DC24 Preparation Workshop – Summary

WLCG MB Meeting
2023-11-21

Christoph Wissing (DESY), Mario Lassnig (CERN)

Some workshop statistics

- Workshop held at CERN November 9th & 10th
- Great attendance across the community
  - We had 89 registrations
    - However for remote participation we did not require registration
    - Nevertheless quite some registrations just for remote participation (Indico instructions were a bit fuzzy)
  - Typically 50+ attendees in the room
  - Around 30 attendees via video conference
  - Participation from experiments, middleware providers and network experts
  - Scheduling the workshop together with pre-GDB on tape evolution and GDB was a good choice
- Lively and constructive discussion atmosphere
  - Positive feedback from participants
  - Workshop recognized as very useful
- Great support by Catharine Noble (logistics & administrative tasks) -THANKS!
- Coordinates
  - Indico agenda
    - Presentations are public, most recordings are already up on https://videos.cern.ch
    - Video recordings restricted to e-group wlcg-doma@cern.ch

Outline of the Summary

- Introduction
  - Recap on WLCG data challenges
  - DC21
- Experiments
  - Status and plans
- Storage & middleware
- Monitoring & tools
- R&D Topics

Recap: Data Challenges for HL-LHC

- WLCG has been mandated to execute data challenges for HL-LHC
  - Demonstrate readiness for expected HL-LHC data rates
  - Increasing volume/rates
  - Increase complexity (e.g. additional technology)
  - A data challenge roughly every two years

- DOMA is the coordination and execution platform
  - Agreements across the LHC experiments and beyond
    - Suited dates
    - Reasonable targets
    - Functionalities
  - Help in orchestration

- Dates and high level goals always approved by WLCG MB

Recap of (initial) modelling & resulting rates

ATLAS & CMS T0 to T1 per experiment

350PB RAW, taken and distributed during typical LHC uptime of 7M seconds

- 50GB/s or 400Gbps

Another 100Gbps estimated for prompt reconstruction data tiers (AOD, other derived output)

1Tbps for CMS and ATLAS summed

ALICE & LHCb T0 Export

100 Gbps per experiment estimated from Run-3 rates

WLCG data challenges for HL-LHC - 2021 planning
https://zenodo.org/records/5532452

Minimal Model

Sum (ATLAS,ALICE,CMS,LHCb)*2(for bursts)*2(*overprovisioning*) = **4.8Tbps for the expected HL-LHC bandwidth needs**

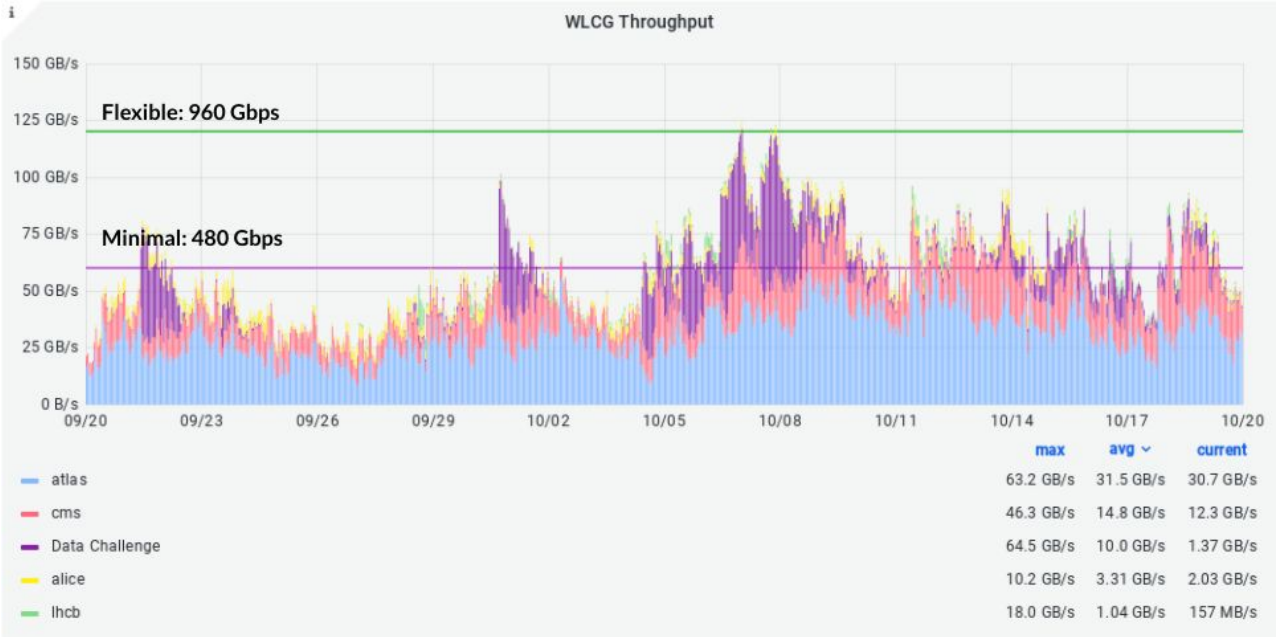
Flexible Model

Assumes reading of data from above for reprocessing/reconstruction in 3 month (about 7M seconds)
Means doubling the Minimal Model: **9.6Tbps for the expected HL-LHC bandwidth needs**
However data flows primarily from the T1s to T2s and T1s!

Data Challenges target: **50% filling of expected** HL-LHC bandwidth needs

DC21

However, we managed to fill 100% of the (minimal) DC21 target!



Network Data Challenges 2021 wrap-up and recommendations
https://zenodo.org/records/5767913

DC24 Timetable

- Dates: **February 12th (Mon) to February 23rd (Fri)**

- Proposal to distribute different exercises over the challenge days
  - Note: All DC exercises run **on top of ongoing production**
  - Common program mostly for ATLAS and CMS
  - Data taking scenario (T0 to T1s)
  - Production like scenarios
  - Ramp up to reach target of flexible scenario
  - Squeeze in dedicated technical tests (e.g. special TCP setup at selected sites)
  - Some contingency to repeat something

- Even with 2 weeks the schedule is tight

DC24 Communications

- Proposed: A short(!) daily call among experiment operators
  - E.g. at 16:00 every working days
  - Briefly discuss status, issues and plans for the next day
  - Not "everyone" needs to attend each of these meetings
  - Open for anyone interested

- We have a Data Challenges Mattermost channel in WLCG team
  - https://mattermost.web.cern.ch/wlcg-gdb/channels/wlcg-data-challenges
  - Expect to use this as the main communication platform
  - Feedback from DC21
    - Don't split discussions among different channels

Experiments

ALICE & LHCb

- Focus on CERN to T1s
  - By far most dominant network traffic

Site	shares	Data written (TB)	Export speed	Staging Speed (GB/s)	Staging duration (hours)
CERN		2117.00	14.00	2.57	48.39
CNAF	14.61%	309.36	2.05	1.60	53.77
GRIDKA	19.56%	414.01	2.74	1.66	69.13
IN2P3	10.93%	231.38	1.53	1.20	53.77
NCBJ	7.30%	154.64	1.02	0.89	48.39
PIC	3.64%	77.13	0.51	0.40	53.77
RAL	28.26%	598.29	3.96	2.40	69.13
RRCKI	0.00%	0.00	0.00	0.00	0.00
SARA	8.24%	174.40	1.15	0.80	60.49
Beijing	7.45%	157.79	1.04	0.63	69.13
Total Tier1s	100.00%	2117.00	14.00	9.58	

- LHCb
  - 1st week: T0 Export to T1s ("DT mode")
  - 2nd week: T1 staging ("AD mode")
  - Both exercises **involve tapes**
    - writing (DT), reading(AD)

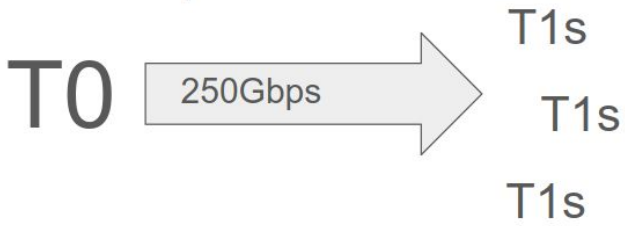
- ALICE
  - Still archiving of HI data from 2023 run
  - If archiving is complete by DC24, synthetic load T0 -> T1s

Centre	Target rate GB/s	Achieved rate GB/s
CNAF	0.8	2 (250%)
IN2P3	0.4	0.8 (200%)
KISTI	0.2	1 (500%)
GridKA	0.6	2 (300%)
NDGF	0.3	0.4 (133%)
NL-T1	0.1	0.9 (900%)
~~RRC-KI~~	~~0.4~~	~~0.53 (128%)~~
RAL	0.1	0.7 (700%)
CERN	*10*	*20 (200%)*

Target 2.5GB/s (T1s) + 10GB/s (T0)

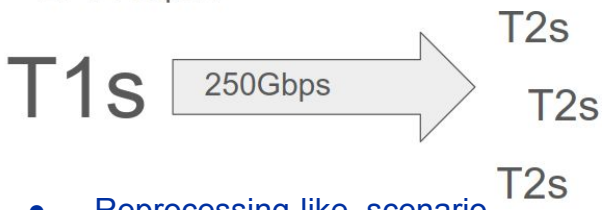
CMS - Main Scenarios

1. “T0 export”



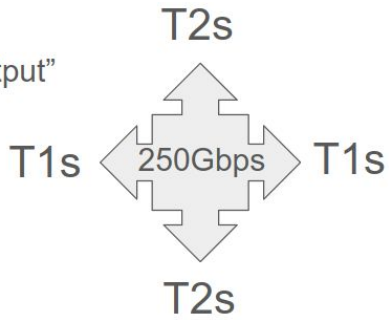
- Rather well modelled
- Numbers derived from DAQ TDR and LHC uptime assumptions

2. “T1 export”



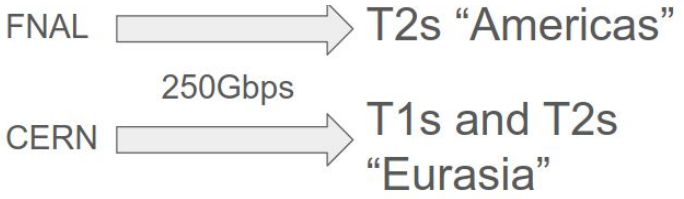
- Reprocessing-like scenario
  - HL-LHC approach not fully developed
- Data rates still somewhat uncertain

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 - Draft plan

Day of challenge	1	2	3	4	5	6	7
Day of week	Monday	Tues	Wed	Thur	Fr	Sat	Sun
Scenario	T0 export	T0 export	**Mixed**	T1 export	**Mixed**	**Mixed**	**Mixed**
			T0 export		T1 export	T1 export	T1 export
			T1 export		Prod. output	Prod. output	Prod. output
Mode	"Data taking"	"Data taking"	T1 read+write	T1s -> T2s	T1s <-> T2s	T1s <-> T2s	T1s <-> T2s
T0->T1s	31	31	31	0	0	0	0
T1s->T2s	0	0	31	31	31	31	31
T2s->T1s	0	0	0	0	31	31	31
AAA	0	0	0	0	0	0	0
Total rate (GB/s)	31	31	62	31	62	62	62
Total rate (Gb/s)	248	248	496	248	496	496	496

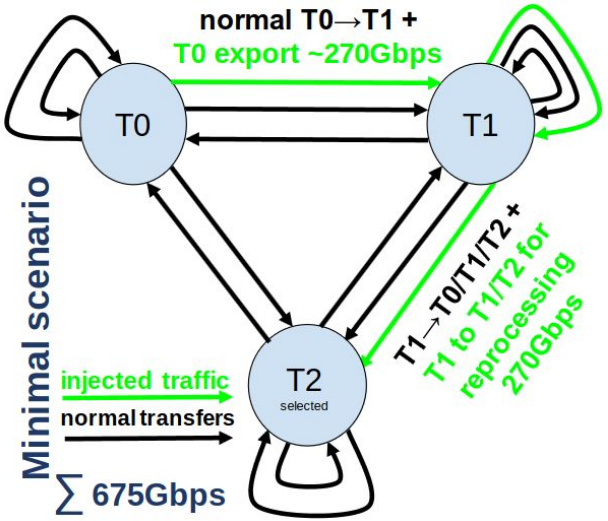
Day of challenge	8	9	10	11	12
Day of week	Mon	Tues	Wed	Thur	Fri
Scenario	AAA	**"Max throughput"**	**"Max throughput"**	Contingency	Contingency
		T0 export	T0 export		
		T1 export	T1 export		
		Prod. output	Prod. output		
			AAA?		
Mode	CERN/FNAL to	Everything	Everything	?	?
	T1s + T2s				
T0->T1s	0	31	31		
T1s->T2s	0	31	31		
T2s->T1s	0	31	31		
AAA	31	0	31		
Total rate (GB/s)	31	93	124		
Total rate (Gb/s)	248	744	992		

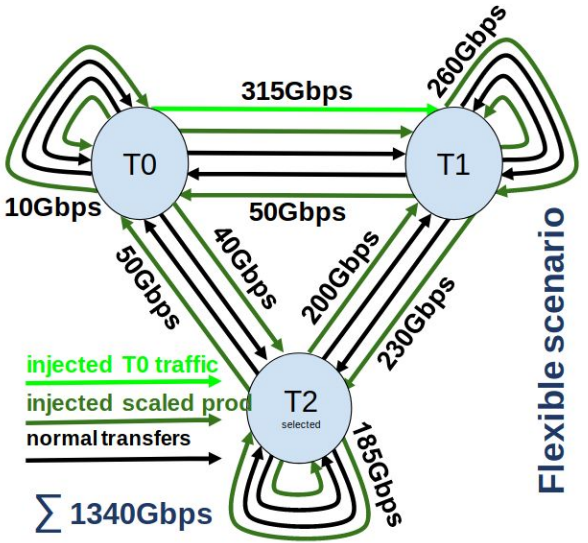
ATLAS - Planning

(preliminary version: 20231103) — Deletion rates are calculated from ingress bandwidth assuming 3GB average filesize)

Table: DC24 (src: ingress / egress)

Site	Tier	Cloud	Site WAN (Gb/s) Total (GB/s)	Site WAN (Gb/s) Usable by ATLAS	DC24 minimal scenario T0 Export	DC24 minimal scenario Total Gb/s & bandwidth Σ ingress	DC24 minimal scenario Total Gb/s & bandwidth Σ egress	DC24 minimal scenario Space [TB/24h] (deletions/hour)	DC24 flexible scenario T0 Export	DC24 flexible scenario Total Gb/s & bandwidth Σ ingress	DC24 flexible scenario Total Gb/s & bandwidth Σ egress	DC24 flexible scenario Space [TB/24h] (deletions/hour)
CERN-PROD	T0	CERN	2100	911	270.0	27.9	291.3	0 (0k)	270.0	93.1 - 112.2	363.1	884 (13k)
T0 summary					270.0	27.9	291.3		270.0	93.1 - 112.2	363.1	
BNL-ATLAS	T1	US	400	400	80.0	82.2	60.0	764 (11k)	80.0	107.5 - 119.6	120.0	1089 (15k)
FZK-LCG2	T1	DE	400	162	32.0	61.7	32.0	431 (6k)	32.0	86.3 - 100.3	64.0	911 (13k)
IN2P3-CC	T1	FR	200	93	33.0	53.3	33.0	413 (9k)	33.0	81.6 - 95.6	66.0	861 (12k)
INFN-T1	T1	IT	300	81	24.0	39.5	24.0	319 (5k)	24.0	54.8 - 64.0	48.0	588 (8k)
NDGF-T1	T1	ND	200	157	16.0	30.7	21.8	151 (2k)	16.0	77.9 - 96.6	32.0	842 (12k)
SARA-MATRIX	T1	NL	400	291	15.0	30.4	15.0	192 (3k)	15.0	54.4 - 66.0	30.0	604 (9k)
pic	T1	ES	200	89	13.0	21.4	13.0	170 (2k)	13.0	29.1 - 34.4	26.0	319 (5k)
RAL-LCG2	T1	UK	400	196	39.0	60.6	39.0	464 (7k)	39.0	88.5 - 100.1	78.0	861 (12k)
RRC-KI-T1 (no active T0 export)	T1	RU	200	73		13.4	8.0	109 (2k)		15.1 - 17.2	16.0	160 (2k)
TRIUMF-LCG2	T1	CA	100	100	30.0	45.9	30.0	403 (6k)	30.0	60.8 - 69.7	60.0	643 (9k)
T1 summary					270.0	429.3	275.8		270.0	655.9 - 763.8	540.0	
CA-VICTORIA-WESTGRID-T2	T2	CA	100	100		5.8 - 7.5	1.5 - 1.5	24 (0k)		3.6 - 13.4	1.5 - 11.0	104 (1k)
Australia-ATLAS	T2	CA	20	20		0.1 - 0.2	0.4 - 0.4	0 (0k)		0.1 - 2.4	0.4 - 2.6	25 (0k)
CA-WATERLOO-T2	T2	CA	40	40		2.0 - 2.4	1.2 - 1.2	7 (0k)		1.3 - 9.8	1.2 - 9.3	90 (1k)
CA-SFU-T2	T2	CA	100	100		5.9 - 7.7	5.7 - 5.7	45 (1k)		43.0 - 61.9	41.4 - 41.4	616 (9k)
praguelcg2	T2	DE	100	100		6.9 - 8.9	2.3 - 2.3	50 (1k)		16.9 - 22.9	15.5 - 15.5	197 (3k)
MPPMU	T2	DE				2.6 - 3.3	1.3 - 1.3	10 (0k)		1.7 - 9.4	1.3 - 9.1	82 (1k)
wuppertalprod	T2	DE	10	9		4.6 - 5.9	1.8 - 1.8	32 (0k)		9.9 - 10.0	6.4 - 8.8	74 (1k)
DESY-ZN	T2	DE	40	40		6.4 - 8.4	1.9 - 1.9	48 (1k)		14.3 - 19.2	12.4 - 12.4	163 (2k)
DESY-HH	T2	DE	100	100		9.1 - 10.0	1.9 - 1.9	49 (1k)		9.9 - 10.0	5.4 - 7.2	48 (1k)
UNI-FREIBURG	T2	DE				2.6 - 3.3	1.7 - 1.7	9 (0k)		1.8 - 11.3	1.7 - 11.6	101 (1k)
CYFRONET-LCG2	T2	DE	10	10		2.6 - 3.2	1.2 - 1.2	9 (0k)		1.7 - 9.9	1.2 - 9.4	90 (1k)
GoeGrid	T2	DE				5.2 - 6.6	1.2 - 1.2	20 (0k)		3.3 - 11.5	1.2 - 9.0	86 (1k)
IEPSAS-Kosice	T2	DE				1.1 - 1.3	0.4 - 0.4	4 (0k)		0.8 - 3.7	0.4 - 3.3	31 (0k)
LRZ-LMU	T2	DE				2.4 - 3.0	1.8 - 1.8	8 (0k)		1.6 - 12.9	1.8 - 12.5	120 (2k)
CSCS-LCG2	T2	DE	100	100		5.6 - 7.2	3.0 - 3.0	22 (0k)		3.6 - 22.3	3.0 - 21.3	198 (3k)
FMPhI-UNIBA	T2	DE				0.9 - 1.0	0.5 - 0.5	2 (0k)		0.7 - 4.2	0.5 - 4.0	37 (1k)
SAMPA	T2	ES	9	9		1.1 - 1.5	0.9 - 0.9	6 (0k)		0.6 - 8.1	0.9 - 7.3	79 (1k)
UAM-LCG2	T2	ES	10	10		0.7 - 0.9	0.4 - 0.4	4 (0k)		3.2 - 4.5	2.9 - 2.9	42 (1k)
ifae	T2	ES	200	200		2.7 - 3.4	0.7 - 0.7	11 (0k)		1.6 - 5.7	0.7 - 4.8	44 (1k)
NCG-INGRID-PT	T2	ES	9	9		0.5 - 0.7	0.2 - 0.2	2 (0k)		0.4 - 2.3	0.2 - 1.9	20 (0k)
IFIC-LCG2	T2	ES				4.1 - 5.2	2.0 - 2.0	17 (0k)		2.5 - 14.2	2.0 - 13.2	124 (2k)
EELA-UTFSM	T2	ES	10	10		0.2 - 0.2	0.3 - 0.3	1 (0k)		0.1 - 2.3	0.3 - 2.3	23 (0k)
TOKYO-LCG2	T2	FR	40	40		16.5 - 21.7	5.5 - 5.5	127 (2k)		30.0 - 39.7	29.8 - 29.8	317 (5k)
RO-07-NIPNE	T2	FR	100	100		4.3 - 5.4	2.6 - 2.6	29 (0k)		18.7 - 26.3	18.4 - 18.4	249 (4k)
BEIJING-LCG2	T2	FR	20	20		0.0 - 0.0	0.2 - 0.2	0 (0k)		0.0 - 1.5	0.2 - 1.3	15 (0k)
HK-LCG2	T2	FR				0.1 - 0.1	0.3 - 0.3	0 (0k)		0.1 - 2.3	0.3 - 2.3	23 (0k)
GRIF	T2	FR	100	100		7.2 - 9.4	4.2 - 4.2	32 (0k)		4.1 - 36.1	4.2 - 33.2	339 (5k)
IN2P3-LPC	T2	FR	100	100		2.4 - 3.0	1.5 - 1.5	14 (0k)		1.6 - 10.0	1.5 - 8.9	117 (2k)
IN2P3-LAPP	T2	FR	20	20		4.8 - 5.9	2.7 - 2.7	27 (0k)		16.1 - 19.9	13.6 - 15.1	174 (2k)
IN2P3-CPPM	T2	FR	100	100		2.5 - 3.2	1.6 - 1.6	17 (0k)		10.0 - 10.0	7.3 - 9.9	89 (1k)
INFN-MILANO-ATLASC	T2	IT	10	10		2.1 - 2.7	1.7 - 1.7	9 (0k)		1.2 - 10.0	1.7 - 10.0	94 (1k)
INFN-NAPOLI-ATLAS	T2	IT	100	100		3.8 - 4.8	2.2 - 2.2	24 (0k)		15.7 - 21.9	14.9 - 14.9	205 (3k)
INFN-ROMA1	T2	IT	10	10		2.5 - 3.2	1.1 - 1.1	17 (0k)		7.7 - 9.9	6.6 - 7.0	88 (1k)
INFN-FRASCATI	T2	IT	10	10		2.1 - 2.6	1.0 - 1.0	7 (0k)		1.5 - 8.6	1.0 - 7.9	75 (1k)
SE-SNIC-T2	T2	ND				0.0 - 0.0	0.0 - 0.0	0 (0k)		0.0 - 0.1	0.0 - 0.1	1 (0k)
UNIBE-LHEP	T2	ND				0.0 - 0.0	0.0 - 0.0	0 (0k)		0.0 - 0.0	0.0 - 0.0	0 (0k)
NIKHEF-ELPROD (no tape)	T1	NL	1000	1000		6.8 - 9.1	3.1 - 3.1	32 (0k)		3.7 - 21.5	3.1 - 21.8	188 (3k)
TECHNION-HEP	T2	NL				4.0 - 5.0	1.5 - 1.5	13 (0k)		2.8 - 13.6	1.5 - 11.6	115 (2k)
TR-10-ULAKBIM	T2	NL	9	9		0.2 - 0.3	0.9 - 0.9	1 (0k)		0.2 - 7.6	0.9 - 7.1	79 (1k)
JINR-LCG2	T2	RU	100	100		1.9 - 2.4	0.8 - 0.8	8 (0k)		1.1 - 6.1	0.8 - 5.6	53 (1k)
RU-Protvino-IHEP	T2	RU	20	20		0.4 - 0.5	0.4 - 0.4	1 (0k)		0.4 - 3.5	0.4 - 3.2	33 (0k)
UKI-LT2-RHUL	T2	UK	10	10		0.9 - 1.1	0.3 - 0.3	3 (0k)		0.6 - 1.9	0.3 - 1.6	14 (0k)
UKI-NORTHGRID-MAN-HEP	T2	UK	40	40		8.4 - 10.8	2.7 - 2.7	61 (1k)		19.0 - 25.6	18.3 - 18.3	217 (3k)
UKI-SOUTHGRID-RALPP	T2	UK	20	20		0.8 - 0.9	0.6 - 0.6	2 (0k)		0.6 - 5.1	0.6 - 4.7	48 (1k)
UKI-SCOTGRID-GLASGOW	T2	UK	20	20		2.5 - 3.2	1.3 - 1.3	10 (0k)		1.6 - 10.2	1.3 - 9.3	92 (1k)
UKI-LT2-QMUL	T2	UK				7.2 - 8.8	2.9 - 2.9	23 (0k)		5.0 - 19.7	2.9 - 19.7	156 (2k)
UKI-SCOTGRID-ECDF	T2	UK				0.8 - 1.0	0.5 - 0.5	3 (0k)		0.6 - 3.7	0.5 - 3.6	33 (0k)
UKI-NORTHGRID-LANCS-HEP	T2	UK	40	40		5.5 - 6.8	4.4 - 4.4	18 (0k)		3.8 - 35.5	4.4 - 32.0	336 (5k)
UKI-NORTHGRID-LIV-HEP	T2	UK				0.7 - 0.9	0.4 - 0.4	3 (0k)		0.5 - 3.2	0.4 - 2.8	29 (0k)
Taiwan-LCG2 (no tape)	T1	TW	20	20		3.5 - 4.1	1.7 - 1.7	9 (0k)		2.6 - 12.2	1.7 - 10.8	101 (1k)
NET2	T2	US	10	10		0.0 - 0.0	0.0 - 0.0	0 (0k)		0.0 - 0.0	0.0 - 0.0	0 (0k)
SWT2_CPB	T2	US	100	100		9.7 - 12.1	8.5 - 8.5	59 (1k)		58.8 - 83.7	60.7 - 60.7	815 (12k)
AGLT2	T2	US	100	100		9.9 - 12.7	7.0 - 7.0	70 (1k)		47.4 - 67.0	49.3 - 49.3	642 (9k)
OU_OSCER_ATLAS	T2	US	100	100		1.2 - 1.6	0.6 - 0.6	5 (0k)		0.7 - 5.4	0.6 - 4.8	49 (1k)
MWT2	T2	US	200	200		24.2 - 32.0	9.9 - 9.9	193 (3k)		60.9 - 82.0	67.2 - 67.2	720 (10k)
BU_NESE	T2	US				2.8 - 3.7	2.5 - 2.5	14 (0k)		1.5 - 16.7	2.5 - 17.7	161 (2k)
BU_ATLAS_Tier2	T2	US				0.1 - 0.1	0.3 - 0.3	0 (0k)		0.1 - 0.4	0.3 - 0.6	4 (0k)
T2 summary						213.1	107.2			574 - 759	420 - 732	
Summary						680.4	674.2			1323 - 1635	1323 - 1635	

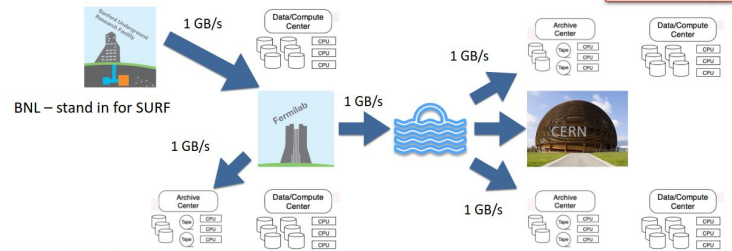




- Rather detailed planning exists
- Rates are mainly scaled values from measured Run-3 values
- Sites are already informed about expected rates

DUNE Involvement in WLCG Data Challenge 24

"FD" Raw Data
archival storage



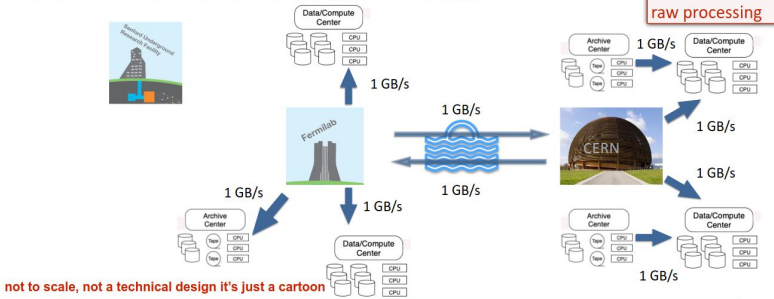
not to scale, not a technical design, it's just a cartoon

- Simulate the archival of 25% of the raw data rate from the Far Detector
  - translates to 1 GB/s from SURF to FNAL
  - replicate that "FD" raw data to archival storage facilities around the world
  - replicate the "FD" raw data to disk storage elements around the world for prompt access from compute elements
- Both job submission and RSE to RSE w/ token authentication/authorization

Brookhaven National Laboratory DUNE

DUNE Involvement in WLCG Data Challenge 24

"FD" Raw Data
raw processing



not to scale, not a technical design it's just a cartoon

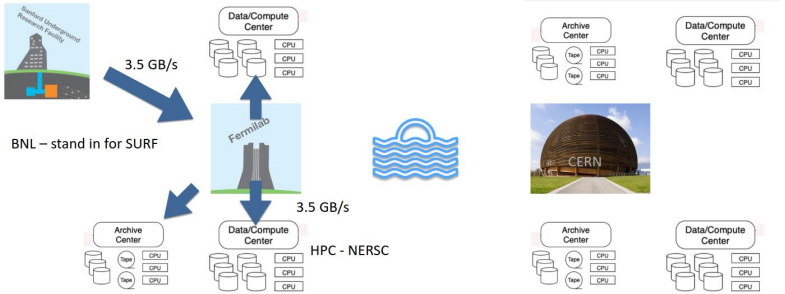
- Maintain continuous processing workload at distributed sites commensurate with 25% "FD" raw data rate
  - utilize compute elements across the WLCG and OSG
  - match the locality of jobs with locality of data at nearby RSEs
- **Both job submission and RSE to RSE w/ token authentication/authorization**

Brookhaven National Laboratory DUNE

- Rates are typically low compared to ATLAS and CMS
- Flows often in the opposite direction with respect to WLCG
- Verify that things continue working during increased load from DC24

DUNE Involvement in WLCG Data Challenge 24

SuperNova Raw Data
rapid transfer & processing



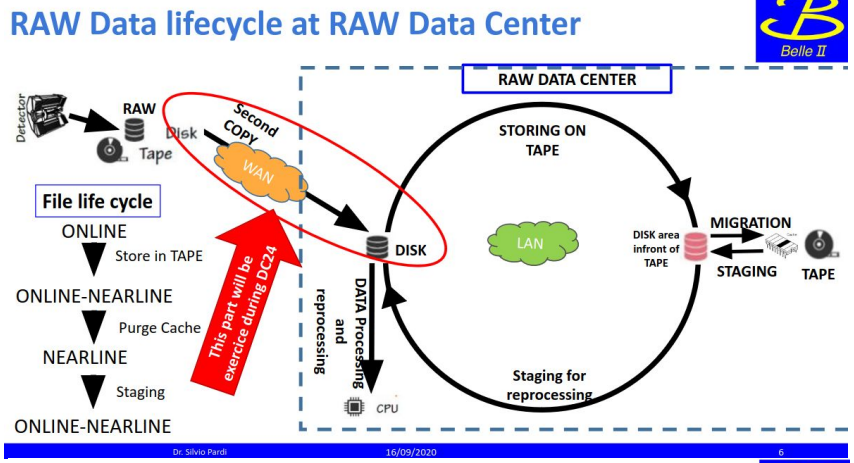
not to scale, not a technical design it's just a cartoon

Note: maximum network rate from SURF – 100 Gbits/sec (12.5 GB/s)

Brookhaven National Laboratory DUNE

Belle II

- Focus is data export
- KEK to RAW data centers
  - BNL (30%), CNAF(20%), IN2P3(15%), UVic (15%), DESY (10%), KIT (10%)
  - All sites support LHC experiment(s)
  - Many (educational) network are shared
- Scale certainly below WLCG exercises
  - Still increase of load at WLCG sites



Belle II Data Challenge 2024

What should be exercised during DC24:

Technology that can be stressed: Network, DDM (RUCIO server at BNL) FTS, Storages, Monitoring System, Protocols, IAM

Main goal: Emulate data transfer conditions in a Belle II high-lumi scenario

Our current estimation for such scenario is 40 TB per day.

Transfers from KEK to raw data centers according to our distribution schema (30%BNL, 20%CNAF, 15% IN2P3CC, 15%UVic, 10%DESY, 10%KIT)

Considering that the average speed needed to transfer 40TB/day is 3.7Gbit/s in outbound at KEK vs all the Raw Data Centers.

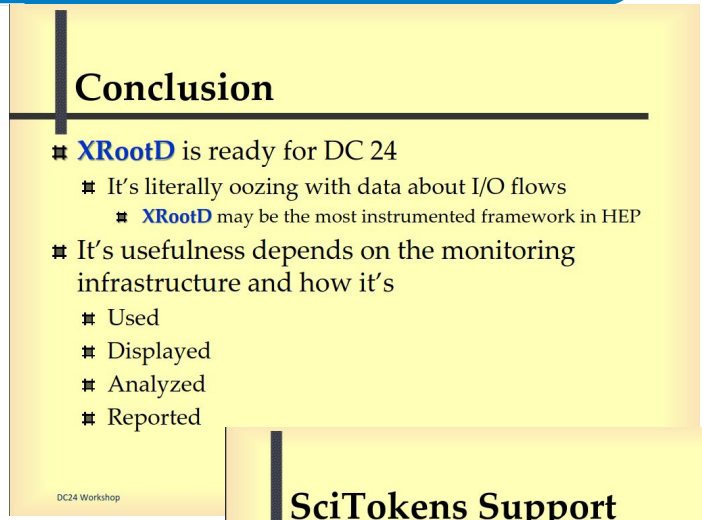
- Hypothesis 1 - The target speed to achieved is 5x3.7Gbit/s = **18.5 Gbit/s**
- Hypothesis 2 - The target is sent **5x40TB in one day**, 5 times bigger than the expected amount of data

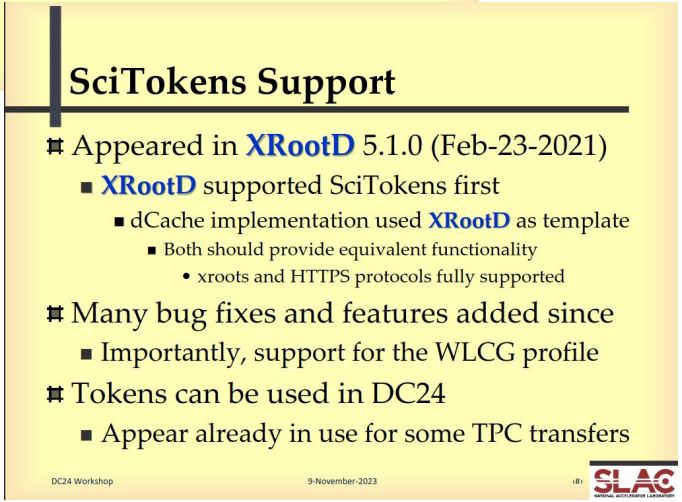
8

Middleware

XRootD

- XRootD is ready for DC24
- Token support is there
- Monitoring capabilities
  - Packet marking and SciTags available in most recent release
  - Discussion:
    - Monitoring of multi VO instances
    - Traffic can be identified on VOMS information using X509, when properly configured (tbc)
    - Less obvious how to proceed with tokens





StoRM

- StoRM ready for DC24
- Token support is there
  - Some open items (links in the talk)
    - Should be no blocker
- Monitoring capabilities

StoRM Deployments @ CNAF-T1 - Token support

- WLCG JWT profile scope-based authorization is enabled on:
  - ATLAS, CMS, DUNE disk storage areas
  - CMS tape storage area
  - Looking forward to LHCb input
- More than other 30 experiments have (non WLCG) token-based authorization enabled
  - BELLE, JUNO, CTA-LST, HERD, etc.
  - Mixed AuthZ based on capabilities and/or identities/groups

14

StoRM plans for DC 24

- Solve StoRM WebDAV thread contention which manifests as load unbalancing between deployed instances
- Support packet marking
- Provide requested metrics (if any)

24

dCache

- dCache ready for DC24
- Token support is there
  - CMS just launched a deployment campaign
- Very recent Golden Release 9.2 is recommended
  - Enables new XRootD monitoring framework
- Discussion item:
  - Bulk deletion without SRM
    - Feature only provided via SRM presently
    - Use REST API?
    - Explore capabilities via HTTPS?

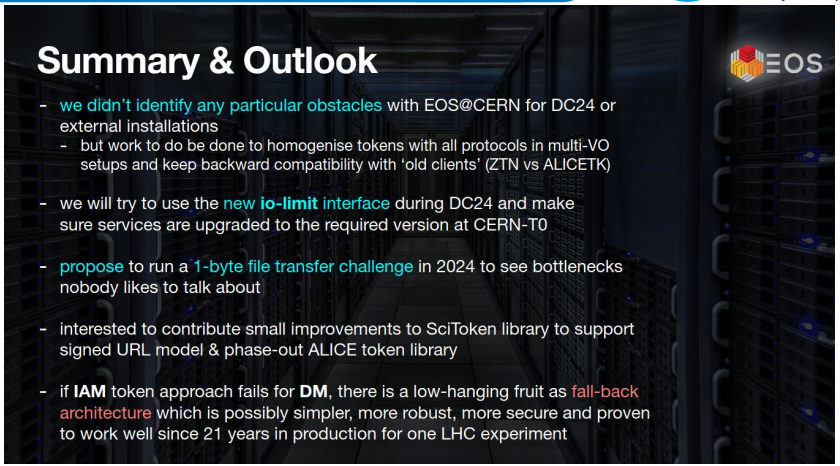
Summary

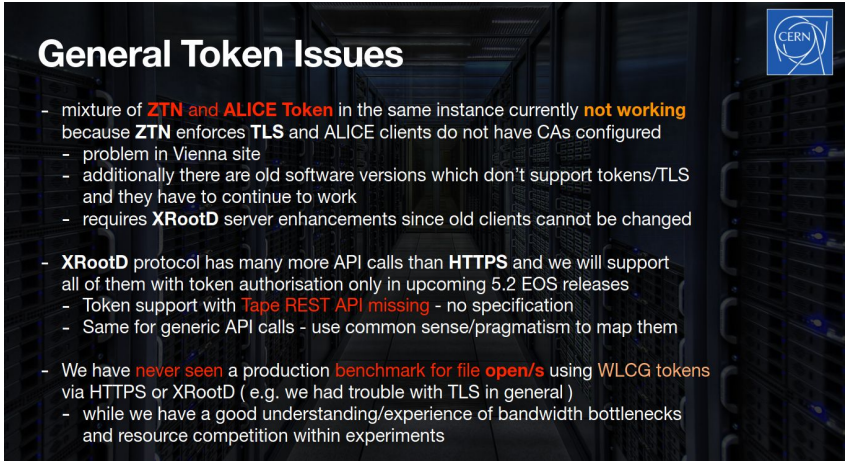
- dCache dev team sees DC24 as an opportunity for large-scale data transfer test (no tape-api tests ☹)
- Token support is available for all token capable protocols: HTTP, REST (TAPE-API), Xroot
- We recommend sites to run dCache version 9.2 to benefit from the latest developments
- We recommend sites to collect as much log information as possible for later analysis

2023-11-10 dCache for DC24 15/16

EOS

- EOS ready for DC24
- Upgrades to recent release
- One token related issue for multi-VO instances
  - ZTN and ALICE token support can presently not work together in one instance
  - Affects T2 in Vienna

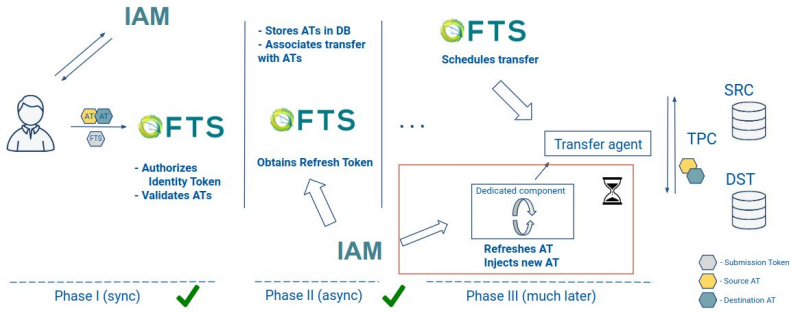




FTS & Tokens

- FTS pilot instance ready
  - First token based transfer between EOS public and CMS EOS at CERN during workshop week
- Full token lifecycle
  - To be fully integrated with Rucio for ATLAS and CMS
  - Details of token usage (lifetime, scopes audience) to be clarified
  - Load on IAM instances remains a concern

Tokens: lifecycle management



Ending remarks

- Close development between FTS and IAM teams
  - Many thanks to CNAF team and other IAM actors involved
- Prototype up-and-running on FTS3-Pilot
  - Looking for early adopters: clients, different token providers, SEs, etc...
- Looking forward to validating Rucio ↔ FTS interaction
  - ...and discovering new errors (from clients and SEs alike)
- FTS development well underway to meet DataChallenge'24

Storage & Tokens

- Reference configurations
  - ATLAS (one per major storage middleware)
    - dCache (Prague)
    - StorRM (INFN T1)
    - EOS (CERN)
  - CMS
    - Plain XRootD well advanced (12/14)
    - dCache (3/32), deployment campaign
    - StoRM (3/3)
    - EOS to start

SE token support summary

- ATLAS
  - 3 site with token support
  - GGUS campaign for sites participating in DC24
- BelleII
  - 3 sites with token support (33%)
- CMS
  - 18 out of 49 sites (37%) already support tokens
  - deploy / configure tokens on each SE
- DUNE
  - would like to use tokens, tokens supported at FNAL
- LHCb
  - would like to use tokens, 0 sites

- Other experiments
  - Interested in using tokens by DC24

- Complete Readiness for tokens is NO prerequisite for DC24

Tools & Monitoring

Monitoring

Overview for the Data Challenges

- Dashboard used for 2021 DC keeps working and can be reused/extended
- Site network monitoring campaign in progress to bring in all T1s and some T2s
- New components required for XRootD monitoring are ready
  - We will need to select some specific extra sites for testing purposes to confirm issues seen in RAL
  - Main focus will be to cover CERN (XRootD) and FNAL (dCache)
    - Other sites will be added progressively/in parallel
- Dependency on Scitags task from the "Research Network Technical Working group"
  - Until we get VO as part of the packet marking, we won't be able to assign the proper one to "multi vo sites"
  - Activity label will also be needed to tell apart "Data Challenge" related transfers as it was done with FTS

What is missing from the experiment or various project perspective!? Let's Discuss!

About 50% of sites reporting in site network monitoring

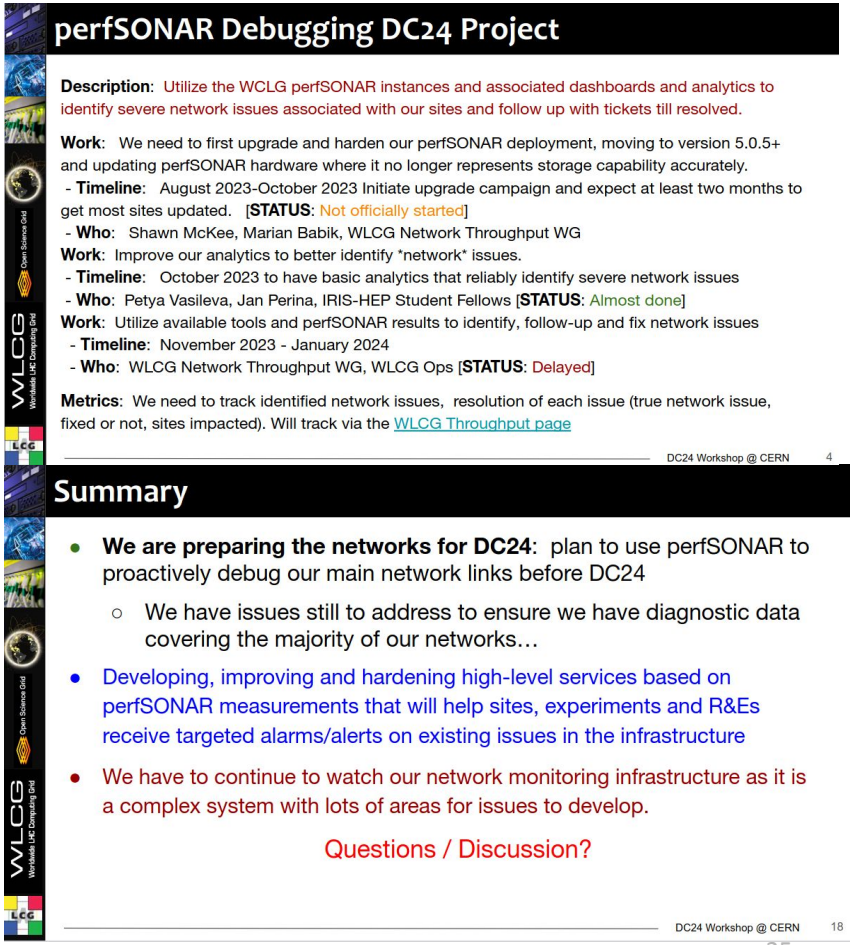
dCache 9.2 needed for new XRootD monitoring

Identification of VO traffic on multi-VO instances to be addressed

14

PerfSONAR

- Comprehensive status report
- Deployment campaign for recent version
- Try to debug network links prior to DC24



perfSONAR Debugging DC24 Project

Description: Utilize the WCLG perfSONAR instances and associated dashboards and analytics to identify severe network issues associated with our sites and follow up with tickets till resolved.

Work: We need to first upgrade and harden our perfSONAR deployment, moving to version 5.0.5+ and updating perfSONAR hardware where it no longer represents storage capability accurately.
- **Timeline**: August 2023-October 2023 Initiate upgrade campaign and expect at least two months to get most sites updated. [**STATUS**: Not officially started]
- **Who**: Shawn McKee, Marian Babik, WLCG Network Throughput WG

Work: Improve our analytics to better identify *network* issues.
- **Timeline**: October 2023 to have basic analytics that reliably identify severe network issues
- **Who**: Petya Vasileva, Jan Perina, IRIS-HEP Student Fellows [**STATUS**: Almost done]

Work: Utilize available tools and perfSONAR results to identify, follow-up and fix network issues
- **Timeline**: November 2023 - January 2024
- **Who**: WLCG Network Throughput WG, WLCG Ops [**STATUS**: Delayed]

Metrics: We need to track identified network issues, resolution of each issue (true network issue, fixed or not, sites impacted). Will track via the WLCG Throughput page

DC24 Workshop @ CERN 4

Summary

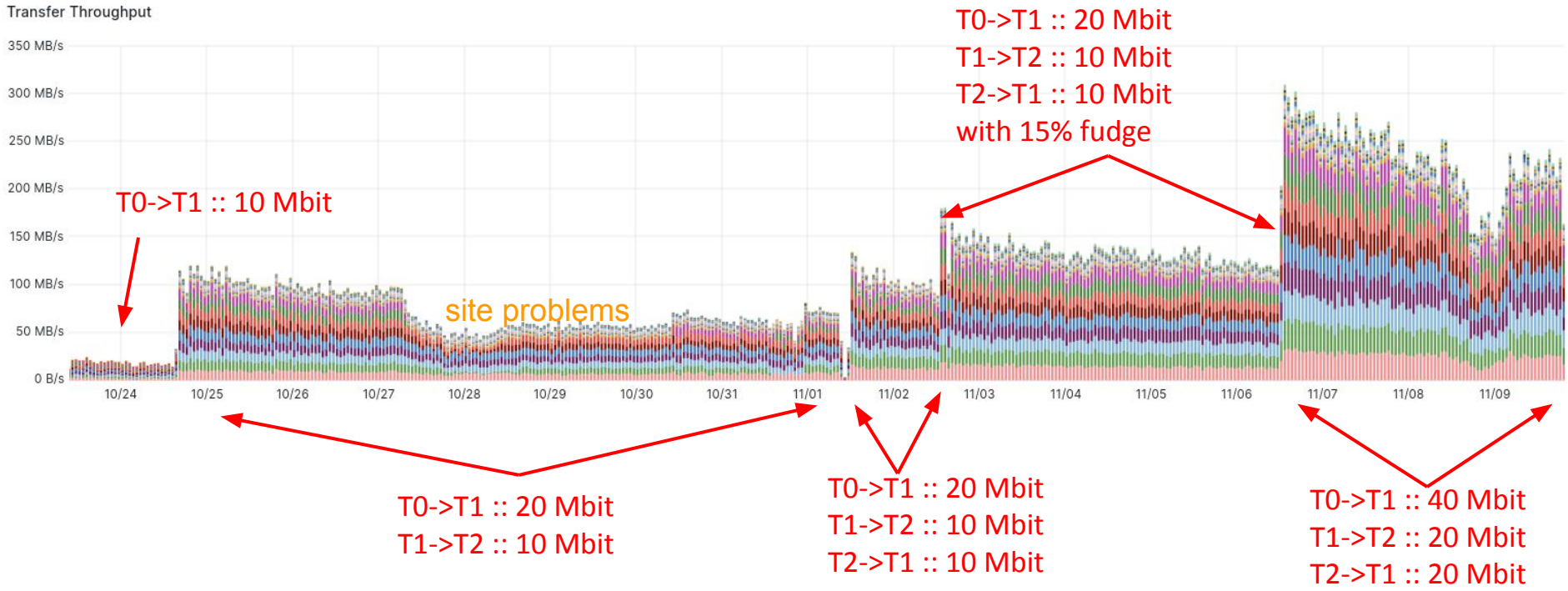
- **We are preparing the networks for DC24**: plan to use perfSONAR to proactively debug our main network links before DC24
  - We have issues still to address to ensure we have diagnostic data covering the majority of our networks…
- Developing, improving and hardening high-level services based on perfSONAR measurements that will help sites, experiments and R&Es receive targeted alarms/alerts on existing issues in the infrastructure
- We have to continue to watch our network monitoring infrastructure as it is a complex system with lots of areas for issues to develop.

Questions / Discussion?

DC24 Workshop @ CERN 18

Rucio injection tool

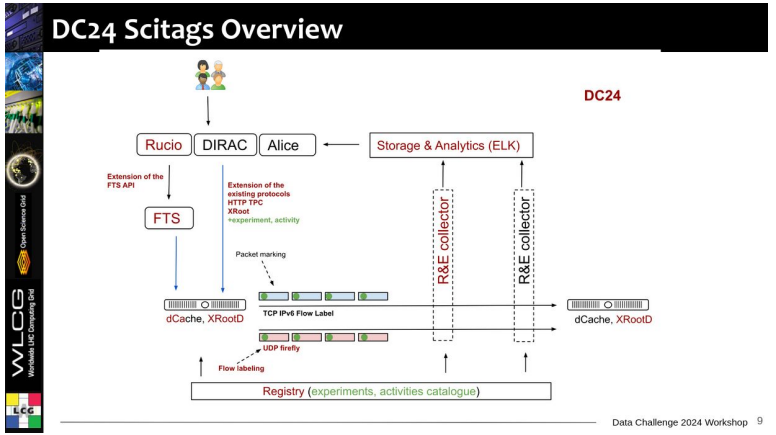
- Injecting extra transfers on top of regular experiment traffic
- Available at https://gitlab.cern.ch/atlas-adc-ddm/dc_inject
- Address experiences from DC21 (injection phases, deletion, rate attenuation, …)
- In use for pre-challenge tests by ATLAS and CMS (dedicated talk on US-CMS network tests)

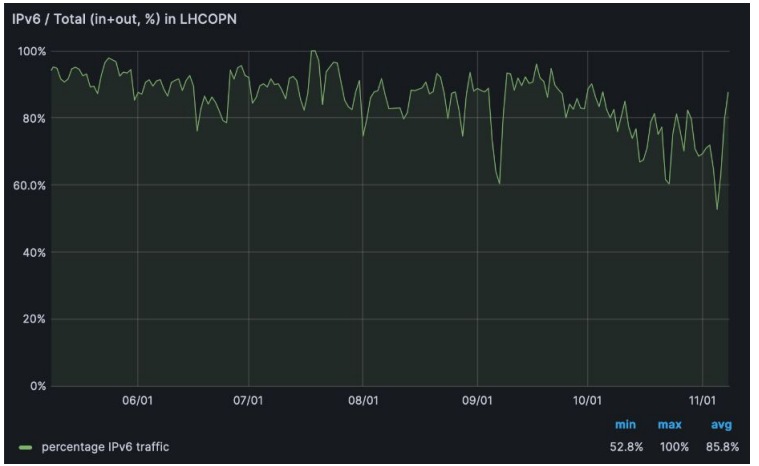


R&D Topics

Network R&D: IPv6-only, Paket Marking, Jumbo Frames, BBRv3

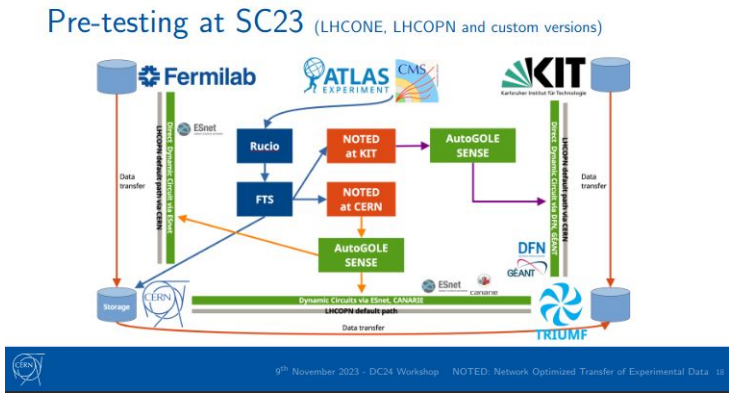
- Several talks on network R&D topics
- IPv6-only
  - Try to move LHCOPN towards IPv6-only
  - "Monitoring exercise" during DC24
  - IPv6 is prerequisite for other technologies
- Packet marking & SciTags
  - Enable where possible during DC24
- Paket Pacing
  - Via TC (Traffic Control) - Linux kernel module
  - Via TPC BBRv3 - Not yet part of WLCG Linux distros
  - Tests on dedicated links planned for DC24
- Jumbo Frames
  - Tests on dedicated links planned for DC24

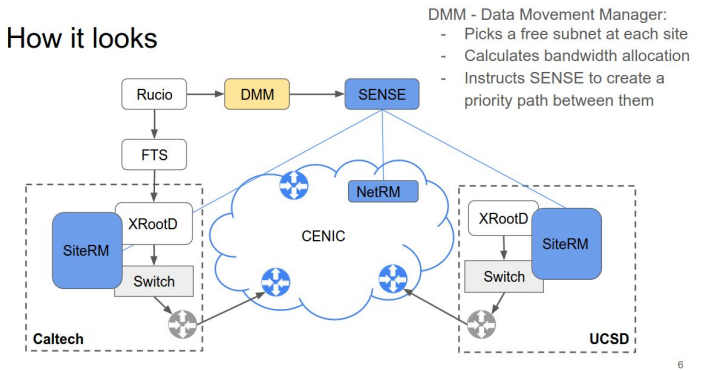




R&D Topics: NOTED and SENSE with Rucio/FTS

- Dynamic use of software defined networks (SDNs)
- Demonstrators during Super Computing 2023
- NOTED
  - Observe links in LHCOPN & LHCONE at CERN
  - Dry-run mode: no real actions during DC24
  - For heavily congested links NOTED could take SDN actions (e.g. re-routing)
- SENSE with Rucio/FTS
  - Dedicated setup in the US by US-CMS
  - Brings own FTS & special Rucio instance
  - Tests rather independent of other DC24 exercises, but run in parallel



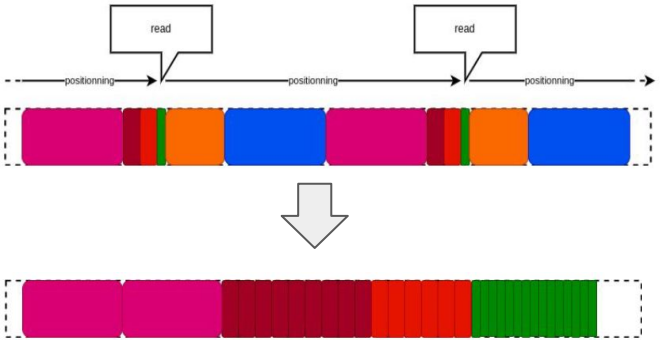


HPCs

- Contribution about HPC integration by CERN Openlab
- HPC usage will increase towards HL-LHC
- Data ingest towards HPC centers is a major challenge
  - Different tooling & protocols
  - Network access restrictions
- **No dedicated exercise planned for DC24**
- Interesting discussion about HPC access (in general) during the workshop

Archival metadata for tape

- Tape exercises not in focus of DC24
  - Topic presented during the workshop with relevant community attending
- Technical proposal developed by CTA / dCache / Rucio teams
  - Based on clearly defined scheduling and collocation hints in JSON format
  - Allow more efficient organization of files on tapes
- To be followed on a separate timeline
  - Will require experiments to define hierarchy of collocation
  - Decide on appropriate format
    - Add to DOMA-BDT, small working group?



Next steps

- DC24 Run Spreadsheet
  - Coordination of each day of running across all experiments (ongoing)
- Overleaf document
  - Initial investigations and discussions
  - Workshop outcome & ramp-up challenges
  - DC24 results
  - Aim to be published as a journal article (Springer CSBS)
- Follow-up in DOMA general meetings
  - Dates (tbc): Dec 6th, Jan 31st

Bing Image Creator: "Worldwide LHC Computing Grid, Data Challenge Workshop, Happy Mood"

The End

(for today)

Bing Image Creator: "Worldwide LHC Computing Grid, Data Challenge Workshop, Serious Mood"