EPJ Web of Conferences **295**, 07039 (2024) https://doi.org/10.1051/epjconf/202429507039
CHEP 2023

Preparing for the next WLCG Network Data Challenge: Site Network Monitoring

Shawn McKee[1,*] *Julia* Andreeva[2] and *Alessandra* Forti[3]
for the WLCG Monitoring Task Force

[1]Physics Department, University of Michigan, Ann Arbor, MI, USA
[2]European Organization for Nuclear Research (CERN), Geneva, Switzerland
[3]Physics Department, University of Manchester, Manchester, UK

Abstract. During the first WLCG Network Data Challenge in fall of 2021 (DC21) we identified shortcomings in the monitoring that impeded our ability to fully understand the results collected during the data challenge. One of the simplest missing components was site-specific network information, especially information about traffic entering and leaving any of the participating sites. Without this information, it is very difficult to understand which sites are experiencing bottlenecks or might be misconfigured or under-used based on their capacity. The WLCG Monitoring Task Force, formed at the end of 2021, was tasked with three main work areas, one of which was site network monitoring. We will describe the work carried out by the task force to enhance our knowledge of network use for WLCG by enabling site network documentation and use, the status of the deployment, and the implications for the next data challenge.

1 Introduction

The WLCG Monitoring Task Force[1] has been charged with investigating, prototyping, and enabling a number of new monitoring capabilities for the next WLCG Network Data Challenge in 2024 (DC24[2]). Visibility into a variety of resources and their use is critical to understand how our infrastructure is operating and what bottlenecks we may be currently experiencing. One of the primary goals of the WLCG data challenges is to identify where our infrastructure is hitting limits so that we can remove those bottlenecks and increase our capacity as we move towards the high-luminatiosity LHC[3] era.

In this paper, we describe one specific aspect of the monitoring work focused on providing better visibility into each WLCG site's achievable network rate, both into and out of the site, so that we have a more complete understanding of the challenge results and limitations. We will motivate the need for this information in the next section and describe our goals, plans, and future evolution in the following sections.

2 Motivation

High-Energy Physics (HEP) collaborations rely on networks as one of the critical components of their infrastructure. To help plan and evolve the effective use of all collaboration resources

*Acknowledgements (TBA).

(especially the network), the WLCG has implemented a series of network data challenges, designed to incrementally increase the target network use in each challenge, ultimately reaching a high-luminosity scale in the last challenge. Each challenge will have multiple experimental collaborations running both their normal workloads and additional injected work targeting the challenge target rates.

There are a vast number of monitoring tools associated with WLCG and individual experiments to monitor the components and applications of the end site, but they don't monitor the network traffic. The WLCG data challenges are targeting the end-to-end use of the network at appropriate scales, and we need to ensure that we understand what each site is able to contribute during both normal production and for the two-week period of the data challenge. This will be critical in identifying where limitations and bottlenecks may be occurring, either within the network or at individual sites.

Understanding how effectively experiments can use their associated computing and storage sites requires us to measure how those sites are using the set of research and education networks. These networks are used to interconnect participating sites, data centers, and scientific instruments within laboratories and throughout the world. The network traffic generated by the HEP and WLCG experiments is split between purpose-built networks such as LHCOPN and LHCONE, as well as research and education networks (R&E networks) operated by campuses, research organizations, national research and education networks (NREN) and general public networks. Although we have access to these backbone networks monitoring data, without understanding the traffic flow to and from all of our sites, interpreting the data we have becomes very difficult when we want to understand how all the infrastructure components operated during the data challenge and ultimately during normal operations.

3 The WCLG Data Challenge Needs

During DC21[4], one of the most problematic aspects was to correctly monitor data challenge traffic and match it to what sites were seeing in their monitoring. There are multiple aspects to this problem.

- Not all sites monitor their network. Therefore, not all sites can provide network plots.
- Site network monitoring, if it exists, is often not publicly accessible. So even if the network monitoring existed the DC21 operators had to rely on screenshots
- Not all sites use the same visualization tools. Even if a screenshot was provided, it was difficult to match it with the DC21 shapes reported by the DC21 dashboard.
- Site traffic is not generated solely by LHC experiments, so the shape and rates reported may really be different

A major recommendation, after analyzing data from DC21, was to implement a more centralized and complete site network monitoring by the time of the next challenge that would make quantitative comparisons between what the rates experiments measure and the rates the sites see.

One other deficiency was noted that operators of the data challenge (and WLCG support in general) have insufficient information about the architecture of a site's network, the components used to build the network and how the site connects to global research and education networks. Understanding these details can be critical to interpreting any monitoring results, as well as for providing support, debugging, and future upgrade advice.

4 Site Monitoring Goals

During the time since the DC21 analysis was completed in early 2022, the WLCG Monitoring Task Force has been considering how best to respond to the DC21 recommendation about site network monitoring. The Task Force spent some time designing and prototyping a possible solution based upon the following guiding principles.

- Regularly monitor the site's TOTAL network use, both incoming and outgoing.
- Ensure that the monitoring system is reliable and secure.
- Minimize the effort required for a site to deploy and maintain monitoring.
- Give site administrators control over their network monitoring and description.
- Provide guidance for describing and publishing the site network details.
- Utilize the WLCG CRIC (Computing Resource Information Catalogue)[5] as the mechanism to publish the site network information

In the next section, we will describe the architecture and choices we have implemented, given the above goals.

5 Implementation Details and Plans

As the Task Force began to design the site network monitoring system, we needed to have a place to host the instructions, example configuration, code, and possibly site network information. Since CERN had deployed Gitlab and WLCG DOMA[6] already had an area set up, we chose it to host our project, rather than some commercial service and created the WLCG DOMA / Site Network Information[7]. We did note that access to the repository would require a CERN account, but we anticipated that most users would meet this criteria. We will monitor if this location limits sites or possible maintainers from easily participating.

5.1 Site Network Information

Once we had a location to host the project, we worked on creating the templates to describe the needed site network information. Templates organize information coherently across sites and are split into mandatory and optional information.

The template is accessible from the Gitlab location, but a quick overview of what is Mandatory and Optional is provided here. Site managers are expected to copy the template file to to a dedicated instance named after their site official WLCG RCSITE (Resource Center Site) name. Then they can fill the information of the Mandatory sections and, ideally, also of the Optional sections. The Mandatory sections are:

- **Network Overview:** brief description of the site topology (single or multiple data centers), site bandwidth, shared or dedicated uplink etc
- **Network Monitoring:** provide link to the json produced by the monitoring
- **Network Monitoring Link Into WLCG-CRIC:** provide link to the WLCG-CRIC Netsite pages

The information in these sections can be brief, but must be provided. The Optional sections are:

- **Network Description:** more detailed description of the site network both internal to the data centre, machine room and connection to the university or other lab.
- **Peering Description:** networks a site belongs to for example LHCOPN, LHCONE, others
- **Network Equipment Details:** description of the equipment switches and routers models and type of interfaces wiring used.

- **Network Diagrams:** network diagrams visualizing the information above.

The optional components are more detailed and are designed to provide information to experts who can better advise and support debugging for sites. Guidance for site administrators is documented in the same location. For the site template, the summary is:

Once the information is filled it needs to be put in an accessible place. As we mentioned gitlab is not accessible to everyone and the network experts in each country may not have any association with CERN. The most common solution is to put a pdf on a site web server. Finally to make this discoverable the URL needs to be registered in the NetSite page which groups the network information of the RCSITE in CRIC.

5.2 Network Monitoring

The needed site network metrics for the total traffic flowing in and out of a site are complicated for a number of reasons that we discussed previously. Often the WLCG site administrators do NOT have access to the network devices. Although some sites may have monitoring, the systems providing the data can be diverse, having different measurement cadences and monitoring characteristics we are not interested in. Often, sites have access to a graph, but it is not easy to extract the relevant metrics from this. The graph may have been produced by different tools which allow only for a qualitative comparison.

The Task Force decided that the most straightforward path would be to provide a simple Python3 based script that could be used to gather the exact data we wanted. The script, plus information on how to deploy, configure, and make available the data, would be accessible from the Gitlab repository.

To deploy this monitoring the network administrator needs SNMP access credentials to all relevant network devices. Several discussions with system administrators indicated that the best place to monitor is switches or routers that aggregate their clusters traffic and not upstream devices which might supply information on other type of traffic. Once installed the script would provide properly formatted data in JSON format that needs to be made available so that it can be consumed by a dedicated collector.

The Task Force made arrangements with the CERN MONIT[8] team to centrally collect data from ALL sites that properly register their monitoring URL in WLCG CRIC following a similar procedure to the site information described above.

The script was designed to use Python3 as the programming language and, for ease of use, requires the EasySNMP[9] software package. Examples assuming that multiple switch/router devices would need to be accessed are provided. The instructions show how to configure the script and enable it as a systemd service. We want to have fine-grained visibility into site network use and assume installers will follow our recommendation to query all interfaces every 60 seconds.

The script was developed and tested on three different sites to verify that the script and instructions would work for our target sites. Testing and prototyping were very useful, exposing some issues that were not found on the original development host.

An example of the python3 script JSON output is shown below for RCSITE AGLT2:


```
{
  Description: "Network statistics for AGLT2",
  UpdatedLast: "2023-04-05T19:59:01.691317+00:00",
  InBytesPerSec: 1294612738.7737598,
  OutBytesPerSec: 1023097622.4124134,
  UpdateInterval: "60 seconds",
  MonitoredInterfaces:
```

https://doi.org/10.1051/epjconf/202429507039

```
    [
      "aglt2-rtr-1.local_Ethernet1/48",
      "aglt2-rtr-1.local_Ethernet1/51",
      "aglt2-rtr-1.local_Ethernet1/52",
      "aglt2-rtr-2.local_Ethernet1/51",
      "aglt2-rtr-2.local_Ethernet1/52"
    ]
  }
```

The CERN Gitlab project has more details in various README and documentation files.

6 Deployment and Visualisation

Once we had a fully fleshed-out project, code, and instructions, the Task Force began a campaign to get sites to instrument their networks. The first question to answer was which sites we should include in the campaign, since there are a large number of WLCG sites. In a presentation at the WLCG Operations Coordination meeting [10] the Monitoring Task Force proposed that we include all Tier-1 sites of WLCG and all Tier-2 sites that provide more than 250 million HEPscore23 hours for the last 90-day period. This resulted in the identification of 53 Tier-2 sites to include and was adopted as the basis for the campaign. We currently have about 39 out of 67 sites instrumented. We hope to significantly improve this deployment ratio as we get closer to DC24 in February 2024.

The CERN MONIT team has prepared a dashboard[11] to visualize the site network monitoring that is publicly accessible. the dashboard displays plots aggregated by RCSITE and NetSite. As previously mentioned the RCSITE name is the official WLCG site name typically of a Tier1 or Tier2 which can be composed by multiple data centres with their independent networking. It is therefore useful to have a plot that displays the RCSITE aggregate traffic and a plot that shows the split of that traffic into its NetSites components. Shown in Figure 1 is an example of a plot aggregated by NetSite where you can see some of the sites split in their University sub-sites.

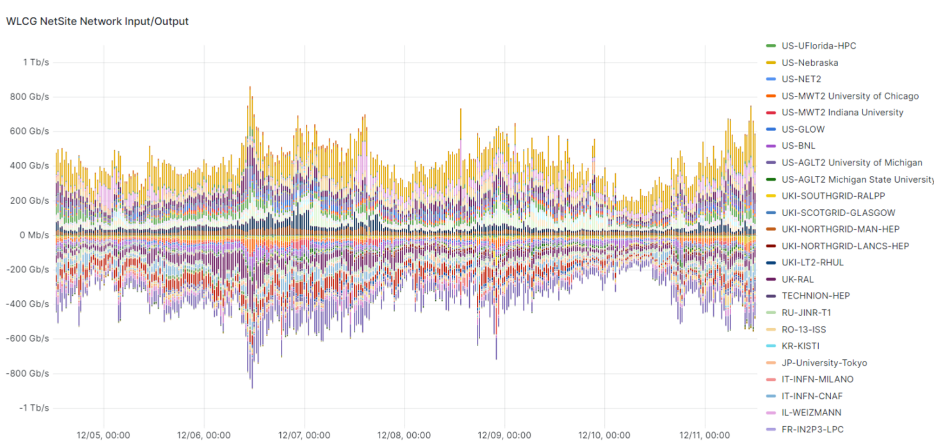

Figure 1: Current MONIT Dashboard showing site network monitoring aggregated by NetSite name where positive values indicate incoming traffic valume and negative values indicate outgoing traffic volume.

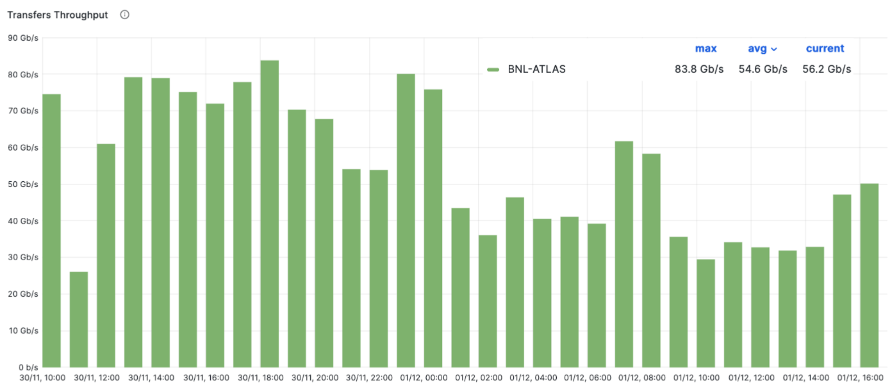

Figure 2: FTS dashboard: BNL rates during the pre-DC T0 export exercise. The rates are calculated dividing the volume transferred by the size of the plot bins. Smaller bin is 1h.

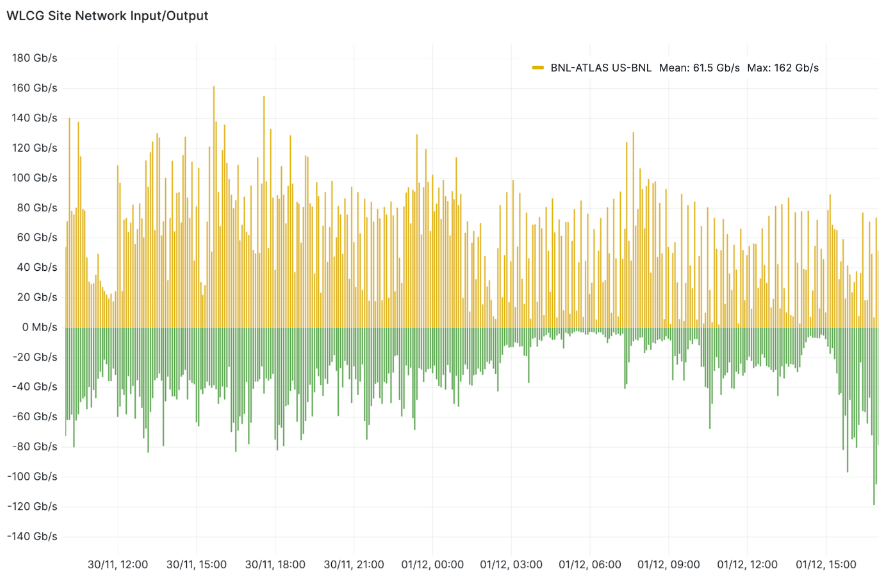

Figure 3: Network monitoring Dashboard: BNL network traffic during the pre-DC T0 export test. Rates pulled every minute and averaged over 5 mins bin give a better picture of what is happening during the transfers. In yellow the incoming traffic which is the object of investigation.

7 Example of Use

The usefulness of this monitoring was demonstrated during a data export test exercise carried out in preparation for DC24. The test consisted in injecting data from the Tier0 at CERN to each WLCG Tier1 every 15 minutes at the rates foreseen for DC24 for T0 export. Usually to monitor transfers we use data from FTS (File Transfer Service) [12] which execute the transfers and monitors how much data it has transferred. The rates are calculated roughly by number of GB transferred divided by the size of the plot bins. The smallest bin usable for

FTS data is 1h. This is an approximation that usually works well but doesn't give the full picture. In the specific case we wanted to investigate why BNL during this pre-DC exercise had yes a constant rate, as seen in Figure 2 but lower than expected.

Looking at the new Networking plots, which can display data with 5m bins, it was clear that the lower average seen in the FTS plots was due to the site absorbing the data too quickly. The plots in Figure 3 shows a clear comb teeth pattern corresponding to the 15 minutes injections that once average over 1h lower notably the throughput displayed. The experiments as a consequence will investigate further injection intervals and file sizes to inject. This type of analysis was previously possible only if the site had network monitoring in place and could share screenshot.

The use of grafana as common visualisation tool allows to align plots. If we use 1h bin also in the networking plots we can see in Figure 4 the traffic displayed is very similar with the networking plots reporting a slightly higher average which accounts for non-FTS traffic.

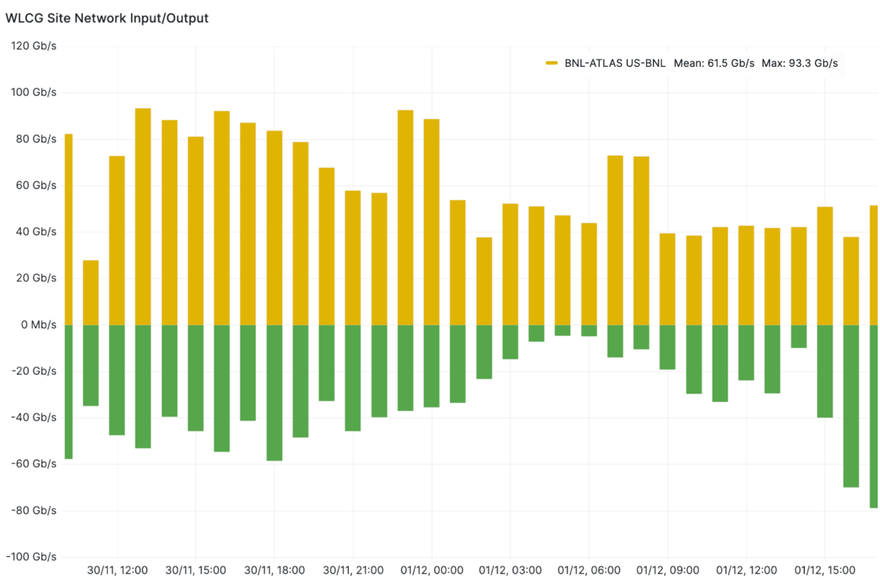

Figure 4: Network monitoring Dashboard: BNL network traffic during the pre-DC T0 export test. Rates pulled every minute and averaged over 1h show a similar pattern to the FTS plot but with higher averages due to non FTS traffic. In yellow the incoming traffic which is the object of investigation.

8 Next Steps and Future Work

We note that this data is very useful beyond just meeting DC24 needs and our plan is that this will become part of the regular production monitoring available for WLCG.

We have already identified a number of issues. The central MONIT infrastructure shows dropouts for some sites on occasion, as well as periods where the data collection was not operating due to MONIT or underlying service issues. We intend to work with the MONIT team to improve the resiliency and robustness of the service.

One weakness we have identified with our framework is that if MONIT or central services are down, we have no way to recover data from the down period. A future update of our script will likely create a series of JSON outputs, kept in a local buffer able to hold approximately 1 week of results. This buffer could be explicitly called to allow a MONIT recovery process to gather data from downtimes.

One more future activity may enhance the framework to provide additional monitoring about WLCG specific traffic instead of just the total traffic. This might be especially useful when sites support more than one WLCG collaboration or have significant non-WLCG traffic.

9 Acknowledgements

We want to explicitly acknowledge the support of the National Science Foundation, which supported some of this work through OSG: NSF MPS-1148698 and IRIS-HEP: NSF OAC-1836650 grants. Furthermore, we acknowledge our collaborations with the CERN IT, WLCG sites, and LHCONE / LHCOPN communities, who also participated in this effort.

References

[1] *WLCG Monitoring Task Force*, https://twiki.cern.ch/twiki/bin/view/LCG/MonitoringTaskForce, [Online; accessed 20-September-2023]

[2] *Data Challenge 2024 Workshop*, https://indico.cern.ch/event/1307338/overview, [Online: Accessed 22-September-2023]

[3] *HL-LHC: High Luminosity Large Hadron Collider*, https://hilumilhc.web.cern.ch/, [Online; accessed 20-September-2023]

[4] *WLCG Network Data Challenges 2021: wrap-up and recommendations*, https://doi.org/10.5281/zenodo.5767913, [Zenodo DOI accessed 22-September-2023]

[5] A. Anisenkov, J. Andreeva, A.D. Girolamo, P. Paparrigopoulos, V. A, *CRIC: A unified information system for WLCG and beyond*, in *EPJ Web of Conferences* (Budker Institue of Nuclear Physics, Novosibirsk State University, CERN and PIC, 2018), p. 8, [Published online]

[6] *WLCG DOMA*, https://gitlab.cern.ch/wlcg-doma, [Hosted on CERN Gitlab; accessed 22-September-2023

[7] *WLCG Site Network Monitoring Project*, https://gitlab.cern.ch/wlcg-doma/site-network-information, [Hosted on CERN Gitlab; accessed 22-September-2023

[8] A. Aimar, A. Aguado, P. Andrade, J. Delgado, B. Garrido, E. Karavakis, D. Marek, L. Magnoni, *MONIT: Monitoring the CERN Data Centres and the WLCG Infrastructure*, in *EPJ Web of Conferences* (CERN, 2019), p. 8

[9] *Welcome to Easy SNMP*, https://easysnmp.readthedocs.io/en/latest/

[10] S. McKee, *WLCG Site Network Monitoring Campaign*, https://docs.google.com/presentation/d/1sB4xPJPLoLbNnfV0mfwfXmHXzX8lZ8VtC4TSqDXWUMw/edit#slide=id.p1

[11] MONIT Team, *WLCG Site Network Monitoring Dashboard*, https://monit-grafana-open.cern.ch/d/MwuxgogIk/wlcg-site-network?orgId=16&from=now-7d&to=now&var-site=All&var-binning=1m

[12] A.A. Ayllon, M. Salichos, M.K. Simon, O. Keeble, J. Phys.: Conf. Ser **513**, 032081 (2014)