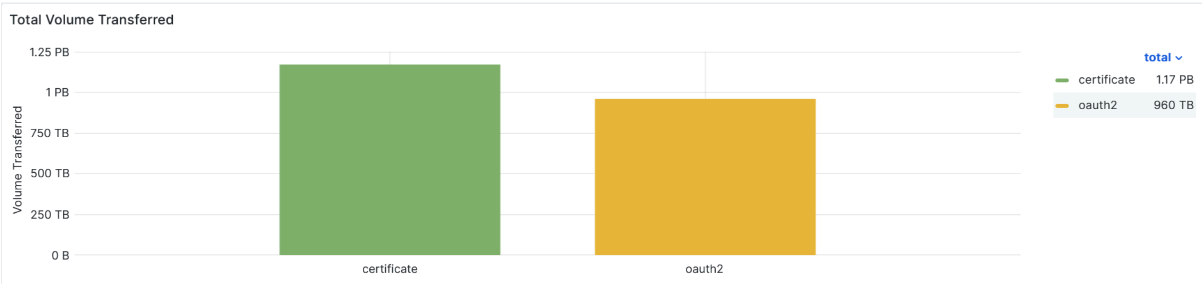


Figure 3.6.7a: Total volume transferred

During the challenge (its writing part) there were some problems with token authentication, namely, the IAM server got overloaded and was not able to issue tokens fast enough. This resulted in slow transfer submissions, as for every FTS transfer, at least two storage tokens were required (refer to the scope description above). Consequently, the submission rate dropped to approximately 0.5Hz on average at the beginning of the writing phase, when the number of submissions was the highest. The overloaded server also contributed to numerous FTS transfer failures due to token expiration. Since we had submitted (almost) all transfers at the beginning of the challenge, some of them spent considerable time in the queue, requiring token refreshment. FTS attempts to refresh tokens before transfer commencement, but this process is performed on a best-effort basis. Due to the overload, many token refreshment operations timed out, leading to numerous FTS transfers initiating with short-lived or expired tokens, resulting in failure. All sites using tokens were affected by this issue, with the most significant impact observed at GridKA and NCBJ.

Also, DIRAC's FTS3 agent (the entity that submits and monitors FTS jobs) got stuck several times. These issues appear to be attributable to the new token support code, as similar problems were not observed before the agent was patched. Below is the efficiency plot of token-based vs certificate-based transfers.

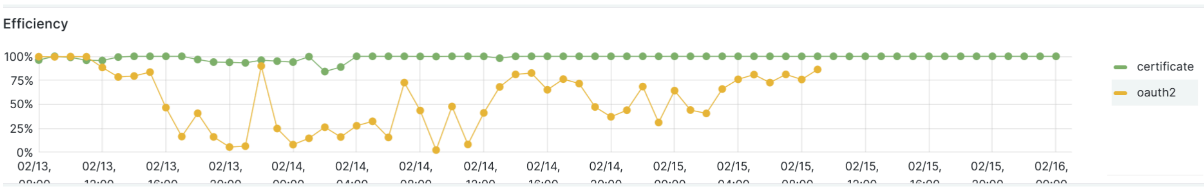


Figure 3.6.7b: The efficiency of token-based transfers is significantly lower than the efficiency of certificate-based ones.

3.6.8 Conclusion

In summary, the Data Challenge was successful for LHCb. We achieved the desired cumulative rates for both the writing and staging phases. Additionally, we successfully tested the new authentication mechanism (tokens). The Data Challenge highlighted significant issues and bottlenecks, such as IAM server overload. These problems will be addressed in due course.

4 Middleware

4.1 FTS

4.1.1 Overview

Data Challenge 2024 marked a significant success for the File Transfer Service (FTS). The service sustained over 20,000 concurrent transfers for 17 hours. The following plot illustrates that this new record surpassed the previous one of just under 9,000 concurrent file transfers by more than double. Both records were achieved with the fts3-atlas.cern.ch instance.

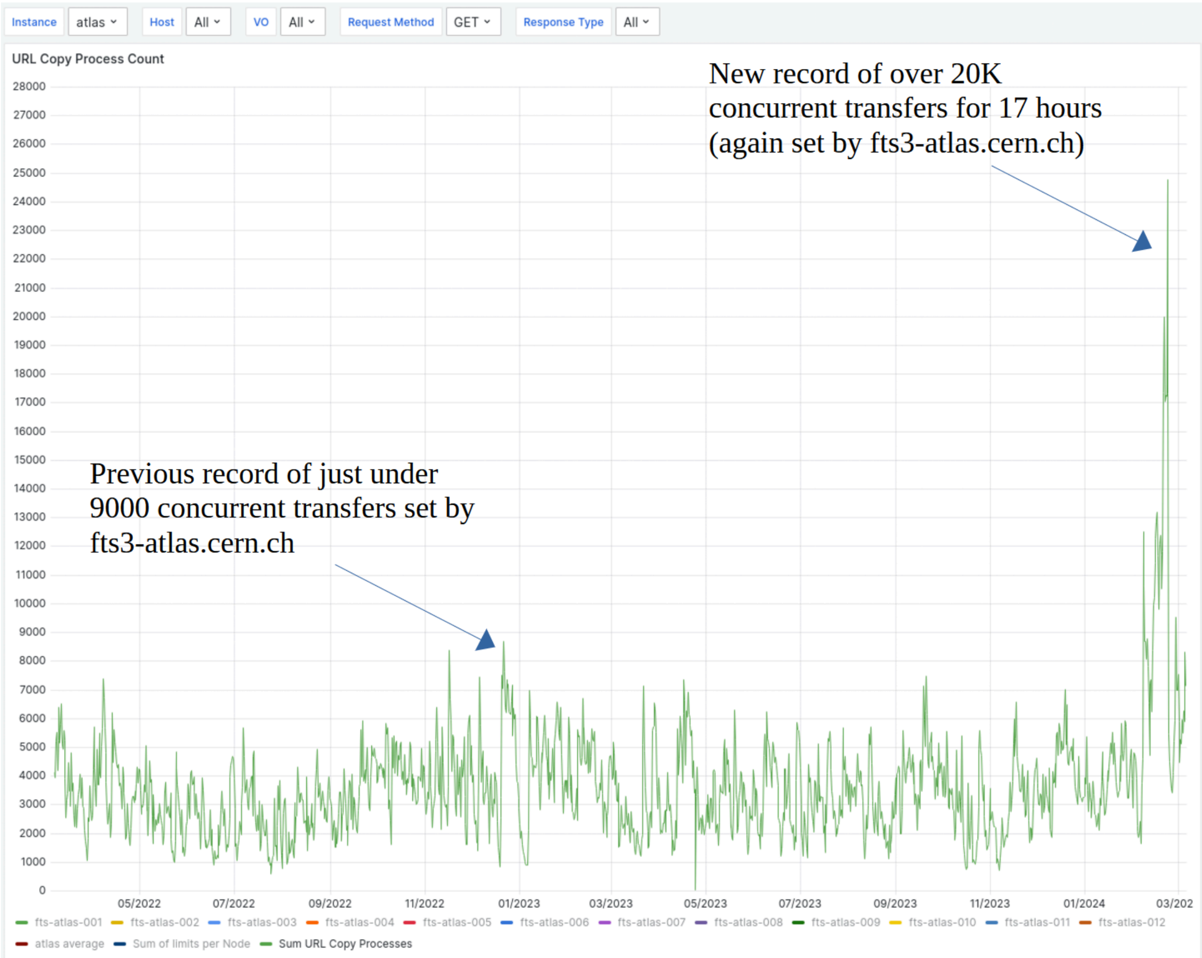


Figure 4.1.1a: Record number of concurrent FTS transfers

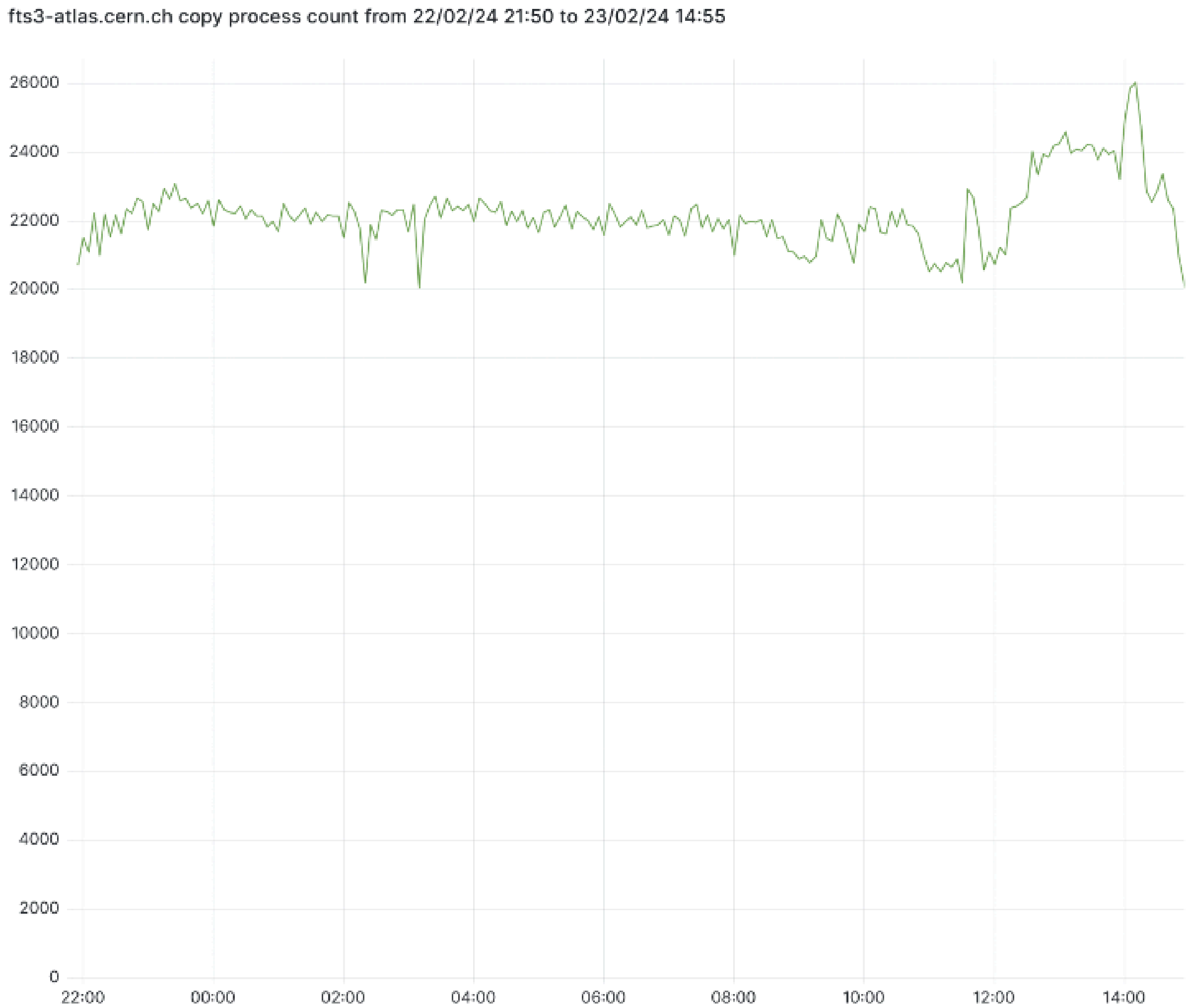
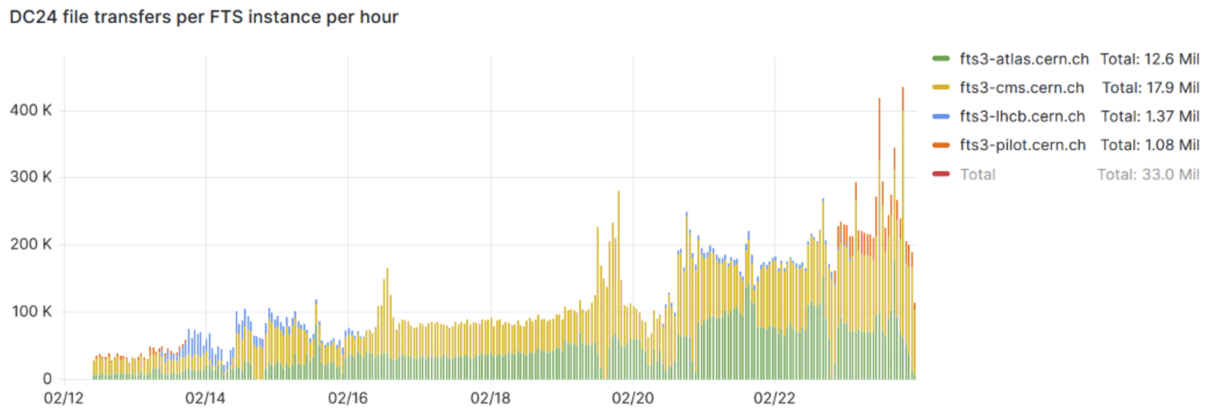


Figure 4.1.1b: 17 hours of sustained transfer concurrency during DC24 on fts3-atlas.cern.ch

Figure 4.1.1c: These two plots show that the CMS and LHCb FTS instances also did very well. The fts3-cms.cern.ch instance reached over 20K concurrent transfers.



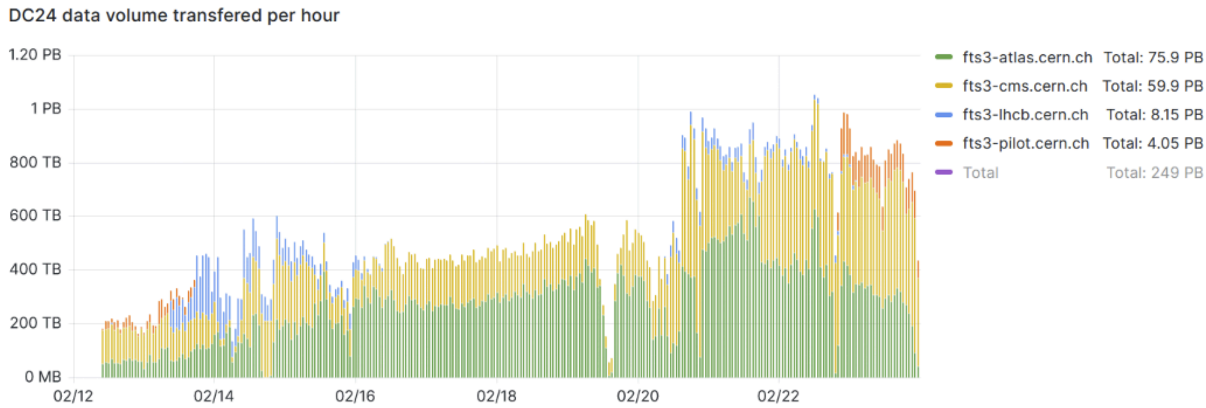


Figure 4.1.1d: These two plots show the file-transfer counts and data volume per hour for DC24.

With the combination of the ATLAS, CMS, LHCb and Pilot FTS instances, the service transferred 33 million files totalling 249 PB of data.

4.1.2 Challenges/Issues

FTS encountered two main challenges/issues during DC24:

1. An overloaded database.
2. A difference between the metrics being maximised by FTS and DC24

There was significant troubleshooting required behind the scenes due to the overload of the fts3-atlas.cern.ch database. To address this issue, the database on-demand service promptly increased the RAM size from 80 to 120 GB in order to alleviate the strain. The FTS team then proceeded to migrate the fts3-yale.cern.ch database, which was not used during DC24, from a high-performance 120GB RAM database server to a more suitable environment aligned with its performance requirements. Subsequently, the FTS team migrated the fts3-pilot.cern.ch database into the now available high-performance database server. This rearrangement of databases enabled the ATLAS experiment to use the fts3-atlas.cern.ch and fts3-pilot.cern.ch databases, each with 120GB RAM. This adjustment was essential to enhance the performance of the FTS service and allocate necessary spare resources to non-DC24 production transfers managed by fts3-atlas.cern.ch.

The primary cause of the database overload was the token-refreshing component of FTS. This component had not yet been optimised during DC24 for two reasons. Firstly, it adopted the approach of polling the FTS database for near-to-expire tokens, and secondly, it utilised SQL that was not optimised. These shortcomings were compromises knowingly accepted before DC24 to ensure working token support by November 2023, ready for DC24.

The following plot from the database-on-demand service illustrates the overloaded fts3-atlas.cern.ch database, which eventually became blocked on Monday, 19th February. This issue arose due to a poor interaction between high database activity and the regular defragmentation operation, scheduled every Monday at 10:00 am. The resulting interference did not manifest until after 13:00.

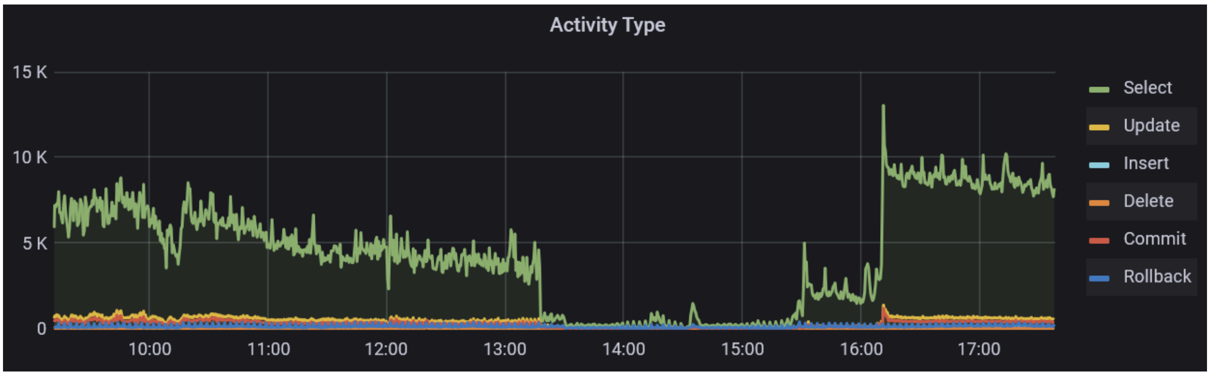


Figure 4.1.2a: The overloaded fts3-atlas.cern.ch instance

DC24 highlighted the disparity between the metrics maximised by FTS and DC24. FTS focuses on controlling the number of concurrent file transfers between different source and destination storage endpoints, ensuring efficient management of endpoints and links without overloading storage endpoints. Conversely, DC24 aimed to maximise the data throughput of these links. While these metrics are related, they remain distinct, and we believe this is the primary reason why DC24 did not sustain its most ambitious throughput target for 48 hours. FTS successfully reached the configured concurrency limits of the storage endpoints it was managing. However, these same limits prevented the desired data throughput from being sustained for 48 hours.

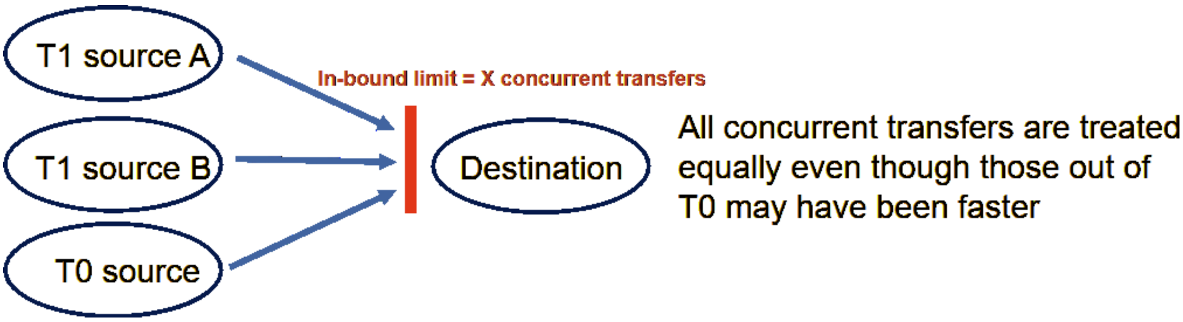


Figure 4.1.2b: Diagram of activity selection bottleneck

FTS manages concurrent file transfers per source and destination link. FTS does not directly manage data throughput. FTS treats all links with the same activity (as was the case for DC24) with equal priority. FTS saturated all of its configured destination endpoints in terms of inbound concurrent file transfers. Links transferring files faster were treated the same as those transferring files slower. All links had the same number of concurrent transfers, regardless of the speed of each transfer.

Another lesson learned pertains to the visual metrics provided by FTS. The FTS Web Monitoring displays links, and often attention is focused on the number of transfers scheduled on a given link (and why not more). Operational experience indicates that most of the time, the system brakes are not the link limits but rather the Storage Endpoint limits (source or destination). Even during DC24, FTS was reaching the destination endpoint limits, although this was not immediately apparent. The team initiated a project to monitor these metrics and develop dashboards for storage endpoint saturation, which will indicate how many parallel transfers are scheduled inbound and outbound for a given storage.

Despite FTS not directly managing data throughput, experiments are still encouraged to review their link and endpoint configurations within FTS in the hope of potentially increasing some of their upper limits.

The FTS Optimiser

Issues were encountered with the FTS Optimiser of the fts3-atlas.cern.ch instance during DC24. Although these issues are not significant enough to warrant their own bullet point as the main cause, they certainly contributed to exacerbating the overall situation.

The FTS Optimiser is the component responsible for determining the number of parallel transfers that should be on a link, within a range from MIN to MAX (default is 2 to 200). For example, if A → B has an Optimiser decision of 170 transfers and there are currently 160 transfers ongoing, then the Scheduler evaluates these two values and identifies that it can schedule 10 more on that link. Whenever the throughput or success rate on that link changes, the Optimiser recalculates the decision. A typical run of the FTS Optimiser on the fts3-atlas.cern.ch instance takes anywhere between 6 to 12 minutes, as it has to assess each pair and compute throughput metrics. The time it takes has been observed to be a linear function of the size of the queues stored in the database. During DC24, it was noted that a full Optimiser run would require around 3 hours to complete. Consequently, the Optimiser decision would only be updated every 3 hours, which negatively impacted the system's ability to adapt to current throughput and success rates. Additionally, when a link configuration is altered, the Optimiser value is recomputed based on the newly configured limits. However, due to the extended time required to complete an Optimiser cycle, changing the configuration would no longer have an immediate impact on the system. To compound matters, a housekeeping cron job restarts the FTS Server process every hour to prevent the rare but critical scenario where certain threads die silently. However, this hourly restart of the FTS Server process prevented the Optimiser component from completing its full cycle. Consequently, the FTS system was scheduling transfers based on a "frozen" snapshot from the last complete Optimiser run.

The FTS Optimiser problems emerged late in the data challenge, during the final 2 days. There was no workaround that could be implemented in time. However, the FTS team has already investigated potential improvements. The first one is to convert the Optimiser from a sequential iteration through all the database pairs into a multi-threaded process. This alone is expected to bring a significant improvement, potentially up to a 10-fold increase in efficiency. Further along the development roadmap, the throughput computation can be transformed from a database-intensive query into an event-based and on-the-fly process. This enhancement will reduce the FTS load on the database and transform the FTS Optimiser into a component that operates on real-time feedback. Lastly, we recognise the necessity of having a mechanism to completely disable the FTS Optimiser (and revert to fixed scheduling sizes), which would have been beneficial during DC24.

4.1.3 Lessons learnt about tokens

The following plot shows that FTS transferred a total of 33 million files during DC24 with half of them being transferred using token-authentication and half using x509 certificates.

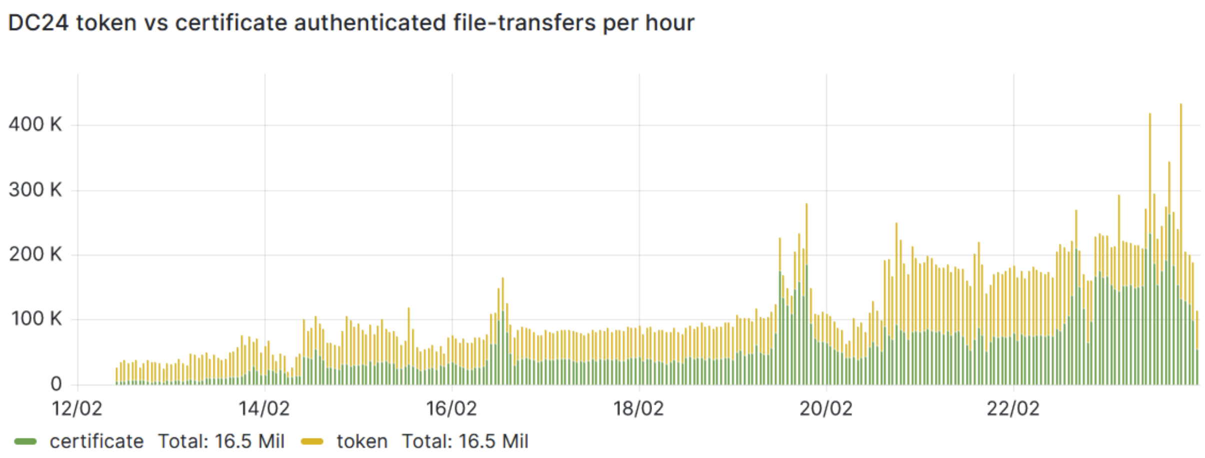


Figure 4.1.3a: Split between certificate and token transfers

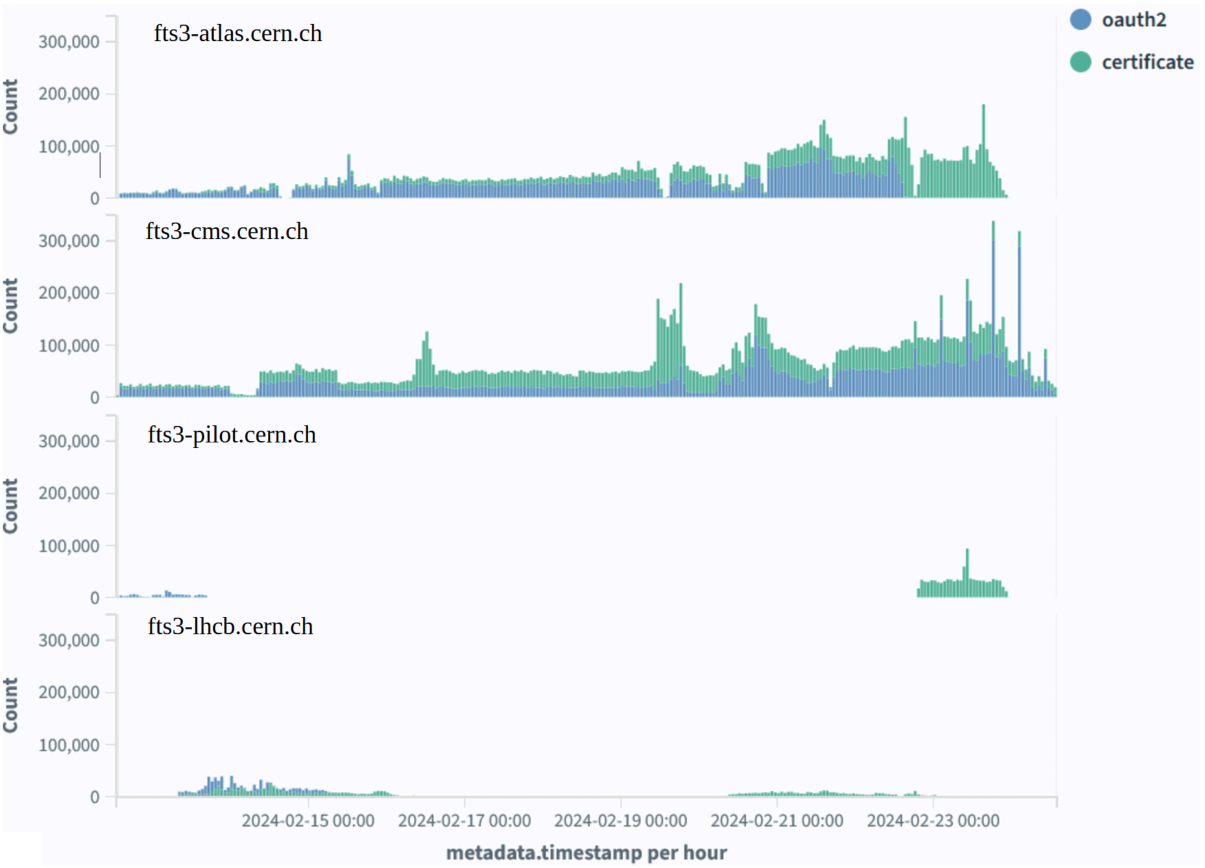


Figure 4.1.3b: These four plots show token versus certificate authentication per experiment plus fts3.pilot.cern.ch

DC24 taught us 4 main points about tokens:

1. Incorrectly used tokens are not secure.
2. Too much time is spent discovering tokens.
3. FTS could not handle intense token tests.
4. FTS did not know (yet) its token limits.

Tokens were leaked during the pre-DC24 tests, being displayed by FTS in redirect URLs and error messages received from file transfers and storage endpoints. A filter was promptly added to FTS to prevent this occurrence during DC24. While FTS may not encounter the same problem again, it serves as a cautionary reminder for other systems to take preventive measures.

There was limited time to thoroughly test FTS token support by DC24 (November 23), leading to numerous unwelcome surprises. Single-use refresh tokens were discovered on the job, causing disruption when FTS went out of sync due to a bug. Fortunately, the IAM development team provided a means to configure this single-shot behaviour to be disabled. Confusion arose from refreshed tokens not sharing the same validity period as their older originating tokens, particularly when 10-hour tokens from the experiments were converted into 1-hour tokens by the "exchange an access token for a refresh token" logic of FTS and IAM.

The inherent insecurity of token misuse prompted the ATLAS and LHCb experiments to request "one-token-per-file" tests. One set of tests was conducted with ATLAS before DC24, and another was carried out with LHCb during DC24. FTS imposed a heavy load on its databases during these tests. Following the initial ATLAS test, FTS had to be modified in two ways. Firstly, its token-refreshing cron-jobs were replaced with daemons to prevent overlapping jobs when IAM was slow, which was depleting the number of connections to the FTS databases. Secondly, the heavier housekeeping tasks of the FTS token refresher daemons had to be separated into dedicated daemons, thereby reducing database load by allowing less frequent execution of housekeeping tasks than token refreshing. Among other tasks, the housekeeping included deleting no-longer-used tokens from the FTS databases.

DC24 clearly highlighted the need for FTS to implement more backpressure mechanisms. Once FTS began to experience meltdown, the situation worsened. Consequently, the FTS team requested a cooldown period with no new DC24 jobs just before the fts3-atlas.cern.ch and fts3-pilot.cern.ch databases were each upgraded to 120 GB RAM.

4.1.4 FTS suggestions regarding tokens

We should continue testing the feasibility of employing one token per file-transfer, particularly in the specific use-case of using modify-tokens (not necessarily for read or create tokens). Achieving this could minimise the impact of token leakage and avoid complications for tape transfers and associated clean-up logic.

There should be a designated "token" person responsible for overseeing both development, deployment, and integration. Having a single point of contact facilitates communication compared to dealing with multiple individuals with different scopes and priorities.

We should officially prohibit the use of single-shot refresh tokens from the WLCG token lifecycle. This feature was enabled at the outset of FTS token tests. If there was any failure in the FTS token refresher to store the latest refresh token, FTS would fall out of sync with IAM, rendering entire sets of tokens unrefreshable.

We need to determine whether FTS should refresh access tokens for WLCG workflows. Can fresh tokens be pushed into FTS akin to how X509 proxy certificates are handled today?

We should establish a consensus on how to include VO information in a token. FTS had to be adjusted to map tokens to VOs. The VO values must align for both tokens and certificates.

IAM performance is crucial, and in many instances, FTS had to implement workarounds to address certain issues. Can IAM token providers incorporate a "reset button" or a "DB purge script" to erase data related to "one-token-per-file" tests?

All storage endpoints should ensure they are configured to use the "dteam" IAM token provider. This configuration will enable the FTS team to conduct token-related tests effectively.

4.1.5 Work already done and a look to the future

Since DC24, FTS has:

- Continued to carry out token tests with the CERN FTS instances at the request of experiments.
- Continued to work with “relaxed” but “risky” modify-tokens.
- Decoupled the parallelism of the token-refresh protocol from access to the FTS database.
- Optimised the SQL used by the token refresher and token housekeeper daemons.
- Started to use the *dteam* IAM provider to carry out tests with disk storages at CERN.

FTS plans to:

- Add a back pressure mechanism – The RUCIO team kindly offered to switch on their FTS back pressure if necessary.
- Allow one-token-per-file- tests on fts3-pilot.cern.ch (but not on fts3-atlas.cern.ch).
- Refresh tokens on demand (start of transfer, checksum, etc.) as opposed to polling the database for near-to-expire tokens.
- Improve the performance of an Optimiser run.
- Allow the Optimiser to be switched off.
- Provide a better way to show the saturation of destination storage-endpoints.
- Provide token-support for tape-transfers once support is finished for disk-transfers.
- Rewrite the fetching of data necessary for scheduling so that FTS no longer needs “everything” in memory.
- Introduce priorities between source and destination transfer-link. This would have helped during DC24 to prioritise fast links over slower ones and it would help in general to prioritise tape-transfers over disk-transfers where necessary.

4.2 Rucio

Overall, the communities utilising Rucio reported success. Rucio demonstrated its ability to scale and meet the demands of Data Challenge 2024. The observed issues were deemed minor, as they did not expose any inherent limitations and do not appear to necessitate significant effort to resolve.

It's important to note that the injection mechanism (reuse of datasets, short replication-rule lifetimes) is somewhat artificial and contributed to the observed issues. These issues are unlikely to manifest with the same severity under expected production conditions.

In preparation for this Data Challenge, the only specific development area addressed by the Rucio team was the implementation of tokens (instead of certificates) for transfers and deletions. Consequently, an entire section is dedicated to that topic.

4.2.1 Remarks on scalability

The components of Rucio most relevant for the Data Challenge are referred to as daemons. Each set of daemons is designed with horizontal scalability in mind: when a new instance is started, it handles a different partition of work compared to the existing instances. Insufficient instances typically result in a backlog of work for that particular daemon, which can be monitored internally through Rucio's monitoring system.

All communities reported the need to significantly scale up some of the daemons during the Data Challenge. There are no specific recommendations regarding the number of instances; operators had to rely on trial and error. However, no community reported reaching a threshold beyond which adding more daemon instances did not improve performance.

One point of discussion among the communities was whether effort should be invested in automatic scaling. With the recommended deployment method being Kubernetes, it could be possible to leverage its capabilities for this purpose. Ultimately, this option was not deemed worthwhile due to three main reasons. Firstly, extreme bursts of work are not common under normal conditions, so operators can manually scale up as needed. Secondly, once scaled up, the Rucio communities did not find significant incentives to scale down again. And thirdly, backlogs may be caused by rare complex issues that cannot be addressed solely through scaling (one such example is described in the next section).

4.2.2 Incidents and future work

Before discussing the observed issues, it may be helpful to provide a brief overview of some Rucio terms. A replication rule, in simple terms, instructs Rucio to transfer data to a destination (when necessary) and keep it there. A rule may have a limited lifetime; when it expires, the rule is deleted and existing transfers may be cancelled. The Submitter daemon handles the submission of transfer requests to FTS. The Cleaner daemon manages expired replication rules, while the Reaper daemon deletes files without any replication rules from storages.

At one point, ATLAS observed a growing Submitter backlog coupled with poor submission performance. Despite the expectation that daemon instances of the same type should not interfere with one another, under certain circumstances, contention with other types of daemons can occur. This was indeed the case. The Submitter prioritises oldest transfer requests first. Following a period of FTS outage, the top of the queue primarily consisted of requests from expired replication rules. Consequently, both the Submitter and the Cleaner attempted to process the same set of requests: the Submitter tried to submit them to FTS, while the Cleaner aimed to remove them before submission. This issue was reported on GitHub as issue #6505. ATLAS operators had to develop and deploy a patch to address this, resulting in the restoration of submission performance and processing of the backlog. An evolution of that patch was incorporated into Rucio 34.1.0.

A second incident involved contention between two daemons, namely the Poller and the Cleaner. Similarly to before, the Poller attempted to update already-submitted transfer requests while the Cleaner tried to cancel them. This contention arose from exceptionally large datasets (50K files) that were mistakenly selected, increasing the likelihood of contention as the number of requests grew. This incident was reported on GitHub as issue #6511.

One issue reported by CMS was poor deletion performance at some storages. The Reaper daemon is structured to operate in chunks: a Reaper instance updates a set of file replicas with a timestamp, attempts to delete each replica from the storage individually, then removes all successfully deleted replicas at the end. During this "being deleted" state, for a time window after the timestamp is set, no other instance may handle that chunk of replicas. Both the chunk size and the length of the time window are configurable, but per Reaper instance. A larger chunk size is recommended for improved database efficiency. Once the time window elapses, other instances may include them in their own chunks. This serves as a safety measure in case a Reaper instance unexpectedly exits.

The reported issue affects storages where deletion operations require a significant amount of time. For instance, if a chunk size is set to one thousand replicas, the time window is ten minutes, and each deletion operation takes one second, it's inevitable that different Reaper threads will end up with overlapping chunks. This isn't inherently problematic for Rucio, but it means that replicas may be attempted to be deleted multiple times, thereby reducing deletion efficiency (same amount of effort from Rucio, but fewer unique replicas from the storage side).

CMS operators had to dedicate effort to isolating slower storages to dedicated Reaper instances with tighter chunk sizes or more generous time windows. This issue was reported on GitHub as issue #6512.

4.2.3 Tokens

Rucio has supported tokens since version 1.22, which was released in 2020. However, the initial implementation was coarse, resembling certificates in terms of capability and not fully leveraging the security and isolation benefits inherent to tokens. Following extensive discussions and planning with numerous other projects and stakeholders, the Rucio team introduced a refined implementation in version 33, released at the end of 2023. Although advertised as a technology preview, it was actively used by ATLAS and CMS. However, due to the late release timing and the commencement of Data Challenge 2024, there was insufficient time for comprehensive testing of how the implementation interacted with those from storage providers, FTS, and IAM, especially under the unprecedented rates achieved during the Data Challenge.

Despite these challenges, the use of tokens was successful, and the set goal was achieved. The following plot illustrates the use of tokens over certificates by ATLAS and CMS, specifically for the 'Data Challenge' transfer activity.

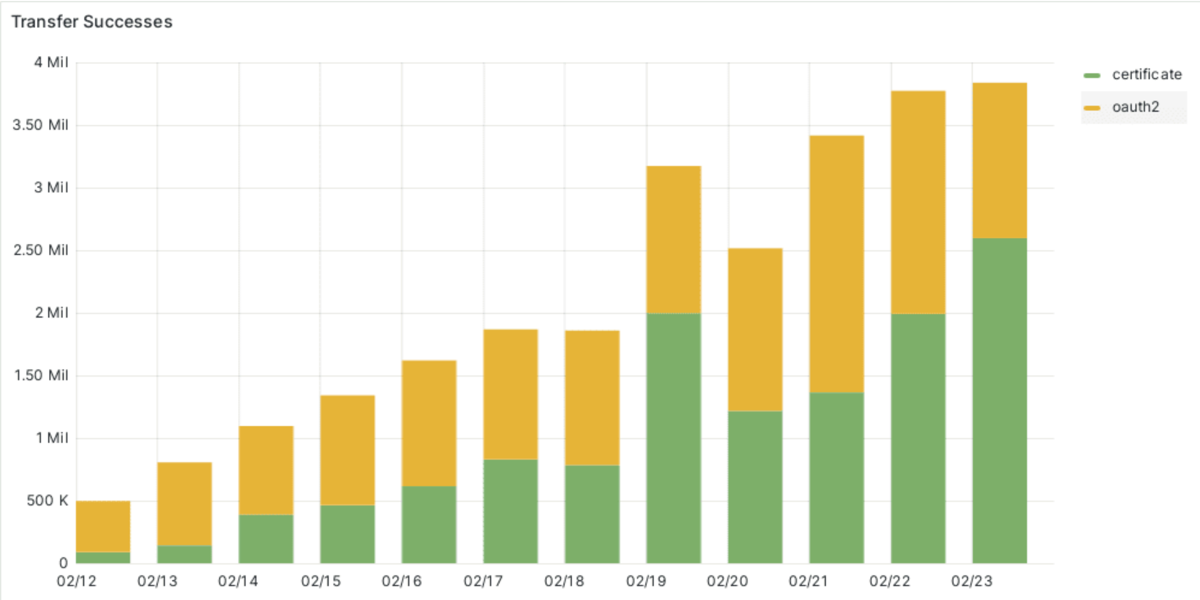


Figure 4.2.3a: Successful transfers using either certificates or tokens

However, this success came with additional costs for the operations teams, requiring increased effort in monitoring, debugging, and coordination. Moreover, the goal for the next Data Challenge is anticipated to be more ambitious, raising concerns about the scalability of current token workflows. Further development and testing will be necessary, along with a need for more token-specific monitoring.

Of particular concern to the Rucio team is the distribution of tokens capable of deleting large directories. Due to their usage, tokens are inherently more susceptible to leaks than certificates. For this specific use case, file-specific tokens may be more appropriate. ATLAS conducted a one-day test prior to Data Challenge 2024, concluding that it is premature to consider them. Nevertheless, more effort is required to strike a balance between security, robustness, and performance.

The complexity of tokens is further compounded by the level of flexibility they afford. Despite efforts to design token workflows, numerous questions remain unanswered. Consequently, the Rucio team supports the proposal to appoint a person or body responsible for overseeing the token effort across all projects, offering concrete and practical advice on future development.

4.3 Monitoring

For the Monitoring service, Data Challenge 2024 served as a deadline for several improvements identified during Data Challenge 2021. These improvements included enhanced Site network monitoring, improved XRootD Monitoring, and, importantly, harmonised data schema to enable unified visualization.

Overall, all deadlines were met, and all components necessary for the enhanced monitoring were implemented before the commencement of the Data Challenge events. As a result, this achievement is considered a success. However, there was significant firefighting required during the exercise, which helped identify several areas for improvement. These areas will be followed up as part of the WLCG Monitoring task force.

4.3.1 Site Network Monitoring

One of the areas identified as lacking in the previous Data Challenge exercise in 2021 was the ability to compare transfer throughput with actual network readings from the sites. To address this, a procedure was established for sites to register their network topology in CRIC and provide a monitoring URL containing JSON-formatted metrics to be scraped by the MONIT infrastructure and stored in a central location. This facilitated the creation of network plots, enabling comparison between transfer throughput and network metrics.

https://monit-grafana.cern.ch/d/MwuxgogIk/wlcg-site-network?orgId=16&from=1707692400000&to=1708901999000&var-site=All&var-bin=1d

At the time of the exercise, a pick of 51 sites was registered and reporting network metrics.

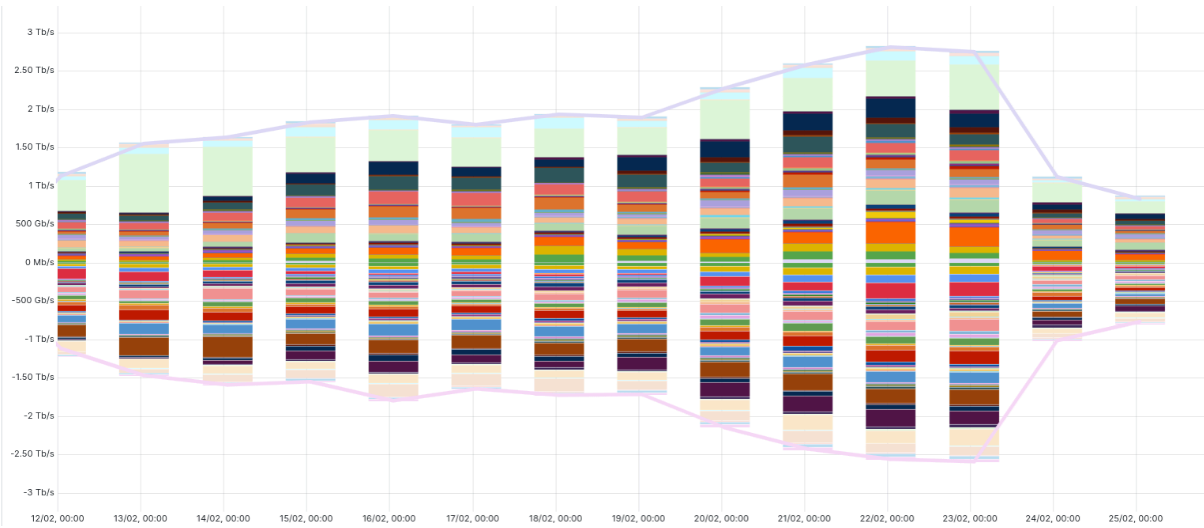


Figure 4.3.1a: Overview of the network ingress & egress transfers split per site

One of the issues identified with the Site Network Monitoring flow was occasional instances where some sites reported excessively high numbers, rendering the dashboard unusable. To address this, a fix was implemented to prevent negative numbers from being reported. Additionally, an attempt was made to limit the reported numbers to the registered WAN value in CRIC. However, this approach was reverted as it resulted in discarding valid values, suggesting that some sites may have had a lower WAN value registered in CRIC than their actual network capacity. The numbers of the network dashboard should be validated by the sites.

4.3.2 FTS Monitoring

Although not initially contemplated as something to look into, some issues appeared and raised some concern during the DC. During the Data Challenge exercise, the following issues were identified and addressed:

1. Remote access filter always set to true: Initially, this setting was adjusted to always be true. However, it was changed to be based on the source and destination site, ensuring more accurate filtering.

2. Incorrect transferred bytes for failed transfers: Initially, transferred bytes for failed transfers were based solely on file size, resulting in inconsistent results. This was rectified to report transferred bytes only for failed transfers, while retaining file size for completed transfers.
3. Some endpoints were not correctly enriched: This issue predominantly affected LHCb and required collaboration between CRIC and MONIT services to rectify and ensure correct enrichment of the endpoints.

4.3.3 XRootD Monitoring

Initially, this work was intended to encompass the replacement of the old GLED collectors with the Next Generation (NG) monitoring, which relied on two new components developed by OSG: the shoveler and the collector. However, the requirements expanded when the need arose to monitor dCache deployments with XRootD doors, particularly targeting FNAL and the CMS pileup data.

During the Data Challenge, approximately 30 sites were monitored using the new flows. The targeted list initially included only T0-T1 for new XRootD and FNAL for dCache. However, some additional sites volunteered to participate in the exercise.

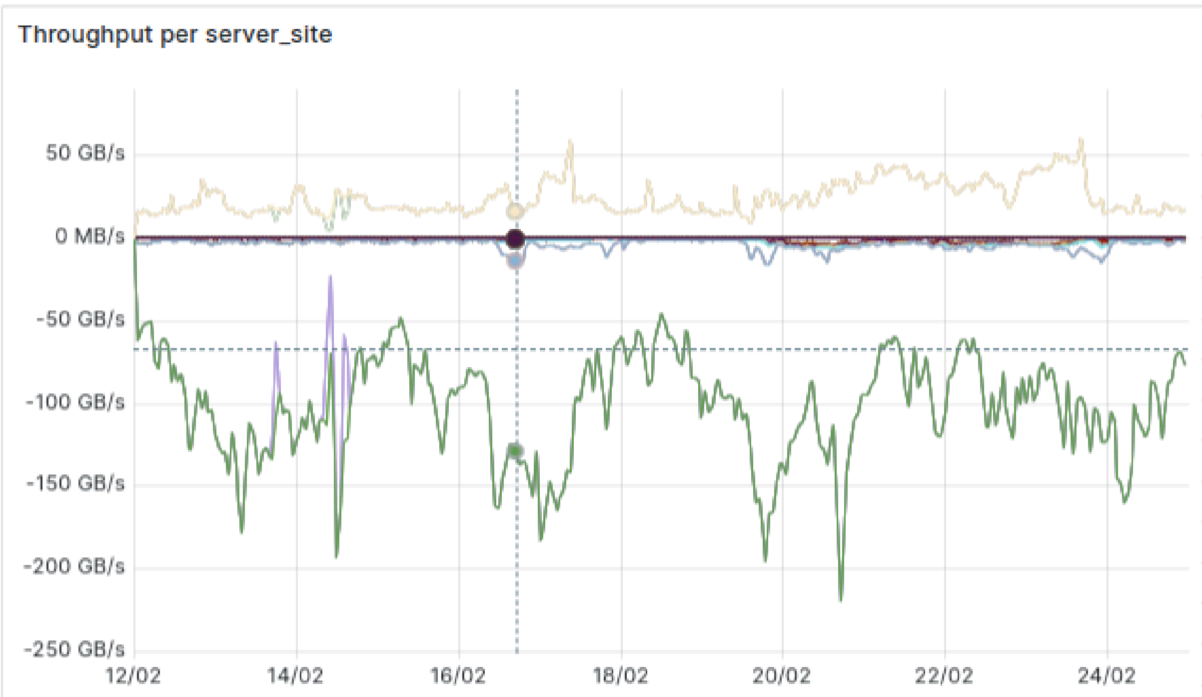


Figure 4.3.3a: Total throughput across all monitored XRootD servers

Both flows were established with sufficient time before the Data Challenge; however, many sites began to join just before or even during the exercise. This revealed various flaws in the new infrastructures that needed to be addressed in real time. During the exercise for the NG XRootD flow, the following two issues were identified:

1. Wrong FQDN resolution: This resulted in incorrect domain resolution, preventing proper data enrichment and plotting.

2. Swapped remote access flags: This led to incorrect data presentation based on the query by this field.

Moving forward, future work includes addressing reported instabilities in the new components. Additionally, efforts will be made to ensure more sites join this new flow, as the old ones will need to be stopped together with the CC7 end of life.

4.3.4 Data Challenge dashboard

A new Data Challenge dashboard was put in place adapting it to some of the new requirements:

- Adding non LHC VOs (Dune and Belle II).
- Provide information regarding the new XRootD Monitoring flows together with the existent FTS one.
- Add new filters based on authentication mechanisms.
- https://monit-grafana.cern.ch/d/d3543f53-950b-4a60-b353-16611bf7f5f7/dc-2024-draft?orgId=20&from=1707692400000&to=1708815599000
- Exclude site internal traffic
- The DC24 dashboard proves to be highly useful, yet it presents certain caveats. It amalgamates disparate data sources with inconsistent parameters for selecting and grouping the data. Consequently, it becomes challenging to consistently modify the plots. This issue, inherited from DC21, remains unresolved. For DC26, there should be a **requirement** for a minimal set of parameters from all participating experiments to be added to the dashboard. Additionally, the monitoring task force should collaborate to establish a common data source, addressing this challenge more effectively.
- Discrepancies between experiment dashboards and DC24 or its next iteration have been noticed and should be understood.

4.3.5 General overview

In conclusion, the expected work prepared by the Monitoring Task Force was available in time for the DC 2024 exercise, and overall, operations proceeded effectively despite encountering various challenges. A major issue identified pertains to the lack of expert feedback before and, in many cases, during the challenges, which would have facilitated proactive preparation and response to some of the data-related issues. Nevertheless, the DC exercise has once again proven invaluable in identifying new problems that must be addressed before the real HL scenario.

4.4 IAM

During DC24, the IAM service token endpoints for ATLAS, CMS, and LHCb experienced heavy usage by Rucio, DIRAC, and FTS. While all IAM services remained available, their databases became overloaded with tokens, resulting in slow response times. Consequently, client timeouts and transfer failures occurred. Throughout the challenge, the services were closely monitored, and several measures were implemented to enhance performance, recognizing that the ways tokens were used were suboptimal in various respects.

The three experiments employed tokens differently. ATLAS and CMS tokens had broad scopes, enabling extensive reuse, while LHCb used file-based scopes. Notably, the token request patterns differed significantly between ATLAS and CMS. Table 4.4a summarises the numbers:

Experiment	Issued tokens	Max. number of tokens in DB	Peak token request rate	Typical token request rates
ATLAS	2.6 M	1.03 M	5 Hz	3 Hz (12 days)
CMS	2.7 M	0.97 M	200 Hz	60 Hz (6h), 20 Hz (10h), 1 Hz (11 days)
LHCb	3.4 M	1.65 M	120 Hz	25 Hz (2 days), 1 Hz (10 days)

Table 4.4a: Token counts across experiments

Several factors contributed to the accumulation of tokens:

- Token lifetimes were relatively long, up to 30 days, which slowed down their cleanup process by IAM during DC24. However, they were automatically cleaned up afterwards.
- Database cleanup processes were slow, which was addressed by increasing the connection pool size.

The introduction of an index on the refresh token tables helped expedite all operations involving those tables. Additional lessons learned include:

- Modification of the token management engine (a third-party product) is necessary to prevent access tokens from being stored in the database. This will enhance the speed of all token-related operations.
- The extensive use of refresh tokens may not be optimal. Discussions are planned among Rucio, DIRAC, FTS, and IAM experts to explore more efficient methods of orchestrating token usage for large-scale data transfers.
- IAM performance tests need to focus more closely on latency measurements.

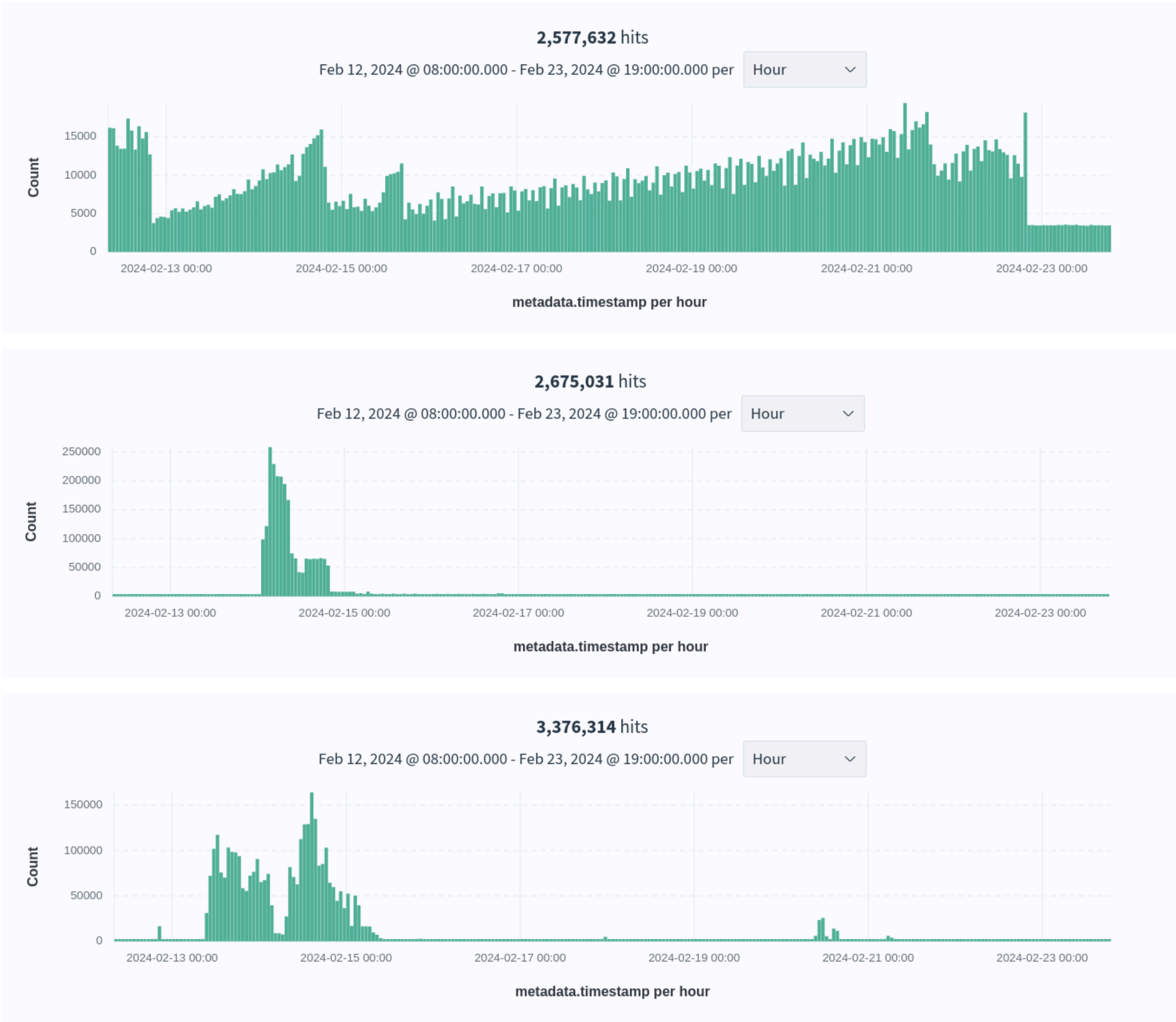


Figure 4.4b: These three plots show the token request counts from the experiments, in order from top to bottom ATLAS, CMS, LHCb.

5 Networks

5.1 Research and education networks

ESnet implements a routing policy characterised by a "cold-potato" approach for managing traffic. This strategy involves the acceptance of incoming traffic into ESnet as close to its source as feasible, while concurrently seeking to deliver the traffic as proximate to its intended destination as possible, thereby maximising the duration of traffic within the ESnet infrastructure. This routing policy is orchestrated through human coordination and formal agreements, leveraging Border Gateway Protocol (BGP) Multi-Exit Discriminator (MED) attributes to denote relative costs associated with routing decisions.

Conversely, GEANT adopts a routing policy that prioritises routes originating from research and education networks over those emanating from Internet Exchanges or commercial peers, accomplished through the manipulation of local preference settings. Additionally, GEANT advertises

MEDs to its BGP neighbours, with these MEDs derived from Interior Gateway Protocol (IGP) metrics. By disseminating these MEDs, GEANT ensures that incoming traffic from neighbouring networks enters its infrastructure at a location nearest to its eventual destination. Notably, neighbouring networks may opt to honour the announced MEDs if they so choose. Furthermore, GEANT allows peers to select between hot-potato and cold-potato routing strategies, or alternatively, to employ a hybrid approach combining elements of both methodologies.

During the Data Challenge, there was a specific focus on optimizing transatlantic paths, including those beyond the ESnet network, especially in the Europe to US direction. However, the delayed delivery of the Amitie cable raised some concerns regarding the potential impact of two 400G LHCOPN circuits on LHCONE traffic.

The primary objective of traffic engineering during this period was to enable GEANT to handle LHCONE traffic originating from London and Paris to the US via their transatlantic infrastructure, thereby relieving pressure on ESnet's links. It was essential to maintain the existing traffic flow between GEANT and ESnet at Amsterdam and Geneva unchanged, although adjustments were prepared if required.

Notably, ESnet did not transmit Border Gateway Protocol (BGP) Multi-Exit Discriminator (MED) attributes for LHCONE peerings with GEANT. This approach proved successful, facilitating the optimal utilization of infrastructure capabilities while achieving the desired traffic engineering objectives.

5.2 Flows and Packets tagging

The Research Network Technical Working Group (RNTWG) has been actively working to enable the deployment of flow labelling and packet marking, aiming to make visible the ownership and associated activity of R&E flows across the network. For DC24, our focus was on sending "firefly" packets, which are UDP packets containing a JSON-formatted syslog schema identifying the owner, activity, and associated flow information such as source and destination addresses, ports, and protocols.

During DC24, we achieved over 80% participation from the CERN EOS CMS instance and the USCMS Nebraska Tier-2 in sending fireflies for their respective data flows. To enable these sites to send fireflies, it was necessary to ensure that the full DDM stack (comprising Rucio, FTS, and XRootD) passed the experiment and activity information all the way to the endpoint responsible for sending the data. New versions of Rucio and FTS were installed prior to DC24 for both ATLAS and CMS to support this initiative.

5.2.1 Results

We confirmed the capability to propagate Scitags all the way to the storages (for both ATLAS and CMS), that fireflies were sent from XRootD and EOS, that fireflies were captured at ESnet collector, that ESnet setup a live dashboard, and that with limited deployment we were able to get valuable insights into flow durations, their characteristics (splits by exp/activity), sources of IPv4 traffic (split by applications) and potential impact of new TCP congestion algorithms (performance correlated with flow data).

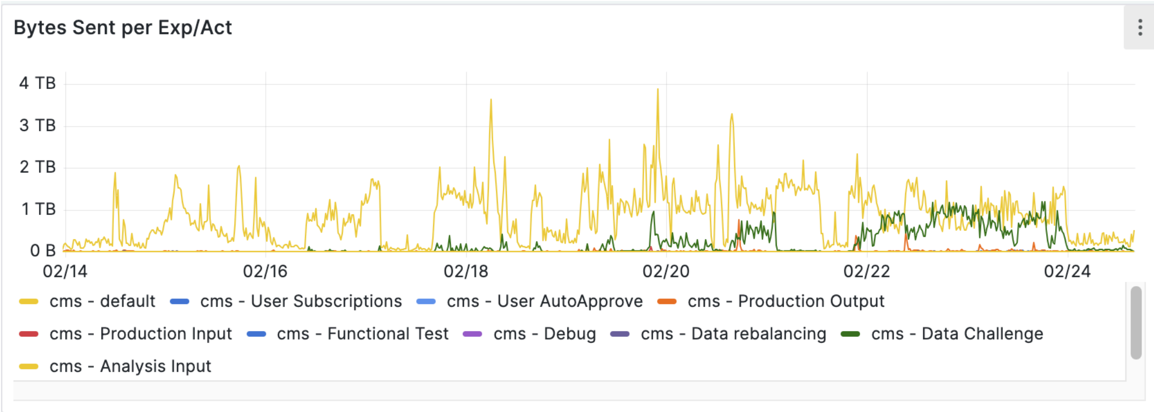


Figure 5.2.1a: University of Nebraska (UNL) shows non-FTS traffic was dominant

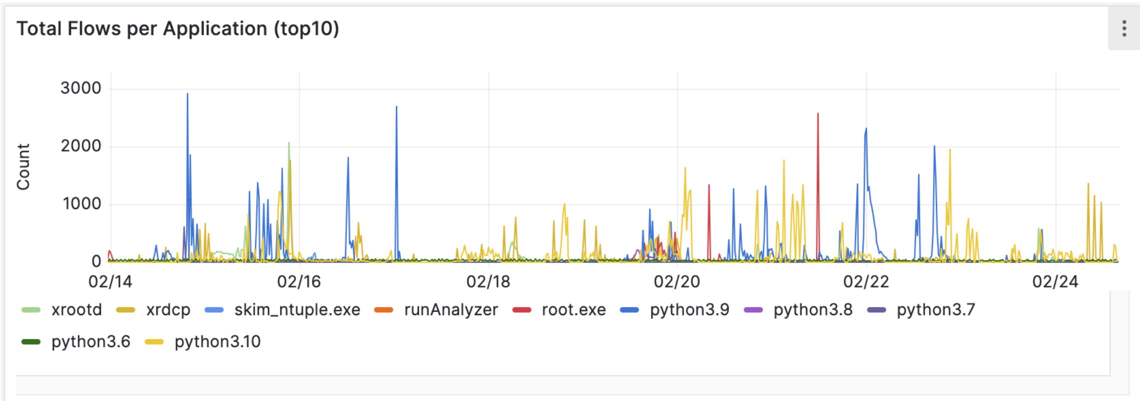


Figure 5.2.1b: Capability to show a split by application as reported by XRootD

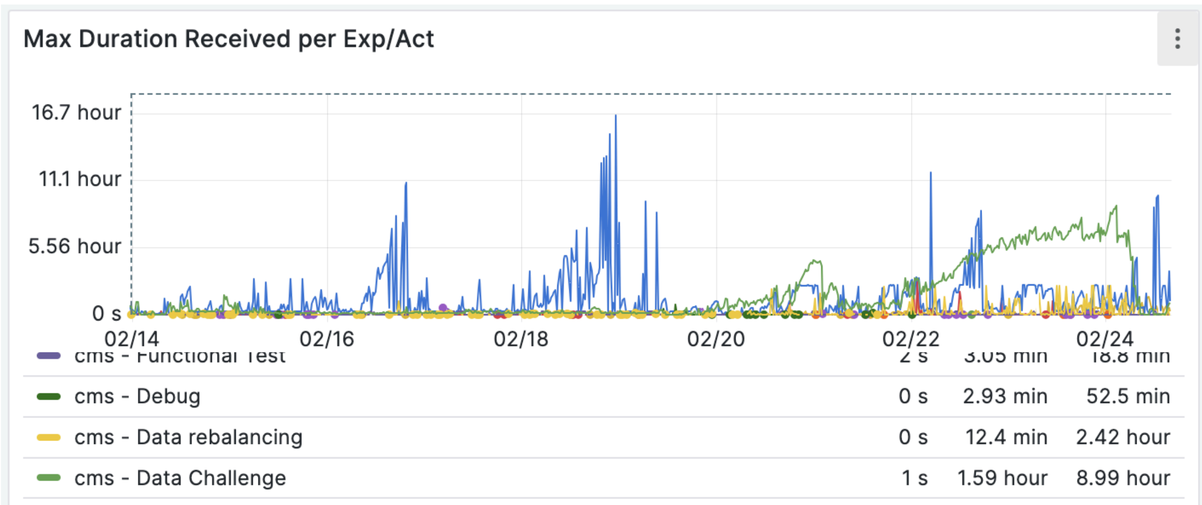


Figure 5.2.1c: CERN EOS CMS max duration of flows split by Exp/Activity.

5.2.2 Issues

We hit an issue with XRootD crashing when receiving scitags http headers. This had impact on ATLAS testing and availability of the ATLAS XRootD storages. The issue was fixed quickly but we were unable to rollout (as DC was already running).

5.2.3 Next steps

These are preliminary results, and we will need more time to look in more detail, specifically: round-trip times that were reported in the flows and can be used to analyse how same destinations performed over time, compute transfer rates as we have both durations and volume, as well as the correlation between applications and IP versions and destinations.

We need to look at the possibility to expand data in the fireflies. This includes reports on retransmits, TCP congestion algorithm, etc., i.e., metrics that would indicate that the network is under stress.

We need to review existing coverage. At CERN we were missing some flows as we didn't have any fallback configured. At UNL there were many flows marked in default activity. We need to understand where they're coming from and how we could instrument them.

For the collectors, the site-collector at CERN was setup to drop fireflies with “local” traffic (within CERN). This might be better fixed in XRootD. However site-collector will likely be useful anyway as we could store and analyse local data easily.

5.3 NOTED

We tested a new NOTED version. The start actions were triggered by link saturation alarms generated by the CERN network monitoring system (CERN NMS) and post confirmation by the presence of large data transfers in FTS to the WLCG site. The stop actions were generated based on the decrease of FTS transfers below a certain throughput threshold.

During DC24, the link saturation alarms in CERN Network Management System (NMS) were set at approximately 50% of the link capacity, as saturation was rarely observed. The Network Observation and Traffic Engineering Daemon (NOTED) ran throughout DC24. For the first 10 days, it did not take any action on the network but instead triggered a warning when additional bandwidth was needed. In the last 3 days, NOTED began taking real actions on the network by performing load balancing for CA-TRIUMF, ES-PIC, and DE-KIT, and it operated in dry-run mode for the rest of the Tier 1 sites. For CA-TRIUMF the load-balancing traffic was routed over the primary and backup link. For ES-PIC and DE-KIT the load-balancing was routed over the LHCOPN and LHCONE networks.

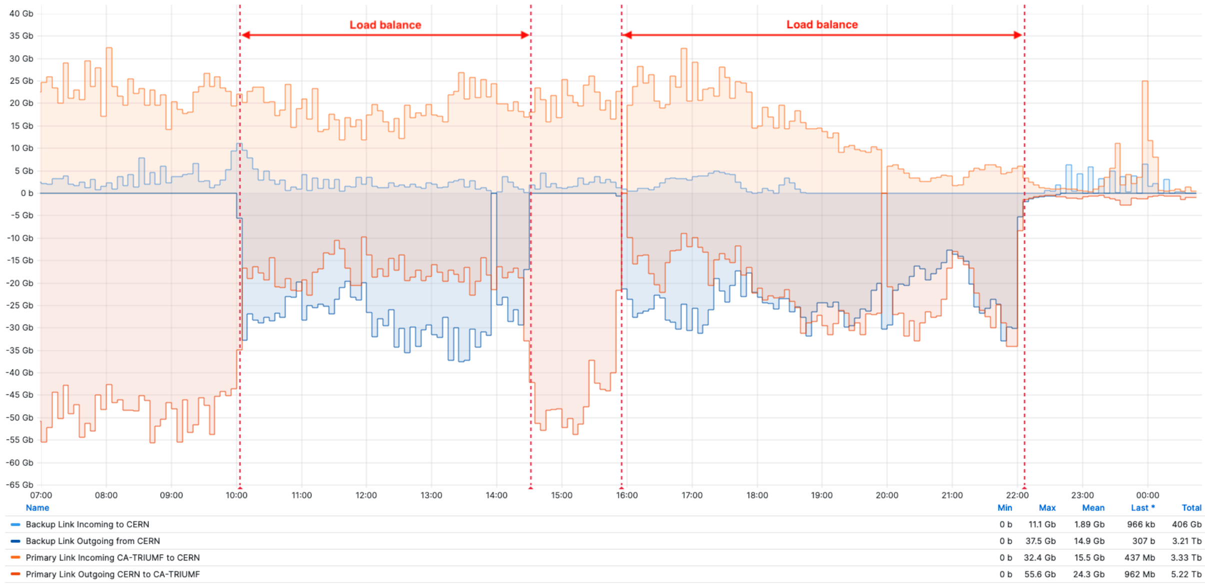


Figure 5.3a: Loadbalancing on CA-TRIUMF link

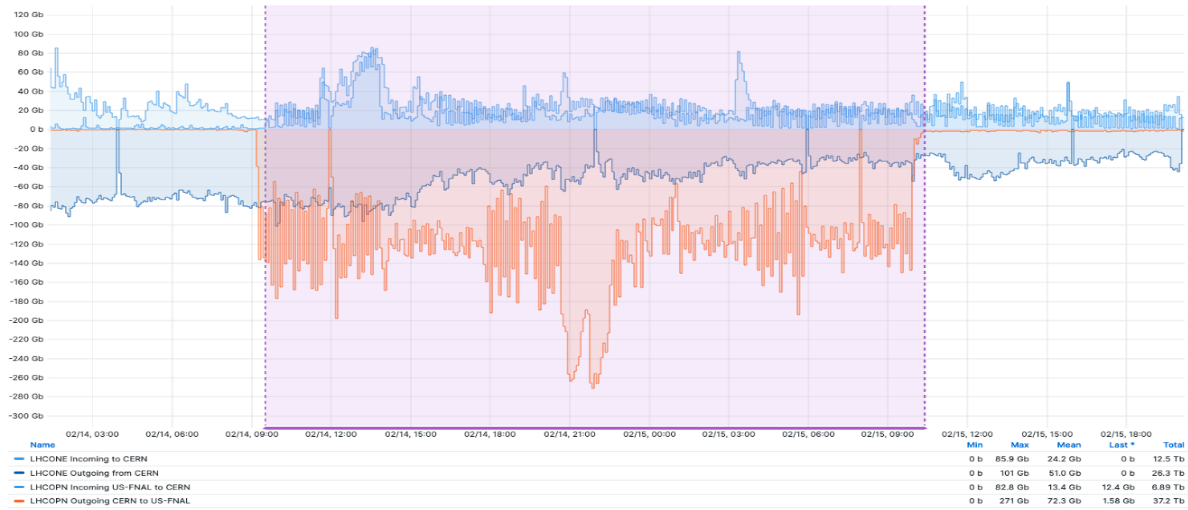


Figure 5.3b: Dry-run on US-FNAL LHCOPN link

The exercise proved to be valuable in addressing various NOTED-driven router configuration issues, such as implementing load-balancing routing for sites with multiple LHCOPN links and ensuring full BGP configuration on both LHCOPN routers. Additionally, it was beneficial for implementing additional checks on FTS. Overall, the results were satisfactory, with the large transfers detection functioning correctly in most cases.

5.4 SENSE

The primary objectives of the Rucio/SENSE activity during DC24 were as follows:

1. Demonstrating the overall system, namely Rucio-DMM/SENSE-FTS-XRootD, can manage multiple Rucio-triggered data flows between at least 2 different pairs of sites.
2. Showcasing the modify feature of DMM, which allows bandwidth allocation to change dynamically to adapt to changing requirements.

3. Testing the real-time API-driven tuning of FTS active/max transfers to synchronise data transfer rate with underlying network resource availability.

These tests used production network resources at FNAL and ESnet, along with prototype deployments of Rucio and site XRootD clusters. The Rucio instance used the CERN FTS-pilot server for the third-party copy operations.

5.4.1 Initial test

We aimed to demonstrate our capability to create and manage bandwidth allocation shares among three data flows: one from UCSD to Caltech and two additional flows, one from UCSD to Caltech and another from FNAL to Caltech, utilizing Rucio-DMM-SENSE/FTS/XRootD.

During this test, we observed that the aggregated bandwidth of all three data flows was notably lower than the available 100 Gbps at Caltech. We hypothesized that our hard Quality of Service (QoS) policies did not align with XRootD's load balancing mechanism across the heterogeneous nodes available, resulting in underutilization of bandwidth in some nodes.

5.4.2 Modifications and improvements

With the aim of achieving better control over bandwidth allocation shares, we tested three different configurations of Network Quality of Service (QoS) policies:

1. Implementing hard QoS only at the Data Transfer Nodes (DTNs) level.
2. Allowing QoS overcommit.
3. Enabling QoS overcommit at the DTNs level and implementing hard QoS at the network level.

Additionally, we transitioned to using FDT instead of Rucio/FTS/XRootD to generate our data flows more easily. Finally, we constrained the available bandwidth at Caltech to 80 Gbps to accommodate other services sharing the same path. This adjustment facilitated a more successful test, as depicted in Figure 5.4.2a.

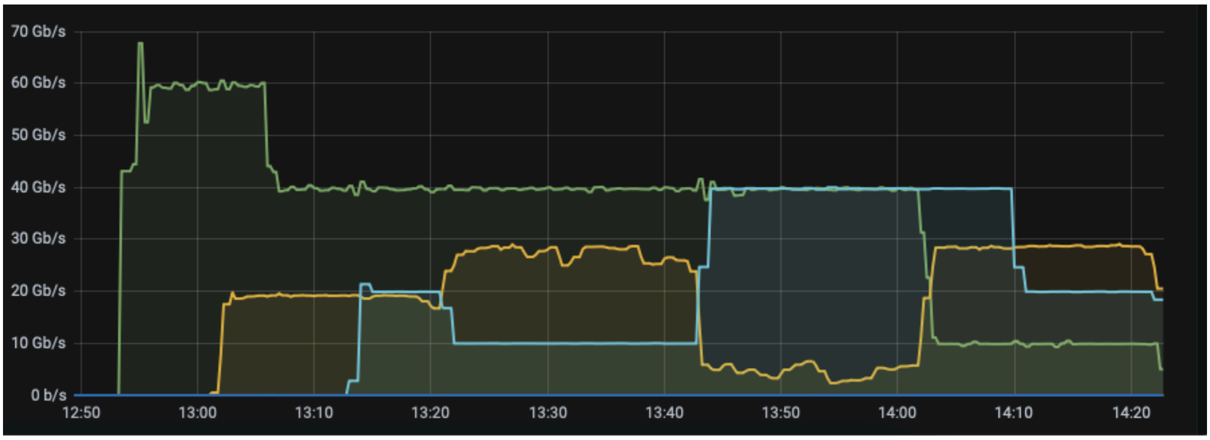


Figure 5.4.2a: Constrained bandwidth test at Caltech

Each color represents one of the three data flows sharing the 80 Gbps of available bandwidth. The allocation shares are more stable, and the usage is significantly improved compared to the initial test. Additionally, bandwidth modifications are observable when there are abrupt changes in the breakdown of bandwidth among the flows.

5.5 Use of IPv6

During DC24, IPv6 traffic on LHCOPN accounted for an average of 86.9% of the total traffic. Figure 5.5a is slightly higher than the average value observed over the last year, which stood at 80.7%.

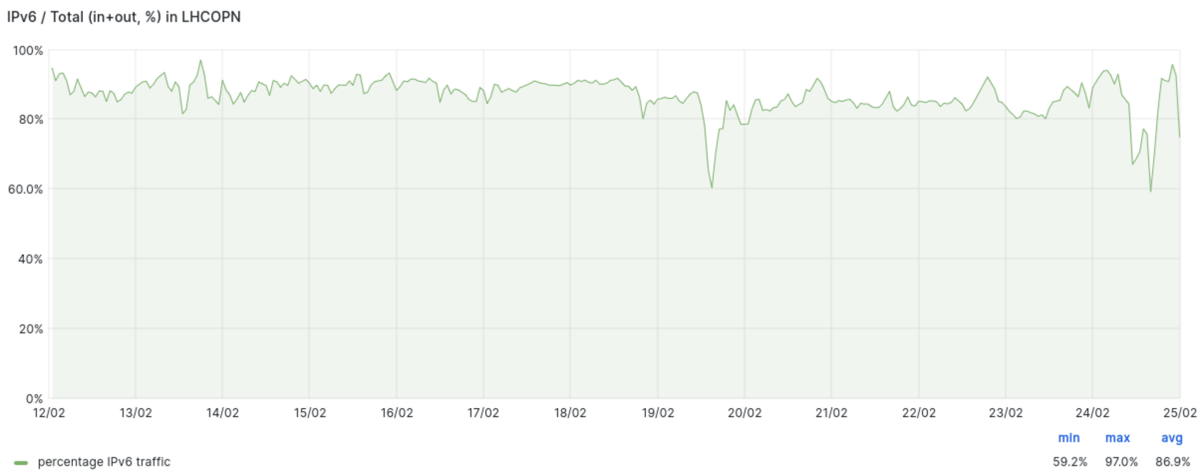


Figure 5.5a: Percentage of IPv6 traffic over the total (IPv4+IPv6) on the LHCOPN links measured on the CERN LHCOPN routers

5.6 TCP congestion protocols: BBRv1 vs. CUBIC

During the last 3 days of the event, 43 EOS servers at CERN, used by both ATLAS (23 servers) and CMS (20 servers), were alternating their TCP congestion protocol every 2 hours (switching every 6 hours starting from Thursday).

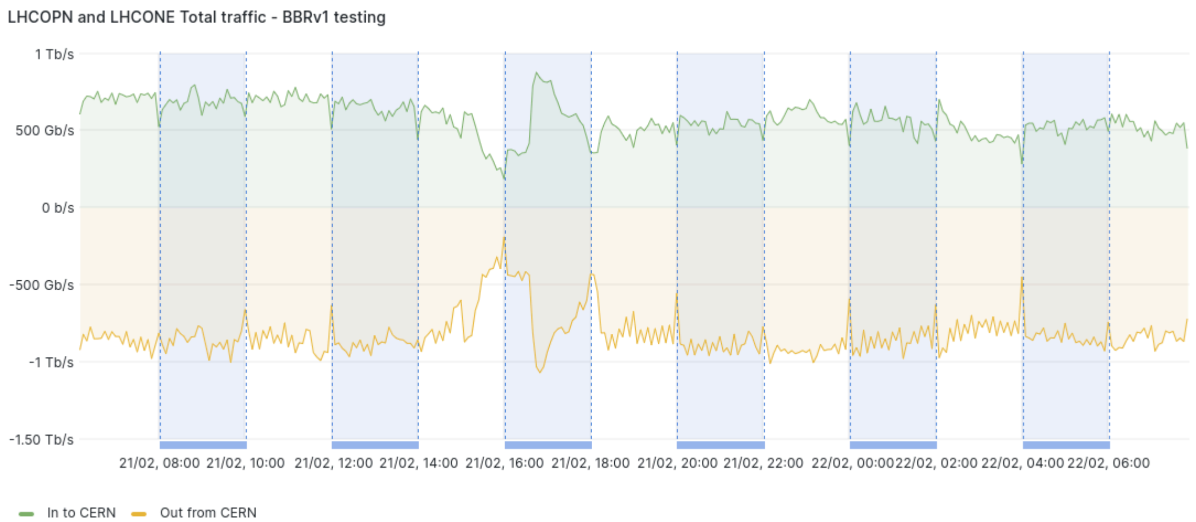


Figure 5.6a: TCP Congestion variants. Blue: BBRv1 active. Otherwise: Cubic

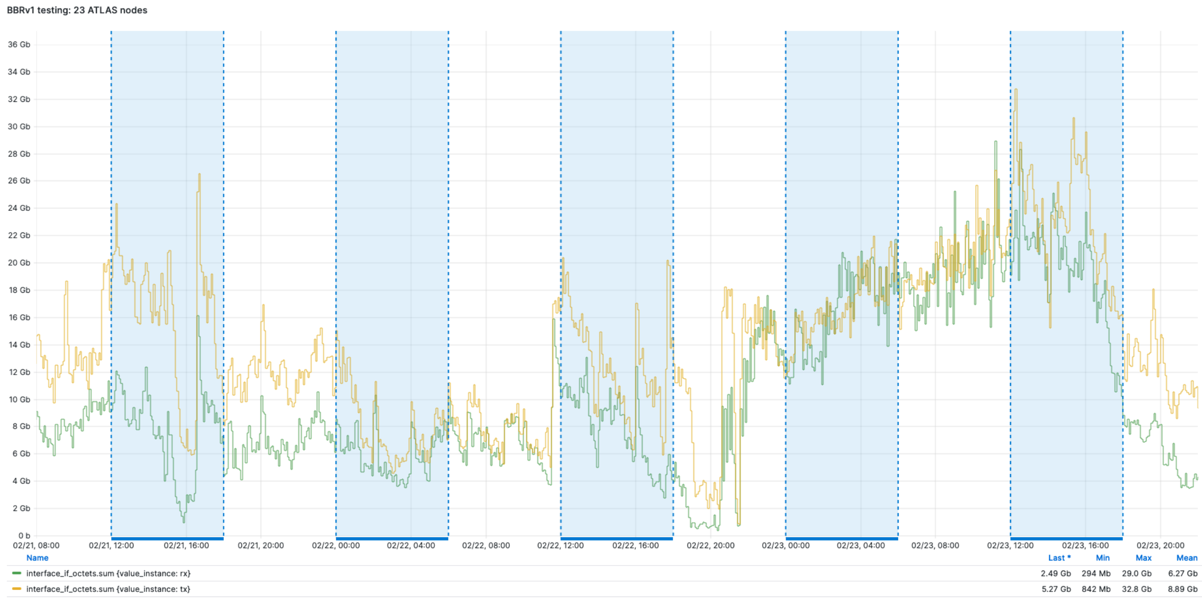


Figure 5.6b: Evaluation on ATLAS nodes

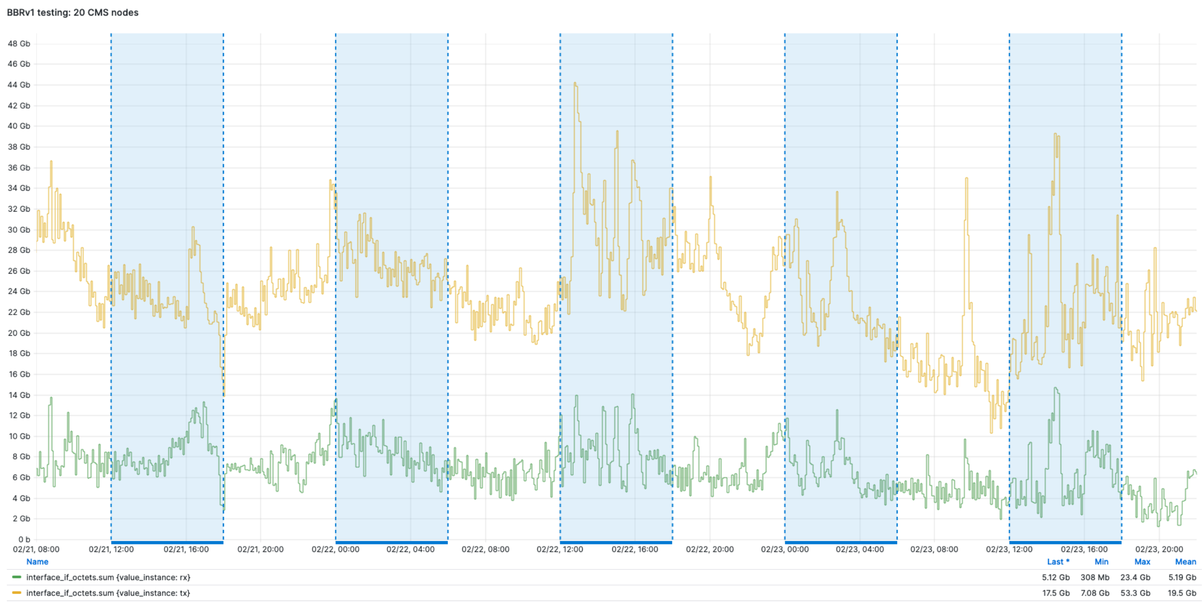


Figure 5.6c: Evaluation on CMS nodes

There was no evidence of gain or loss observed when using BBRv1. This lack of discernible impact is likely because BBRv1 tends to demonstrate advantages in congested network conditions, which were not prevalent during the testing period.

5.7 Networking at CERN

The CERN network operated flawlessly throughout the entire period. Both the LHCOPN and LHCONE links were used at high rates, yet neither of them experienced congestion. Additionally, the general internet access through the CERN Internet firewall maintained a load consistent with normal days. This positive outcome indicates that the majority of heavy traffic was contained within LHCOPN and LHCONE, without impacting CERN's general services.

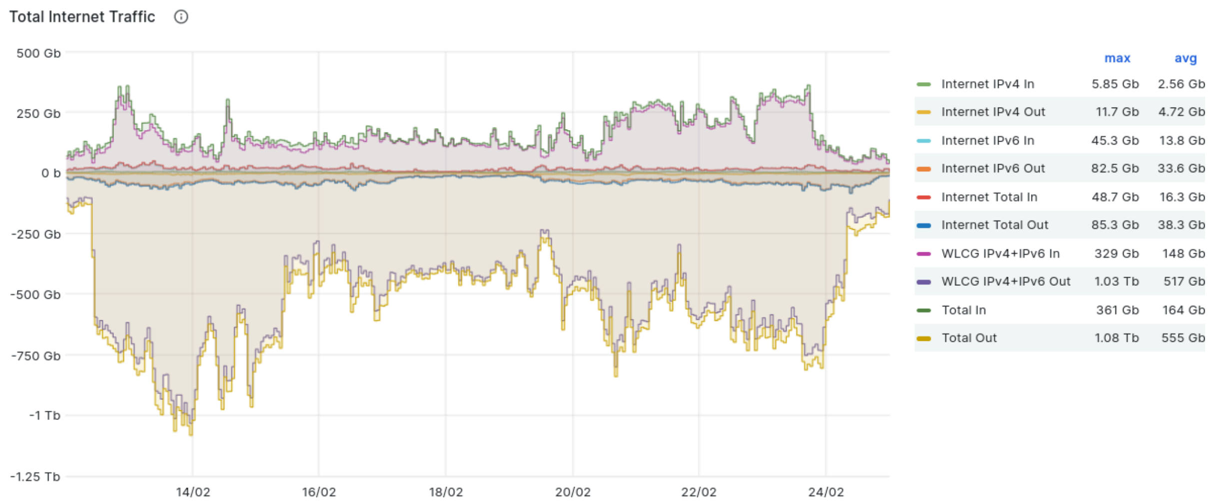


Figure 5.7a: CERN total Internet traffic statistics, including General Internet, LHCOPN and LHCONE

5.8 LHCOPN

The LHCOPN network performed exceptionally well throughout the entire period, with no indications of prolonged congestion. However, during the first three days of DC24, the link to UK-RAL was unavailable due to a fiber cut in the Channel. Fortunately, traffic was seamlessly rerouted to LHCONE, ensuring continuous connectivity and minimal disruption.

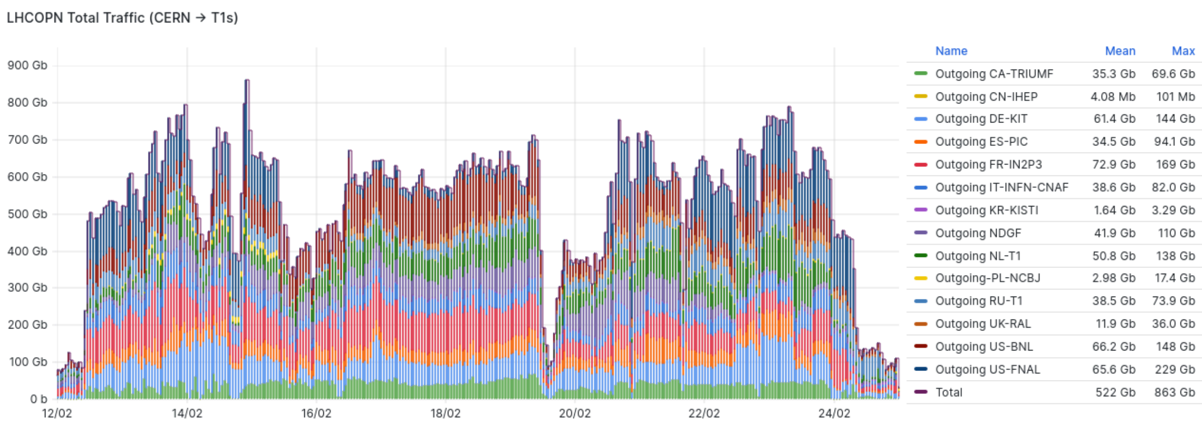


Figure 5.8a: Aggregated LHCOPN statistics collected from CERN LHCOPN border routers: T0 to T1s

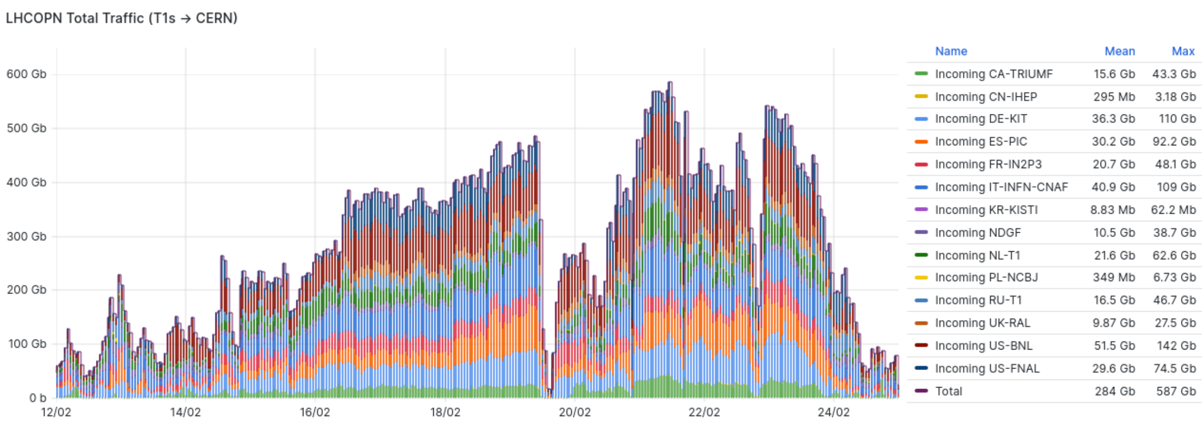


Figure 5.8b: Aggregated LHCOPN statistics collected from CERN LHCOPN border routers: T1s to T0

6 Sites

6.1 BNL

During DC24, the target rates for BNL were not met. Subsequent tests have demonstrated that the site is capable of achieving throughput rates well beyond those specified in the DC24 flexible scenario rates, with the potential to reach its design limit of 1.6 Tb/s. High-resolution monitoring of network throughput reveals a comb-like shape during T0 expert tests, which is potentially attributed to low transfer request frequency and intercity factors. No bottlenecks were observed in our storage system during DC24, and the compute operations of the site remained largely unaffected by the event. A detailed report can be found at here.

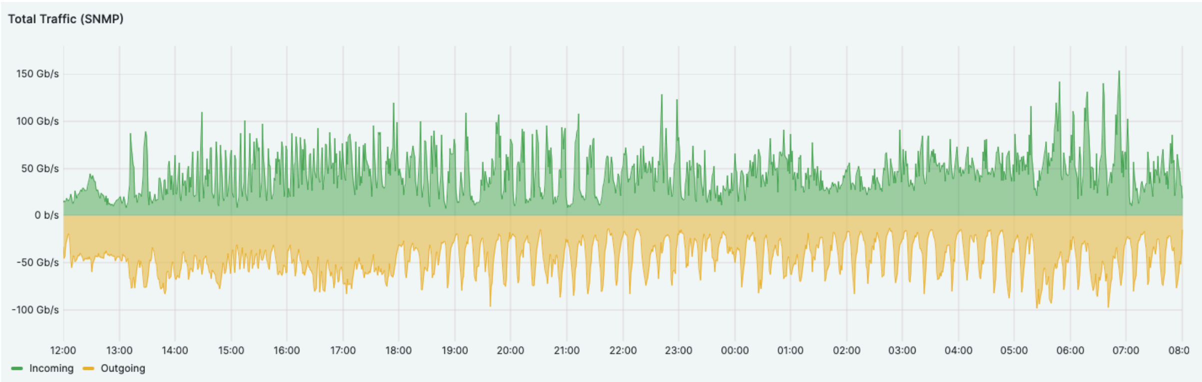


Figure 6.1a: BNL total SNMP measured traffic

During this Data Challenge, BNL did not experience any indications of network bottlenecks on our inter-switch links, whether LAN or WAN. The peak values of ingress (blue, 265 Gb/s) and egress (brown, 300 Gb/s) remained well below the BNL network capacity.

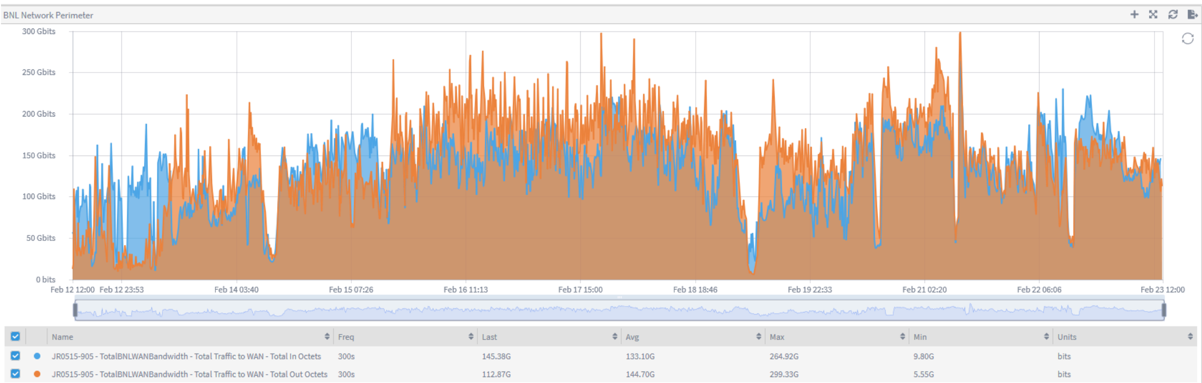


Figure 6.1b: BNL ingress and egress overlaid

No issues were identified with storage hardware regarding CPU, memory, or load during the Data Challenge. Comparing available monitoring tools and understanding discrepancies in their dashboards can be a challenging and time-consuming task. The outcome of DC24 was discussed at BNL during a workshop that brought together diverse USATLAS managers, ATLAS computing experts, storage and network specialists, including the core dCache team. During this session, the network configuration and deployment model of dCache at BNL were thoroughly elucidated.

6.2 CERN

To ensure the stability, availability, and performance of the EOS Open Source storage system for seamless data exports to Tier 1 facilities, preliminary tests were conducted by the EOS team. These tests, conducted on ATLAS, CMS, and LHCb EOS instances, aimed to assess the system's capability to handle throughput levels exceeding the predefined thresholds. On January 23rd, 2024, the platform demonstrated a throughput exceeding 550 GB/s and reaching 600 GB/s without disrupting ongoing production traffic.

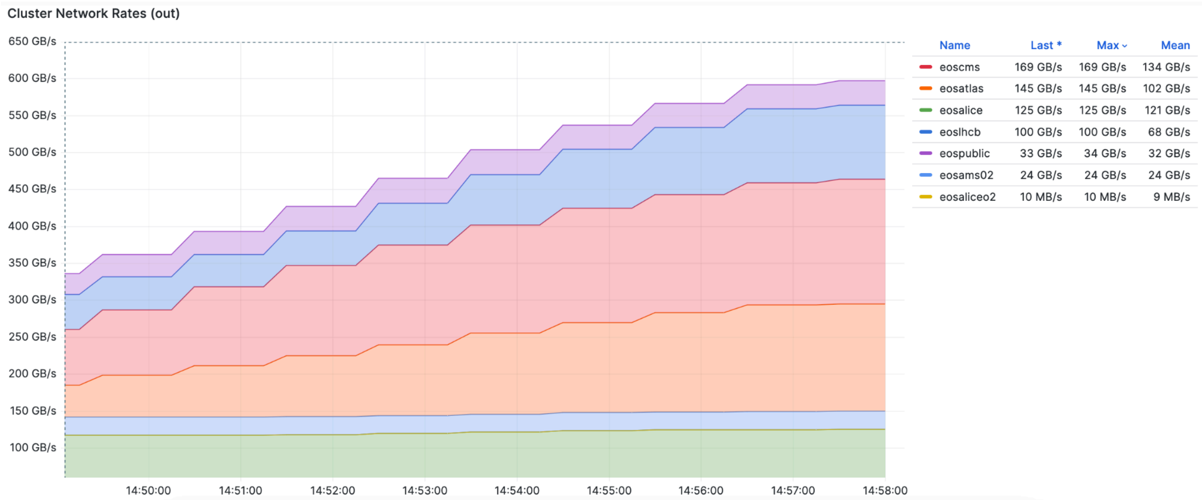


Figure 6.2a: CERN network throughput evolution

Subsequently, during the data challenge period, network performance metrics illustrated consistent and stable throughput from the CERN site to Tier 1 facilities, with no discernible issues stemming from the CERN EOS instances.

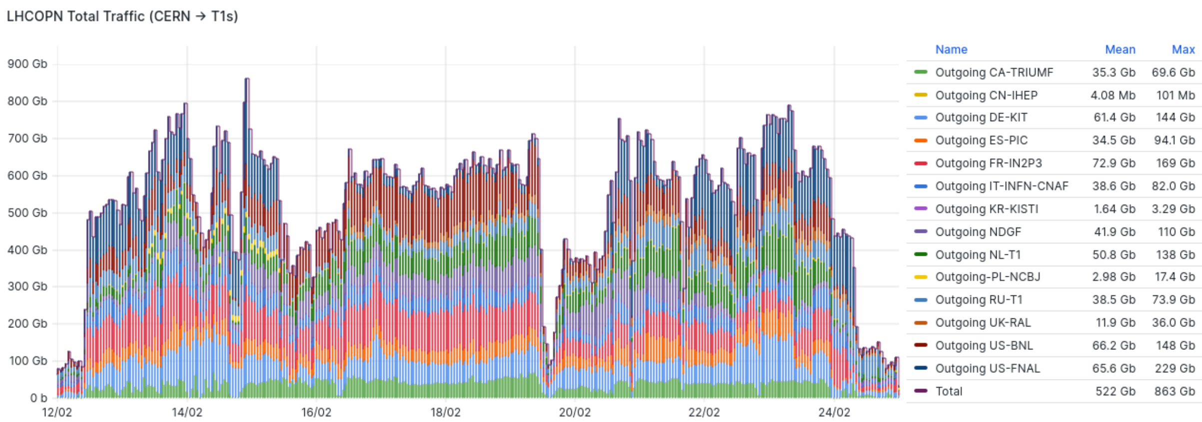


Figure 6.2b: CERN to Tier-1s traffic over LHCOPN

Notably, during the second week of the challenge when experimental data throughputs increased, the EOS cluster performed as expected, reaching a peak throughput of 450 GB/s. This throughput level, albeit lower than that achieved during internal tests, suggests potential capacity for experiments to further push their data processing and transfer limits.

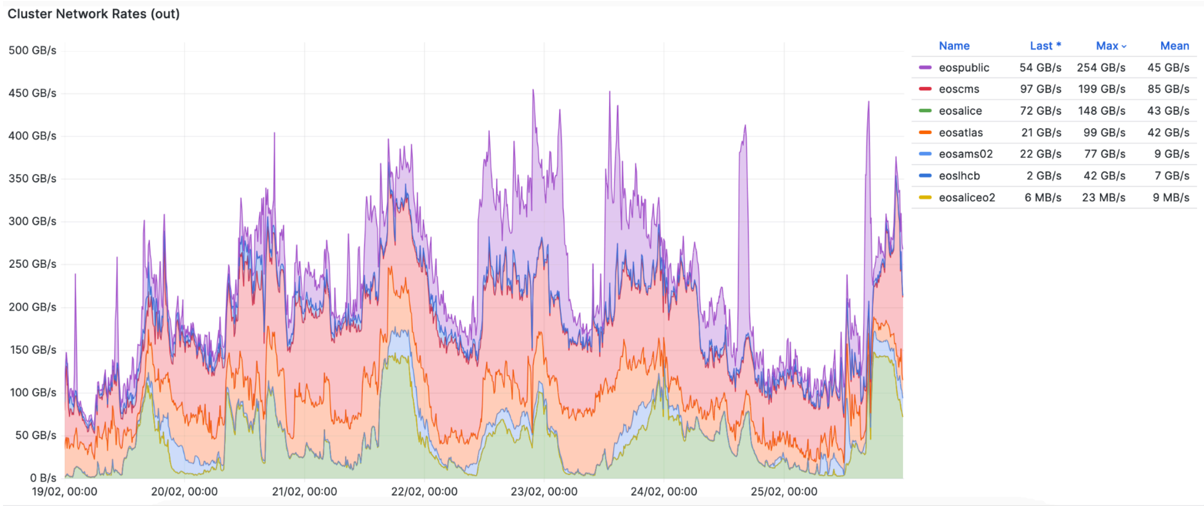


Figure 6.2c: Outgoing network rates

The following plot illustrates that the volume of HTTP traffic was notably lower compared to the production traffic handled by XRootD. Even so, the combined total of both was less than the throughput achieved during our internal tests, suggesting that the experiments could still accommodate additional export transfers.

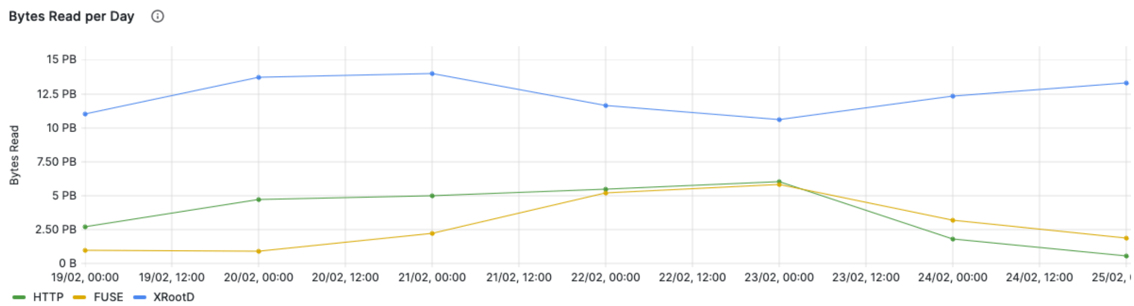


Figure 6.2d: Bytes read per day with various access methods (HTTP, FUSE, XRootD)

Moreover, the deployment of HTTP tokens across all experiments at the CERN site facilitated the testing of this secure access to EOS instances. The effective utilization of HTTP tokens is evident in data transfer service metrics, as depicted in the provided plot, with the green line representing token usage.

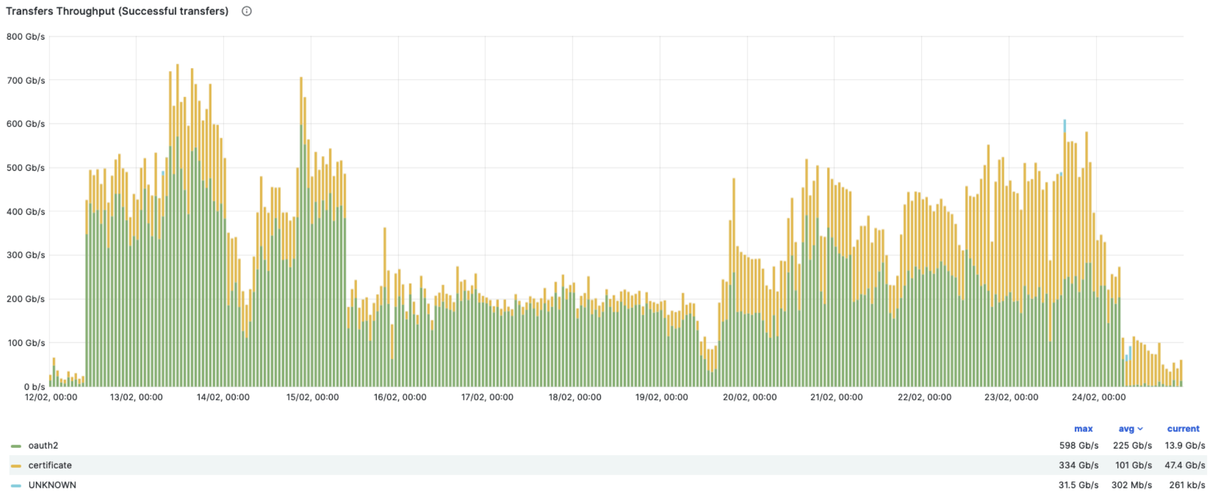


Figure 6.2e: Aggregated throughput of successful transfers by certificate or token

Simultaneously, HTTP has now fully replaced all the transfers previously conducted using the deprecated GridFTP protocol. It is imperative to acknowledge that GridFTP facilitated multistream capability for single transfers, whereas HTTP currently lacks this feature, necessitating the submission of multiple transfers to attain equivalent outcomes as previously achieved.

Furthermore, the deployment of scitags, a packet marking configuration, for the CMS experiment was instrumental in monitoring various activities, as evidenced by network activity plots.

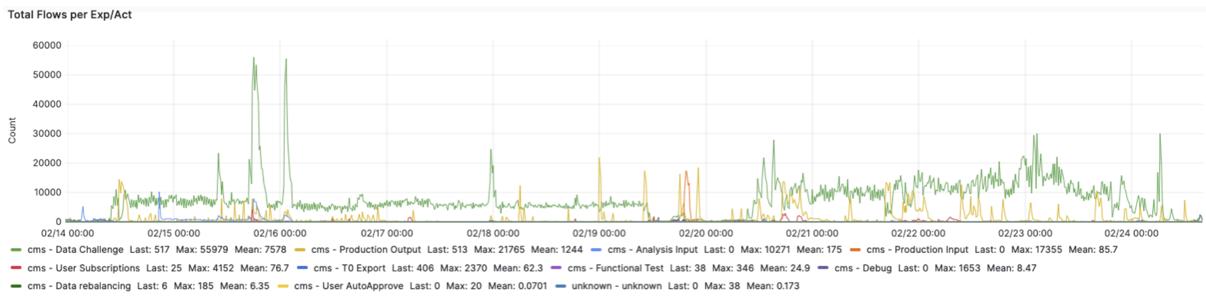


Figure 6.2f: Total number of flows per Exp/Activity

Lastly, during the DC24 period, the EOS team at CERN facilitated the testing of congestion control algorithms (BBR versus Cubic) for the ATLAS and CMS experiments. This involved alternating between the BBRv1 and Cubic congestion control algorithms every six hours across designated storage nodes, already upgraded to the Alma9 operating system. This activity enabled a comprehensive evaluation of algorithmic performance for the network team. The network team noted the satisfactory performance of BBR upon its activation, alleviating initial concerns stemming from its higher aggressiveness compared to Cubic. However, the absence of congestion during DC24, attributed to the exceptional network performance, precluded the acquisition of definitive findings. Consequently, the network team is currently exploring strategies to introduce latency for future testing phases.

6.3 CNAF

General context

During DC24, the INFN-CNAF Tier1 participated in supporting ALICE, ATLAS, CMS, LHCb, and Belle II experiments.

ATLAS, CMS, LHCb, and Belle II experiments used StoRM for their data management needs. Each of these experiments had dedicated StoRM WebDAV endpoints, except for Belle II, which shared resources with multiple other experiments.

StoRM WebDAV endpoints supported token authentication and authorization (AuthN/Z) for ATLAS, CMS, and Belle II. However, it could not be configured for LHCb due to ongoing discussions regarding paths within token scopes, as documented in GGUS ticket 165048.

On the other hand, ALICE used dedicated XRootD endpoints for its data management.

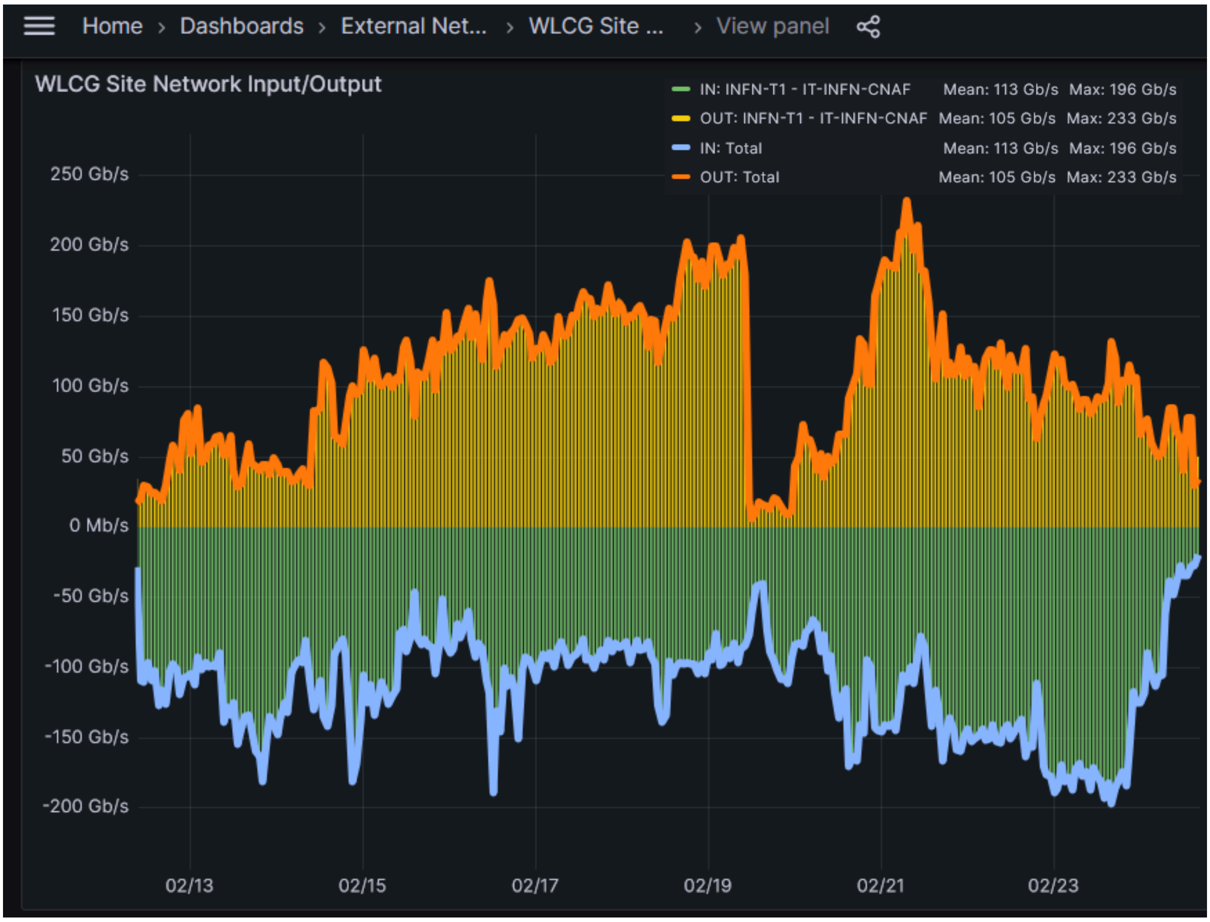


Figure 6.3a: A view from the WLCG network monitoring during DC24 for INFN-T1 site. Drop of pressure between 19th and 20th February is due to central services issues.

Specific feedback

For ALICE, the target rate was achieved and surpassed. There was a 24-hour interruption of transfers on February 21st to prevent the buffer from filling up, as the experiment had been writing at a much higher rate than expected (alongside concurrent reads for md5sum calculation). SSD disks were incorporated into the buffer filesystem to manage such intense I/O, and ALICE adjusted the number of streams to maintain a consistent rate of 1-1.5GB/s thereafter.

For ATLAS, the T0 export rates were not attained in the second week, potentially stemming from prioritisation issues in FTS. Indeed, rerunning the T0 export yielded significantly improved results (GGUS:165526). Overall, the rates surpassed the anticipated values, and we noted transfer failures along with significant disparities between internal monitoring and FTS monitoring (GGUS:165355).

For CMS, adjustments to FTS parameters were made prior to DC24 to ensure smooth testing, particularly during the first week. We met and exceeded the anticipated rates, although storage endpoints became saturated in the second week, leading to CMS SAM tests registering as red.

For LHCb, we attained and surpassed target rates in the initial phase of the exercise (EOS-CNAF_disk, CNAF_disk-CNAF_tape), despite experiencing numerous transfer failures. Recently, we repeated the exercise with a maximum of 50 FTS transfers instead of 200, resulting in a significant improvement in success rate. In the latter part of the DC24 exercise (CNAF_tape-> CNAF-disk), we did not achieve the target rate. Therefore, we are presently rerunning this test to eliminate monitoring issues (e.g., rates computed over periods with no transfers) and enhance site performance.

For Belle II, the target rates were comfortably achieved. The experiment reported transfer failures on February 13th due to concurrent GridFTP activity from the experiment involving the same endpoints.

General feedback

Exercises like DC24 prove invaluable in pinpointing bottlenecks within sites, and we strongly advocate for conducting such exercises along with preparatory tests and test repetitions. However, the DC constitutes a stress test that significantly impacts sites and strains storage endpoints. Therefore, we believe that the DC24 timeframe should be excluded from A/R computation (as is currently the case, as indicated in GGUS ticket 165509), and SAM tests should acknowledge ongoing DC activities.

Furthermore, we observed a substantial production load during the challenge, including staging activity from ATLAS. This activity occasionally had no impact but, in other cases, heavily burdened the infrastructure. To prevent overloading storage endpoints, the mechanism by which FTS regulates injection based on success rate would have been highly beneficial. Additionally, we noted an imbalance in requests between T0-T1 and T1-T1 for ATLAS, which adversely affected site performance and warrants improvement within FTS.

Lastly, the throughputs reported by FTS monitoring for our site are notably lower than our observed values. Unfortunately, our monitoring cannot differentiate between DC traffic and production load, but this does not appear to be the sole cause of the observed discrepancy (as illustrated in GGUS ticket 165526). Consequently, we propose using and comparing different metrics, such as transferred terabytes per day.

Recommendations

We plan to standardize the StoRM WebDAV instances dedicated to CMS, as we observed increased load and higher failure rates in servers with fewer CPU cores.

We will reconsider the hardware configuration for LHCb to better accommodate their workflow. This includes addressing the fact that the @INFN-T1 tape buffer and disk share the same filesystem and are managed by the same endpoints.

Recently, we upgraded StoRM WebDAV across all endpoints to enhance efficiency, resolving the issue described in GGUS ticket 165355. Additionally, we intend to implement performance markers in StoRM WebDAV for further improvement.

6.4 FNAL

The maximum injection rates for FNAL occurred on the final day of the challenge, reaching 26.15 GB/s for writes and 37.37 GB/s for reads. Conversely, the peak rates achieved were 31 GB/s for writes and 39 GB/s for reads, indicating the successful completion of the exercise for the T1.

Another noteworthy metric illustrated in Figure 6.4b is the sustained rates of 24.3 GB/s for writes and 36.6 GB/s for reads over 21 and 12 hours, respectively.

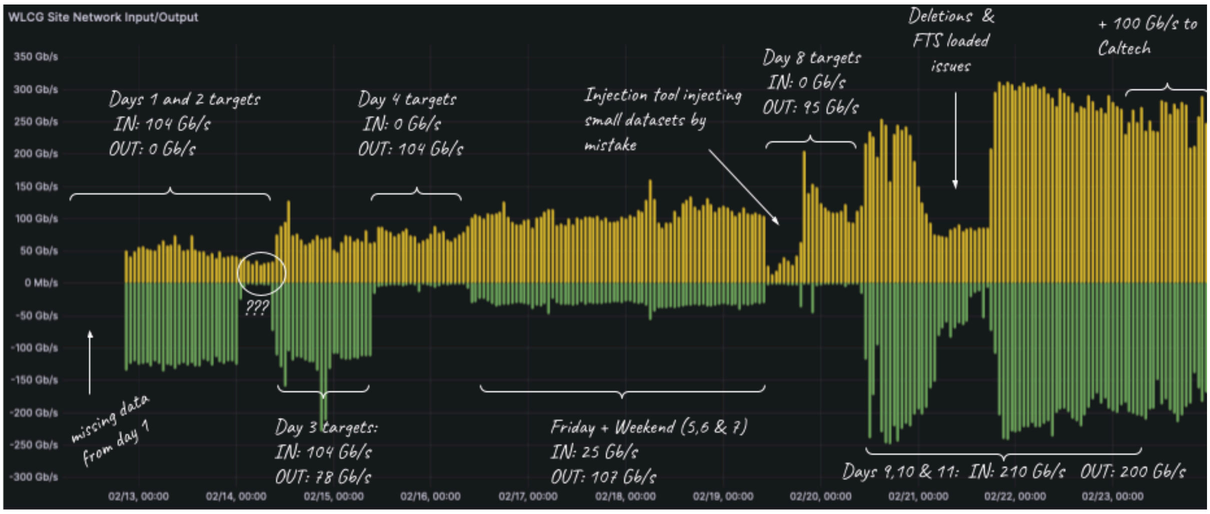


Figure 6.4a: Observed rates at FNAL

Figure 6.4b: Maximum sustained rates

6.5 IN2P3

The IN2P3 Tier1 participated in DC24 for five experiments: ALICE, ATLAS, CMS, LHCb, and Belle II. The LHCONE and LHCOPN links each had 200 Gb/s of bandwidth. ATLAS, CMS, and LHCb share a dCache infrastructure for storage, while Belle II uses another dCache infrastructure. ALICE storage, on the other hand, is managed through a dedicated XRootD infrastructure. The dCache services were configured to be token compatible.

No issues were identified on the network infrastructure, and there were no reported issues with the storage services. ALICE, CMS, and Belle II reported no issues with the IN2P3 site, and all targets were successfully achieved.

We summarise the following specific feedback:

For LHCb, the usage of the Tape REST API was successfully tested. During the LHCb staging process, a buffer overflow occurred. This issue was resolved by increasing its size.

For ATLAS, due to an overload of storage connections, the ATLAS hammer cloud mechanism initiated testing at the site, causing a halt in ATLAS production on-site. To resolve this, ATLAS had to decrease the number of connections on FTS, resulting in reduced throughput rates.

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

Table 6.5a: The flexible (T0 -> Tier1) and flexible (T0+T1 -> T1) phases highlight poor results for all Tier1s, with particular emphasis on IN2P3 (highlighted in orange), compared to other Tier1s.

Our assessment suggests that this difference is indeed linked to the reduction in throughput rates implemented to address the Hammer Cloud auto-exclusion.

6.6 KIT

DC24 did not reveal any storage issues at DE-KIT. The problem caused by the narrow dCache queue setting was resolved by adjusting it to unlimited. The observed issues in dCache and XRootD for ALICE were not directly caused by the DC24 workload.

DE-KIT is equipped with 2 x 100G LHCOPN links to CERN. Additionally, DFN provided a temporary 100Gbps link between DE-KIT and the DFN PoP in Frankfurt. This link was then transferred from Frankfurt to the Géant PoP in Geneva (CERN) by Géant. The capacity of this third 100Gbps link was load balanced across the two 400G uplinks between Géant and CERN. This extra 100Gbps capacity was integrated into LHCOPN and was only accessible during DC24. Immediately after the deployment of this additional capacity, 280Gbps were used, as demonstrated in the following graphs.

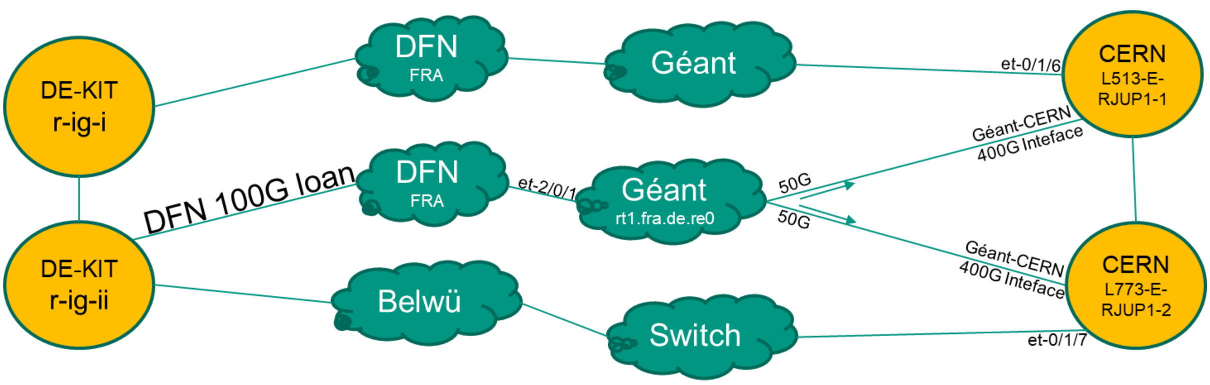


Figure 6.6a: Network schema

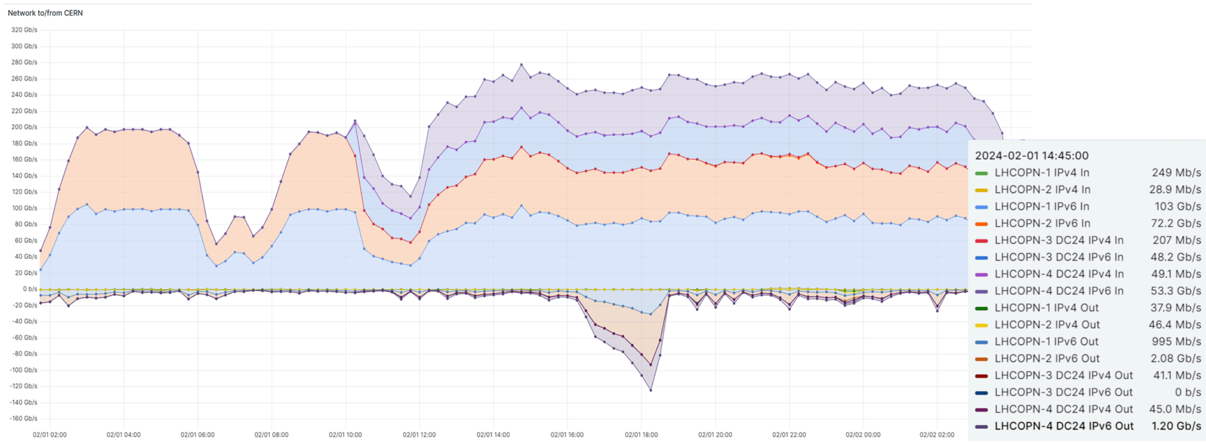


Figure 6.6b: Measured throughput at KIT by network channel

6.7 NDGF

An unfortunate combination of a minor dCache 9.2 bug and our haproxy configuration resulted in HTTPS 3rd party copies failing under load, affecting both production and DC24 operations. Consequently, we were occupied with debugging this issue and were unable to devote much attention to the ATLAS portion of DC24. Despite this, the accumulated FTS retries did uncover some network bottlenecks in the long-haul network between CERN and NDGF, although these were operating at or above the full HL-LHC data rate (20-30Gbyte/s).

During the same period, the ALICE tape data migration proceeded without issue.

A subsequent Data Challenge test conducted a couple of weeks later proceeded smoothly, and the increased load did not pose any problems for networking or storage at NDGF.

6.8 PIC

PIC contributed to DC24 for four experiments: ATLAS, CMS, LHCb, and DUNE and participated in the challenge with both a Tier-1 and Tier-2 for ATLAS. All experiments used the PIC dCache infrastructure for storage, with dedicated servers allocated to each VO to ensure secure data housing and prevent conflicts from concurrent VO activities. Prior to the challenge, both the dCache service and HTCondor-CEs were pre-configured to ensure token compatibility. LHCb's habilitation request was accommodated shortly before DC24 (GGUS #165055).

PIC successfully completed the DC24 tests, consistently meeting and exceeding combined transfer rate targets without encountering bottlenecks. However, during the second week of the challenge, the site experienced lower-than-expected pressure for some LHC experiments. It is important to note that these scale issues were not directly attributed to the site itself, but rather to poor performance at scale by FTS or Rucio, or issues with token usage. Some tests were repeated after DC24, and PIC demonstrated its capability to meet the targets of the flexible model.

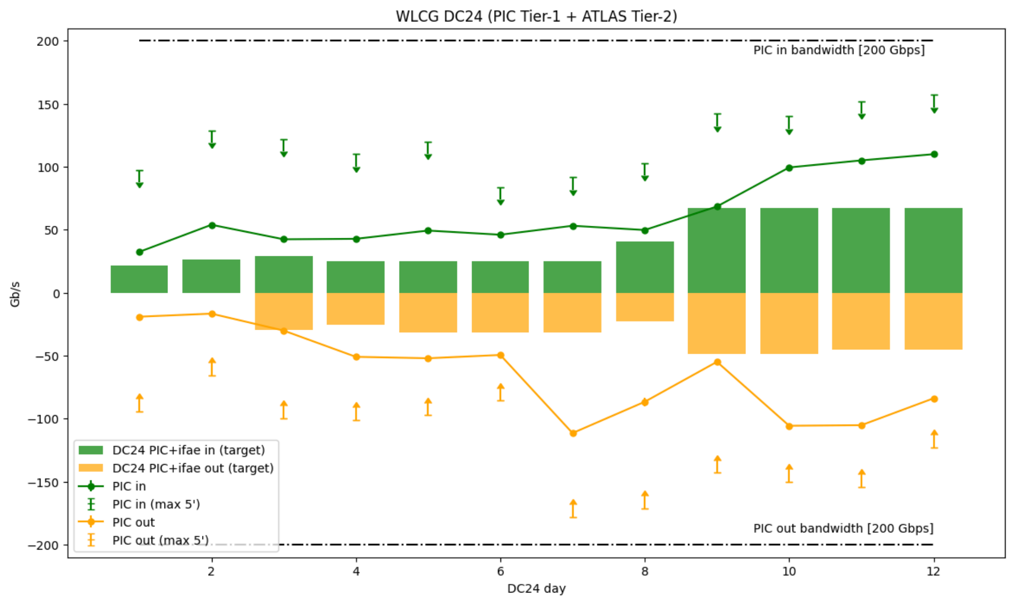


Figure 6.8a: Daily incoming and outgoing targets for both Tier-1 and Tier-2

Throughout the DC24 test days, PIC maintained an average incoming (outgoing) traffic of 63 (64) Gbps, with peak values reaching 157 (178) Gbps, respectively, within specific five-minute intervals. Figure 6.8a illustrates the daily incoming and outgoing targets for both Tier-1 and Tier-2 at PIC (excluding DUNE targets), alongside the average and maximum daily traffic values recorded for both incoming and outgoing traffic at PIC. The traffic displayed represents the total PIC traffic as measured in the WAN network infrastructure. The values obtained during DC24 were typically five times higher than what we experienced in the last months of 2023 and the beginning of 2024 at PIC, prior to the challenge.

No issues were detected within the network infrastructure, and storage services operated without any identified major issues. However, specific feedback from the site includes:

Observation of sporadic error messages during the copying of LHCb files from local tape storage to local disk storage due to dCache's handling of requests (GGUS #165228). The issue arises when a file is requested twice, causing the second request to block the first. Once the initial request completes its first stage, the subsequent part of the job attempts to copy the restored file to a disk location. However, dCache detects another pending stage request and lacks direction on where to redirect it, as one disk pool already has a recall pending. Although this behavior has been reported to dCache, these errors disappeared when the first requests were completed and the targets were fulfilled.

Occurrence of several micro-cuts on both LHCOPN and LHCONE circuits by our NREN provider did not affect PIC performance. However, on the night from the 20th to the 21st of February, a significant maintenance cut occurred on the LHCOPN. Traffic was diverted to LHCONE.

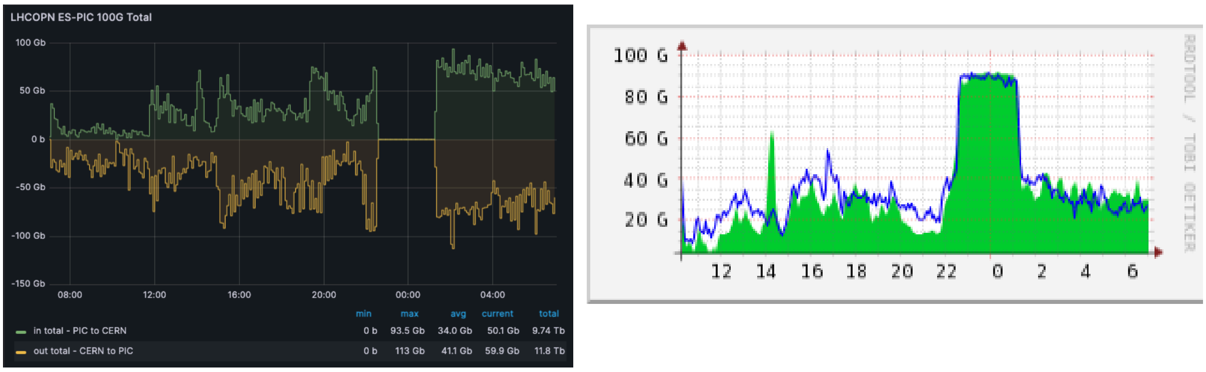
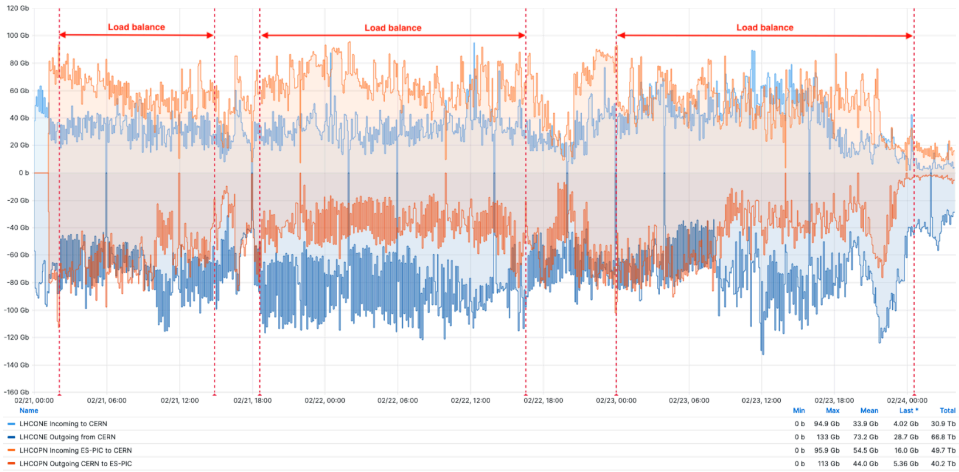


Figure 6.8b: Two plots showing successful activation of NOTED at PIC, operating as a DEMO site, to implement load balancing between LHCOPN and LHCONE from February 21st to 23rd, 2024.



- ES-PIC load balancing between LHCOPN and LHCONE (from 21st to 23rd of February 2024)

Figure 6.8c: DUNE data streaming rates from PIC to European sites

DC24 traffic helped uncovering that the LHCOPN and LHCONE shared the same physical path. The NREN was contacted and separate paths were established after DC24.

6.9 RAL

Overall, DC24 was very successful from RAL's perspective. Various issues were debugged, fixed, and optimised in real time, with further improvements noted. By the end of the challenge, the throughput was 2–3 times higher than we have ever achieved before. We identified a new set of bottlenecks well above Run 3 normal load, which can be worked on over the coming few years.

In the run-up to DC24, transfer tests were conducted against RAL, identifying an asymmetric routing issue over the LHCOPN. The problem was identified and scheduled to be fixed shortly before the Data Challenge; however, the LHCOPN link was cut under the sea between the UK and the Netherlands at the end of January. It wasn't fixed until Friday, 16th February, which was at the end of the first week of the Data Challenge. After the fix, it took a further few days to address and tune other issues that were discovered. During the first week, the link failed over to using the LHCONE; however, this was a relatively untested mechanism and had never been tested under high load, and required significant effort to tune correctly. On February 14th JISC notice that the link (300Gb/s) between them and GEANT was saturated. The throughput in the first week was therefore low.

As the second week of the Data Challenge progressed, the throughput steadily increased as we implemented various tunings. We started to reach the physical limits of the underlying hardware and put extensive effort into balancing the load across the gateway servers. In the final three days of the challenge RAL consistently exceed the minimal targets for both ingress and egress. While we didn't reach the flexible scenario this was expected as only Echo was being tested.

At RAL, we have entirely separate Disk (Echo) and Tape (Antares) services. When Antares was initially designed in 2019, it was assumed that the VOs would use Echo as the buffer and only access Antares via multi-hop FTS transfers. In practice, VOs would rather interact directly with Antares, and we have therefore redesigned it to be publicly accessible. We expect VOs will export raw data to it directly during the remainder of Run 3 and beyond. However, during DC24, Antares was not included in the data export tests, leaving a significant amount of available throughput untested.

Antares was tested by LHCb in their data staging test. The write rates (Echo -> Antares) were lower than the target (2.64 GiB/s out of 3.96 GiB/s), while the read rates exceeded the target (3.28 GiB/s out of 2.40 GiB/s). After the Data Challenge, ATLAS repeated their data export test, and we asked them to test Antares. This was the first major stress test of Antares since entering production, and were able to achieve a throughput of 3.4GiB/s (72% of the target) to Tape. This was a very informative test and there are plenty of areas we can improve.

The largest, unexpected issue identified during DC24 involved the deletion rate. As the load increased deletes started to take much longer to process (>20 seconds). More space was provided to the experiments to ensure that it didn't impact the challenge and once the load dropped the deletes quickly caught up. We have not fully understood the underlying cause but there is clearly some race condition at high load.

Other issues that were identified include the following:

Load Balancing: We gained significant experience in attempting to balance the load across the gateways.

Gateway Hardware: We encountered the 25Gb/s network limit on our gateways; consequently, we have procured a 100Gb/s capable gateway.

Tokens: Tokens had been enabled before the Data Challenge on our storage endpoints; however, for various reasons, the VOs did not use Tokens against RAL during the challenge. Most of the improvements will come from the VOs, but some XRootD patches are needed.

As a multi-VO site, we are aware of the contention between each VO's FTS instances, and further tuning may be necessary.

6.10 SARA

For LHCb, we encountered some issues with writing to tape. We identified two issues: the tape write pool group for LHCb was slightly too small, and there is a bug in dCache that sometimes prevents pools from writing to tape. (Reference: https://github.com/dCache/dcache/issues/7511). Internal transfers (SARA to SARA) for LHCb encountered a limit: transfermanagers.limits.internal-transfers, which defaults to 150. We increased it to 1500. We had previously increased transfermanagers.limits.external-transfers, which serves a similar purpose but for external TPC.

For ATLAS, we observed impressive throughput, reaching 21 GB/s for 5 consecutive days. It was limited by a SARA internal uplink of 200 Gbit/s. Transfers from SARA to RAL for LHCb were hindered because this uplink was saturated (https://ggus.eu/index.php?mode=ticket_info&ticket_id=165359). After the "normal" DC24, we conducted a network test in collaboration with ATLAS and Nikhef, utilizing a temporary 1200 Gbit/s connection with Nokia equipment. With this connection, we achieved transfers from the ATLAS EOS to the SARA dCache of up to 45 GB/s, while simultaneously conducting normal production activities. For an hour, we reached 530 Gbit/s, with a peak of 661 Gbit/s. Reports of this experiment will be forthcoming.

6.11 TRIUMF

Overall, DC24 was successful, serving as a valuable exercise to identify issues at sites. No network bottleneck was found, and no major upgrades are needed for disk. Tape operations meet ATLAS computing model requirements, but not at such high loads. Token authentication worked well.

No network bottleneck was found. Thanks to CANARIE, BCNET, and SFU data centre for making the new 400G link available.

Overall, disk performance during DC24 was very good. However, a large number of deletions led to temporary high load on the namespace. Dark data cleanup backlogs caused load balancing issues among pools (due to different cleanup paces) and some transfer errors. (Reference: https://ggus.eu/index.php?mode=ticket_info&ticket_id=165343) No manual intervention was required. It is worth noting that the large number of HSM staging requests/pins also significantly contributed to the namespace high load. Token authentication forced FTS to use only external WebDAV doors, resulting in fewer available movers for increased concurrent transfer tests. Some SAM tests timed out. No upgrade is needed for disk.

The Data Carousel was an unexpected aspect, not part of DC24, but worth mentioning. Together with the staging requests from PanDA jobs, 2.2PB of data were staged during DC24, exceeding the tape total buffer size of 1PB and far surpassing the Staging CAP setting. It needs to be discussed if future PanDA job staging requests will go through Rucio/FTS. The HSM setup meets ATLAS computing model requirements, but not for this high load. There was high load on most of the hsmpools, resulting in many failed transfers. Each failed transfer issued a pin for 7 days, which was only released after the transfer successfully finished. A large number of 7-day pins (65k) were left in the PinManager, occupying tape buffer spaces, which resulted in less buffer for staging. Repeat pins caused repeat reads, with about 1/3 of the files being repeat tape reads. We had to lower the number of movers for transfers to reduce the chance of overload and timed out some transfers. (Reference GGUS ticket: https://ggus.eu/index.php?mode=ticket_info&ticket_id=165364)

As followup, we suggest that FTS stages files to local DATADISK, not to remote RSE, except for short latency sites like BNL. Also, the TRIUMF Tape system now checks file existence before passing the staging request to tape. The short-term plan is to increase I/O capacity and use old decommissioned storage units. A hardware refresh is planned for next year, with hopes of securing a better deal within the limited budget.

6.12 AGLT2

Overall, our site was not stressed during DC24. Pre-challenge testing had a much larger impact (see plots below).

Networking was not a bottleneck for DC24, but pre-DC24 testing reminded us about a 2x40G "bottleneck" at the University of Michigan site, which will be removed once new campus switches/routers are put into production by the end of the year. Network capacity was not a problem for our site during this test, and the traffic was overwhelmingly IPv6. As a reminder, ESnet plots reverse the incoming/outgoing labels from our site's point of view.

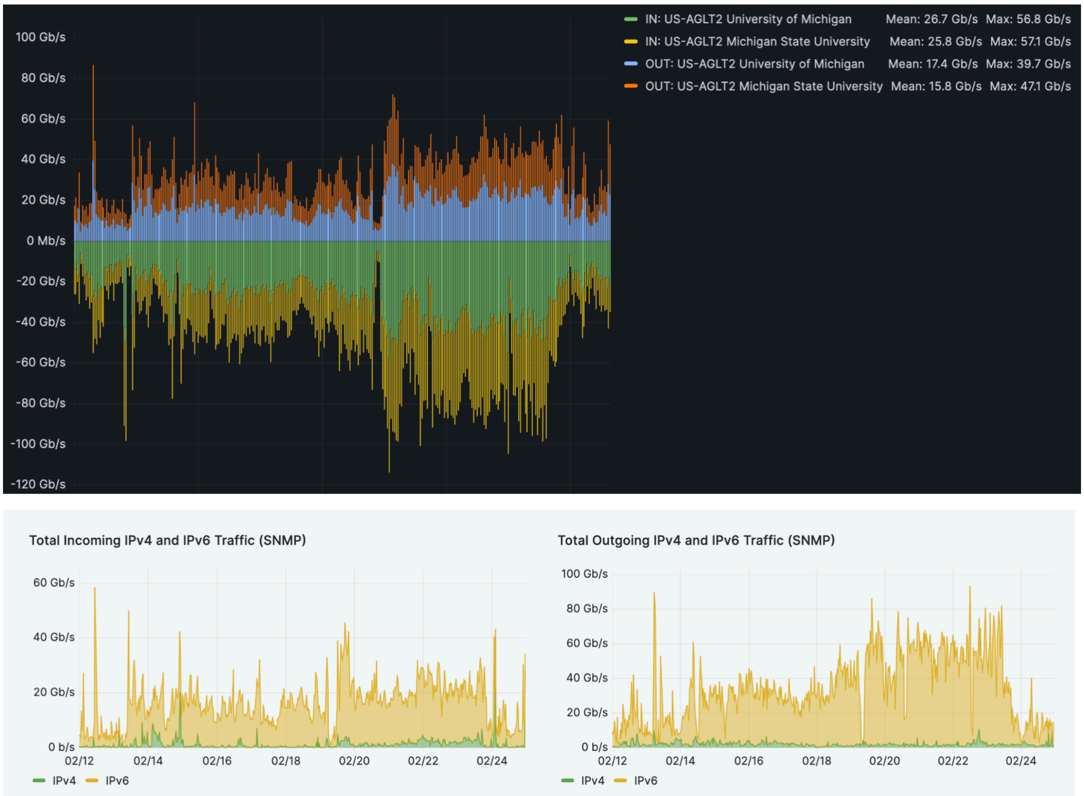


Figure 6.12a: Three plots showing the ESnet rates. Top plot colour coded by institute, bottom plots colour coded by IP version

A 2x40G bottleneck currently exists between AGLT2-UM and UM campus, which was occasionally made visible during Hiro's Jan 2024 higher rate pretest. This bottleneck will be removed before 2026/29, most likely before the end of 2024.

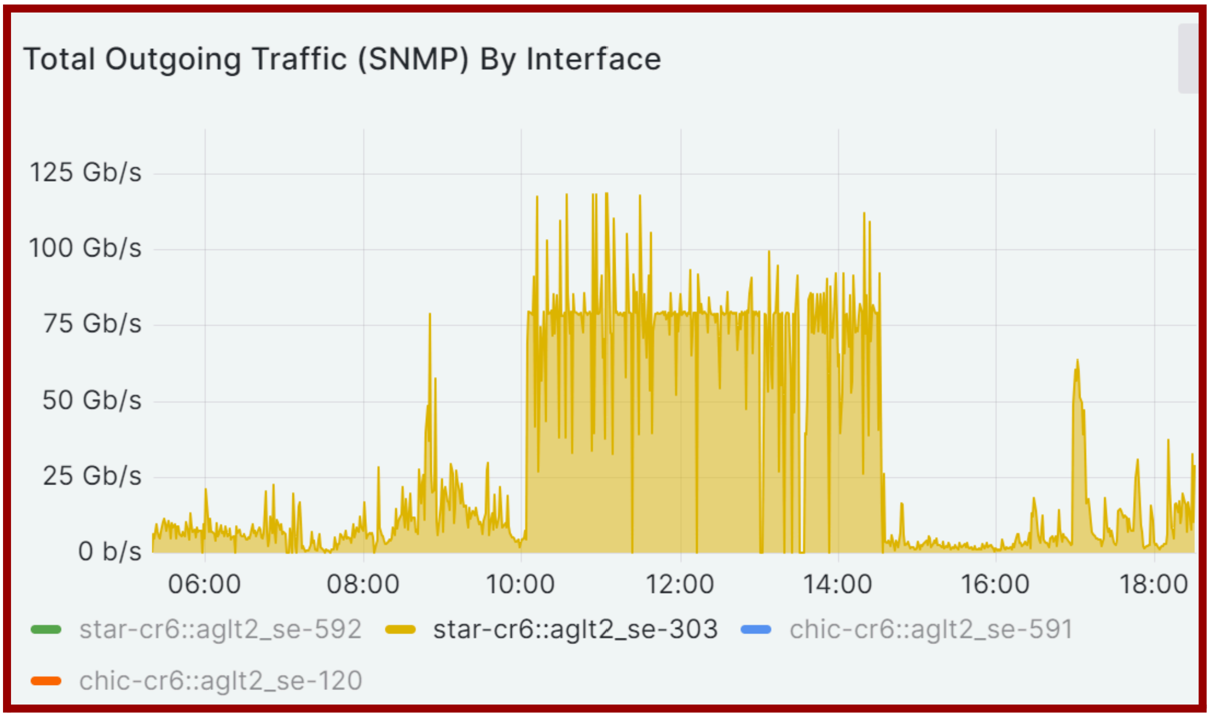


Figure 6.12b: Total outgoing traffic by interface

The dCache pool servers msufs03 and msufs04, equipped with 2x60 8T disks on MD3460+MD3060e, were overloaded, with occasional loads exceeding 500, while msufs02, with 60x 4T and 60x 8T disks, was not significantly overloaded, with load peaks around 30. We were already planning on downsizing all such systems to just one MD3460 shelf and only 60x 8T disks per server as part of our 2024 storage purchase plan. This test, highlighting the difference between msufs02 and msufs03/04, suggests that our current plan will provide significant headroom for all three of these servers. Additionally, all 8T disk servers are expected to be retired by 2026/29.

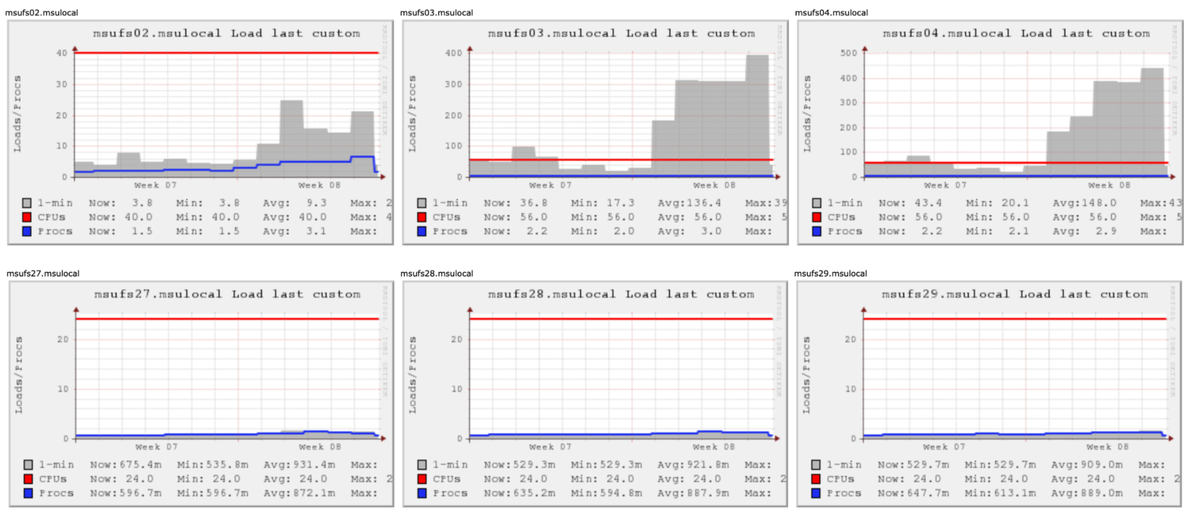


Figure 6.12c: All our newer dCache pool systems (R740xd2) were not stressed. dCache doors, zoo nodes, and head nodes were not overloaded.

Once dCache supports firefly flow labelling, we plan to enable it at AGLT2. With our transition to EL9, we will also be able to enable flowd packet marking, given the use of a 5.14+ kernel.

AGLT2 also plans to participate in network utilisation optimisation via 'tc' (packet pacing), jumbo frames, and new protocol testing (BBRv3 and others).

As next steps, we suggest the following. We were already aware of the current performance limitations of dCache Pool servers msufs02 and msufs03. This is addressed with our 2024 storage purchase plan, currently awaiting funding. The network bottleneck between AGLT2-UM and UM campus will be addressed with new campus routers currently planned for summer 2024.

6.13 MWT2

Overall, the site was not stressed except for 6 older Dell MD3460 high-density storage servers, which had trouble keeping up with the additional load.

In the last few hours of the Data Challenge, the free disk space in our storage pools dropped far lower than usual: we reached below 200 TB of free disk space out of our 21500 TB, approximately 1% free disk space. Interactions with Dimitrios Christidis (Rucio/DDM) and Petr Vokac (ATLAS DC24 coordinator) around 11 pm CERN time on Friday, February 23rd, eventually assured us that the deletion procedure would be fast enough to prevent completely filling the MWT2 storage elements. However, since the site was in full production, there were transfers from the production system proceeding independently, so we were still concerned because we had never previously allowed the free space to go so low. Given the lateness of the hour on Friday night at CERN, Judith Stephens (lead MWT2 sysadmin) and I (Fred Luehring / Physicist) were quite worried about having a weekend crisis.

The WAN network held up fine, and the network bandwidth never approached the network capacity at the site holding our storage: the University of Chicago. At the time, the University of Chicago site had 200G external connectivity, and as can be seen on the plot showing the entire month of February, bandwidth consumption reached a maximum of about 150 Gb/s.

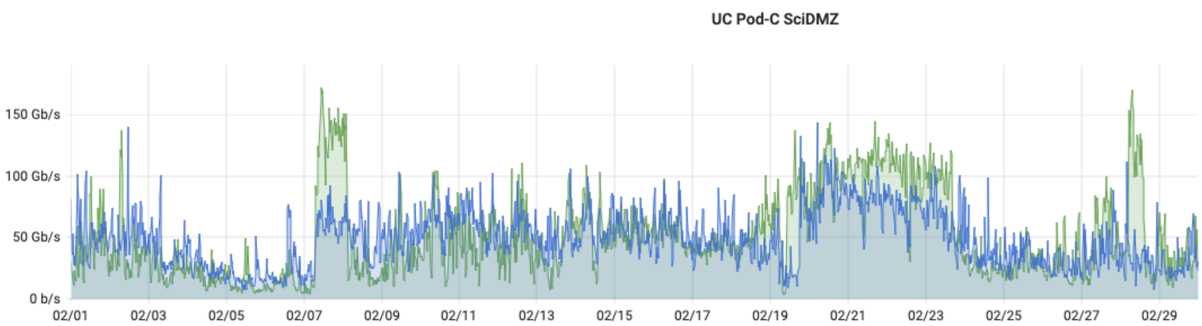


Figure 6.13a: The period of peak bandwidth consumption starts before the nominal beginning of the data challenge (13 Feb) and continues right to the end of the challenge on 23 Feb.

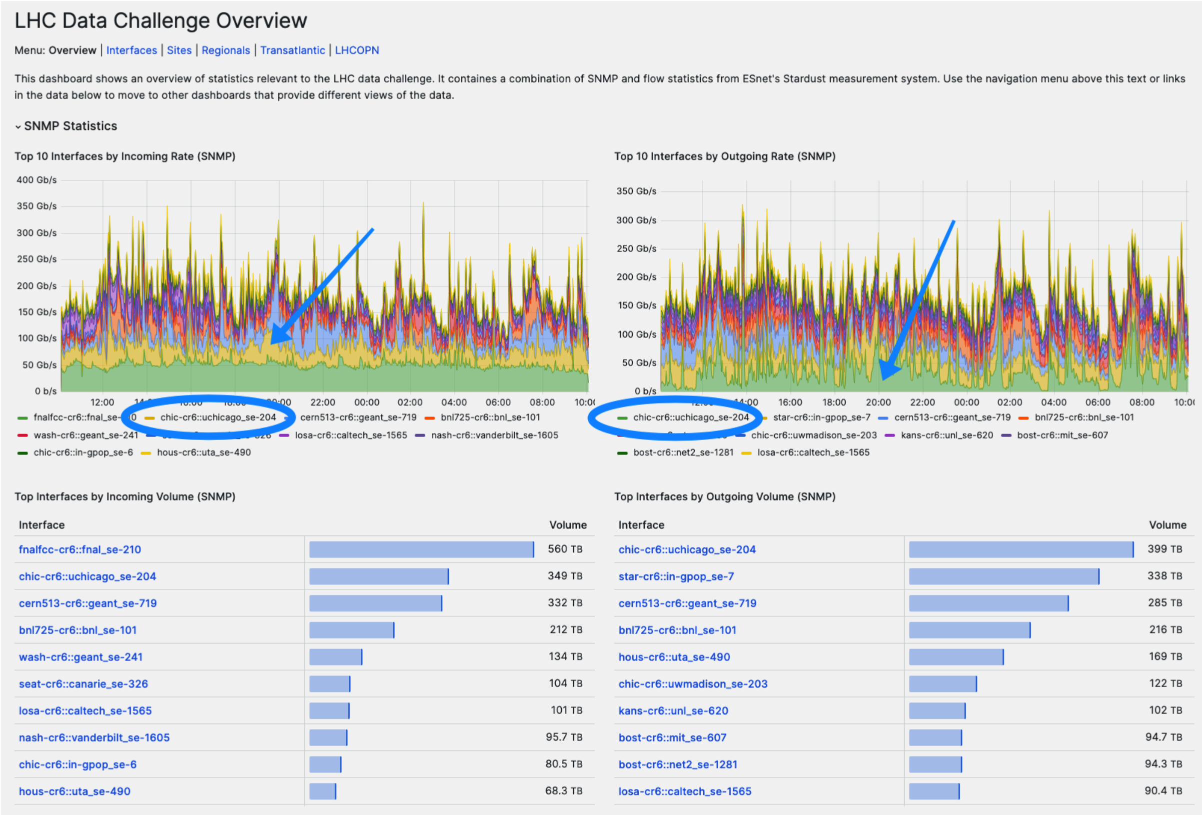


Figure 6.13b: MWT2 performed well compared to other LHC sites.

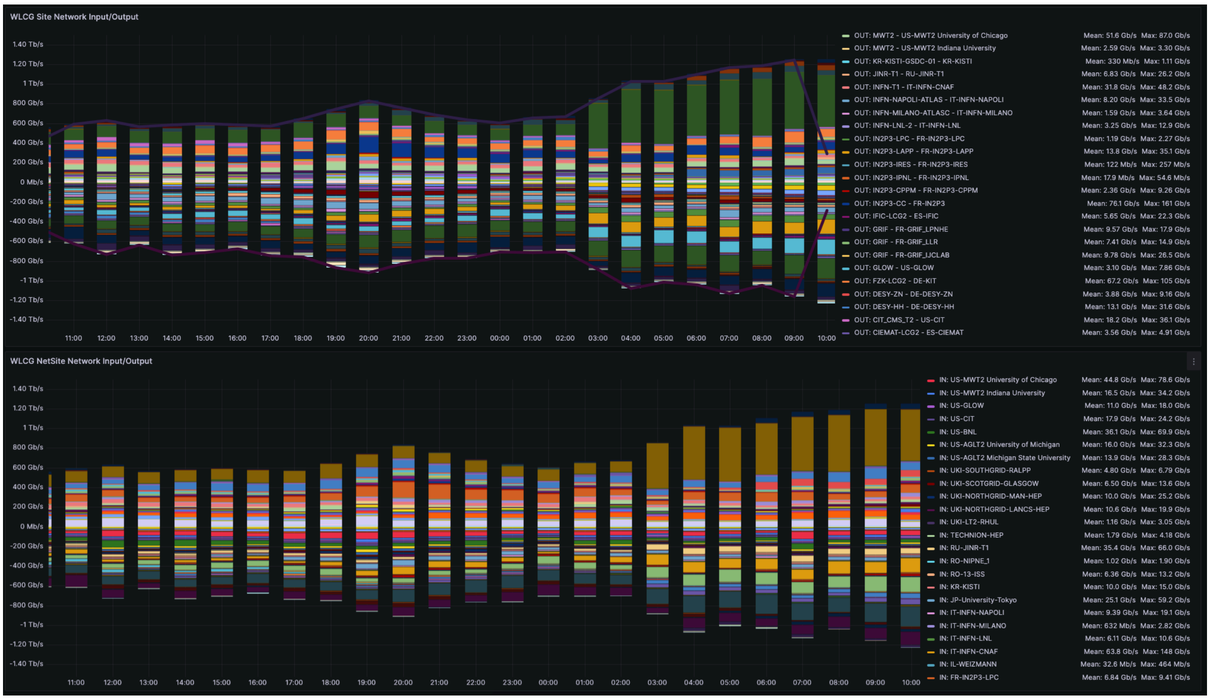
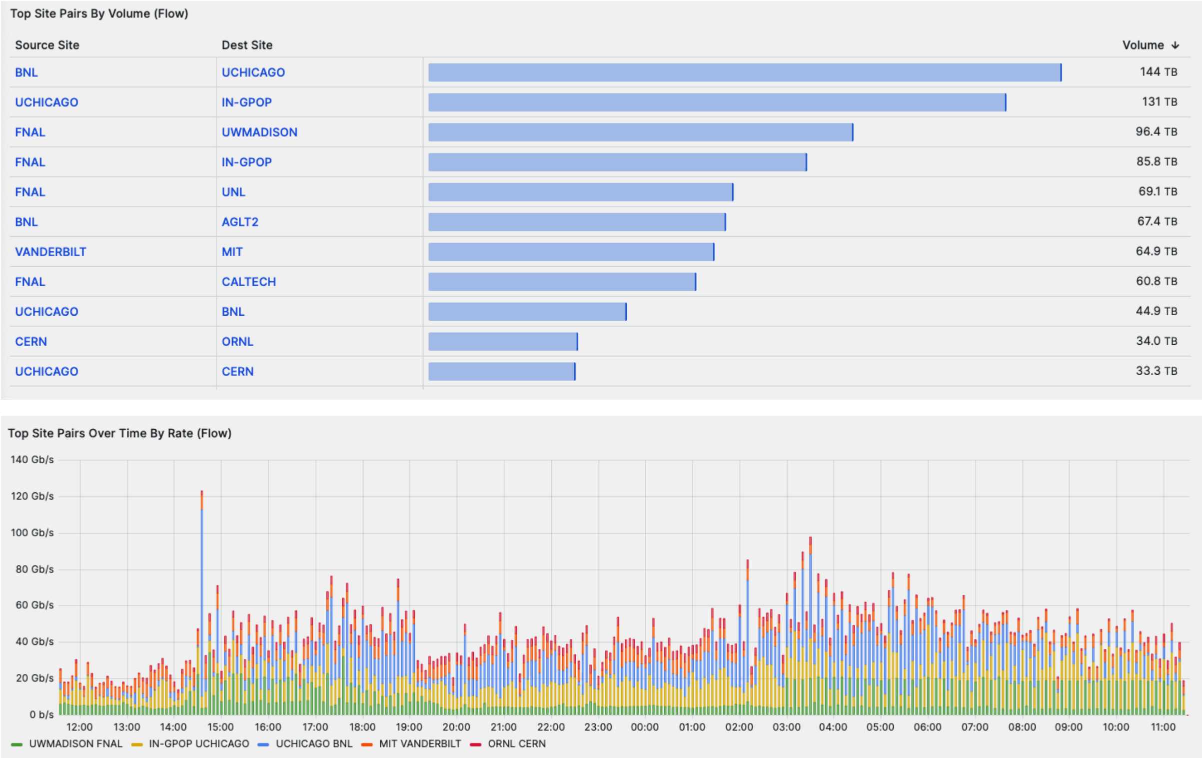


Figure 6.13c: The two plots indicate transfers dominated by DC24 (BNL to UCHICAGO) while also maintaining normal job-related production between Chicago and Indianapolis, MWT2_IU (UCHICAGO to IU-GOP).


Top Site Pairs By Volume (Flow)

Source Site	Dest Site	Volume
BNL	UCHICAGO	144 TB
UCHICAGO	IN-GPOP	131 TB
FNAL	UWMADISON	96.4 TB
FNAL	IN-GPOP	85.8 TB
FNAL	UNL	69.1 TB
BNL	AGLT2	67.4 TB
VANDERBILT	MIT	64.9 TB
FNAL	CALTECH	60.8 TB
UCHICAGO	BNL	44.9 TB
CERN	ORNL	34.0 TB
UCHICAGO	CERN	33.3 TB

Figure 6.13e: These two plots show that two of MWT2's links were among the top pairs during some time intervals

We discovered missing network reporting from one of the three-site MWT2 federation, the University of Illinois. This has since been fixed. The Inbound transfers were reasonably successful even under heavy load.

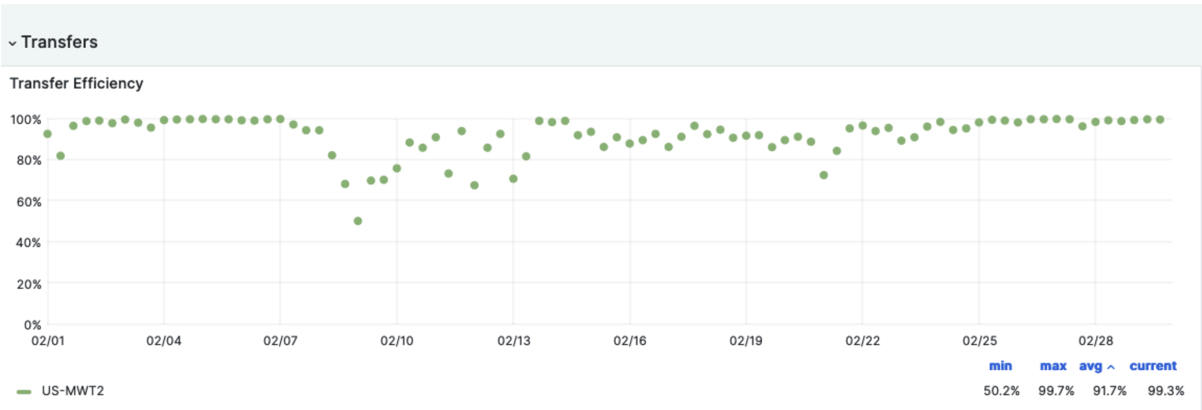


Figure 6.13f: Transfer efficiency plot of the MWT2 federation

The higher rate of inbound transfer failures during the beginning of the Data Challenge are caused by the storage issues which were partially mitigated by reverting a dCache tuning that was applied to all storage on the advice of OSG / dCache experts. See the storage section below.

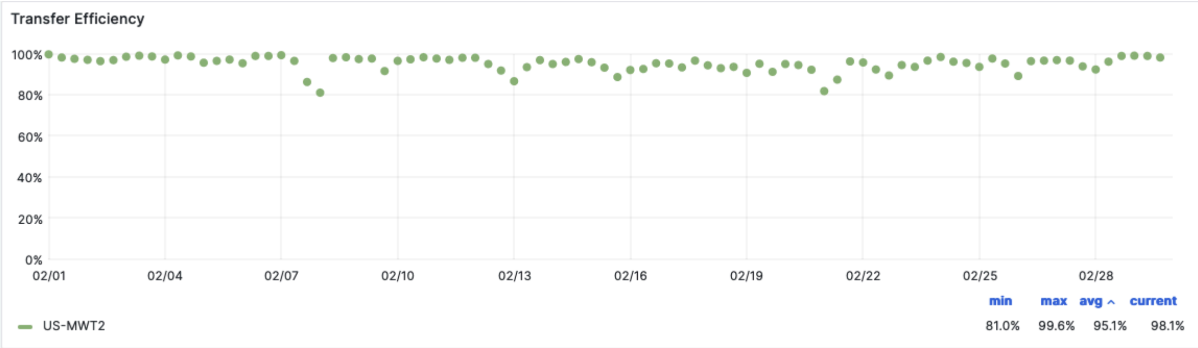


Figure 6.13g: Outbound transfer efficiency

Both of the above plots do include internal transfers within the MWT2.

A number of I/O issues on our older high density MD3460 pools were observed. Interference with normal dataset deletions and transfer errors caused by older dCache pools were observed. We plan to retire these in our upcoming storage purchase. In the meantime, we have reduced the number of pool movers on these pools. After the MD3640 pools struggled at the beginning of the challenge, we reverted the number of dCache movers from 2000 to the default value of 100 on the afternoon of Wednesday, February 14. The value of 2000 was recommended to us by experts and did improve performance on the newer storage servers. However, the MD3460 servers could not handle having that many simultaneous transfers. This plot shows the wall-time weighted job success efficiency for production jobs only during February.

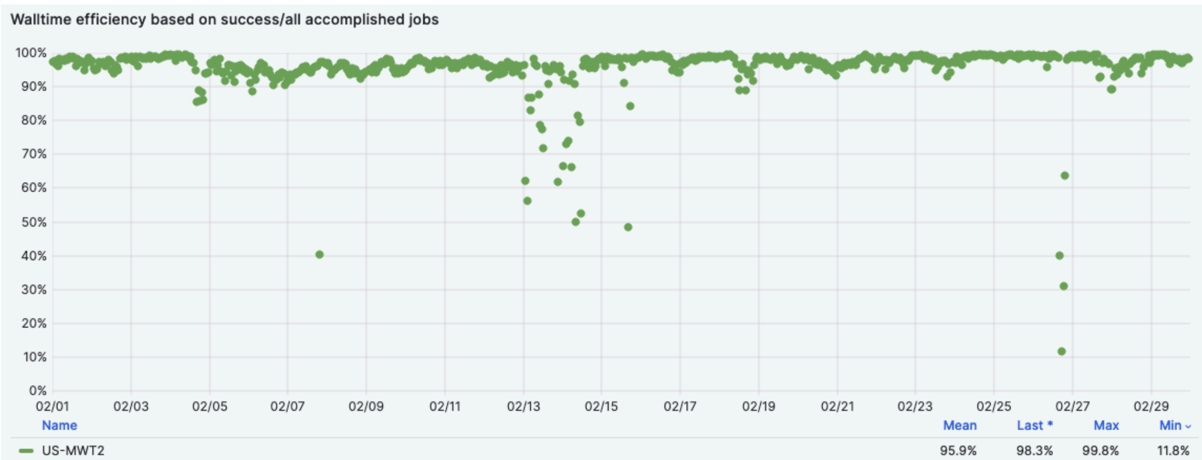


Figure 6.14h: dCache tuning was reverted because the job success rate increased significantly during the day of February 14.

We observed network lockups on our s-nodes with 8 MD1200 pools. We updated these to AlmaLinux 9 after the Data Challenge and have not seen this issue since the upgrade. All of the storage servers purchased after the 3 servers bought with MD1200 shelves have been 2U Dell servers with 24 spinning disks. The most recent storage servers have been the Dell PowerEdge R740xd2 model with 20 TB disks. This configuration provides a raw capacity of 480 TB. Formatted as RAID6, this yields approximately 400 TB of space, which will not suffer data loss even if 2 disks fail. These servers have

2 x 25G uplinks and performed well during the Data Challenge. The Grafana monit-atlas plots show that file deletion efficiency was good, in disagreement with what was observed locally.

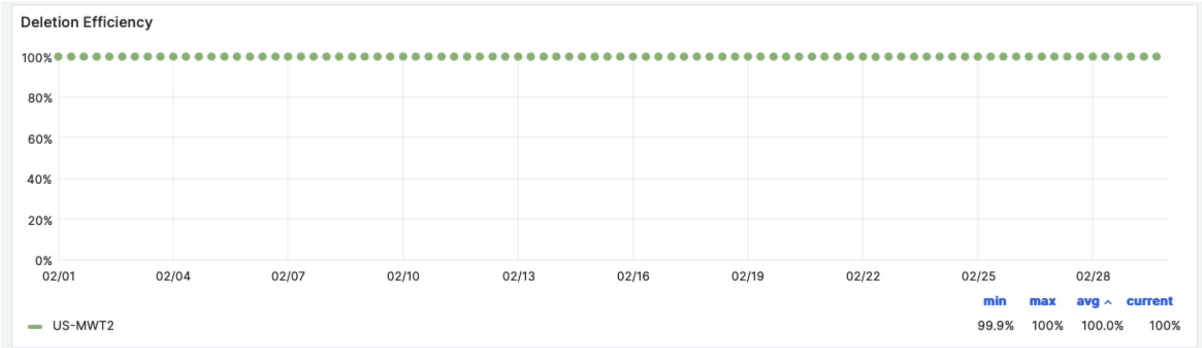


Figure 6.14i: File deletion efficiency

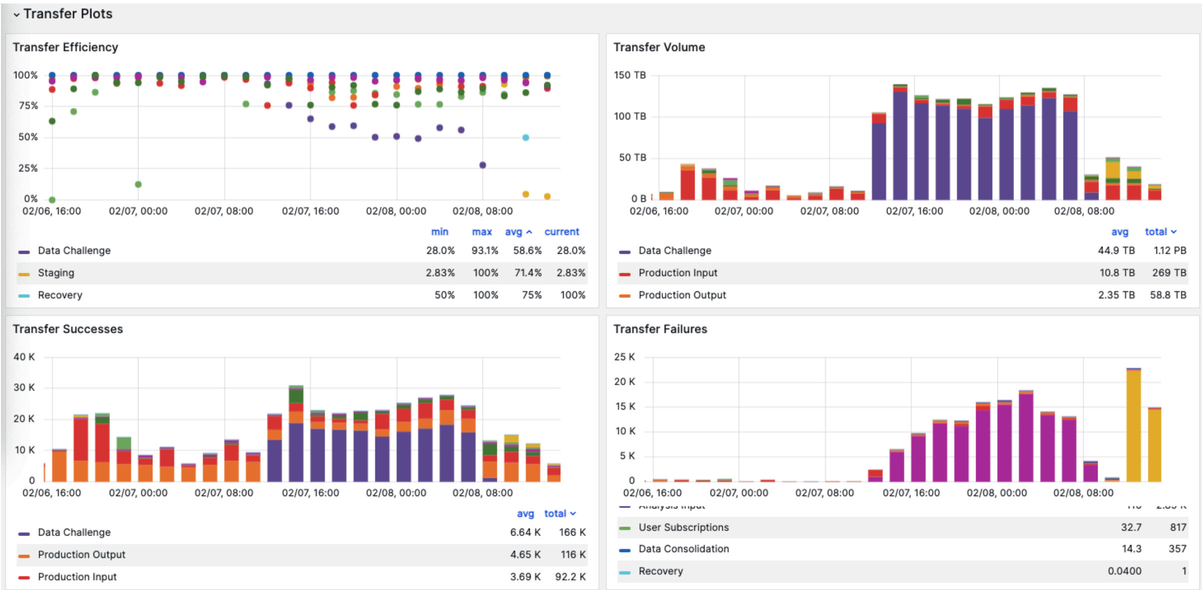


Figure 6.14j: Indicating transfer failures during DC24

As feedback for the future, it would be good to designate on-call people during the next data challenge. We were fortunate that both Dimitrios and Petr were still awake when it appeared that the site was about to run out of space. MWT2 is also planning to replace the older high-density Dell storage MD3460 as soon as possible. The 3 servers with Dell MD1200 shelves are next in line for replacement.

6.14 NET2

During the DC24 tests, the NET2-LHCONE connection speed was 10Gb/s. DC24 saturated the NET2 connection with ESNet during the entire testing period. The majority of the traffic was done over IPv6. Inbound production traffic had already saturated our connection a little over a month before the challenge.

Since the site was already operating with a saturated connection, some of the DC24 transfers had low efficiency since they were competing with production transfers. The efficiency was worse when the DC24 coordination team increased dramatically the allowed number of transfers per site. The low efficiency persisted until a couple weeks after the ending of the challenge. The conclusion is that the maximum number of concurrent transfers should take into account the maximum bandwidth available.

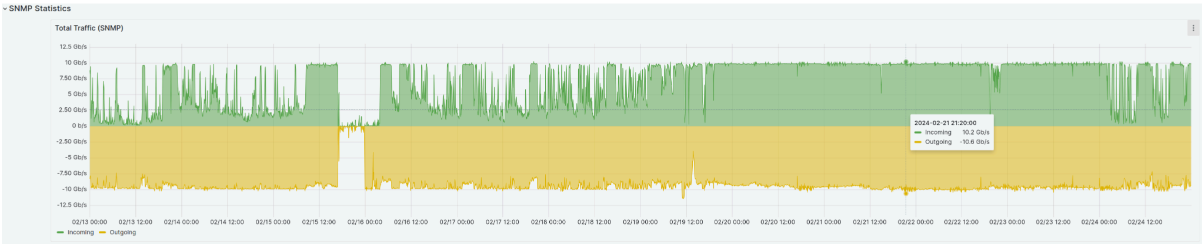


Figure 6.14a: Total throughput as monitored by ESNet, with site inbound and outbound transfers are showed reversed in the ESNet [plot]

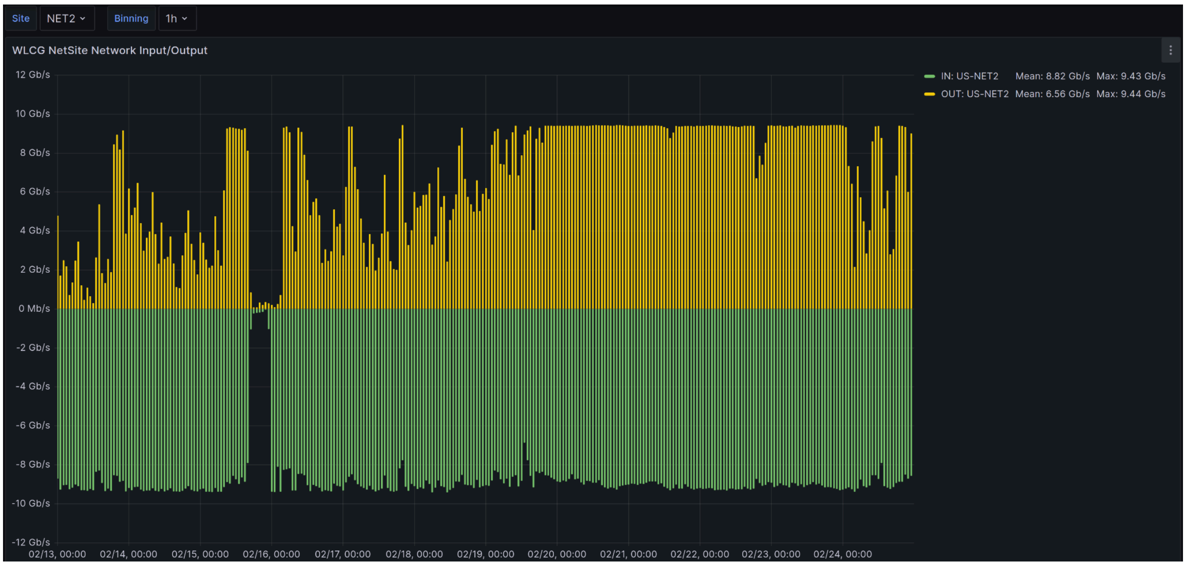


Figure 6.14b: Site network monitoring [plots]

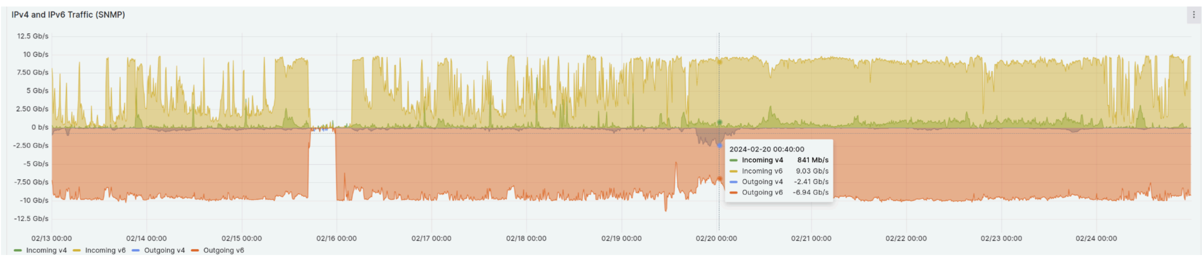


Figure 6.14c: Total throughput plotting IPv4 and IPv6 [separately]

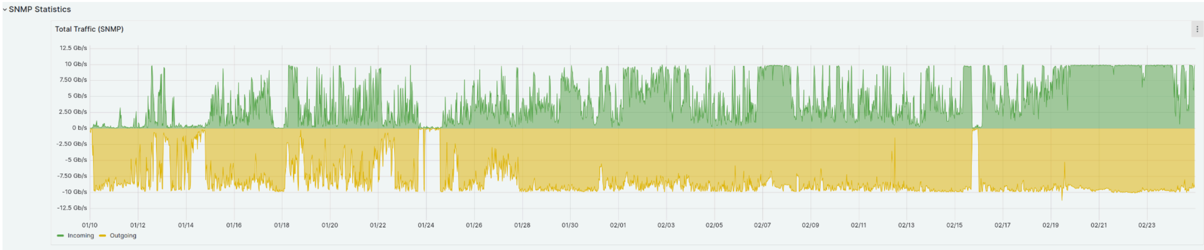


Figure 6.14d: NET2 throughput including the month before DC24 showing how normal production already saturates the inbound connection

We observed the following problems and production interactions:

An urgent and unscheduled network maintenance was required between February 15th and 16th. The downtime was not related to DC24 activities.

https://atlas-cric.cern.ch/core/downtime/list/?id=5865
https://atlas-cric.cern.ch/core/downtime/list/?id=5875

Some transfers failed with a timeout due the large amount of concurrent requests competing for the 10Gb/s connection. The storage system responded well throughout the test period.

As Follow up we suggest the following. The NET2 network worked as expected during DC24. The site is upgrading the connection to LHCONE during the summer of 2024 to multi-200 Gb/s connections.

6.15 SWT2

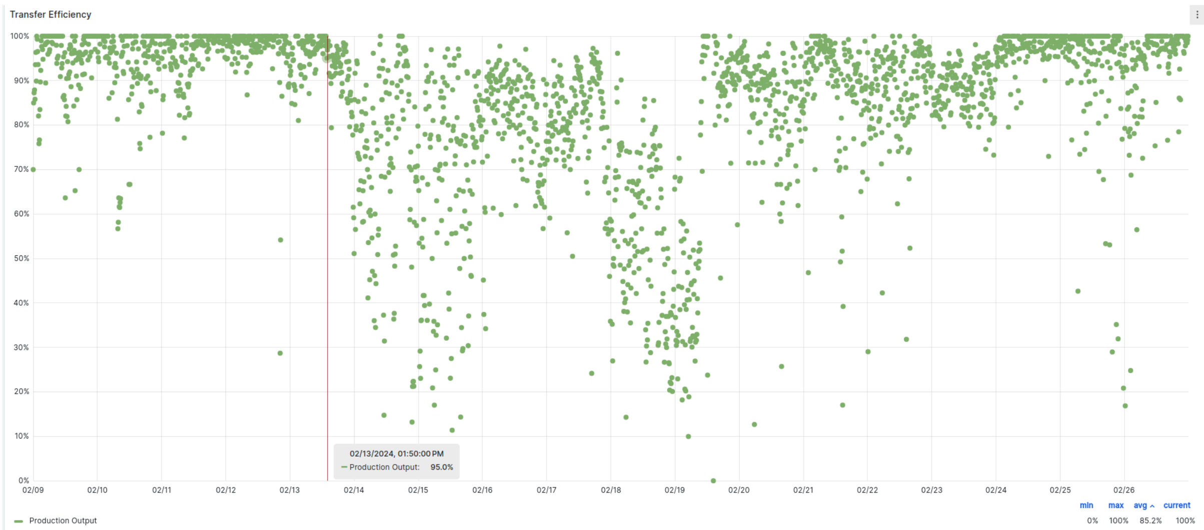


Figure 6.15a: FTS could not keep up with the flow of data.

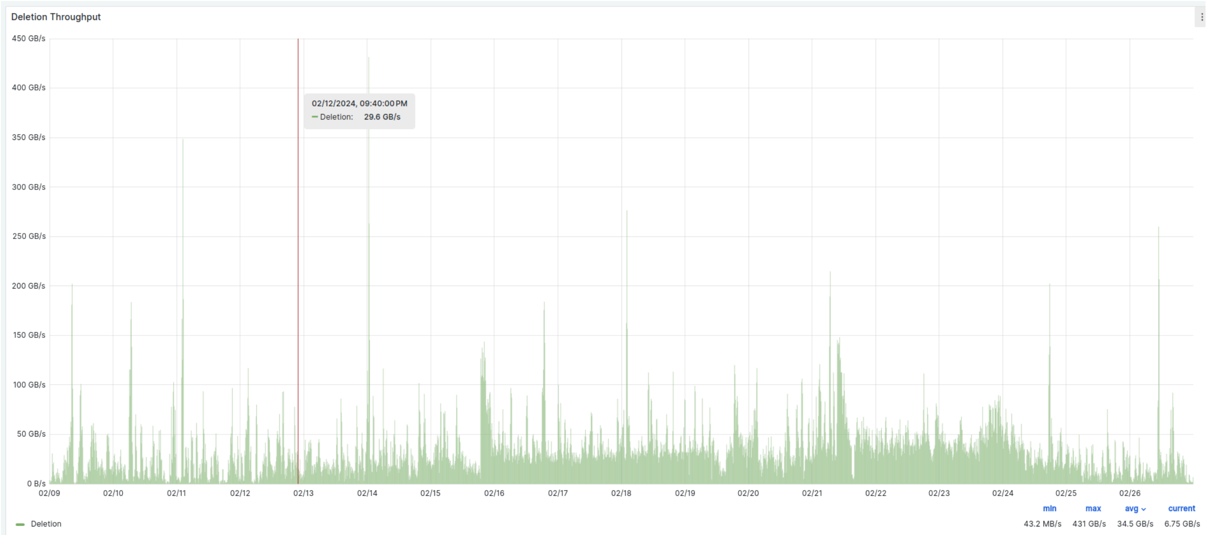


Figure 6.15b: The deletion rate was too slow for the Data Challenge

We never reached the network throughput limit because of FTS and deletion limitations. However, we encountered an unrelated network problem due to an upgrade we performed at our site the week before DC. This was clearly poor timing on our part, and it serves as a lesson learned.

At OU, transfers were capped at about 20 Gbps. This limitation occurred because the dual-25G NIC uses a team driver rather than bond, which apparently doesn't aggregate across NICs very well.

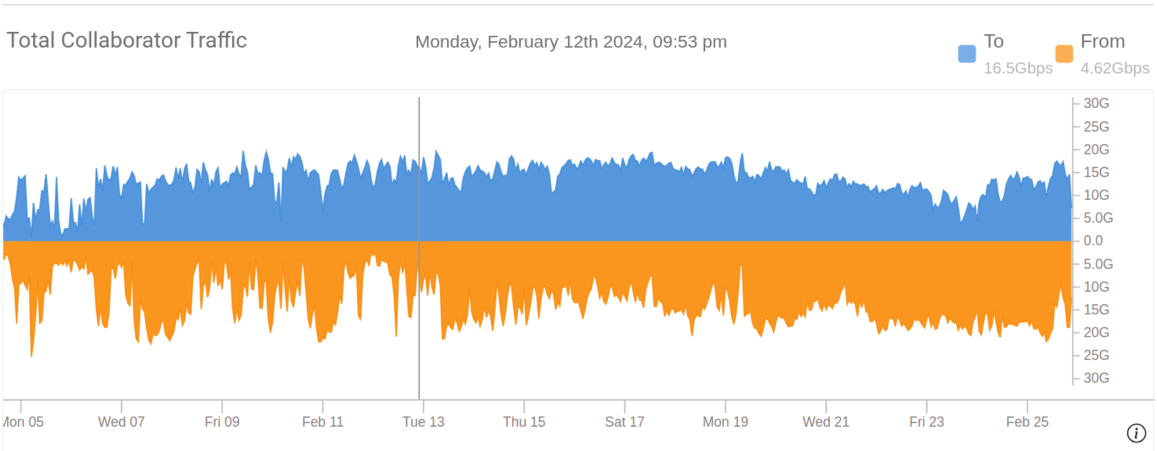


Figure 6.15c: Total collaborator traffic

At OU, 7 old T630 XRootD storage servers were quite taxed during testing, which caused transfer failures. However, this issue is not terribly relevant for the future, as OU's SE will be replaced by new 100G DTN and ceph-based backend storage this summer.

As followup and next steps we suggest to tune and fix FTS and deletion.

6.16 DESY

DESY provides a WLCG Tier-2 for ATLAS, CMS, and LHCb. Together with KIT, DESY is the German RAW data centre of the Belle II experiment. In addition, DESY provides resources for the ILC community and smaller HEP experiments. Beyond HEP, DESY is the host laboratory for the European XFEL and a major contributor to the CTA observatory and other astroparticle physics experiments. The compute and storage resources for distributed computing are located at two sites: DESY-HH in Hamburg and DESY-ZN in Zeuthen, in the Berlin area.

During DC24, the DESY sites were targeted by ATLAS, Belle II, and CMS. No operational problems were observed, nor major impacts on ongoing computing activities. Nevertheless, the 100Gbit WAN connection showed high utilization, which appeared sufficient for DC24 scales but would require upgrades for more demanding future exercises.

The two DESY sites have a special setup, which can presently be fully captured in the WLCG site network monitoring: Site network details and monitoring are only available for layers 3 and higher. In non-vanilla network setups, correlations between sites might not be mapped as expected. For example, in the DESY case with locations Hamburg and Zeuthen, both are also separate WLCG sites (DESY-HH and DESY-ZN) including separate network masks assumed. However, Zeuthen is connected to WAN with a dark fiber through Hamburg, with both sites sharing the same WAN up-/downlinks. Since this correlation on layer-2 is not possible to be indicated in the layer-3 monitoring or the sites' CRIC networking information, transfer throughput and capacity information might not be adequately realised. A suggestion was made that additional tags could be added to UDP fireflies to include site and/or instance information. (Reference: https://github.com/scitags/scitags.github.io/issues/1)

6.17 CZ-Tier2

The WLCG Tier-2 centre in the Czech Republic supports the ALICE and ATLAS Experiments. Only ATLAS transfers were relevant during DC24. The site is distributed across 3 locations with 3 institutions, but ATLAS storage is located only in the main site at FZU. This site is connected by a dedicated 100 Gbps link to CESNET, which is used only for LHCONE transfers. This link continues to the GEANT network with a throughput of 100 Gbps. A link to the public internet is realised with a throughput of 40 Gbps. Only transfers to the FZU computing centre use this link, other traffic to the FZU is routed via different links. The ATLAS DC24 transfers were able to fill the LHCONE link, and we observed maxima above 90 Gbps in one hour intervals during the second week of DC24. The average incoming traffic for the whole site was 56 Gbps during the second week of the DC24, specifically Monday 2024-02-19 10:00 to Friday 2024-02-23 22:00, and 45 Gbps for the outgoing traffic. These values from the local network monitoring via LibreNMS are in a satisfactory agreement with the DC24 dedicated dashboard, which states 52 Gbps for incoming and 44 Gbps for outgoing traffic.

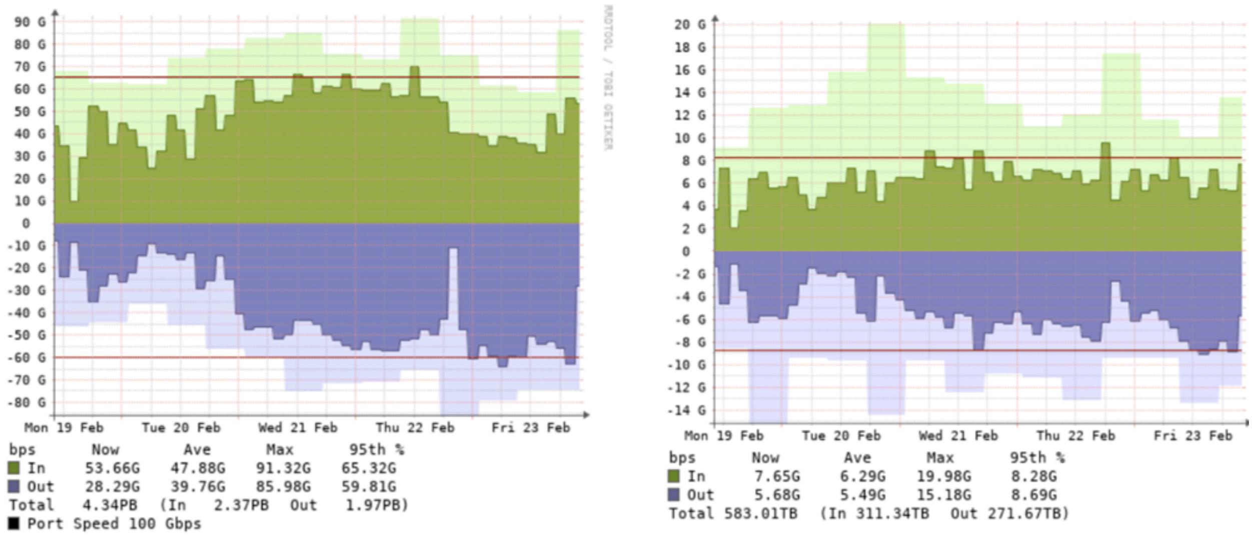


Figure 6.17a: Traffic during the second week via LHCONE link (left) and the public internet (right).

DC24 showed that the site was able to handle traffic close to the LHCONE network capacity. The ratio between the transfers via LHCONE and the public internet was almost the same for incoming, at 86.4 % via LHCONE, and outgoing traffic at 87.5 %. This observation, that most of the WLCG traffic to the Czech Republic goes via the dedicated LHCONE link, instead of the public internet, pleased the Czech NREN provider CESNET. The international network capacity of the generic internet is more limited than the capacity to GEANT, which is used by LHCONE. The Czech connection to GEANT was updated to 2x200 Gbps shortly after DC24. The CZ-Tier2 site connection will be updated to 200 Gbps later in 2024. There were no errors on the link even during the peak, only a relatively small number of drops occurred for outbound traffic, as can bee seen in Figure 6.17a on the LHCONE link.

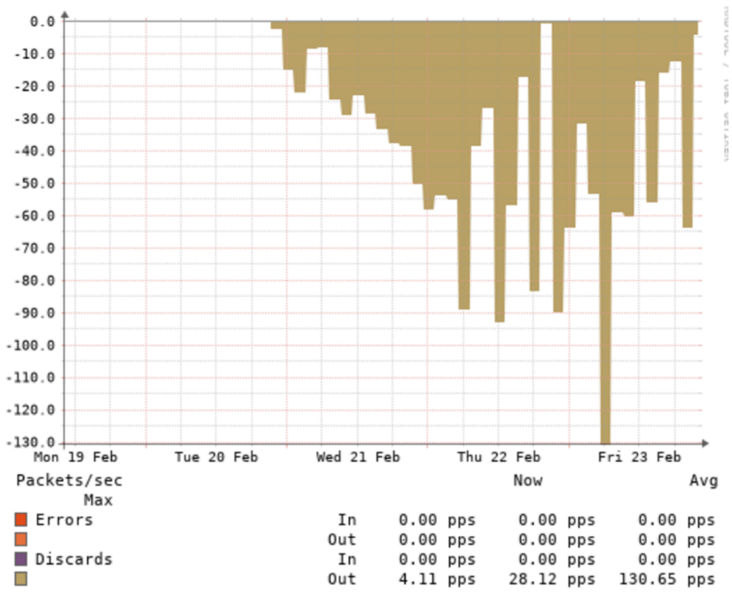


Figure 6.17b: Number of errors was zero even during the peaks, there were some discards for the outbound traffic on the LHCONE link.