USATLAS Data Challenge 2024 Take-aways

Shawn McKee / University of Michigan and
Ofer Rind /Brookhaven National Lab

USATLAS Technical Presentation

https://indico.cern.ch/event/1377561/

February 28, 2024

WLCG Data Challenges

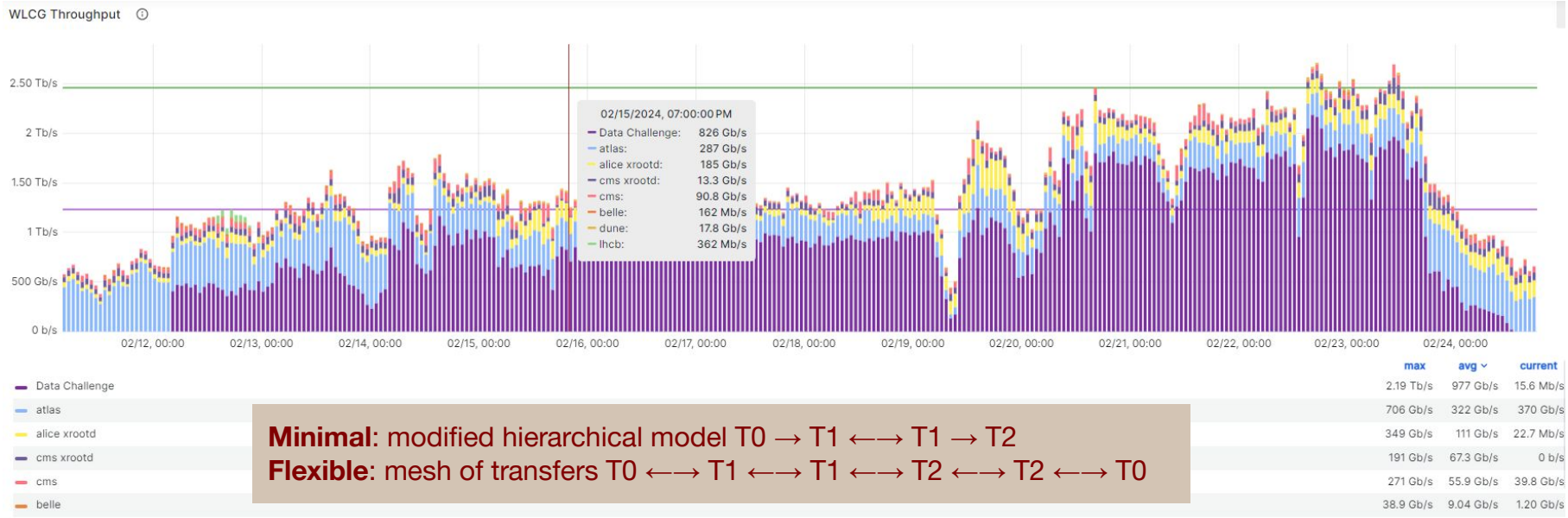
The **WLCG Data Challenges** are a ~biennial series of four increasingly-complex exercises which started in 2021 and are aimed at demonstrating readiness at the HL-LHC scale.

DC24 targeted **25%** of HL-LHC scale and includes T1/T2

Such exercises are only meaningful if there is worldwide, multi-experiment participation; they are planned and managed by the WLCG DOMA (Data Organization, Management and Access) group. **USATLAS was a strong participant**

These data challenges provide many benefits, allowing sites, networks and experiments to evaluate their progress, motivate and validate their developments in hardware and software and show readiness of technologies at suitable scale.

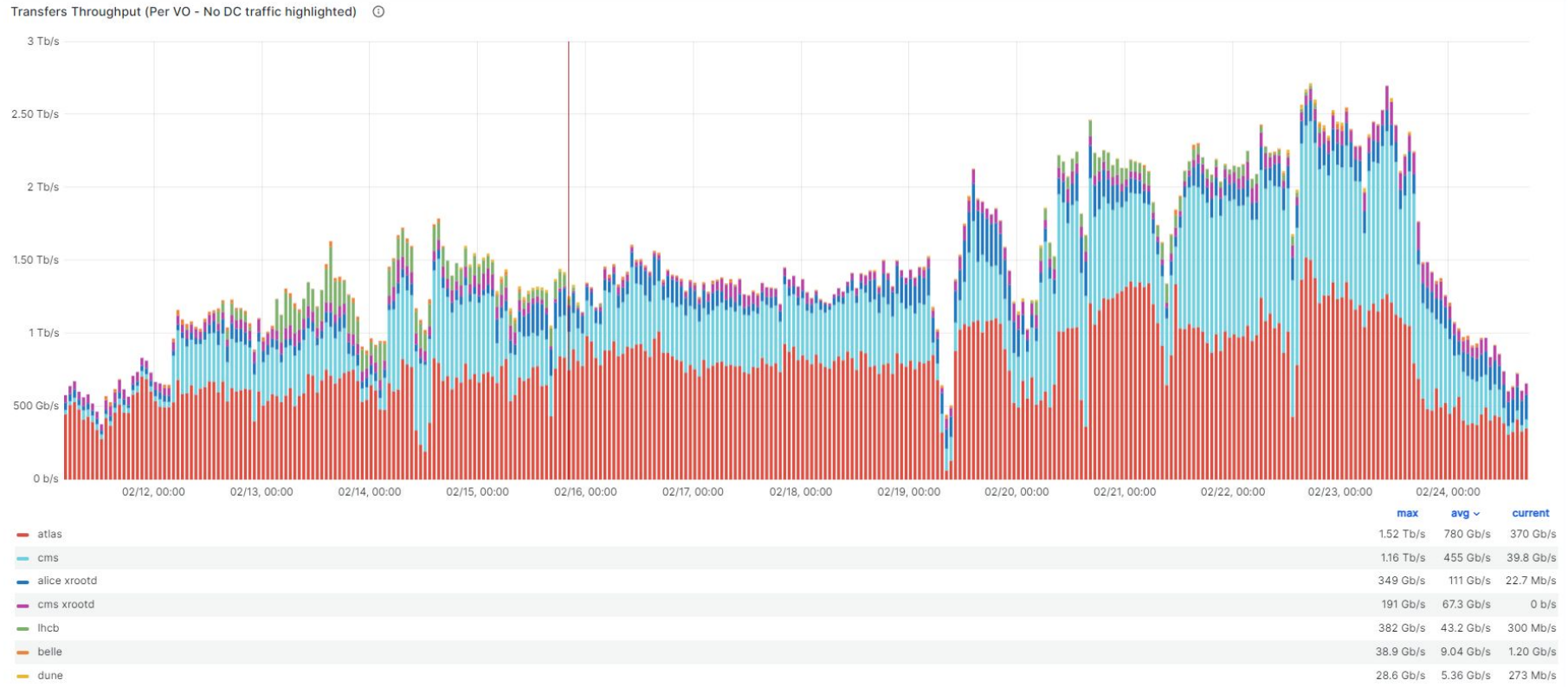
DC24 Dashboard By Activity



Minimal: modified hierarchical model T0 → T1 ←→ T1 → T2
Flexible: mesh of transfers T0 ←→ T1 ←→ T1 ←→ T2 ←→ T2 ←→ T0

Purple line is **Minimal** DC24 target, while green line is **Flexible** target
Goal was to reach both for at least 48 hours…made baseline, missed flexible
We were close and tweaks in Rucio, FTS and our transfer plans could succeed

DC24 Dashboard By Experiment



ATLAS traffic dominated DC24

DC24 by Country or Tier

Source Countries

	total	percentage
USA	22 Mil	23%
Switzerland	17 Mil	17%
UK	8 Mil	9%
Germany	8 Mil	9%
Italy	6 Mil	7%
France	6 Mil	7%
UNKNOWN	4 Mil	4%

Destination Countries

	total	percentage
USA	19 Mil	20%
Switzerland	11 Mil	12%
Italy	11 Mil	12%
UK	8 Mil	9%
Germany	8 Mil	9%
France	6 Mil	6%
Sweden	5 Mil	5%

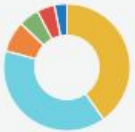
US delivered our % as Source, not quite as Destination

Tier-2s matched Tier-1s for number of transfers as source or destination

Source Tiers

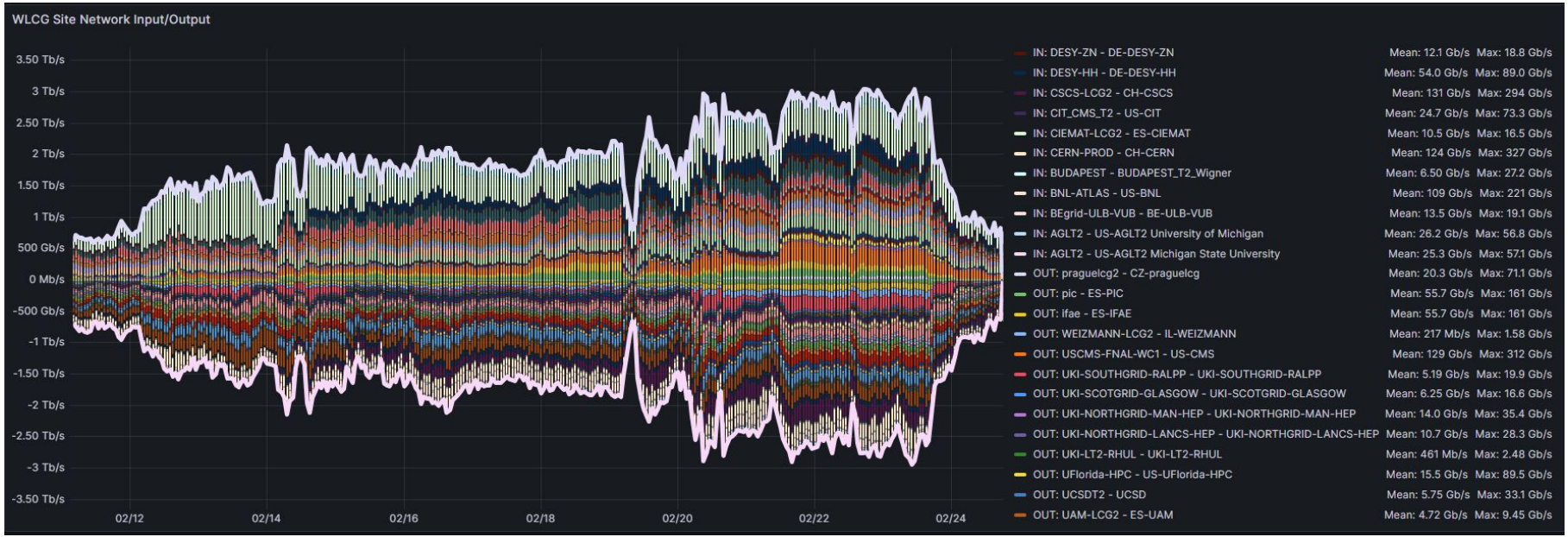
	total	percentage
1	36 Mil	37%
2	35 Mil	37%
0	16 Mil	16%
-1	4 Mil	4%
-1	3 Mil	3%

Destination Tiers



	total	percentage
2	38 Mil	40%
1	37 Mil	39%
0	8 Mil	8%
3	5 Mil	5%
-1	4 Mil	5%

WLCG Site Network Monitoring

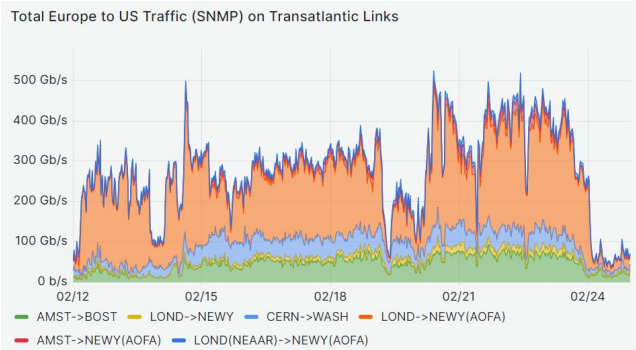


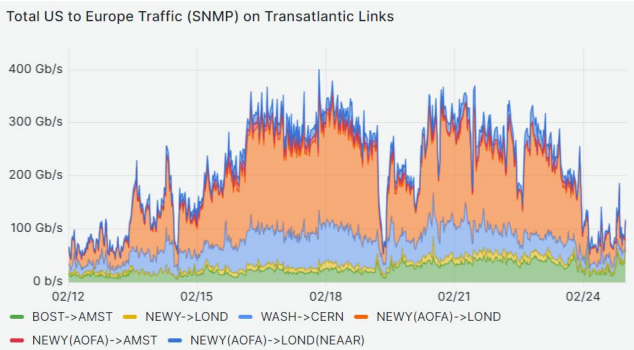
Site network monitoring was useful and tracked **total** network activity from ALL sources for WLCG sites (at least those who enabled monitoring)

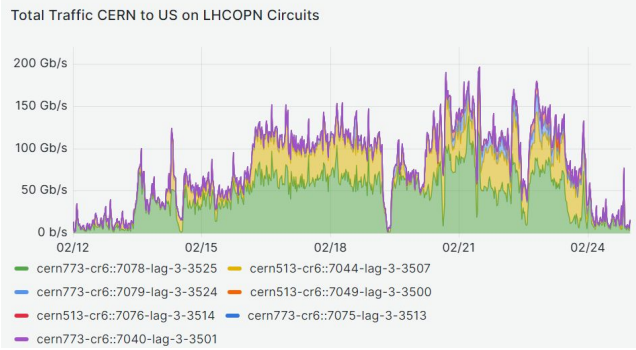
Transatlantic Links (ESnet Stardust Monitoring)

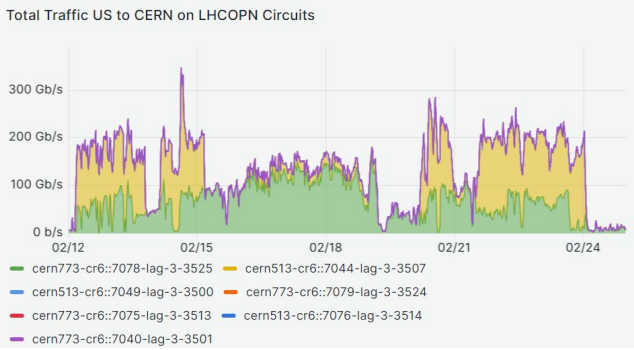
These performed well despite pre-DC24 concerns, with peaks up to 0.5 Tbps

Total capacity during DC24 on ESnet provided links was 0.9 Tbps









USATLAS DC24 Overview

How did DC24 go for USATLAS?

- In general our sites performed well BUT we did identify issues
- **AGLT2**: Not stressed but prior testing showed 80Gbps bottleneck at UM
- **MWT2**: Some issues with dCache performance needs investigation
- **NET2**: Some storage/transfer issues encountered. Only had 10 Gbps
- **SWT2**: Problems with network capacity, issues with OU deletion and storage performance, missing site network monitoring

At this point we need to analyze how our sites and software performed, especially while monitoring systems still have the data at finest granularity.

We also tested new technologies but not uniformly across the whole facility.

Overall ATLAS DC24 Issues Noted

- Low T0 Export to T1 throughput needs to be understood (DDM Monitoring, DC24 Monitoring, LHCOPN Network)
- MANY site issues
  - Slow deletion at RAL - https://ggus.eu/index.php?mode=ticket_info&ticket_id=165358
  - Worsened NDGF-T1 - https://ggus.eu/index.php?mode=ticket_info&ticket_id=164846
  - Timeouts to Milano - https://ggus.eu/index.php?mode=ticket_info&ticket_id=165356
  - SSL errors to CNAF - https://ggus.eu/index.php?mode=ticket_info&ticket_id=165355
  - Timeouts to FZK - https://ggus.eu/index.php?mode=ticket_info&ticket_id=165393
  - **Timeouts at OU_OSCER_ATLAS - https://ggus.eu/index.php?mode=ticket_info&ticket_id=165362, https://ggus.eu/index.php?mode=ticket_info&ticket_id=165379**
  - Expired tokens in the FTS causing problems at DESY-HH - https://ggus.eu/index.php?mode=ticket_info&ticket_id=165397
  - Timeouts at UKI-SCOTGRID-GLASGOW - https://ggus.eu/index.php?mode=ticket_info&ticket_id=163552
  - "Unexpected server error" to NIKHEF during pre-DC24 test - https://ggus.eu/index.php?mode=ticket_info&ticket_id=165263
  - "Unexpected server error" to UKI-NORTHGRID-LANCS-HEP - https://ggus.eu/index.php?mode=ticket_info&ticket_id=165394
  - Dark data caused by heavy load at TRIUMF-LCG2 - https://ggus.eu/index.php?mode=ticket_info&ticket_id=165343
  - IFIC tickets is not blaming DC24 but errors stopping with the end of it - https://ggus.eu/index.php?mode=ticket_info&ticket_id=165395
  - IN2P3-CC being overloaded and HC putting the site into test for lack of storage free connections
    - Cured by reducing the number of connections in FTS
  - **SWT2 large wave of jobs in transferring state (concurrent with a wave of evgen jobs)**
  - FZK QMUL slowed down transfers
  - INFN-T1 - (one) stuck doors while retrieving JWK from IAM https://ggus.eu/index.php?mode=ticket_info&ticket_id=165355
  - Some sites applied storage limits tuning, e.g. FZK (GGUS:165393), TRIUMF (GGUS:165364)…, there were few more and not all communicated with GGUS => for final report we should also ask sites what they observed (e.g. SARA internal? throughput saturated their links (GGUS:165359), INFN also observed huge traffic (GGUS:165355), we don't fully understand much higher throughput on some links,

WBS 2.3 DC24 Work Areas

As part of USATLAS facility we identified areas of engagement for Data Challenge 2024 (DC24)

- perfSONAR site network debugging prior to DC24
- Enabling WLCG Site Network load monitoring
- Deployment of traffic marking for xrootd / dCache
- Demonstration of packet pacing at two or more sites
- Network load testing (see Hiro's talk) to find bottlenecks

How did we do?

perfSONAR Site Network Debugging

We were able to use perfSONAR in advance of DC24 to rule out major network issues.

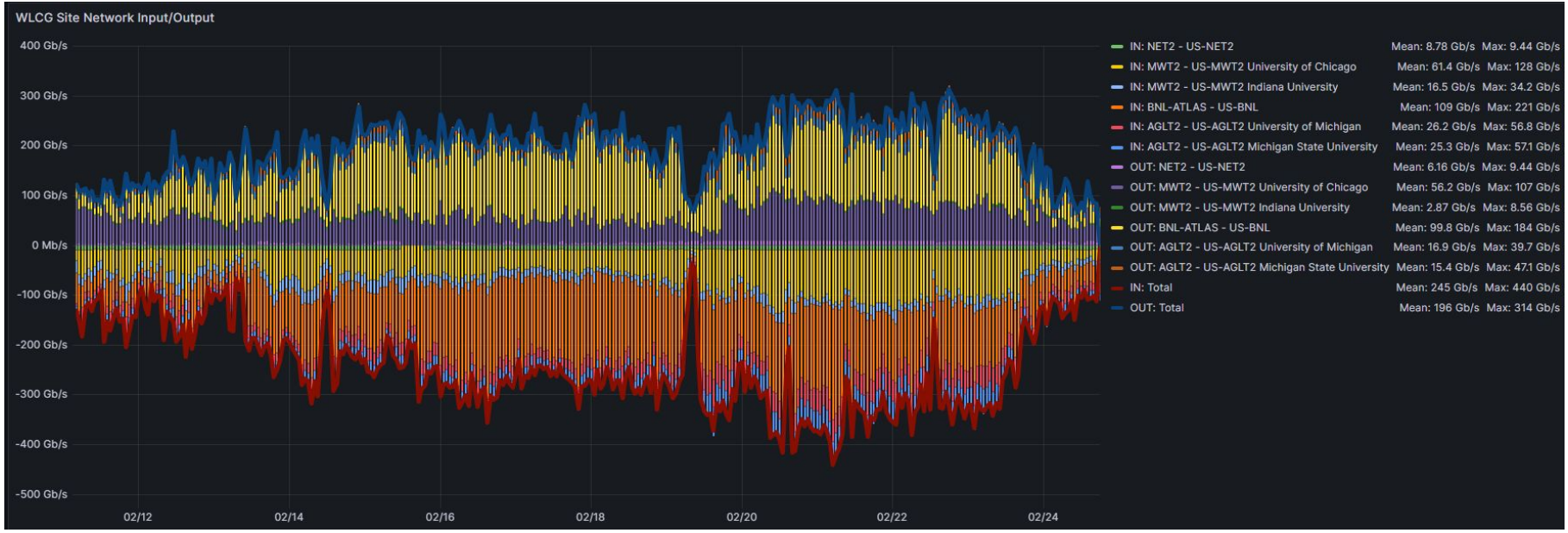
We have noted that many of our sites globally are in need of hardware refreshes and OS updates for their perfSONAR hosts.

For USATLAS, sites should ensure their perfSONAR hosts:

- Are running an EL9 OS
- Have network hardware similar to storage
- Are (auto)updated to the most current perfSONAR
- Are located in the same “network” location as the storage

Enabling WLCG Site Monitoring

We have most of our sites monitored and described in WLCG-CRIC, but missing SWT2



Packet and Flow Marking

CMS was the primary demonstrator of fireflies for DC24

(See https://public.stardust.es.net/goto/dX0DkXTSg?orgId=1)

Sites will need xrootd 5.6.8+ or a future version of dCache to enable fireflies and eventually, packet marking.

Lots of interesting information becomes visible, once we begin consistently marking our traffic.

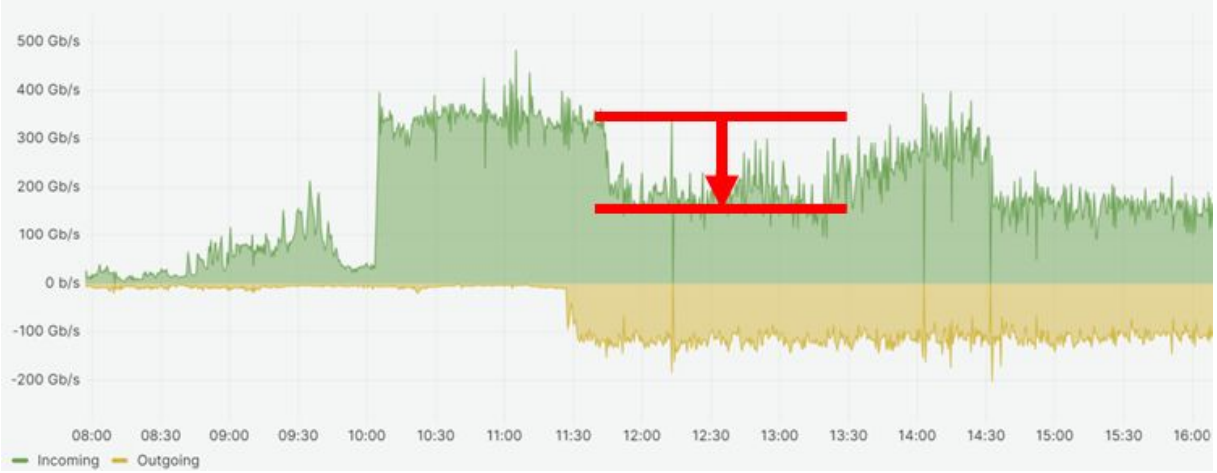
Allows identification and understanding of our R&E traffic anywhere in the network…

Network Load Testing / Bottleneck ID

We had a very successful test within the US prior to DC24

Joint US ATLAS and CMS network challenge

- BNL was able to push data out to MWT2 and AGLT2 at up to ~400 Gbps.
- When the reverse direction data flow (into BNL) was added, there was a drop in the original direction (out of BNL) after some delay.
  - Disk IO issue at source and/or destination?



Next Steps

ATLAS wants to retry before the WLCG workshop in May

Plan to improve transfers and priorities based upon links

USATLAS should plan another test amongst our sites as well since DC24 was not really able to stress most of them.

We should also plan to complete:

- Packet/flow marking: need dCache version or updated Xrootd (5.6.8+) deployed and configured
- Storage tokens enabled and tested at all sites
- Site network monitoring in place (update WLCG CRIC)

Summary

Overall DC24 was successful in identifying where we have bottlenecks or software/hardware issues to address

We need to complete an analysis of DC24 and our sites should try to gather information to document findings and next steps.

Lots of work to do to analyze and update all the various components that make up our USATLAS facility and we need to identify effort to do so…

Question, Comments?

Acknowledgements

We would like to thank the **WLCG**, **HEPiX**, **perfSONAR** 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:

- OSG: NSF MPS-1148698
- IRIS-HEP: NSF OAC-1836650

Backup Slides Follow

Resources DC24 Related

- Mattermost
- Meetings
  - 12.02.2024 - 23.02.2024 - DC24 - Daily Operation Meetings (Indico: 17972)
- Monitoring
  - DC 2024 (draft) (Monit)
  - FTS Status (Monit)
  - FTS Transfers (Monit)
  - FTS Servers (Monit)
  - WLCG Site Network (Monit)
  - LHC Data Challenge Overview (ESNET)
  - LHCOPN CERN
  - Xrootd dCache monitoring (Monit)
  - Experiments:
    - ATLAS:
      - DDM Transfers (Monit, no delay)
      - DDM Throughput (Monit)
      - ATLAS Rucio monitoring (Monit)
    - CMS:
      - Rucio Storage Occupancy (Monit)
      - CMS FTS (Monit, from FTS logs)
      - CMS Fireflies
  - Sites: (note ESnet pages cover all USATLAS and USCMS sites)
    - BNL:
      - LHC Data Challenge LHCOPN Circuit Details (ESNET)
      - LHCOPN
      - LHCONE (ESNET)
    - NET2
      - LHC Data Challenge (Monit)
    - SWT2
      - LHCONE (ESNET)

Documents:

- Global rates
- DC24 - Targeted transfer rates per site
- DC24 - Notes and Observations (to capture plots for talks)
- ATLAS configuration