USATLAS mini-Data challenge

Hironori Ito, Brookhaven National Laboratory

Shawn McKee, University of Michigan

ATLAS TIM at Stony Brook University, 2025/01/23

@BrookhavenLab

Fall 2024 US Mini Challenges

As previewed in the Nov 13, 2024 [WLCG DOMA](), both USATLAS (and USCMS) undertook some capacity mini-challenges, designed to benchmark our current infrastructure.

These were simple load-tests where we wanted to evaluate the capacity limits for our various sites.

We were not trying to identify where we might adversely interact with other activities, as we do when we run the regular data challenges.

# Original plan: USATLAS Mini Data Challenge Fall 2024

- ## Use load test1 from T1 to each T2s at full T2's network capacity
  - To check if there are any changes from the results from the last test.
  - Network capabilities of US T2s: AGLT2(200 Gbps) MWT2 (200 Gbps), NET2 (expected to be 400 Gbps). SWT2 (100 Gbps)
  - Individually as well as simultaneously
    - Simultaneous test might present "choke" point in the path.
- ## T2s to T1 at full wan disk capacity.
  - Not capable to reach the full network capability of BNL at 1.6 Tbps due to the storage layout of T1 storage
- ## T1 Tape staging and readout test.
  - Check the staging throughput and readout throughput of staged data from BNL.
- ## Check and validate the accuracy of the various monitor at the site as well as the central ones at CERN, ESNet, BNL,...

[1]The program is found at the following BNLBox folder https://bnlbox.sdcc.bnl.gov/index.php/s/XGs6LJEGNzf69zK

Brookhaven National Laboratory

AGLT2 Ingestion Testing

- The observed throughput for injecting AGLT2 was about 150Gbps.
- Various monitors were checked against each other to evaluate their accuracies.
- Although all monitor shows the similar number, CERN Dashboard seems a bit higher the other two? However we must note that CERN Site Network is ALL traffic

BNL FTS Monitor

CERN Dashboard

ESNet
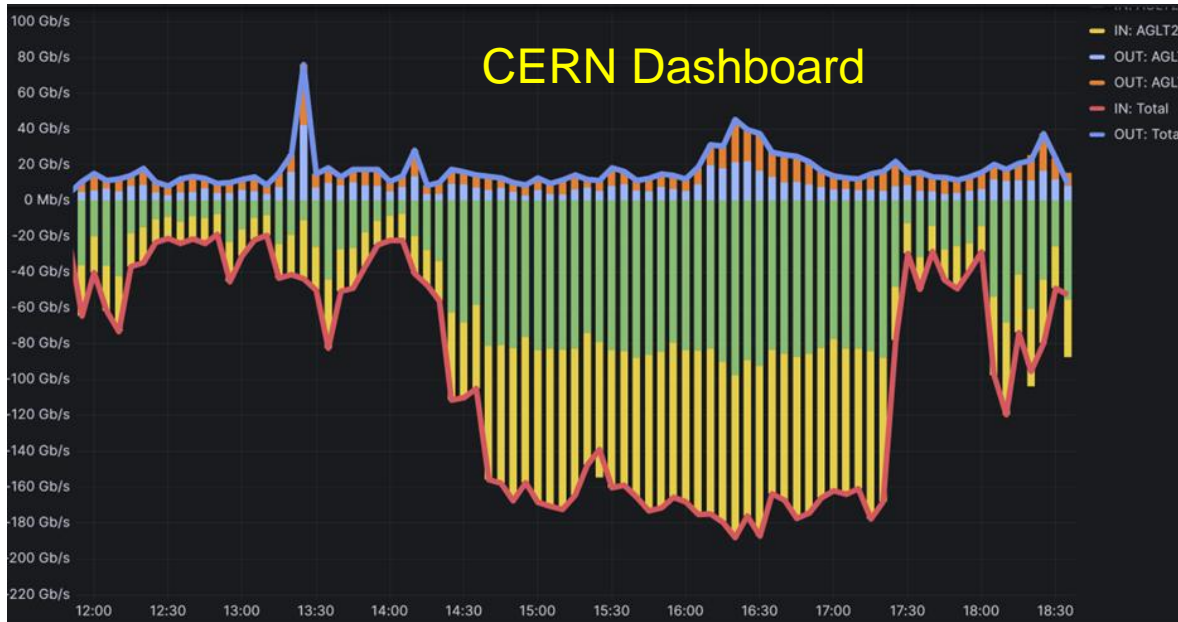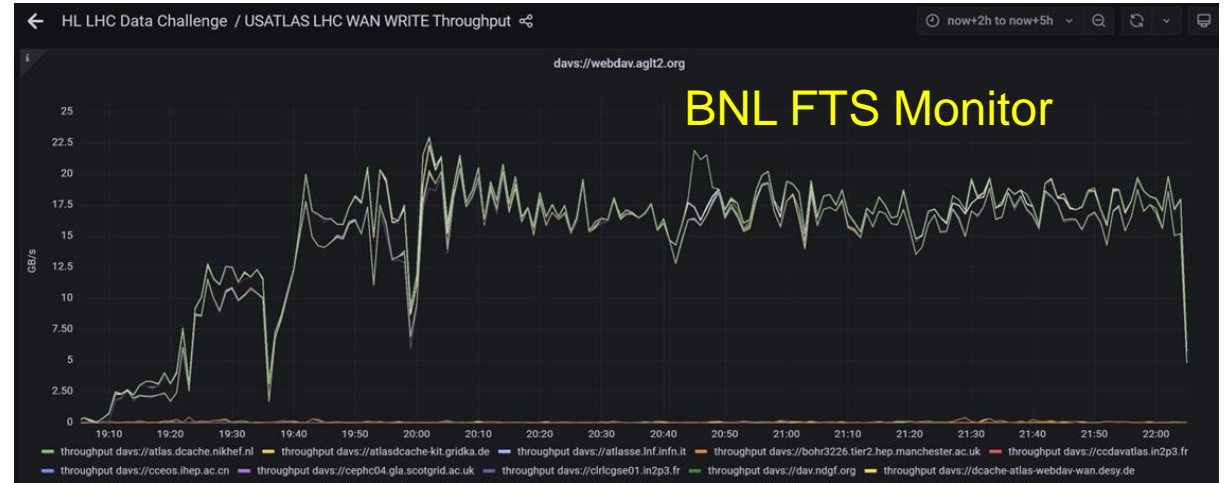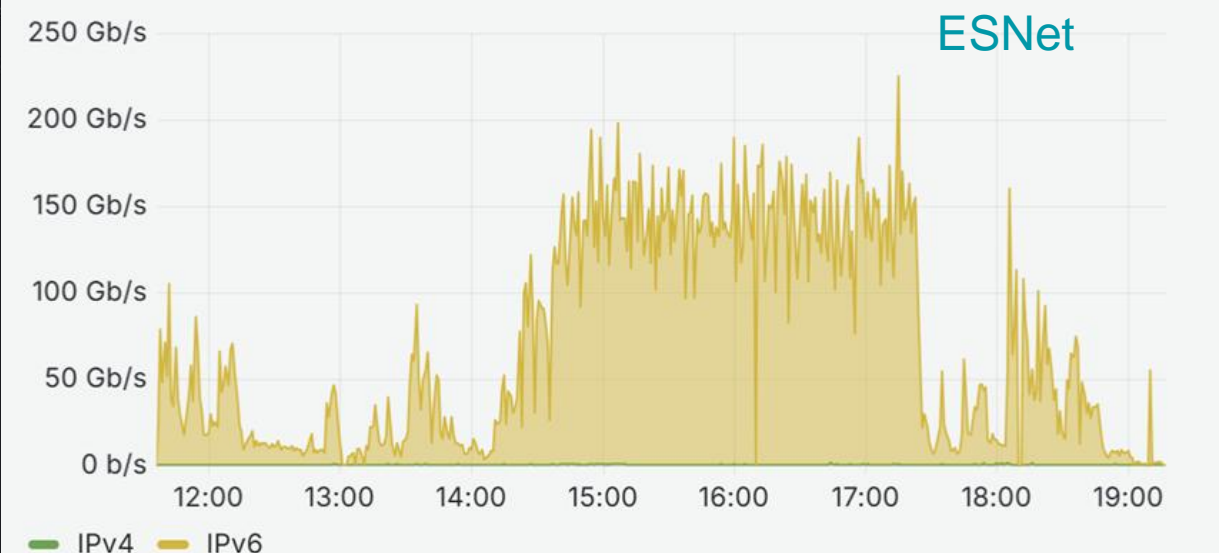
AGLT2 Site Report: Mini DC traffic in 4 hour window
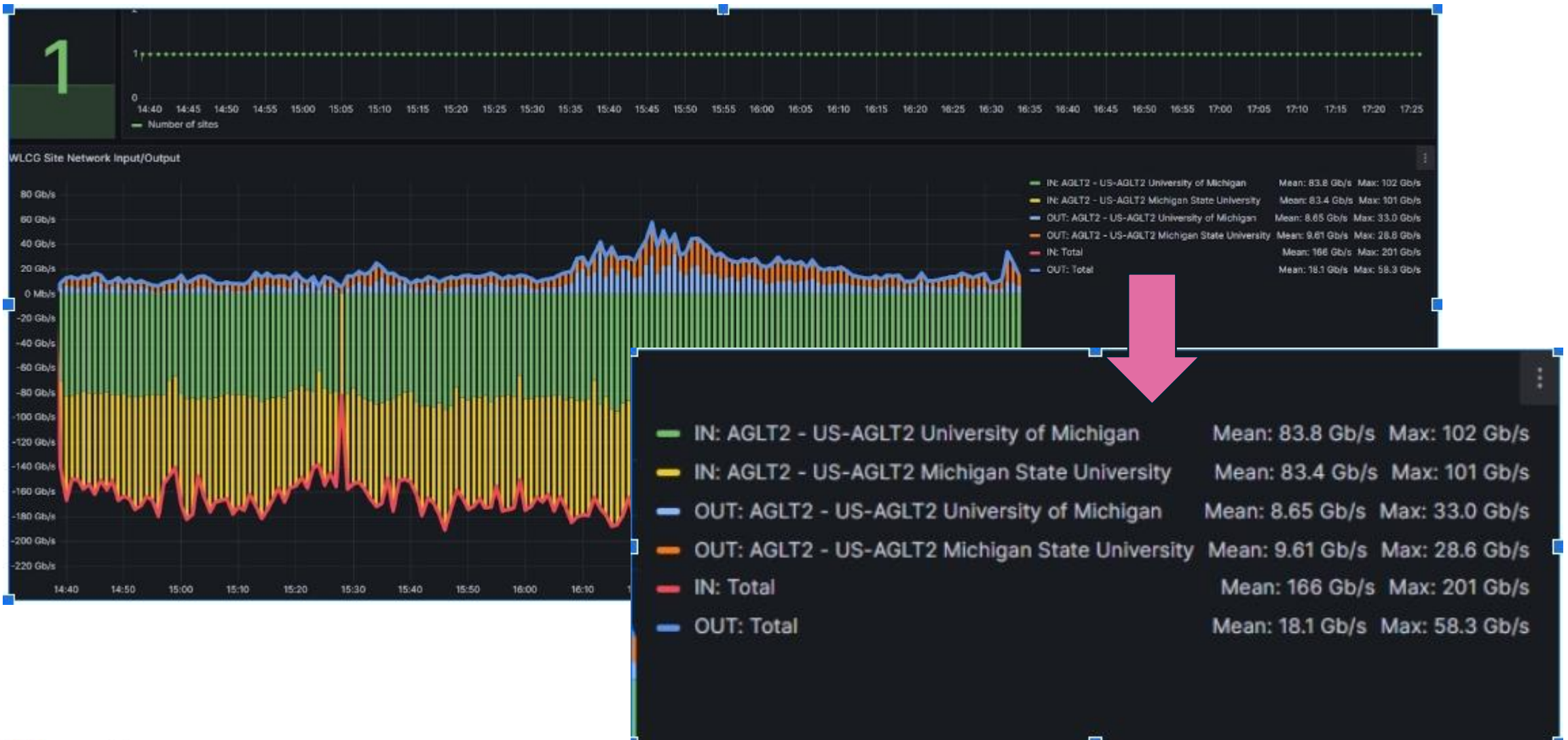
AGLT2 Site Report

- The UM network (80 Gbps) is saturated, but MSU (100 Gbps) is not (84% used)

- **UM site**: No bottlenecks observed from border switches, rack switches, dcache head nodes (postgresql), dCache head/pool nodes (cpu load, memory usage, disk IO performance)

- UM site: 33 pool nodes (R740xD2 and R760xD2) with 66 pools , storage nodes monitor shows DC traffic added an avg of 300MB/s to each node (300MB/s*33*8=79.2Gbps, so the limit comes from the 80 Gbps link)
    - There is still plenty of room for more IO based on historical peak IO. We need **~264 Gbps WAN** in order to saturate the storage nodes' IO capacity.

- **MSU site**: has similar hardware and quantity implying we should be able to source/sink **~500 Gbps WAN** if we had that link capacity available.

- It would be interesting to repeat the test to identify the limiting factor.

Detailed site report available

Brookhaven
National Laboratory
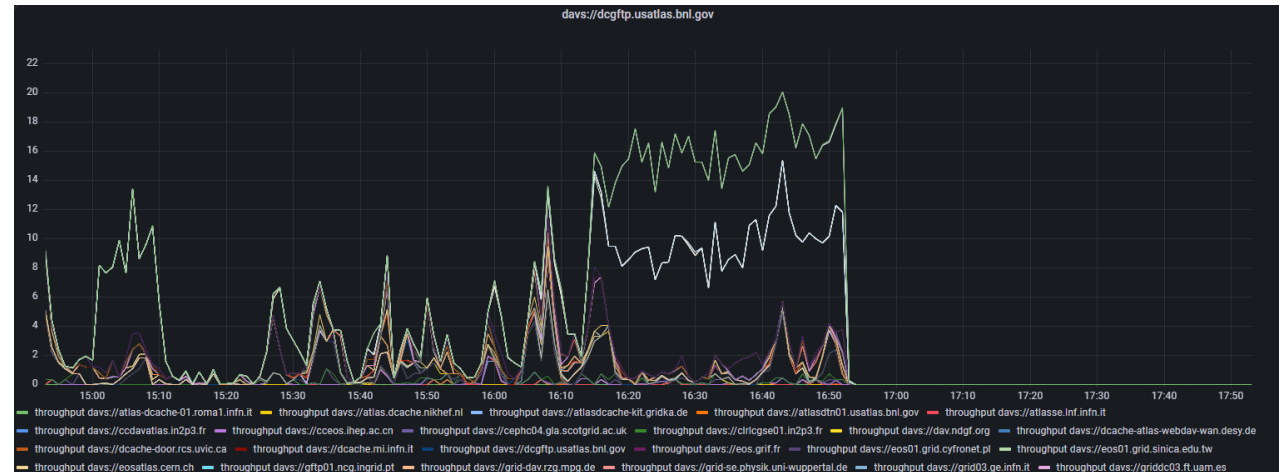
MWT2 Ingestion Testing

- The observed throughput for MWT2 was about 230Gbps.
  - UC 180Gbps (200Gbps limit)
  - IU 50Gbps (100Gbps limit)
- CERN Dashboard shows again a bit higher values.
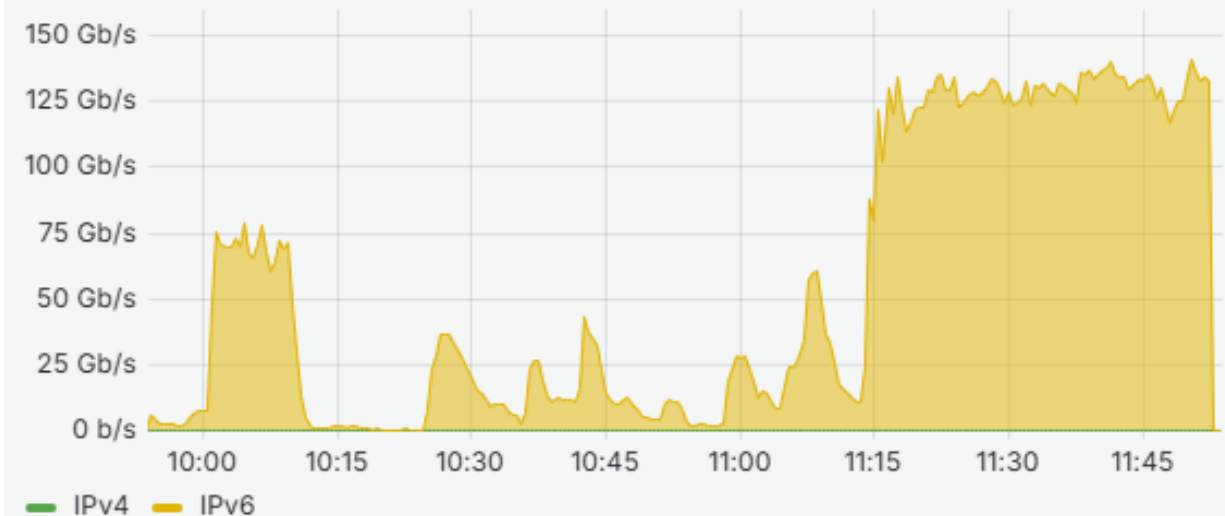- NOTE: ESNet monitor only shows UC.
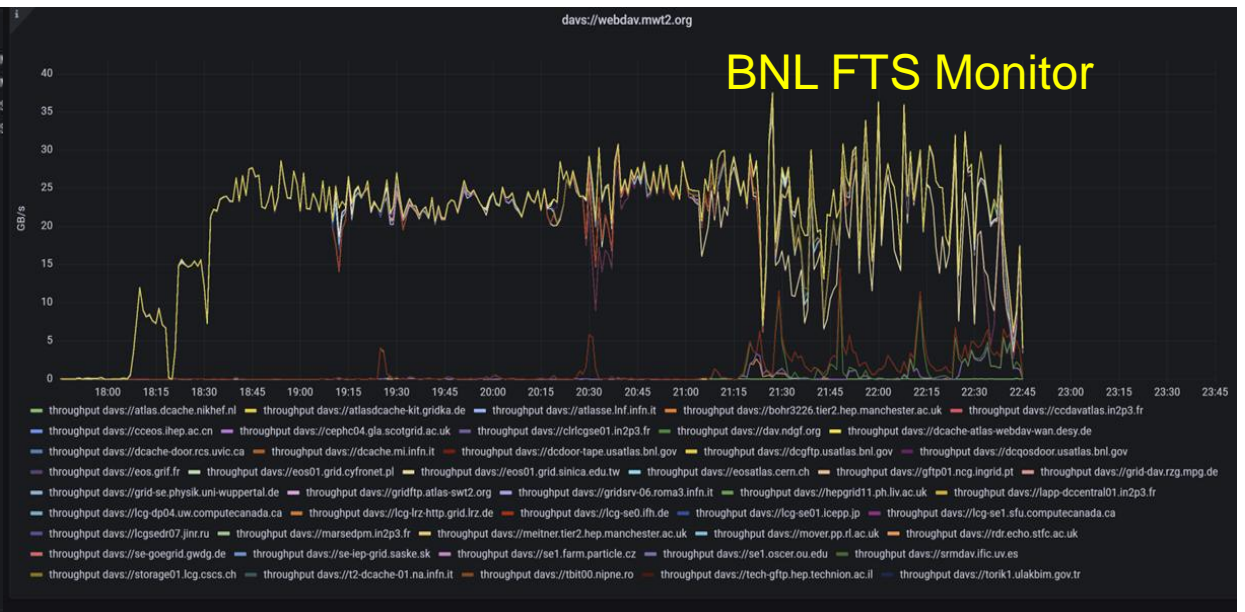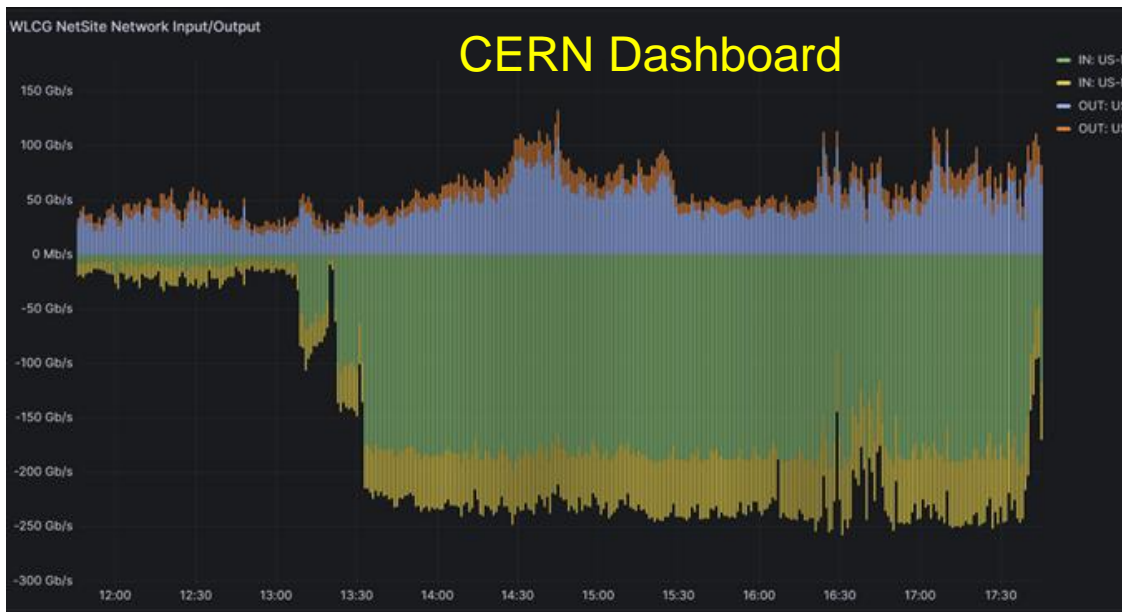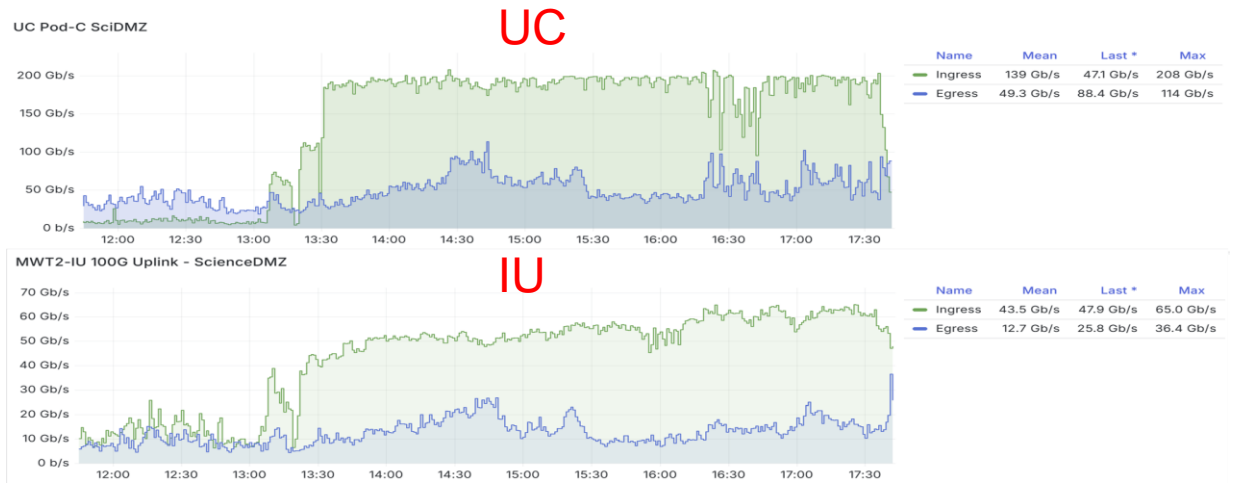
BNL T1 Ingestion Testing

- Complication
  - It requires multiple sites to drive BNL to its bandwidth capacity
  - AGLT2 encountered storage issue at the time of BNL testing.
    - Cause of delay
  - Some shorter testing after AGLT2 became operational.
- It achieved ~125Gbps.
- It requires additional testing to investigate the actual current limitation. (Redo in February?)
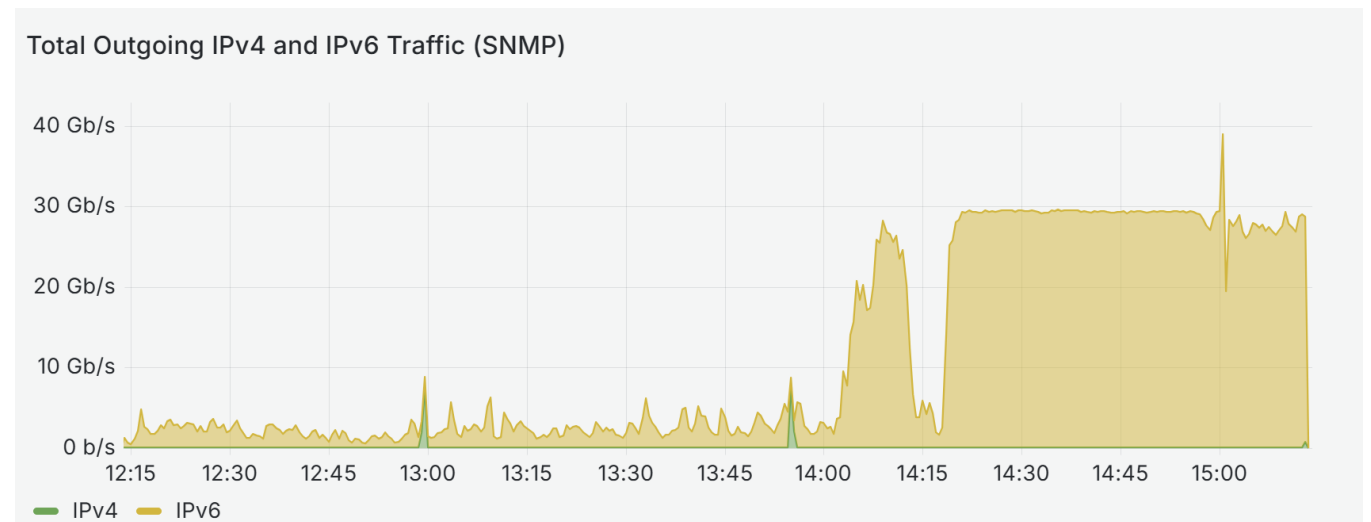

Brookhaven National Laboratory

SWT2

- SWT2 (UTA) has achieved 30 Gbps.
- This is still the limit of the network at the site.
  - The flatness of the plot indicates that it is indeed the network limit.
- Discussion with the network engineer is under way to increase the bandwidth. (Needs both bandwidth and DTN capacity increases)

Total Outgoing IPv4 and IPv6 Traffic (SNMP)

USATLAS Testing Summary

- Summary of injection testing to US ATLAS sites
  - AGLT2 has achieved 150Gbps.
  - MWT2 has achieved 200Gbps.
  - BNL has achieved 125Gbps.
    - In addition to the disk throughput, the analysis of tape system in terms of the staging throughput is on going.
      - Staging test will be conducted soon (the next week?) (see the next presentation about the analysis of files in BNL tape system).
  - SWT2 UTA has achieved 30Gbps.
  - NET2 ~~was not quite ready~~ for the testing.
    - ~~It is waiting for the completion of the network upgrade to 400Gbps~~
    - The site is now connected with 400Gbps as of 2025/01/22.
    - It will be tested as soon as the site is ready (next week?)

Targets on the next Round

- AGLT2:
  - In: >200Gbps,   Out: 200Gbps possible?
- MWT2:
  - In / out: 200-250Gbps
- BNL:
  - In: >200Gbps (NOTE: This is important for "out" test for AGLT2/MWT2)
  - Out: >400Gbps for reading (Use AGLT2 and MWT2 at least)
  - Staging: identify reasonably available limit
    - Requires a bit of additions to the current load generator.
- SWT2
  - >30Gbps when it becomes available.
- NET2
  - In/out >100Gbps.   400Gbps?

Brookhaven
National Laboratory