Data Challenge 2021 Monitoring Review

WLCG Ops Coordination

Borja Garrido & Rizart Dona

on behalf of the WLCG DOMA Network Data Challenges Working Group

11 November 2021

Data Challenge 2021

- Network Data Challenge 2021 goals
  - Commission HTTP-TPC
  - Demonstrate we can fill 10% the bandwidth that is requested at HL-LHC scale
- Relevant presentations given
  - GDB, 13 Oct (https://indico.cern.ch/event/876794/)
  - WLCG Ops, 14 Oct (https://indico.cern.ch/event/1085877/)
  - GDB, 10 Nov (https://indico.cern.ch/event/876795/)

Monitoring Main Goals

- Provide a set of useful dashboards & tools for the Monitoring of the DC activity
- Leverage existing WLCG know-how & integrate new functionality
- Requirements
  - Visualizations should be based in commonalities across experiments
  - Avoid experiment specific visualisations and labels

Documentation

Infrastructure

- Existing MONIT infrastructure used
- Grafana instance, dashboards are under the WLCG Organization
- Data Sources
  - FTS Aggregated data source → represents hourly aggregations of FTS transfers
    - Most panels are based on this data source
    - Historical accounting data
  - FTS Raw data source → represents individual FTS transfers and their state changes
    - Only used in two panels
    - Fit for real time monitoring
  - WLCG Aggregated data source → contains data from the FTS Aggregated data source and the Xrootd Aggregated data source
    - Used to present a general view of the WLCG traffic (thus, xrootd traffic is also included)
- FTS based data sources chosen since they are more complete and capture the traffic from the two main experiments that participated, namely ATLAS & CMS

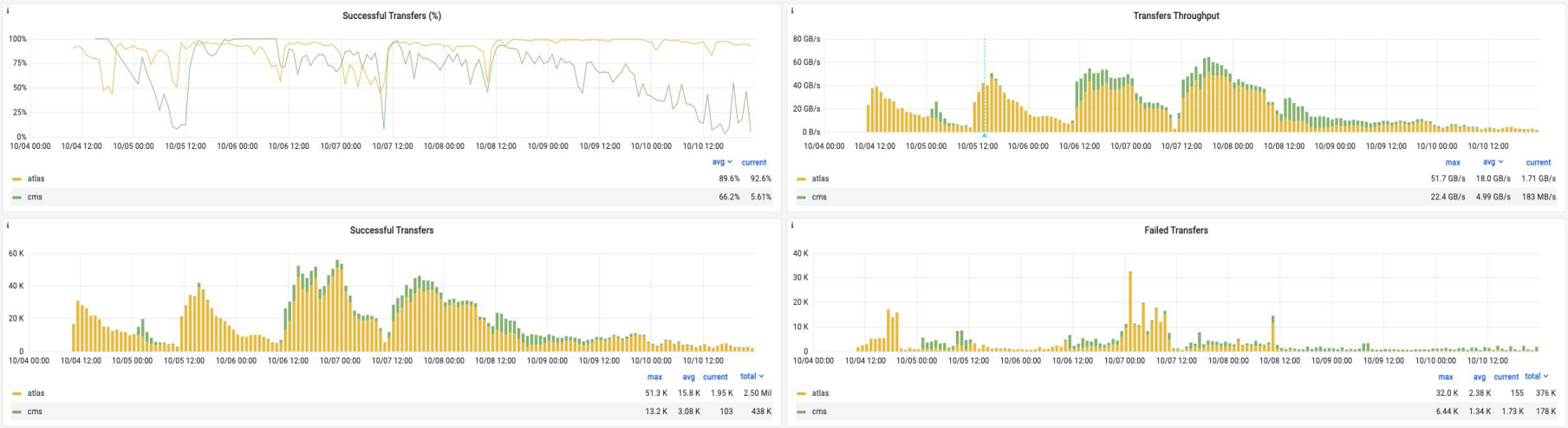
Dashboard (I)

- The dashboard serves as the central monitoring portal for all experiments that participated in the DC21
  - Includes some special panels for tape monitoring
- Multiple panels with varying functionalities, the core ones are:
  - Single value metrics (success rate of transfers, average throughput, volume transferred, etc.)
  - General plots that display the trends of metrics of interest
  - Pie charts that display the transfers breakdown to certain fields of interest
  - Error & failure plots, pie charts and links to the failed transfers log files
  - Tape traffic specific metrics and plots
  - Custom throughput tables for pairs of source/destination sites
  - Efficiency/Volume/Throughput matrices for the ATLAS, CMS & LHCB

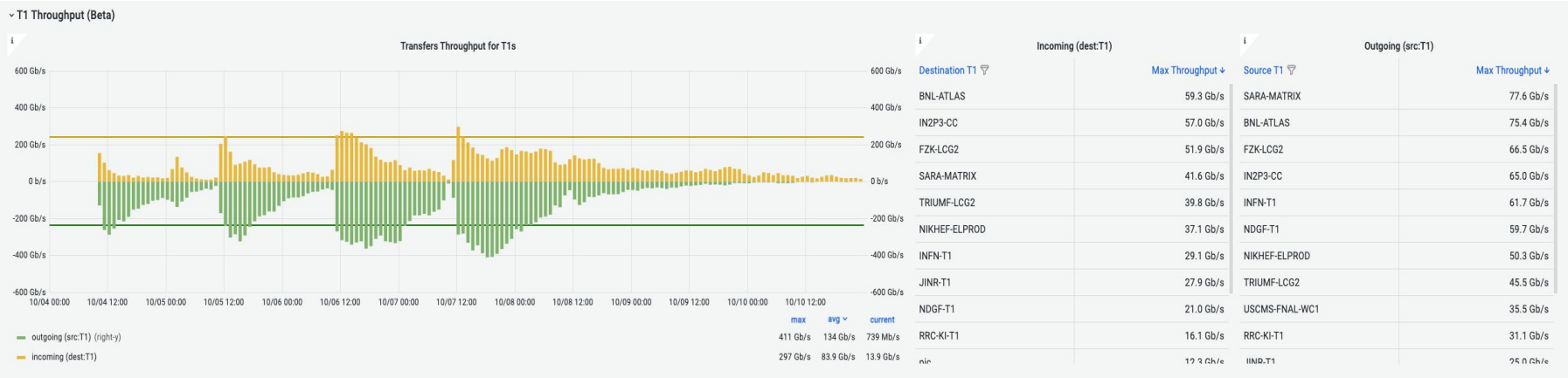
Dashboard (II)

- Additional panels implemented → insight on the network impact of the DC21 activity to WLCG in general, and specifically to T1s that participated:
  - WLCG FTS & Xrootd throughput plot that displays the DC21 activity along with the experiments' activity that is not connected to DC21
  - T1 throughput plot and tables for incoming and outgoing throughput
  - Breakdown of T1 throughput plots for each individual T1 that participated
- Extensive filtering capabilities → limit or expand the views from drop down menus & ad-hoc filtering

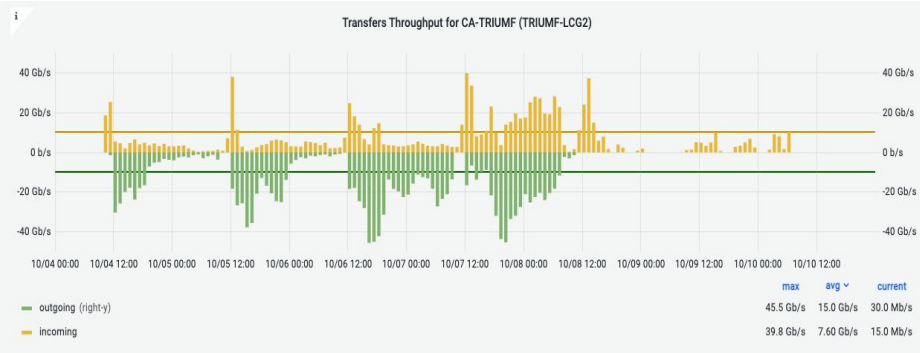
Dashboard (General Plots)

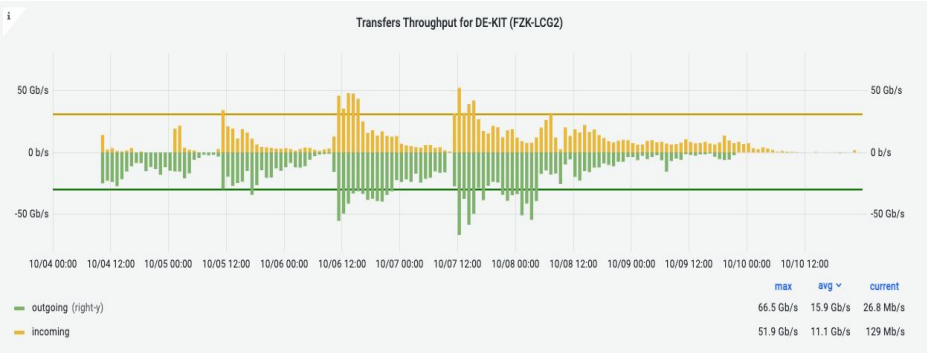


Dashboard (All T1s)

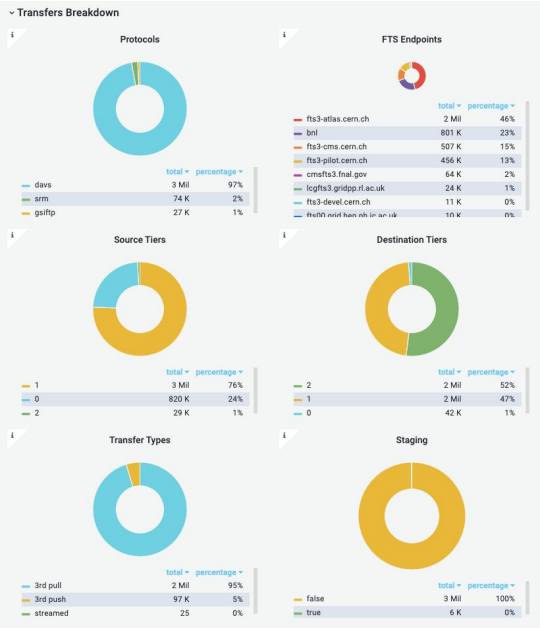


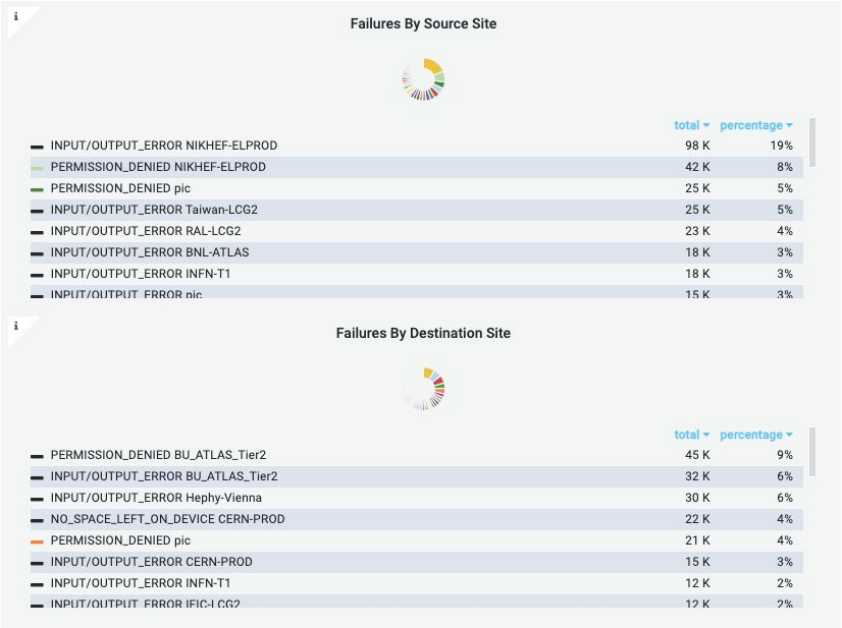
Dashboard (Individual T1s)



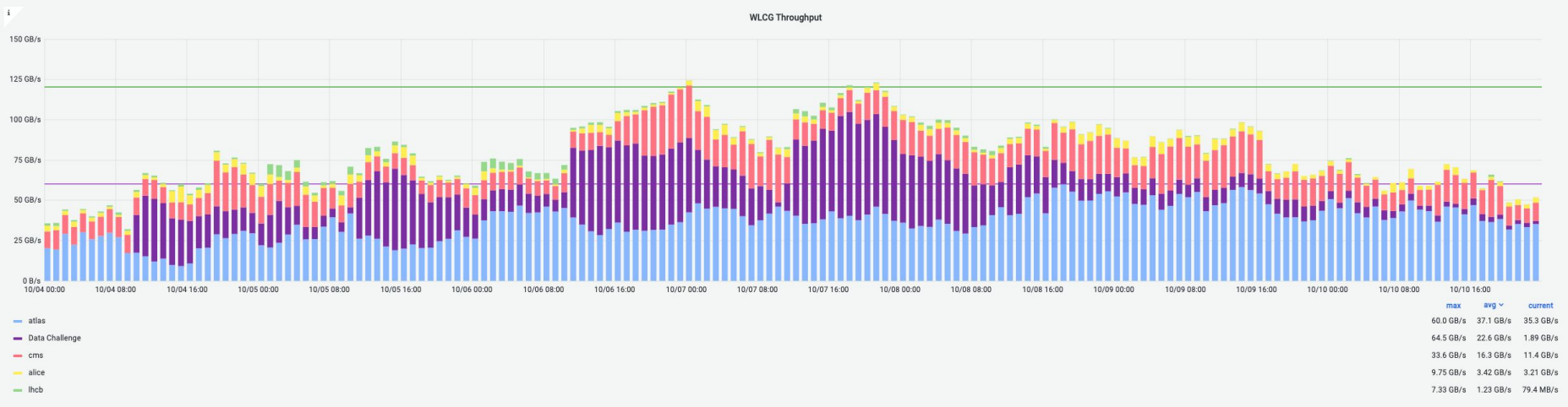


Dashboard (Metrics & Errors)





Dashboard (WLCG & DAC21)



Problems Found

- Throughput calculation for FTS data
  - Approximation based on transfers bytes / visualisation bin
  - Good enough for accounting, confusing for regular operations
- Missing fields for some experiments
  - Labeled as UNKNOWN
  - No tier information for some CMS FTS data
- Incomplete set of labels for LHCb FTS transfers
- Incomplete XrootD data
  - Lots of missing fields, hard to integrate in a FTS based dashboard
  - Not clear what the data represents
- Little/Non existent documentation about existing data sources and their fields

Future Steps - Dashboard

- DC21 dashboard has a lot of functionality/panels → useful for users
- ..but, at the same time it feels too crowded, some fragmentation is needed
  - Migrate the “real time” monitoring panels and build a new WLCG dashboard based on the raw data sources (e.g. FTS team already makes use of this data)
  - Migrate certain sections to new dashboards and interlink them
    - Provide somehow a dashboard-wiki for WLCG
- Generalise panels that are customized for DC21
  - The dashboard views should be useful and easy to analyse independently for all experiments

Future Steps - Data Sources

- Make sure no XrootD packets are lost
  - Currently based model is losing around 60% of packets
  - Work has started in collaboration with OSG to address this
- Identify the minimum required set of labels and make sure they are present
  - For regular operations view
  - For accounting purposes
- Curation of the data to identify missing information

Lessons Learned (I)

- DC collaboration was really useful to identify the Monitoring needs
  - It put working together different roles of people, all of them needed for a good outcome
    - Data experts
    - Tool developers
    - Monitoring managers

- Identifying the requirements needs to be done top to bottom
  - What to visualise (Dashboard)
  - Which filters to have on top
  - Where are the inconsistencies
  - Can these labels be provided
  - How to better integrate within the Monitoring tools

Lessons Learned (II)

- Management required
  - Every user may have different views on what's needed to be visualised
  - There should be some entity (WLCG?) discussing and deciding what needs to be added into the dashboards
  - Small group of people allowed to do changes (otherwise it becomes hellish to manage)

- Operations VS accounting
  - Data may slightly vary in usefulness between both
  - Not a good idea to mix views

Summary

- Overall good outcome of the Monitoring capabilities produced for the DC
- Monitoring can still be squeezed to provide more benefits
  - Will require some clear goals/achievements
- There's still a lot of work to be done
  - Requires involvement of different parties
  - Needs to be properly planned
    - Address first the bigger issues that will provide most benefit

Thanks all for attending!
Questions & Answers