WLCG Data Challenges

Sang-Un Ahn

1 November 2023 @ ATCF7

WLCG Data Challenges

- Data transfer challenge based on the current (2021) understanding of HL-LHC network needs
- Commissioning the readiness of the infrastructure in preparation for HL-LHC
  - Two use cases:
    - Export of RAW data from CERN to the T1s
    - Data Reprocessing

Use Cases (1/2)

- Export of RAW data from CERN to the T1s
  - 4.8 Tbps of network capacity from CERN to the T1s by the time of HL-LHC (~ 7M seconds of data taking per year)
    - Breakdown:
      - Flat traffic from ATLAS & CMS ~ 400 (RAW) + 100 (other) Gbps per experiment
      - Flat traffic from ALICE & LHCb ~ 100 Gbps per experiment
      - x2 for absorbing bursts (peak) of use
      - x2 over-provisioning to avoid running at full occupancy and suffering the operational problems

Table 1: network bandwidth needs per T1 (or region)

T1	%ATLAS	%CMS	% Alice	% LHCb	ATLAS+CMS Network Needs (Gbps) Minimal Scenario in 2027	Alice Network Needs (Gbps) Minimal Scenario in 2027	LHCb Network Needs (Gbps) Minimal Scenario in 2027	LHC Network Needs (Gbps) Minimal Scenario in 2027	LHC Network Needs (Gbps) Flexible Scenario in 2027
CA-TRIUMF	10	0	0	0	200	0	0	200	400
DE-KIT	12	10	21	17	450	80	70	600	1200
ES-PIC	4	5	0	4	180	0	20	200	400
FR-CCIN2P3	13	10	14	15	450	60	60	570	1140
IT-INFN-CNAF	9	15	26	24	480	110	100	690	1380
KR-KISTI-GSDC	0	0	12	0	0	50	0	50	100
NDGF	6	0	8	0	110	30	0	140	280
NL-T1	7	0	3	8	140	10	30	180	360
NRC-KI-T1	3	0	13	5	50	50	20	120	240
UK-T1-RAL	15	10	3	27	490	10	110	610	1220
RU-JINR-T1	0	10	0	0	200	0	0	200	400
US-T1-BNL	23	0	0	0	450	0	0	450	900
US-FNAL-CMS	0	40	0	0	800	0	0	800	1600
(atlantic link)					1250	0	0	1250	2500
Sum	100	100	100	100	4000	400	410	4810	9620

Use Cases (2/2)

- Data Reprocessing
  - Data at the T1 needs to be staged from tape and exported to the T2s for processing (assuming most of that processing happens outside the T1)
  - **Processing 100% of data collected in the year in less than three months (~ 7M seconds) requires the same targets for T1 to T2s traffic**

Challenges

- By the time of HL-LHC, plans to demonstrate capacity to fill ~ 50% of the full bandwidth required in the minimal scenario with production-like traffic (storage to storage, using the third-party copy protocols and data management services used in production by the experiments)
- Milestones: 10% in 2021, 30% in 2023, 60% in 2025 and 100% in 2027
  - 2021 target: commissioning capability to transfer data at a higher rate for periods of Run 3
  - 2024 target: 25% (revised)
- R&D Activities
  - Packet Marking, TCP Tuning, Token based access, HTTP/WebDAV, Tape REST API, etc.

Table 2: data challenge target rates

T1	LHC Network Needs (Gbps) Minimal Scenario in 2027	LHC Network Needs (Gbps) Flexible Scenario in 2027	Data Challenge target 2027 (Gbps)	Data Challenge target 2025 (Gbps)	Data Challenge target 2023 (Gbps)	Data Challenge target 2021 (Gbps)
CA-TRIUMF	200	400	100	60	30	10
DE-KIT	600	1200	300	180	90	30
ES-PIC	200	400	100	60	30	10
FR-CCIN2P3	570	1140	290	170	90	30
IT-INFN-CNAF	690	1380	350	210	100	30
KR-KISTI-GSDC	50	100	30	20	10	0
NDGF	140	280	70	40	20	10
NL-T1	180	360	90	50	30	10
NRC-KI-T1	120	240	60	40	20	10
UK-T1-RAL	610	1220	310	180	90	30
RU-JINR-T1	200	400	100	60	30	10
US-T1-BNL	450	900	230	140	70	20
US-FNAL-CMS	800	1600	400	240	120	40
(atlantic link)	1250	2500	630	380	190	60
Sum	4810	9620	2430	1450	730	240

DC24

- Target Data Transfer rate: 25% of HL-LHC rate (reduced from 30% to match with present LHC schedule)
  - Average over a timeframe of 48 hours, and not only as a peak rate
- ALICE, ATLAS, CMS, LHCb, DUNE, Belle II, JUNO
  - Experiments will test their data transfer frameworks and inject additional traffic on top of production in order to reach the target rates
  - Not only focusing on exporting RAW from CERN to T1s, but also staging data from T1s to T2s (storage to storage, using TPC and DT frameworks)
- Data Challenges Workshop: 9-10 Nov 2023
  - https://indico.cern.ch/event/1307338/

Timeline & Coordination

- Timeline: Feb 12 to Feb 23 2024 (for 2 weeks)
- Ramp-up challenges are on-going
  - Dedicated small-scale tests of environments and features, including the ability to rapidly deploy new version of software if necessary
  - Objective is to ensure that the different parts are in working order and ready to be used at the start of DC24
  - Use of production endpoints instead of separate test endpoints is encouraged to ensure that the experiment's wishes for infrastructure challenge, not operator challenge, are met
- WLCG DOMA General Meetings
  - Last Wednesday of each month
  - Asynchronous discussion, information sharing through minutes, mailing, ongoing technical documents

Technical Content

- Token-based authentication and authorization for disk endpoints
- Functionalities and improvements provided by the SDNs
  - AutoGOLE/SENSE, NOTED and ALTO/TCN - readiness of links and routers required
- HTTP/WebDAV protocols
- New Tape REST/API
- Under discussion on including dataflows for sharing infrastructure, such as AF

Monitoring

- Single (Central) monitoring -
- XRootD/EOS monitoring in preparation
- Monitoring on LAN traffic from sites are encouraged to be published to the central monitoring
  - WLCG site network monitoring:
    - https://monit-grafana.cern.ch/d/MwuxgogIk/wlcg-site-network

ALICE

- Use cases
  - Writing to custodial storage
    - CTF (RAW) from CERN EOS to CERN CTA and various custodial backends at T1s
    - Custodial storage of analysis containers (AODs): low-rate continuous streams from all WLCG sites to T0 and T1s depending on data location
  - Reading from custodial storage
    - CTF re-processing campaigns that requires custodial access will be done during LS3
- During DC24,
  - Active transfers of data accumulated from 2023 Pb-Pb period + some additional data injection to keep the rates stable

ATLAS

- Test coverage
  - T0 Disk to T1 Tape (only to T1 tape buffers, no actual migration to tape)
  - T1 Tape to T2 Disk (fake with T1 disk origins)
- Data movement mechanism
  - Selection of unique data (data selected not presented at destination site)
  - Centralised rule injection per origin at the required maximum rate with short lifetime (~ 24h)
  - Rucio & FTS schedules data movement queues with a low-priority, no impact on experiment's ongoing activity
  - Short lifetime allows quick deletion and repeated use of unique data

CMS

- Similar to other experiments
  - Test the network and disk usage
  - Use of Rucio & FTS + monitoring
  - Token-based AuthN/AuthZ for disk endpoints
  - Participating T1 sites (PIC, KIT, IN2P3, RAL, FNAL, CNAF)
  - Some T2 sites (in discussion)
- Small scale tests are planned for DC24 (not much in details)

LHCb

- DC24 will be in parallel with the usual business, data processing activities + possible reprocessing
- Common monitoring with FTS in the context of Data Challenge
- Possible token access to some SEs
- New T1s: NCBJ and IHEP

Belle II

- Technologies to be tested: DDM, FTS, storage, network, monitoring system, protocols, IAM
- Emulate data transfer conditions in a Belle II HL scenario with the global network highly utilized concurrently by LHC and other experiments (LHCOPN/LHCONE)
  - Current estimation ~ 40TB per day
  - Transfers from KEK to RAW centres with Belle II distribution schema
    - BNL(30%), CNAF(20%), CC-IN2P3(15%), UVic(15%), DESY(10%), KIT(10%)
- Note: DC24 will likely overlaps with Belle II data taking

DUNE

- RSE (Rucio Storage Element) to RSE with token auth
- Compute Node to/from nearby RSE
  - JustIN workflow system triggers data transfers between nearby RSE's and storage-less compute sites
- SURF to FNAL onward
  - DUNE detector data will be transferred to FNAL for archival storage and half of RAW will be sent onward to European datacenters for processing
  - All processed data will be sent back to FNAL for archival storage
  - protoDUNE data challenge: >3.5GB/s CERN → FNAL, >1GB/s SURF → FNAL
- Overlapping with the peak LHC experiments transfer tests

JUNO

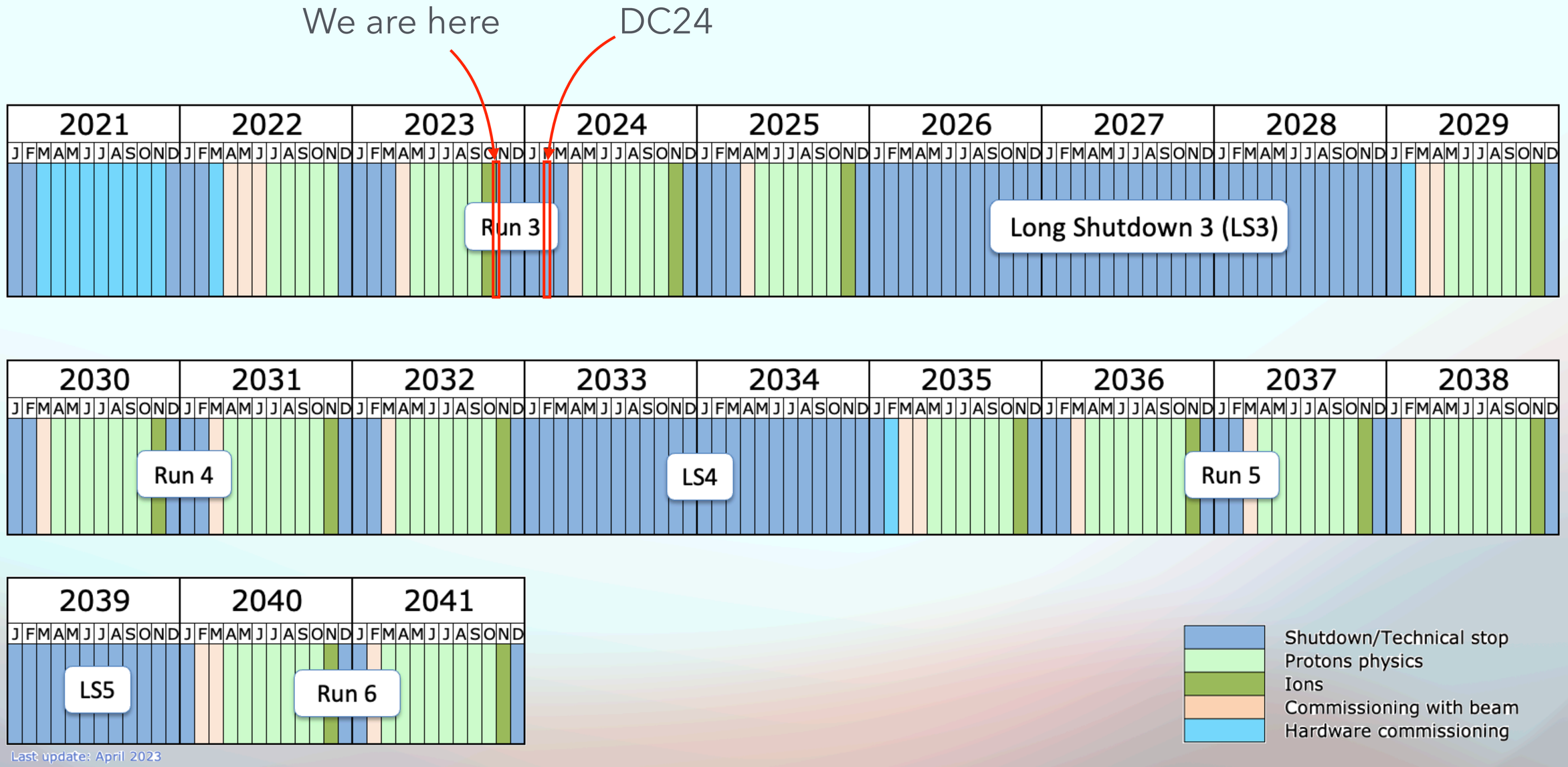
- Test data management services and protocols (DIRAC, FTS, monitoring, IAM, HTTP-TPC)
- RAW replication target shares of three sites: CNAF (100%), JINR (100%), CC-IN2P3 (1/3)
  - Target rate is 4.5 Gbps (1.5 Gbps per site)
- Small-scale challenges
  - Data management system (DIRAC+FTS+HTTP-TPC), TCP tuning
  - perfSONAR, monitoring dashboard
  - Bandwidth evaluation from IHEP to three sites with upgraded links (10 Gbps → 100 Gbps)
  - Verify token support for SEs
- 48 hours of overlapping with the peak LHC experiments transfer tests during DC24

ATCF Perspectives for DC24

- WLCG encourages the participation from T2s for this challenge
  - T1 to T2s transfers with the same target rates of T0-T1s
  - Disk-to-disk transfer with token auth
- DC24 inputs (to be defined) to T2s will be discussed in WLCG DOMA
- We need to identify a person from contributing institutes, particularly playing a major role in the challenge
  - Who participates in Asia and which part of challenges?

References

- DOMA DC24 Document and its references
  - https://indico.cern.ch/event/1225415/contributions/5155042/attachments/2593516/4476291/Data Challenge 2024.pdf
- WLCG DOMA General Meetings
  - https://indico.cern.ch/category/10360/
- DC24 Experiments updates - WLCG DOMA General Meeting (29 March 2023)
  - https://indico.cern.ch/event/1258343/



http://lhc-commissioning.web.cern.ch/schedule/LHC-long-term.htm

Questions?