WLCG Network challenges in preparation of HL-LHC

Notes from a meeting with Simone, Shawn, Frank, Edoardo, Bernd

LHCONE virtual meeting #45
16th of September 2020

Data production in HL-LHC

CMS will produce 364PB of raw data per year, running for 100 days/year

- **The traffic from CERN to all the CMS T1s will be ~42 GB/s for 100days per year**, as we want to export in quasi-real time.

- E.g. FNAL (40% of resources for CMS only) will import at ~150 Gbps of RAW data

- ATLAS has similar numbers

- Considering all the experiments and allowing some headroom, **we would need >1Tbps across the Atlantic by 2027**

Numbers based on the document Use Case #13 of the ESNET planning review

Needed capacity

By the time of HL-LHC, we will need **to demonstrate our capability to transfer ~1Tbps across the Atlantic,** allowing some contingency.

Note that today there is headroom between the transatlantic network needs and its capacity. This headroom will shrink considerably for HL-LHC as the transatlantic capacity will not increase by a factor 50 as the data volume.

Data challenges

Data challenges could consist in demonstrating the capability to transfer an increasing volume of data over the Atlantic, **to reach the target of 1Tbps sustained traffic in 2027**. For example, 150Gbps in 2021, 350Gbps in 2023, 600 Gbps in 2025 and 1Tbps in 2027.

The same should be done between CERN and a subset (or all) of the T1s in Europe, scaled by the size of the T1 with respect to BNL and FNAL.

T1-T2: Reprocessing at T2s

The 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).

The target is to be able to reprocess 100% of the data collected in the year and stored at a specific T1 in less than three months.

For FNAL, this means recalling 2PB/day, so 200Gbps from tape.

The data could be streamed directly to the processing centers or buffered at the T1 and transferred in a burst. This has different implications on the storage needs at T1s and T2s, the balance with CPUs and the network needs.

Reprocessing at T2s – cont.

We assume the T1 will need to commission its capability to stream 1Tbps to the T2s to account for the burst scenario.

The 1Tbps T1 egress capacity is the target for 2027 for a 40% T1 serving only one experiment. The targets for the other T1s can be derived from there.

Also for T1-T2 challenges, intermediate targets should be defined along the same lines as for the previous use case and challenged at the time of the experiment's reprocessing and derivation campaigns (e.g. through the data carousel) in Run-3.

Reprocessing at HPCs

HPC will also be used.

The use case where an HPC would provide an allocation of 5k nodes (128 cores each) for many days capable to process 10kHz of events, implies demonstrating the capability to stream the same 1Tbps of data into a HPC

Intermediate targets should be defined also for this case

R&D on network capabilities - Tagging

The data challenges will happen in parallel to the production activity and in fact will complement it. We need the capability to **mark the traffic for different activities by tagging the packets** (See RNTWG Packet Marking) with the source experiment and application purpose.

We should focus on the scheduled traffic (asynchronous, storage to storage, via FTS) as that will be the bulk of the network utilisation. The focus should be instrumenting the T1s and largest T2s which participate to the challenges. This will be the foundation to understand the possible future needs for traffic shaping and orchestration.

R&D on network capabilities - Shaping

Traffic shaping to is the second achievable target and traffic orchestration is the longer term one.

The HPC use case for example could be one that would benefit largely from traffic shaping and the R&D could look in that direction for the first demonstrator.

R&D on network provisioning

The challenges should follow but also drive the expansion of the network capacity. At the same time, we will need to acquire access to extra network capacity to proceed with the challenges and the R&Ds on dynamic provision of additional bandwidth

The third subgroup of the RNTWG is focused on
Network Orchestration

Other projects like SENSE, NOTED, DTNs, AUTOGOLE… could enable an efficient use of bandwidth or deliver extra bandwidth when more is needed

Bandwidth for Data Challenges

Data challenges could be run using extra bandwidth where already available. For example by pairing LHCOPN links with LHCONE. Or complementing the ESnet capacity over the Atlantic with the one offered by GEANT and others.

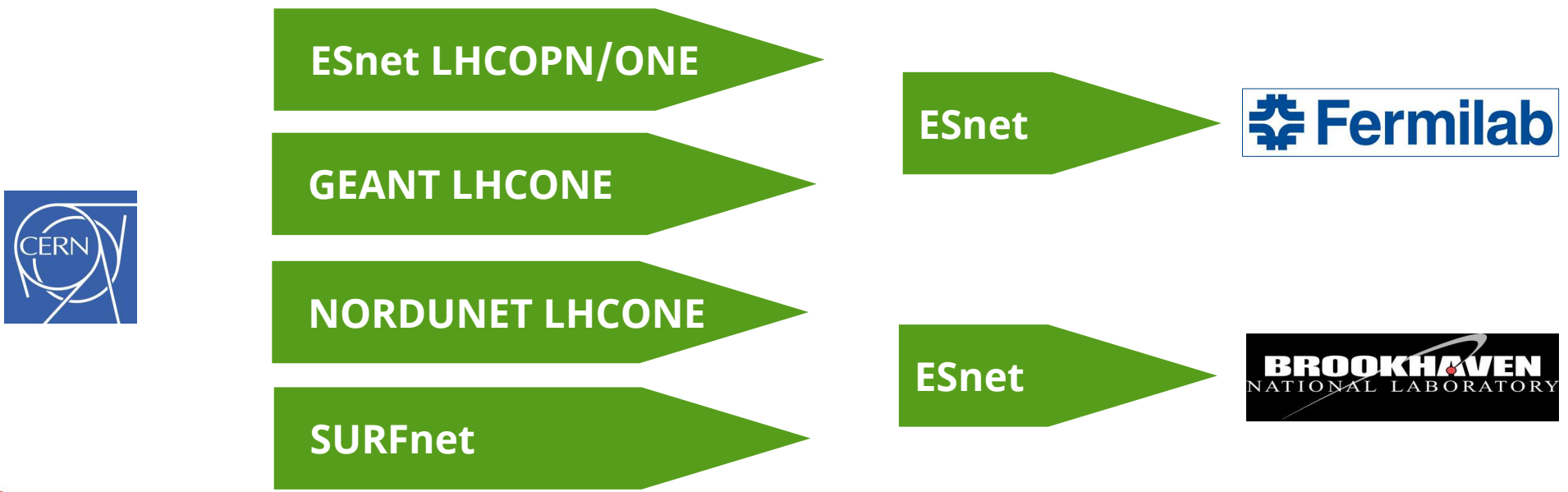
We could also lease network lines of given capacity between T0 and T1s. An example is the CERN-Amsterdam 400Gbps line and we could think about others, discussing how to share the cost of the lease.

At the same time, we should explore the possibility of leasing a network connection offered by large cloud providers and compare the cost with the previous cases. This case would require large bandwidth connections to a Cloud providers on both ends

Transatlantic challenge: T0 to US T1s

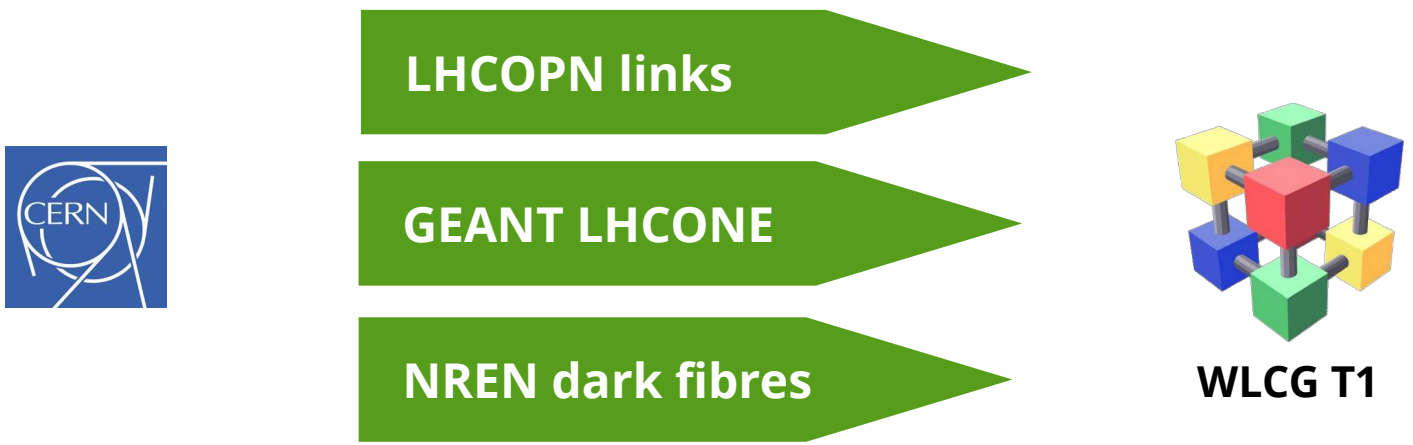
Could be done aggregating several transatlantic links of the different LHCONE providers (7-800Gbps today?)

But what bandwidth can be delivered to the T1s?



European challenge: T0 to T1s

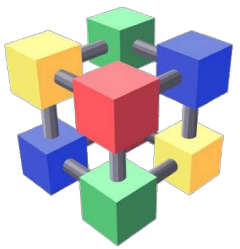
Also for Tier1s in Europe several links could be aggregated, leveraging the large availability of dark-fibres



E.g. soon available 500Gbps between CERN and NL-T1

HPC challenges

Connect HPC centres to R&E networks at large bandwidth



WLCG T1

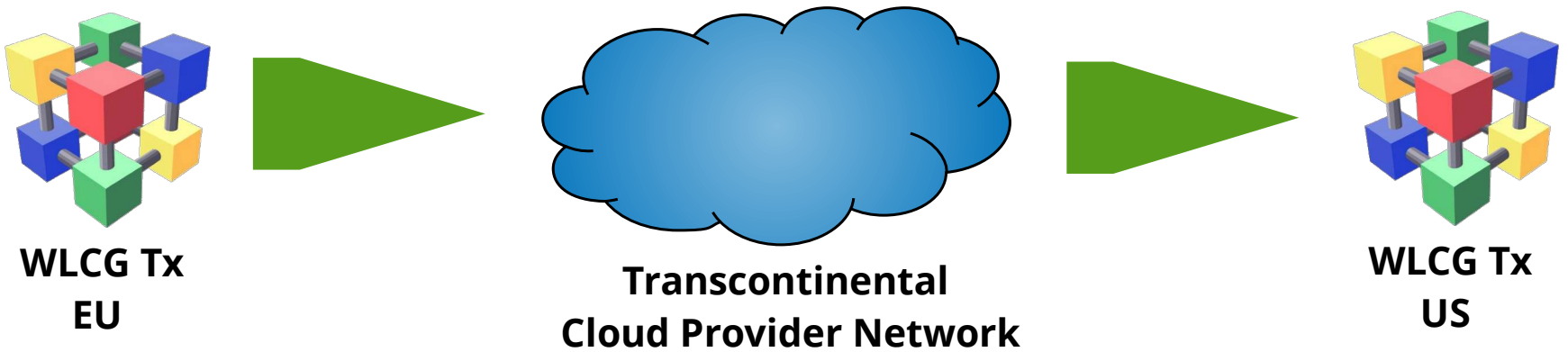
LHCONE

NRENs connectivity

HPC

Bandwidth from cloud providers

Explore the possibility to use the transcontinental networks of large Cloud providers



Questions or comments?

lhcone-architecture@cern.ch