

- Nominate a leader for the discussion, and a note taker. Nominate someone to present the answers to the questions at this afternoon's plenary. Please mix career levels!
- Fill in the questions (see next session)
 - These are meant as prompts - if they aren't relevant, or different ones are relevant, please feel free to adjust.
- The point is not to answer the questions (unless they are crazy easy to answer), but to fully understand the questions
 - Right question?
 - What viewpoints does it need to be considered from (e..g facilities, end-user analysis, network, etc.)

Proposed Questions For Each Breakout Group

Slide Deck to fill in answers for afternoon plenary:

▣ Coordinated Ecosystem for HL-LHC Computing R&D

To be presented at the Plenary Tuesday afternoon.

1. Topic Description
2. Summary of the Discussion
3. What is the current state?
 - a. E.g. Current activities and projects
 - b. Where are the activities and projects in their lifecycle?
4. What are the funded projects in this area (nationally? Internationally?)
 - a. And how should they coordinate?
5. What are possible blueprints and existing forums for discussion?
 - a. And expected outcomes from these discussions
6. What should this topic look like in 3 years?
 - i. What near and medium term actions are needed?
7. How do we know whatever solutions we have will work once deployed?

Parallel 1: Network & Storage System R&D

Topic Description:

- High Performance Storage System to meet our Network Requirements
 - Scalable storage system for efficient data management....
- Data Challenge to include managed networks - with benchmarking the entire stack (3+ year view)

Framing of discussion:

What should the community target for activities during DC23, including integration of both Storage and Networking R&D activities?

Summary of the Discussion:

What exists today?

Action items from Discussion:

1. We need a requirements spec for the total IO that a storage system at each T1/T2 in the USA has to provide.
 - a. Pay attention to whether a differentiation of service between different T2s warrant a differentiation of requirements.
 - b. Pay attention to monitoring requirements, especially if the requirements are hard to meet.
2. We need to have a process for going from where we are to where the specs say we have to be.
 - a. This includes benchmarking and metrics that allow us to assess when we are on track.
 - b. We should utilize both the DCxx set of data challenges, as well as smaller dedicated challenges inbetween that benchmark to meet metrics for parts of what's needed in the DCxx progression.
3. Engage with facilities program in DOE via both the IRI and the post-exascale software sustainability program.
4. Engage with LCCF in NSF in analogy to IRI in DOE.

What is the current state?

Networking:

LAN at T1 and T2 is 100G today.

2023 ESnet and CENIC are routing at 400G

=> by DC23 (in early 2024) there is going to be a path of 1.5Tb across Atlantic, out of which 500G reaches CERN (? check this ?) [[ESNet presentation](#)]

- For shared resources (networking, site-level storage), we need to actively include non-LHC experiments in the tests.

Storage:

- It appears no one has done planning / coordination for total throughput for storage. We have some base estimates (e.g., support the WAN transfers) but LAN is unclear.
 - Need a process to plan storage capabilities for our sites – covering holes identified today.
- We do not have a computing model for analysis facilities. Particularly, there are open questions around what IOPS are needed as opposed to raw throughput.

Monitoring and usage accounting. What's required here?

What is needed from the network to support inference workflows?

What are the funded projects in this area (nationally? Internationally?)

- Rucio/SENSE collaboration. Particularly, establish capability with USCMS as an exemplar and then potentially including USATLAS.
 - CMS contribute to DC23 with a dedicated Rucio instance for Rucio/SENSE integration. I.e. not integrated with global Rucio instance for CMS.
 - Engage ATLAS ops program after DC23 on Rucio/SENSE integration. Until then, Shawn on networking and Rucio team on data management are informed about what's going on on a regular basis, but no ATLAS sites involved in testbed activities.
- Packet marking (includes IRIS-HEP, USATLAS Ops, CERN): Technology is more mature. Not deployed yet but should be targeted for 2024.
 - ESN6 is expected to participate here.
- Capabilities inside the network (FAB, FABRIC): Deployment of ServiceX inside FABRIC at CERN to deliver columns across the transatlantic link.
- Storage specialization based on workflow needs (BNL LDRD). Hierarchical storage between tiers.
- CREST department at FNAL is overhauling the storage strategy. Will need inputs from experiments and coordination with BNL.
- IRIS-HEP: Demonstrate site-to-site WAN transfers at 500Gbps to tune individual technology stacks (e.g., XRootD + FTS + Rucio).
- ASCR/ECP: ADIOS, next-gen HDF5. Mostly focused on access to the storage system at the local site from the compute.
 - Possibly other ASCR investments we need to discover? (ASCR Software Sustainability Program for software ecosystem supporting high-speed I/O and data analytics)
 - Integrate directly with facilities.
- CERN/FNAL: next-gen ROOT files: RNTuples including objects stores backend: DOAS and soon S3.
- FNAL Facilities: Change from Enstore to CTA.
- Computational Storage, VAST, Samsung
- Recoverable lossy compression ROOT (name of company)

What are possible blueprints and existing forums for discussion?

- Site-level design and requirements workshop.
- Engage DOE facilities with the IRI process.
- US-DC23 goals?

What should this topic look like in 3 years?

- Big checkpoint is meeting WLCG Data Challenge goals (DC23).
- Rucio/SENSE: Initial integration during DC23 with USCMS. Present what's accomplished and then decide on goals for DC25.
- Packet marking.

How do we know whatever solutions we have will work once deployed?

