DC27 and (Mini) Data Challenge Planning

Shawn McKee / University of Michigan

USATLAS F2F at HTC26

(https://agenda.hep.wisc.edu/event/2432/contributions/38165/)

June 10, 2026

Overview

We need to discuss and plan for DC27 and our Data Challenge activities between now and then.

- Lots of details and context in previous presentations
  - "Network Planning for Sites", #77 S&C (2024)
  - "ATLAS ADC DDM R2R4 Discussion", #81 S&C (2025)
  - "Transforming ATLAS Sites", Ljubljana #82 S&C week (2025)
  - "Updates from perfSONAR, SciTags and the RNTWG", from HEPiX IPv6 WG meeting (2026)
  - "Data Challenges: Status and Update", ATLAS S&C (2026)

With only 15 minutes, I don't want to rehash context and motivations…

Improving Our Production Infrastructure

Our (Mini) Data Challenges are intentionally run with **production infrastructure**, not special demo hardware.

- The goal is to incrementally improve our production infrastructure, eventually reaching HL-LHC requirements.

Significant capability improvements identified during challenges are intended to move into production.

Bottlenecks in hardware, software and architectures are supposed to be **identified** and **fixed**.

Capabilities that are not improvements should be **deprecated**.

Preparing Technologies and Capabilities

HL-LHC will require more resources than we can currently afford.

- To address this, the experiments are working hard to optimize workflows
- New technologies and capabilities will play a critical role in bridging the gap

The **WLCG data challenges** are designed to regularly test where we are relative to where we need to be for HL-LHC.

Possible technologies to test and, if beneficial, integrate

- **New / improved storage servers** (Gen5/6 PCIe, NVMe, new NICs, etc)
  - Define/document LHC server best practice for hardware and configuration (host tuning)
- **SciTags** (traffic identification anywhere in the network)
- **Traffic optimization** (via Jumbo Frames, pacing, new protocols)
- **Network Orchestration** (SENSE/Rucio, NOTED, GNA-g efforts, etc)
- Improvements (or alternatives) to **WebDAV and Xrootd protocols**
- Improvements to **storage elements** (dCache, Xrootd, STORM, EOS, etc)
- Evolution of **Distributed Data Management** (Rucio, FTS, etc)

Recent Challenges

USATLAS has participated in a number of mini-challenges during the last year (see WLCG DOMA Twiki)

- January 2025 we tested jumbo frames and benchmarked capacity
- August 2025: mini-capacity challenge in the US
  - This was another benchmark in our progress towards DC27
- September 2025: shutdown of IPv4 on LHCOPN for BNL
  - No issues observed. IPv4 traffic shifted to LHCONE.
- In 2026: IPv4 blackout, SENSE/Rucio testing and Host Tuning testing, NET2 -> Prague 400 Gbps testing

Goals for DC27

For DC27, we have some milestones we are targeting:

- All sites should be moving the majority of their data via **IPv6**
- We should have a few **IPv6-only** sites for each experiment
- At least 80% of the traffic should be **identified via SciTags**
- At least 50% of the traffic should be using **jumbo frames**
- **Rucio/SENSE** to be used by few Production sites
- Sites should be able to easily utilize **90% of their declared WAN bandwidth** for an extended period (many hours to days)
- **Network traffic monitoring** should be able to track throughput by network type (LHCOPN, LHCONE, Research & Education, Commodity)
- Site perfSONARs updated (hardware & software) and reliably operating
- At last one example of exploiting ESnet's HighTouch data
- Incorporate Tape as part of the challenge (tape was NOT part of DC24)

All of these areas could benefit from regular testing via mini-challenges

Host-Tuning Evaluation

Quick summary: we wanted to test ESnet Fasterdata recommendations on our production storage hosts.

Goal: Improve the throughput and/or reduce the load per transfer, decreasing time for transfers & improving net use.

As part of the ongoing perfSONAR upgrade campaign, we created a fasterdata-tuning.sh (bash) script to implement the recommendations from the Fasterdata web pages.

The WLCG mini-capability challenge for host-tuning was done Feb 2-3 and involved USATLAS and USCMS sites.

Host-Tuning Challenge Outcome

Summary: we are still working on interpretation.

- We set up to gather a significant amount of monitoring: CheckMK plots, Ganglia plots, ESnet Stardust and various logs
- Problems with conflicting production (merge) jobs at MWT2 made tests on 2nd day fail
- Re-ran two days later (just tuned config) showed best transfers we have seen (MWT2->AGLT2)
  - Other direction exposed MSU outbound issue being investigated.

USCMS saw significant benefit in tuning their **perfSONARs**

- Mid-to-low 10's of Gbps on a 100 Gpbs link went to 98/99
- Less clear on the impact for the full end-to-end system tests

DC27 Site Checklist

- ALL systems are dual-stacked
- SciTags is enabled and configured on storage
- Jumbo frames are in use on storage hosts
- perfSONAR is upgraded (hardware and software)
- Systems are tuned for optimal distributed computing performance (host-tuning + switch and service configs)
- Site network monitoring is capturing all WAN data flows
  - Need to understand how to differentiate traffic by type: LHCOPN, LHCONE, R&E and commodity

Near To Mid-Term Plans

The **host-tuning** will need multiple iterations before DC27

- If the things we tuned are NOT the bottleneck, no improvement
- We need a focused team testing at more than one site and should schedule a new host-tuning test by September
- Tuning clearly helps perfSONAR but storage needs further study

In general, continued mini-challenges for both **capacity** and **capability**.

Not detailed in this talk but very important: getting more advocates to drive capability challenges for ATLAS:

- ATLAS file sizes (increase x10 to understand cost-benefit)
- Broader deployment of Network optimizations (Jumbo frames, pacing, new protocols)
- SDN (SENSE/Rucio), assuming testing shows benefits (Next talks).

Summary

Our facility has a lot of work to do in implementing capabilities and testing our setups.

Ideally, the testing we do (and have done) **should inform our purchasing** by identifying the most beneficial configurations/hardware.

We need a multi-site team working on mini-challenges and especially the host-tuning.

Question, Comments, Discussion?

Acknowledgements

We would like to thank the **WLCG**, **HEPiX**, **perfSONAR, USLHC** and **OSG** organizations for their work on the topics presented.

In addition we want to explicitly acknowledge the support of the **National Science Foundation** which supported this work via:

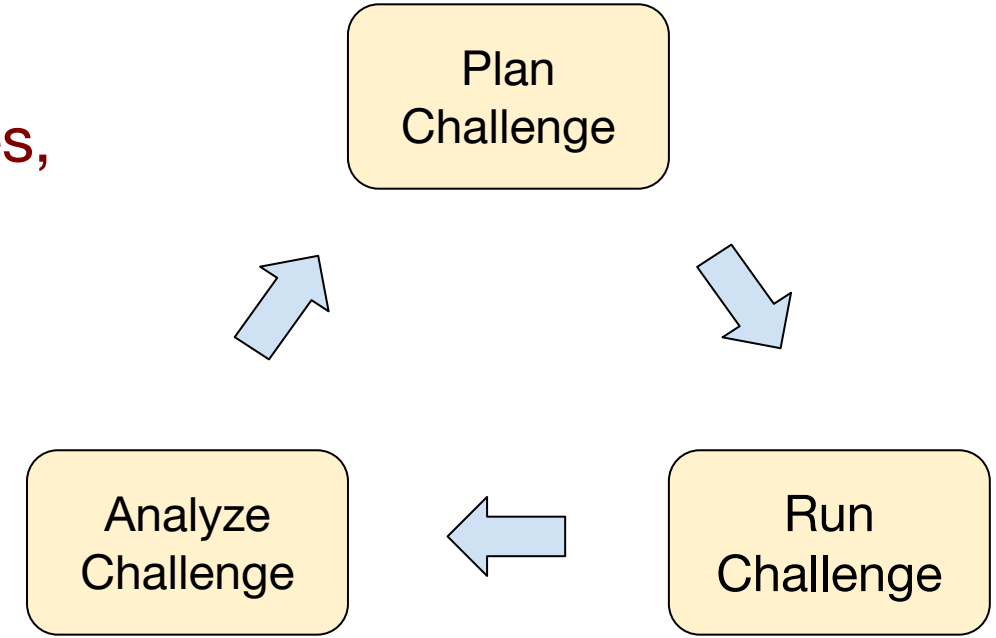
- IRIS-HEP: NSF OAC-1836650 and PHY-2323298

Backup Slides

The Challenge Loop

WLCG has chosen to run a series of Data Challenges to regularly benchmark our sites, software and infrastructure and monitor our progress towards meeting HL-LHC requirements.

The full challenges are too infrequent, thus **Mini challenges** are added



Planning Challenges

WLCG Challenges are major undertakings and planned years in advance: **DC21**, **DC24**, **DC27** and (likely) **DC29** involve significant effort across many experiments

Mini-challenges are intended to be **lighter weight** and should be doable by just a few people or even just a single site.

We run two flavors of mini-challenges: **capacity** & **capability.**

- **Capacity:** tests throughput IN/OUT to identify bottlenecks
- **Capability**: evaluates software, tools, tunings or architectures that may improve our ability to use our cyberinfrastructure.

Will We Transform Our Sites?

In executing the various (mini) data challenges to-date, we have assumed sites will naturally incorporate improvements.

I am not sure that is that case…

WLCG sites have some specific MOU requirements.

Sites supporting ATLAS have a few more requirements.

New beneficial capabilities or identified bottlenecks won't necessarily be broadly deployed or always fixed.

ATLAS Site Requirements

We build sites to meet the MOU requirements from WLCG:
CPU(HS23) and **Storage(TB)** (not network, IOPS, throughput)

- No real performance requirements give broad range of values:
  - Storage quality variation: SMR, SATA, SAS, SSD or NVMe
  - CPU: even the same CPU can give very different result depending upon host memory, disk, bus, etc.
- **Good**: Gives sites flexibility to meet a few basic requirements. Abstraction (SE or CE) allows lots of choices for how needed services are delivered
- **Bad**: Large variations in site performance and how well the services they deliver behave.

My worry: sites optimize cost at the expense of being a good distributed computing component for ATLAS.

Transforming our Sites

The data challenges provide us with an opportunity to evaluate our existing hardware, software and architecture to identify bottlenecks, limitations and misconfigurations.

Given that HL-LHC is ~5 years away, now is the perfect time to re-evaluate our site's hardware configuration and architecture so that we can have a suitable baseline ready for HL-LHC requirements.

- **Five years of hardware purchases can fully replace our current hardware**
- Incrementally transforming sites should allow a smooth transition in capability

It is **critical** that sites understand how they fit into our globally distributed infrastructure so they can meet the HL-LHC requirements and use-cases.

- Mini-challenges are a great opportunity to understand our current capabilities, identify bottlenecks and prototype new technologies.

What Have We Learned from Data Challenges?

The two WLCG Network Data Challenges we have run so far have been very helpful for the experiments, sites, R&E networks and technology innovators.

- **Sites** were able to identify bottlenecks that were not obvious before
- **R&E networks** were able to gain understanding about how the various participants data flows might interact with each other across their topology.
- The **WLCG experiments** and partners were also able to identify where bottlenecks in software, services and architecture exist
- **Technology proponents** were able to do “at-scale” testing to inform software and service evolution

The most recent challenge was a US **capacity** mini-challenge: presentation

Methodology Criticisms

The methodology for our WLCG data challenges can be criticized in a number of ways:

- The “production” equipment now is NOT the production equipment in the HL-LHC era, causing us to potentially evaluate and optimize for the wrong hardware and capabilities.
- The ATLAS operational mode for HL-LHC is not clearly defined…are we incrementally optimizing and testing something totally different than what we will end up with?
- Using the extremely fast impact of AI/ML on science as an example, the iterative, incremental evolution of our facilities, software and architectures may miss out on new thing X.

I think these (and similar criticisms) deserve discussion. What are the alternatives and associated downsides?

Last Note on Monitoring

One very important aspect of all the various challenges we have been undertaking is the critical role monitoring plays.

If you can't monitor it it…you can't manage, debug or understand it

We need to continually verify and validate our monitoring

- It is often not accurate in what is being measured.
- It can be brittle and components fail or disappear over time.
- There may not be existing monitoring for critical items necessary to understand the systems be evaluated.
- New capabilities often require new associated monitoring to be developed.

Our Data Challenges and mini-challenges have improved our monitoring a LOT!