

# Data Challenges Monitoring WS

---

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

GDB

12 May 2021



# DC2021

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



# How to run challenges

- The experiments will carry out their usual activities
- DC will use the experiments frameworks
  - Activity should appear as any other experiment activity
    - Named “Data Challenge”
  - Can make it easier to avoid interfering with production
    - Throttling **an internal activity** easier
- The DC injection of data will be done centrally with coordinated effort
- Build up to the final week over the year
  - We will need volunteer sites particularly T1s



# Resources

- 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
  - Need to discuss access to these resources
    - For now DC dev team
- Use existing rucio01 account already has a certificate
  - Need to be enabled by the experiments to run transfers
    - CMS and ATLAS already on board with this
- 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>



# 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



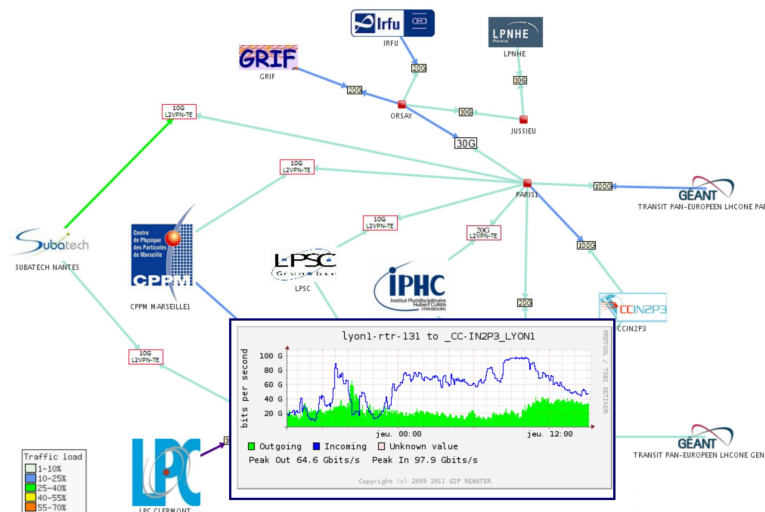
# Sites monitoring (a.i)

- Aim is to prepare the WLCG infrastructure and demonstrate that we can use the bandwidth effectively
- Is the 10% of HL-LHC bandwidth is already covered by the current infrastructure?
  - This 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**
- For reporting we will use gitlab to collect sites information
  - Host site templates (network monitoring web page templates) and provide a guide document for how to flesh out the template.
    - Target is having them ready by the end of next week.
    - [IRIS-HEP/OSG Network milestones](#) doc as a baseline
- Possible to use CRIC for links to monitoring
  - WLCG Networking group already discussing network topology integration in CRIC



# 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](mailto: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





# xrootd Monitoring

- xrootd was highlighted as problematic in MONIT
  - Agreed the information should come from the clients and not the servers
- Affects mostly Alice and CMS
- Two xrootd flows
  - One for ALICE, where we receive some aggregated summary of the transfers from the clients.
  - GLED collectors, where we receive a message per transfer from the servers.
- Information content completely different
  - Alice is essential src,dst,timestamp, MB transferred
  - GLED looks like a dump of anything the servers can report
- Borja Garrido volunteered to restructure this
  - We start with Alice because they already use a client approach
    - Might learn something from it
  - But we will need also CMS input



# Packet Marking

- Activity led by the Networking Technical Working group
- There was a kickoff discussion meeting last week
  - <https://indico.cern.ch/event/1035822/>
- Plan is to integrate the activity in the data challenges if something can be tested by September
  - On selected channels
  - Not interfering with DC itself



# 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](mailto:doma-data-challenges-development@cern.ch)
- <https://mattermost.web.cern.ch/wlcg-gdb/channels/wlcg-data-challenges>
- [IRIS-HEP/OSG sites network monitoring document](#)

