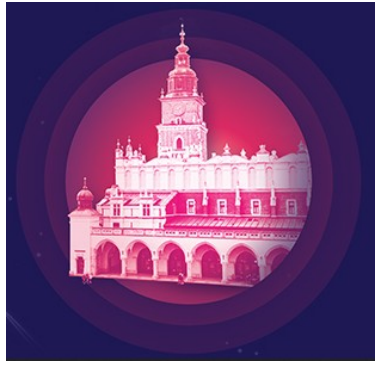
# Data Challenge 24 CMS Activities

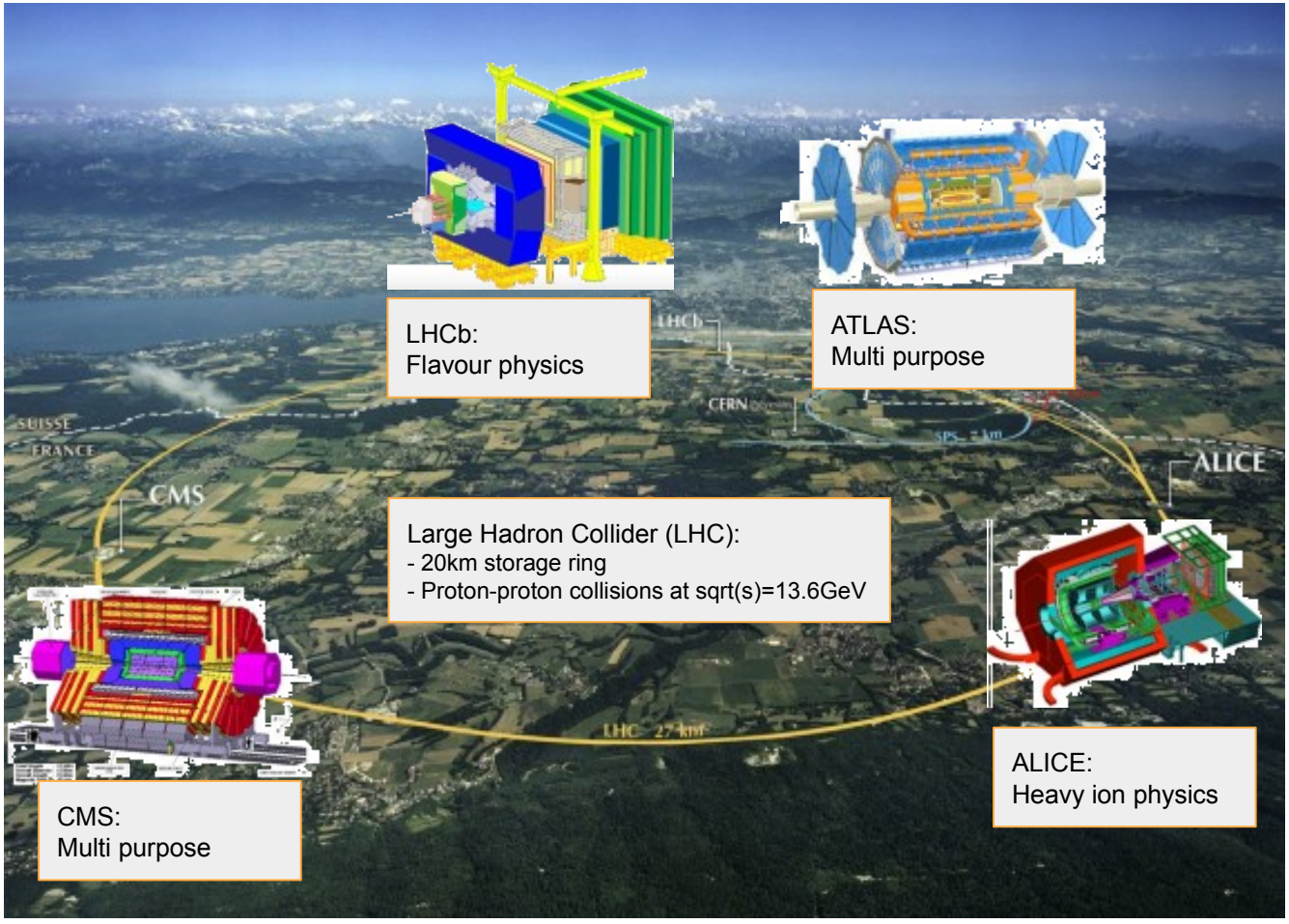
Rahul Chauhan (CERN), Katy Ellis (RAL),
Andres Manrique Ardila (Wisconsin), Hasan Ozturk (CERN),
Panos Paparrigopolos (CERN), Garyfallia “Lisa” Paspalaki (Purdue),
Christoph Wissing (DESY) for the CMS Collaboration



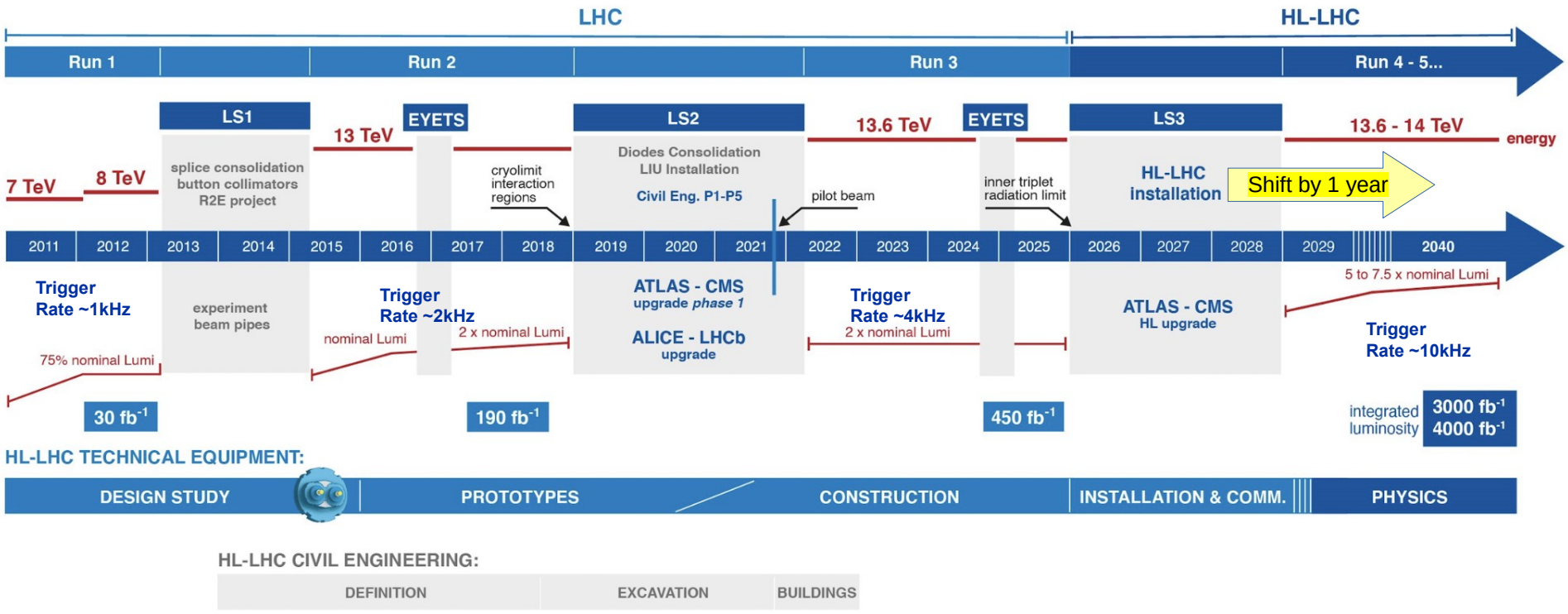
CHEP 2024

21st -25th October 2024

# The Large Hadron Collider & Experiments



# Towards HL-LHC




Data rates @HL-LHC increase by roughly an order of magnitude due to bigger/more complex events and higher logging rate

Figure adopted from:
Zerlauth, Markus & Bruning, Oliver. (2024). Status and prospects of the HL-LHC project. DOI; 615. 10.22323/1.449.0615.

# WLCG Data Challenges

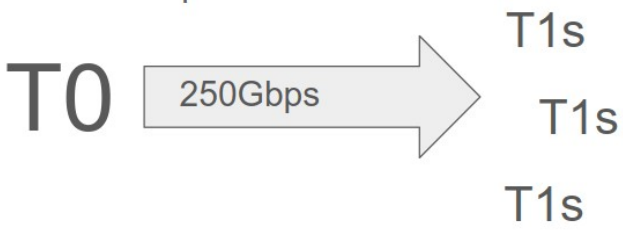
**Demonstrate readiness for the (expected) HL-LHC requirements**

**Series of data challenges with increasing throughput and technical complexity**

- DC21
  - 10% challenge, focus on Tier-0 to Tier-1 export (“minimal model”)
  - Involved experiments: ALICE, ATLAS, CMS, LHCb
- DC24
  - 25% challenge, more complex flows including traffic between T1-T2, T1-T1 and T2-T2
  - Involved experiments: ALICE, Belle II, CMS, DUNE, LHCb
  - Token based authentication for transfers, use of packet marking for net flows, SDN exercises
- Future DCs
  - A DC likely in LS3
    - About 50% of HL-LHC
  - A DC about a year before HL-LHC start
    - (Close to) 100% of HL-LHC

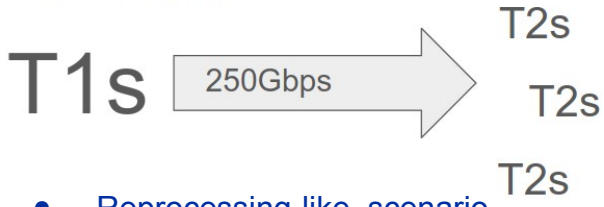
# Traffic Modeling in CMS

1. “T0 export”



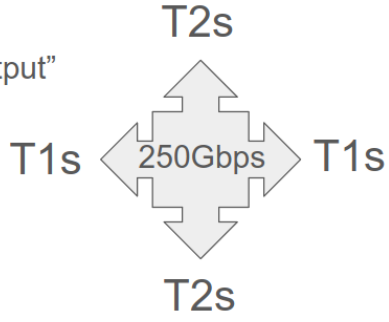

- Rather well modelled
- Numbers derived from DAQ TDR and LHC uptime assumptions

2. “T1 export”



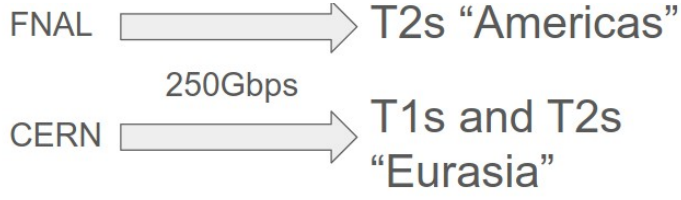

- Reprocessing-like scenario
  - HL-LHC approach not fully developed
- Data rates still somewhat uncertain

3. “Production output”




- MC & derived data scenario
  - HL-LHC approach not fully developed
- Data rates still somewhat uncertain

4. “AAA”




- Unscheduled remote reads via Xrootd
  - Main traffic presently MC premixing served from CERN and FNAL
  - HL-LHC approach not fully developed
- Data rates still somewhat uncertain

# CMS DC24 Menu

| 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 |

- 2 (working) weeks
- Program over the days aligned with ATLAS to run similar things
- ALICE and LHCb run a program mostly focused on CERN to Tier-1 traffic
- Also Belle II and DUNE participated in the challenge

# DC Execution and Main Tools

**Data challenge was executed on the production infrastructure**

- Usual MC production, data processing and analysis jobs continued (unaffected)
- For the challenge additional data transfers get initiated in RUCIO (CMS data management tool)
- Monitoring can distinguish them via “activity” tags

**Main tool: dc_inject.py**

- Suited existing datasets get subscribed with a short life time to destination site(s) to meet a certain target rate on a link
  - Rucio submits bulk transfer requests to FTS (File Transfer Service)
  - Once life time expires data are removed again (to allow for new transfers)
- Initially developed in ATLAS already for DC21
  - Co-developed between ATLAS and CMS during DC24 preparation

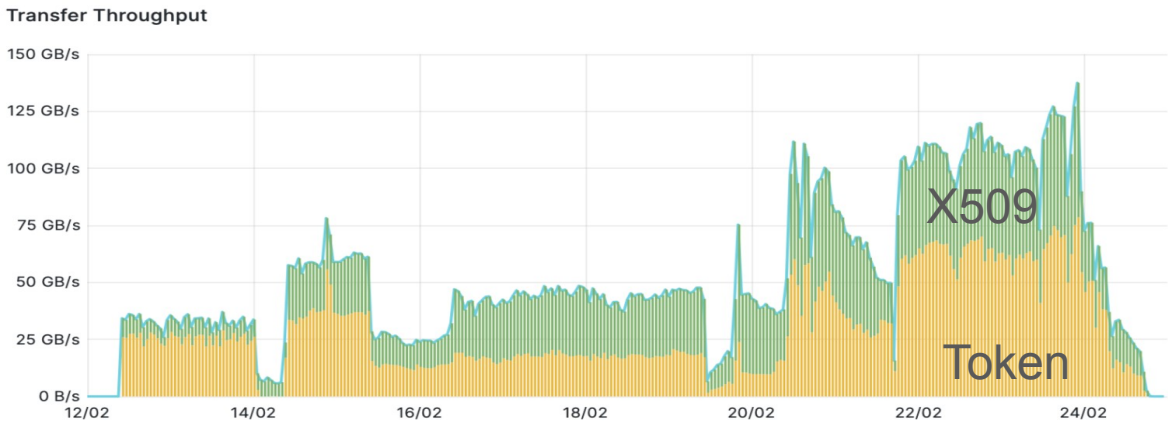
# Getting Sites & Middleware ready for Tokens

## Introduction of tokens for authentication

- Proper support required throughout the full stack:
- Rucio -> FTS -> storage systems (and related monitoring)
- Many options regarding tokens scopes & life time
- No experience using tokens at scale in production before DC24

## Intense collaboration between middleware developers and experiment teams

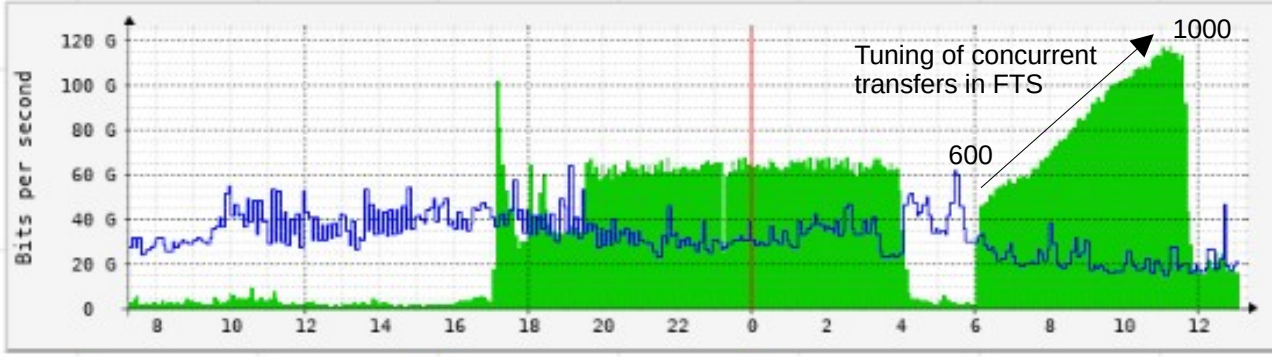
- Work out a simple, but good enough configuration for the DC24
- Tests with “early adopter” sites
- Roll-out to as many sites as possible
- Validation (mainly via ETF/SAM)
- 19 sites ready before DC24
  another 6 ready during DC24



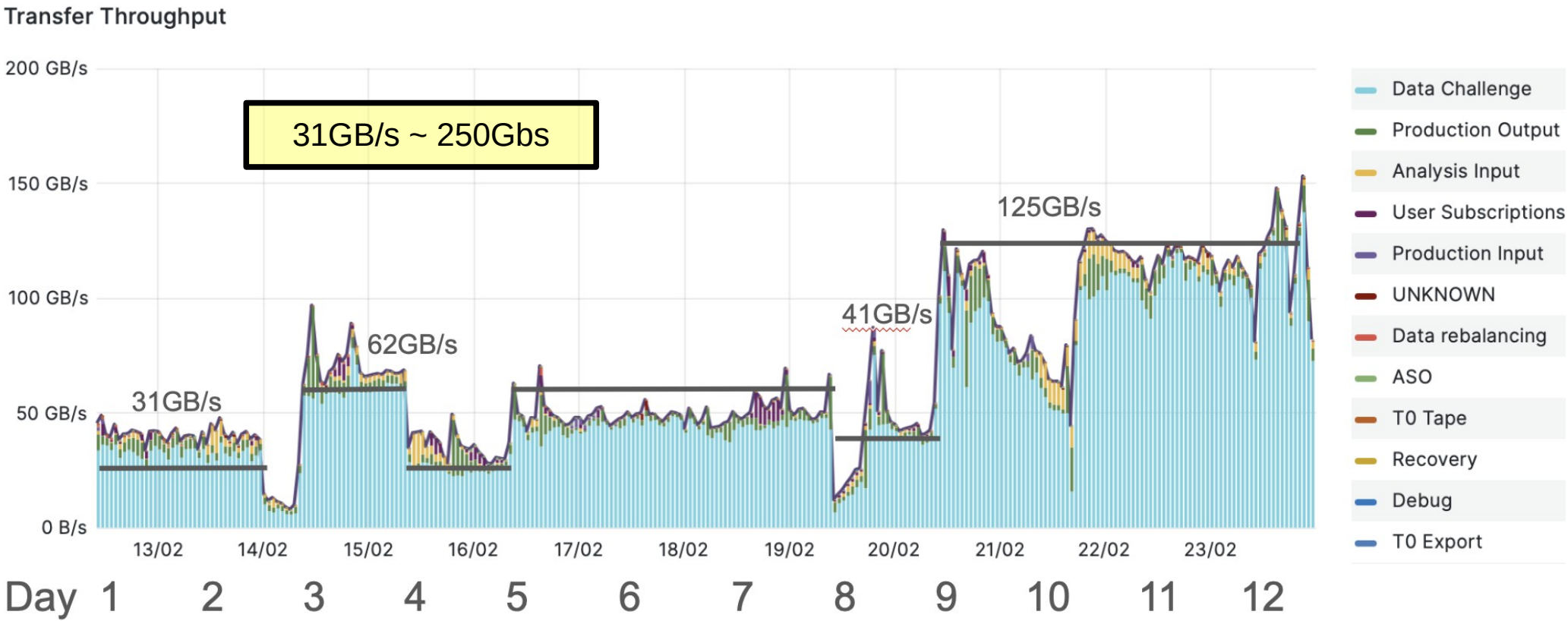
# Pre-Exercises

- Pre-tests were crucial for the success of the challenge
- Opportunity to gain operational experience and develop a concrete plan for running the actual challenge:
  - Improvements of monitoring and expanded the dc_inject tool to suit CMS needs
  - Sites also got a chance to tweak their configuration and internal monitoring
- Regional tests, e.g. in the US and UK
- CERN to individual Tier-1 export, partly coordinated with ATLAS

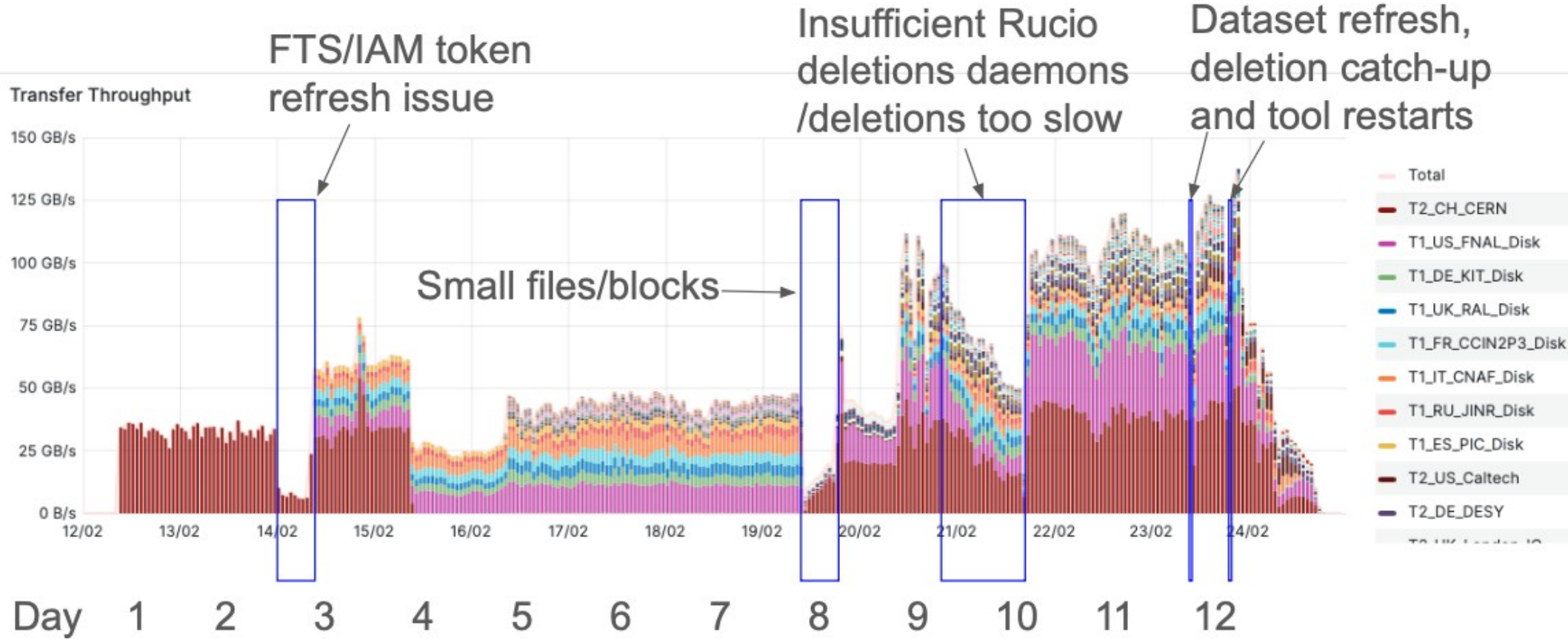
CERN to FNAL pre-tests in Nov 2023



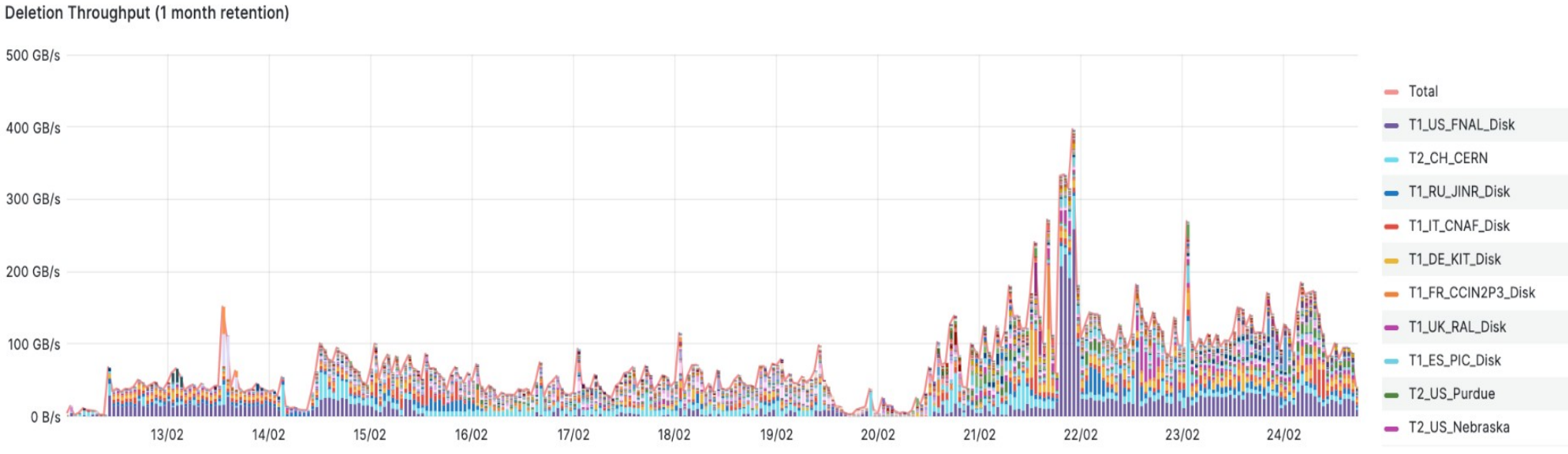
# Results – Throughput



# Results – Operational Items



# Data Challenge - Deletions




- Disk space available to subscribe DC24 data is limited
- Timely removal is key to allow for continuous data influx
- During the challenge monitoring of deletion performance was important
- Need to tune the Rucio reaper daemon

# Comparison: Rates achieved vs. targeted – Tier 1s

| | | 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 |

Ratio = observed/targeted

Green - ratio > 0.9

Yellow – 0.9 > ratio > 0.7

Orange – 0.7 > ratio > 0.5

Red - ratio < 0.5

# Comparison: Rates achieved vs. targeted – Tier 1s

| | | 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. | | | | | 1.28 | 0.93 |
| 4 | T1 Export | N/A | 0.37 | N/A | 0.91 | N/A | 1.12 | N/A | 0. | | | | | 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 | 0.5 | | | | | 0.56 | 0.65 | 0.25 |
| 10 | All | 0.70 | 0.34 | 0.98 | 0.74 | 0.58 | 0.65 | 0.56 | | 0.70 | | .03 | 0.98 | 0.63 | 0.28 |
| 11 | | 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 |

LHC OPN link to RAL disturbed

Tuning of FTS

Typically storage overloaded

Typically storage overloaded

Ratio = observed/targeted

Green - ratio > 0.9

Yellow – 0.9 > ratio > 0.7

Orange – 0.7 > ratio > 0.5

Red - ratio < 0.5

# Comparison: Rates achieved vs. targeted

| Day | T0 Export | T1 Export | T1s ⇔ T2s | T2s ⇔ T2s | AAA: CERN to T2s | AAA: FNAL to T2s | T2 to 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 |

Ratio = observed/targeted

Green - ratio > 0.9

Yellow – 0.9 > ratio > 0.7

Orange – 0.7 > ratio > 0.5

Red - ratio < 0.5

# Further Observations

**Pure network bandwidth was not a limitation**

- Actually various sites asked for additional injection to challenge their LAN link
- Some extra load got injected trying not to overrun FTS

**FTS performance and tuning**

- CMS FTS instance at CERN was quite busy maintaining injections for ~200 links
- Thanks to heroic efforts by the FTS team the service kept running
  With over 1000 links fed for DC24 the ATLAS instance sort of fall apart
- Together with the FTS team CMS data management team learned a lot
- Clear indication for the need of a “back pressure” mechanism

**Interplay of Rucio, FTS, IAM**

- Crucial for token handling
- Valuable experiences how to scale usage of tokens

**Rucio reaper (remote deletion agent)**

- Configuration needed quite some tweaks to achieve sufficient performance
- Further improvements on the road map for future development

# Upcoming Data Challenges

**Oberservations from recent DC are input for future challenges**

**Improved AAA (CMS Xrootd federation) testing**

- DC24 used Rucio injections to create AAA-like load for CERN and FNAL
- Upcoming DC should introduce load using the real system

**Tokens should be established**

- Opportunity to test token approach at new unprecedented scales

**Some incorporation of more recent SDN technology**

- During DC24 some SDN prototypes were tested, but mainly “outside” of the CMS systems
  - E.g. SENSE Rucio or NOTED
- Opportunity to test a deeper integration of such SDN based technology in CMS systems

**Continued DC-like mini challenges**

- Great opportunity to test ongoing R&D on networks and storage
- Some initial test together with ATLAS planned in Dec 2024 and Jan 2025

# Summary: Communalities DC24 & CHEP Travel

Things get planned months ahead

… however sometimes do not run exactly as scheduled

# Summary: Communalities DC24 & CHEP Travel


Photo by Thomas Hartmann (DESY)

Things get planned months ahead
… however sometimes do not run exactly as scheduled,
but finally sort out one way or another

DC was similar to that – quite a successful and useful exercise

- Targets regarding overall throughput were generally met
- Looking in detail a number of potential bottlenecks were observed
  - Scalable handling of tokens
  - Scalability of FTS and Rucio
  - Risk to overload storage systems
- Lessons learned during the challenge guide further developments
- DC24 was a good community effort –
  **CMS wants to thank the middleware developers,**
  **site administrators and other experiments**