

Data Challenges Monitoring WS

A. Forti, M. Lassnig, S. McKee, R. Di Maria, R. Dona

WLCG Ops Coordination

06 May 2021



Data Challenges Monitoring WS

- Monitoring workshop happened on [27 April 2021](#)
 - a. Overview of the monitoring infrastructure and discussion on how we can run the first Data Challenge this year
- Summary of the possible actions
 - a. Short term (before Summer)
 - i. Collect available site monitoring in a single place
 - ii. Make site monitoring available through automatic procedures (either via push or pull)
 - iii. Create cross-experiment Data Challenge dashboard
 - iv. Start early with "low-percentage" Data Challenge traffic
 - v. Conduct cost-benefit analysis for integration of more data sources (esp. MonALISA & LHCbDIRAC)
 - b. Medium term (before DC#1)
 - i. Technical integration of more monitoring data sources
 - ii. Study performance bottlenecks
 - iii. Discuss technical integration of tools like NetSage



Data Challenge

- Data Challenge 2021 should do 2 things
 - Commission HTTP-TPC
 - Demonstrate we can fill 10% the bandwidth that is requested at HL-LHC scale
- DC2021 aimed for last week of September 2021
 - Date may be adjusted to go later if needed but the aim is for September



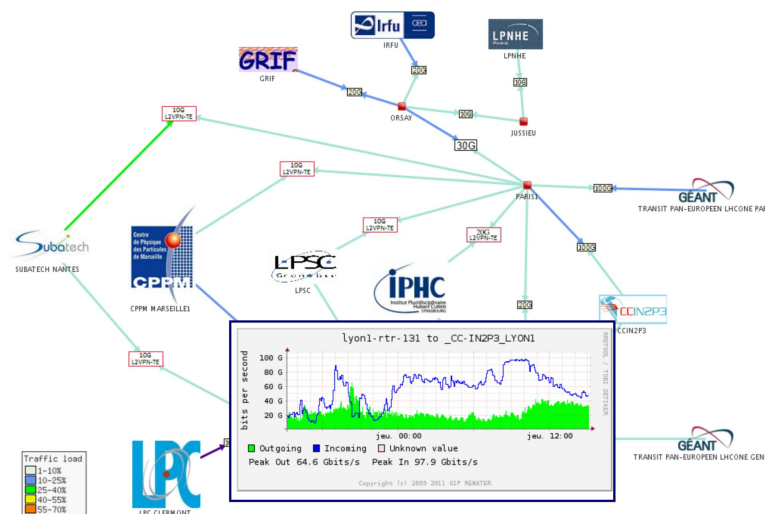
Sites monitoring (a.i)

- The overall aim is to prepare the WLCG infrastructure and demonstrate that we can use the bandwidth effectively
- Whether the 10% of HL-LHC bandwidth is already covered by the current infrastructure or not needs to come out from the monitoring or from appropriate reporting (action a.i. and a.ii.)
 - For sites it is important they help with the information
- We created a repository in gitlab to collect the static information on the sites for a.i.
- Possible to use CRIC need to discuss addition
 - Networking group already discussing network topology integration



Sites Monitoring (a.ii)

- To add metrics in MONIT we need to decide the metrics first
 - At the WS French sites showed their weathermap which has dynamic information already exposed of they ingress and egress
 - They will participate to the creation of the metrics document we are planning not only metrics but also on the best way to expose the information
 - Other sites welcome to participate are welcome to write to
 - doma-data-challenges-development@cern.ch



Dashboard (a.iii)

- At the workshop we agreed that we already have a beefy central infrastructure: MONIT
 - A lot of information is already in it and just has to be accessed and grouped in new dashboards
 - Starting from CMS/ATLAS which have the most similar schema
- Access to WLCG Grafana Org, Data Challenges folder:
 - <https://monit-grafana.cern.ch/dashboards/f/qY7d-gjMz/data-challenges>
 - Starting with FTS based data sources
 - Adding enriched data from rucio



How to run challenges (a.iv.)

- Decided to use the experiments frameworks
- The experiments will carry out their usual activities
- The DC2021 injection of data will be done centrally with coordinated effort
- Requested already resources for DOMA DC
 - WLCG Doma Openstack project created
 - This is going to host the machines that will run the tests etc.
 - Resources are permanent so that DC activities can continue also in between challenges
 - for example to prepare for future challenges and do ad hoc network testing
 - Repo to host testing code:
<https://gitlab.cern.ch/wlcg-doma/data-challenge-2021>
 - JIRA to track the activities
 - <https://its.cern.ch/jira/projects/DOMATPC>



Cost Benefit analysis (a.v.)

- LHCb/Alice need discussion (and work) on how to import their data and massage them to have a uniform dashboard
 - And how to run the challenges from centralized infrastructure
- Further actions:
 - No practical outcome yet



Links

- [HL-LHC network needs and data transfer challenges](#)
- [Data Challenges planning docs](#)
- <https://its.cern.ch/jira/projects/DOMATPC>
- <https://gitlab.cern.ch/wlcg-doma>
- doma-data-challenges-development@cern.ch
- <https://mattermost.web.cern.ch/wlcg-gdb/channels/wlcg-data-challenges>

