

Data Challenge Monitoring Mini Workshop

Alessandra Forti, Mario Lassnig, Shawn McKee

DOMA General

28 April 2021

<https://indico.cern.ch/event/1032952/>



Summary



We had a very productive workshop yesterday

<https://indico.cern.ch/event/1027287/>

Objective Prepare the network infrastructure for the bandwidth required by HL-LHC experiments

Step 0 Survey the status quo

Contributions

Overview of the (data) challenge

Existing experiment dashboards: ALICE, ATLAS, CMS, LHCb

Central CERN MONIT monitoring infrastructure

FTS

NetSAGE

Network R&D working group

Sites input (CERN, IN2P3, RAL)

Data challenge monitoring overview





How to deal with parallel/on-top experiment activities

Interference between data taking and DC#1


Timing of the data challenge (1st LHC beam in September)

How to discriminate regular experiment activity from DC#1 in the monitoring



DCs Monitoring

- Most important part to get right
 - What we do now is the baseline for future challenges
 - Baseline for future data challenges
- There is already a large MONIT infrastructure that experiments use to various degrees
 - Experiments have their own dashboards
 - Use of the central monitoring infrastructure is uneven
 - Even when using same tools might not have the same information included
 -
- Getting a more uniform monitoring is the work that needs to be done
 - Even if not all at once
 - Not all the experiments? not all the information? not yet?



4

Experiments dashboards



Shared MONIT pipeline for ATLAS and CMS

XRootD monitoring (CMS AAA, ALICE MonALISA)

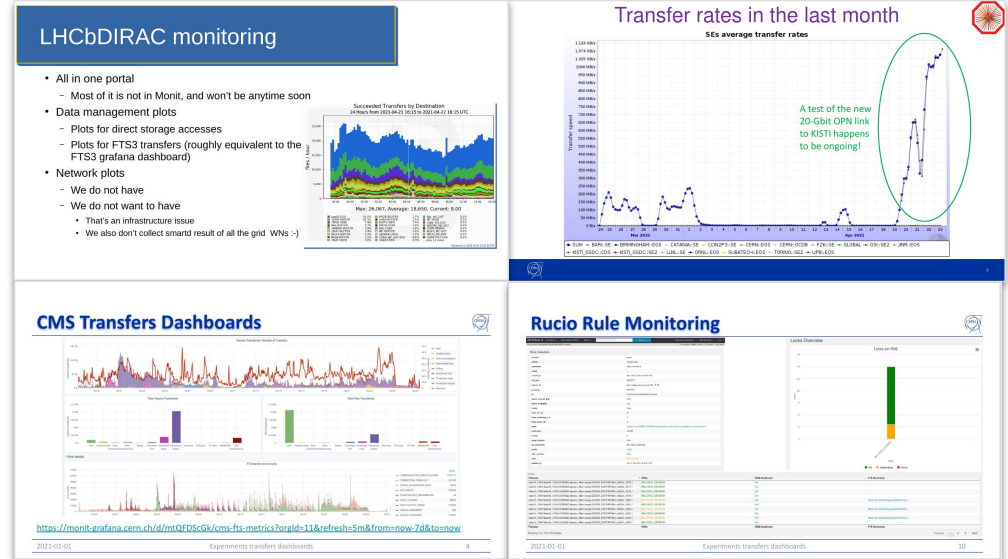
Export of MonALISA and LHCbDIRAC data sources

Tape monitoring vs. Tape buffer monitoring
Data vs visualisation

Delegate network as an infrastructure concern

How to deal with fair share of network

Experiment view of DC transfers can be useful (e.g., Rucio Rules monitoring)



Central MONIT infrastructure



Hosts specific data for ATLAS & CMS

Also includes LHCONe, LHCOPN and perfSONAR metrics

Enriching WLCG monitoring with more experiment details

How to merge FTS and XRootD monitoring

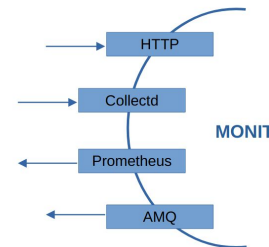
Client-side XRootD monitoring

Server-side monitoring feasibility

Correlate client usage with known network limits

Using MONIT - Ingestion

- **You can send metrics, alarms, or logs**
- **Several endpoints available**
 - You send your data via HTTP
 - You use or write a Collectd plugin
 - You expose Prometheus data for scraping
 - You put your data in a Messaging Broker
 - (from outside CERN only AMQ is supported)
- **Free to choose your client**



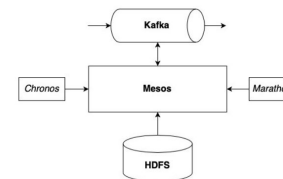
27.04.2021

MONIT Data and Usage

11

Using MONIT - Processing

- **Optional step mainly used for:**
 - Enrichment, Aggregation, ETL
- **Most jobs are Spark based**
 - Can run in streaming or batch mode
 - Few Logstash jobs for simple cases
- **Managed by MONIT or Users**
 - Today MONIT runs jobs for WLCG use cases and large Experiment use cases



27.04.2021

MONIT Data and Usage

14

Information about partial transfers (Waste calculation)

Integrate optimizer decisions into central monitoring

Messages summary

- FTS uses directory queues for recovery & inter-process communication
- When monitoring messages are enabled:
 - A new queue is managed: *monitoring*
 - A dedicated component, handling this queue and sending to ActiveMQ
 - There are 4 types of messages sent in this queue
 - Transfer Start, Transfer Complete, Transfer State, Optimizer



Types of FTS Messages

- **Transfer Start** → `transfer.fts_monitoring_start` (ActiveMQ topic)
 - Created only by the transfer agent at start of transfer
- **Transfer Complete** → `transfer.fts_monitoring_complete`
 - Created only by the transfer agent at end of transfer
- **Transfer State** → `transfer.fts_monitoring_state`
 - Created by QoS and Transfer daemons when a file state is changed ([Transfer Statemachine](#))
- **Optimizer** → `transfer.fts_monitoring_queue_state`
 - Created by the Optimizer when adjusting parameters (# of actives, EMA, throughput)

Complete format in [FTS Messaging documentation](#)



Improve knowledge of actual network topology usage

Identify performance bottlenecks (e.g., by organization, continent, ...)

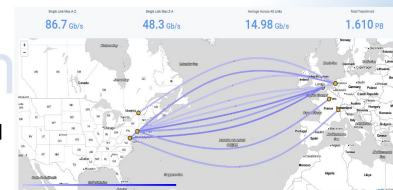
Technical integration with MONIT

Possibility to integrate (majority) of WLCG sites

NetSage ANA:

<http://ana.netsage.global>

- Current and historical SNMP data
- Bandwidth patterns across all ANA links



4

What NetSage Does Best

- Answers questions asked by network engineers and network owners
- Human-readable summaries and patterns
- Gives people the higher-level pattern so they can narrow down a time frame and then use local tools that have more detail
- Simplifies and makes accessible basic data

5

Network R&D



Two aspects

Help understand behaviour of network

Help troubleshoot

Refresh perfSONAR

Investigate network setups (robustness, security, age, misconfigurations, ...)

XRootD Packet marking incoming

Focus on low-hanging fruits, keep low complexity and trustable

Proposed Site Monitoring Page

We need sites to provide us with a site network information we can use to understand the results of our data challenges. My proposal for discussion is that sites create a web page with three sections:

First section should include links to real-time network monitoring that, **at a minimum**, provide the site ingress and egress network rate (MBytes/sec)

- More detailed monitoring is welcome and encouraged
- Adding descriptions of the monitoring technology and characteristics is strongly recommended

Second section should provide a summary of the site network: overview of equipment (vendor/model) and a description of the LAN and connectivity to the WAN.

Third section (optional?) should provide site network diagrams showing how storage and compute are connected to the LAN and how the WAN is reached.

DC Monit-Mini-wkshop 2021 5

Data Challenge Preparation: Possible Areas of Work

Here is a discussion list of possible areas of work to pursue for the DCs:

- Create complete **example** network monitoring page for a site
- Aggressively pursue problems identified by perfSONAR monitoring
 - Firewalls, packet-loss, low throughput, flapping routing, etc
- Develop and deploy script to check and recommend host network tuning
- Work with sites to get appropriate network monitoring page in place
 - Initial focus should be on the biggest sites
- Work with R&E networks to identify suitable monitoring resources
- Centralize data from site and R&E network monitoring
- Prototype and deploy packet marking as possible for this Fall's Challenge
- Begin traffic shaping testing and deployment (likely a focus for 2023).
- Implement and test Network Orchestration capabilities (GNA-DIS)
- Create site best practices documentation regarding network monitoring, tuning and architecture

Others?

DC Monit-Mini-wkshop 2021 9

Next steps



Short term (before Summer)

- Collect available site monitoring in a single place

- Make site monitoring available through automatic procedures (either via push or pull)

- Create cross-experiment Data Challenge dashboard (See Data Challenge update)

- Start early with "low-percentage" Data Challenge traffic (See Data Challenge update)

- Conduct cost-benefit analysis for integration of more data sources (esp. MonALISA & LHCbDIRAC)

Medium term (before DC#1)

- Technical integration of more monitoring data sources

- Study performance bottlenecks

- Discuss technical integration of tools like NetSage