Unified Experiment Monitoring

Ewoud Ketele[1*], *Domenico* Giordano[1**], *Panos* Paparrigopoulos[1], and *Maarten* Litmaath[1]

[1]CERN, European Laboratory for Particle Physics, Geneva, Switzerland

Abstract. The Unified Experiment Monitoring (UEM) is the project in WLCG with the objective to harmonize the WLCG job accounting reports across the LHC experiments, in order to provide aggregated reports of the compute capacity used by WLCG as a function of time. This accounting overview of all LHC experiments is vital for the strategy planning of WLCG and therefore it finds the strong support of the LHC Experiments Committee (LHCC). However, creating common overviews is challenging, due to the different internals of each experiment monitoring system and also due to the long time-scale of the reports to cover at least a decade of data. These monitoring systems evolved largely independently over time, implying that the UEM project has to design and implement different approaches to couple the multiple data sources within the CERN IT monitoring tools which will be used. Last but not least, the different terminologies have to be aligned into a useful and coherent set. This contribution summarizes the motivations of the project, the challenges faced, the design adopted to overcome them, and the presentation of the state of the art.

1 Introduction

The Worldwide LHC Computing Grid (WLCG) [1] collaboration is a vital component in the effort to distribute, store, and analyze the vast amounts of data generated by the four main LHC experiments at CERN. With a mission to support the research of over 12,000 physicists worldwide, WLCG consists of an enormous distributed computing infrastructure, also known as the grid, of approximately 1.4 million CPU cores and more than 2 exabytes of storage. The grid spans over 147 sites [2] and 42 countries. These sites are distributed over multiple tiers. Tier-0 is the CERN computing center which provides about 20% of the compute capacity. Tier-1 sites have sufficient storage capacity to serve as a backup for the CERN computing center and have round-the-clock support for the Grid. Tier-2's are typically universities and other scientific institutes, which can store sufficient data and provide adequate computing power for specific analysis tasks. They handle analysis requirements and proportional share of simulated event production and reconstruction [3]. This infrastructure handles over 2 million jobs per day, and global data transfer rates can exceed 60 GB/s. Monitoring this infrastructure is important to support operations activities in the experiments. This happens at different levels: the LHC experiments monitor the jobs that they submit to the grid; the underlying grid infrastructures EGI (European Grid Infrastructure) and OSG (Open Science Grid) monitor the sites to which jobs are submitted; and WLCG Accounting provides a

*e-mail: ewoud.ketele@cern.ch
**e-mail: domenico.giordano@cern.ch

Table 1. critical metrics that are available in all four LHC experiments operational monitoring

Metric	Note
Number of running cores per hour	How many CPU cores were used by jobs in the interval of an hour
Number of running jobs per hour	How many jobs were running in the interval of an hour
Wall-clock time	Duration of a job *times* the number of cores used
Wall-clock work	Wall-clock time *times* the HEPScore23 score of the CPU used

Table 2. Aggregation levels over which the metrics from Table1 can be aggregated

Aggregation level	Possible values of the aggregation level
Tiers	Tier-0; Tier-1; Tier-2; Tier-3
Activities	Analysis; Monte Carlo simulations; Reconstruction; User jobs; "other"
Resource type	Grid; HPC; public clouds; "other"

global overview of the resources provided and consumed. Due to the decentralized nature of the monitoring, some gaps are left open and some optimizations can still be pursued. Both the LHC Experiments Committee (LHCC) and WLCG Operations Coordination suggested that a more unified approach and a larger adoption of the centralized CERN IT monitoring infrastructure, MONIT [4], would benefit WLCG operations. It is the task of the Unified Experiment Monitoring (UEM) project to assist LHC experiments with migrations to MONIT and to provide a global picture of the resources used by them. This paper is divided into three parts. First, the goals of UEM are explained. A second section explains how the different technical challenges encountered were overcome. Finally, an outlook on the next steps in the project is presented.

2 Unified Experiment Monitoring

When job monitoring is lacking on a WLCG level, questions such as "How many CPU cores were used at Tier-1 sites last year?" cannot be answered easily. The data that is needed to answer such questions surely exists and is available in the monitoring infrastructures of the LHC experiments. However, access to this data was not readily available to WLCG in October 2022. Even with access to the data, creating a unified job monitoring report was hard because a range of definitions is used for common concepts across the four experiments. On top of that, operational monitoring at experiments follows the experiment needs. For example, monitoring dashboards of the CMS [5] and LHCb [6] collaborations do not display resources under Tier-0 or -3 respectively; and only the ATLAS [7] collaboration monitors the jobs per resource type (Grid, High Performance Computing (HPC), public clouds, ...). The UEM project assisted in the migration of the monitoring infrastructure of LHCb and created additional pipelines to ensure access to the ALICE monitoring data sources. It also has defined a set of common-but-critical metrics to be published in the unified job monitoring dashboards. These metrics are shown in table 1. The aggregation levels over which these metrics can be collected are shown in table 2.

In the first phase of the project, the focus was on making available the metrics aggregated by tiers. Example plots are shown in Figures 1 and 2.

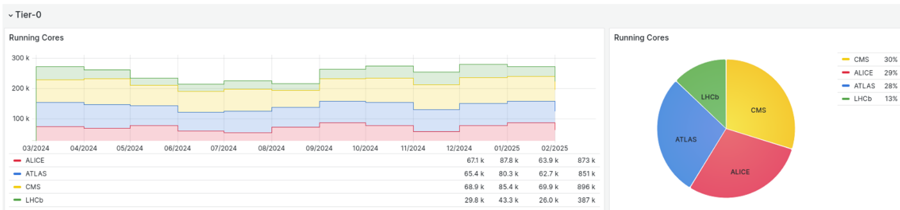

Figure 1. The Grafana visualization showing the number of running cores per hour per LHC experiment for Tier-0. (ALICE: red, ATLAS: blue, CMS: yellow, LHCb: green)

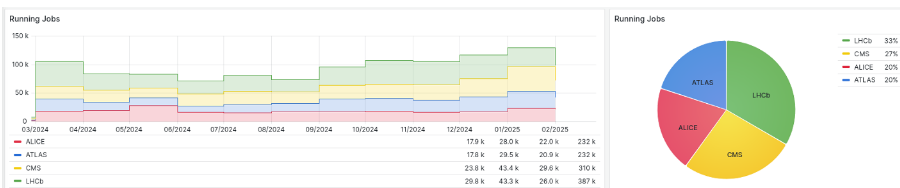

Figure 2. The Grafana visualization showing the number of running jobs per hour per LHC experiment for Tier-0. (ALICE: red, ATLAS: blue, CMS: yellow, LHCb: green)

3 Technical challenges

3.1 Accessing job monitoring data

Getting access to the required data from ATLAS and CMS was straightforward on a technical level since both were already using the MONIT infrastructure for their monitoring operations. However, at the start of the project, ALICE and LHCb were not (yet) using this common infrastructure. To make the required data available, new data pipelines needed to be created.

3.2 Migration of LHCb monitoring into MONIT

LHCb was already planning to make the switch to a Grafana [8] instance provided by MONIT for their operational monitoring. The LHCb OpenSearch [9] and MySQL [10] databases that hold the job monitoring data were added as data sources to the MONIT Grafana organizations for both LHCb [11] and WLCG [12][1] as is shown in Figure 3. This allowed the translation of the existing monitoring dashboards to Grafana equivalents by converting all Lucene and SQL queries used in the earlier monitoring dashboards into Grafana queries. A specific task taken care of by the UEM project was the migration of the dashboards used by grid operations shift crews. On the WLCG side, this migration has allowed LHCb job and core numbers to be obtained automatically for LHC Experiments Resource Review Board (RRB) reports [13]. A dashboard that shows the visualizations for the RRB reports is now also available in the WLCG Grafana organisation [14]. Whereas the RRB report visualizations are only created twice a year, this dashboard is always showing the most recent data due to its live nature.

[1]MONIT Grafana organisations need CERN credentials to access

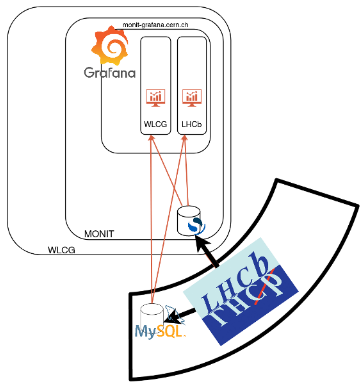

Figure 3. Data pipelines that were created (in red) to make LHCb job monitoring data stored in OpenSearch indices and MySQL tables available in Grafana organizations for LHCb and WLCG.

3.3 Making ALICE monitoring data available in MONIT

Since the ALICE collaboration [15] does not plan to migrate to the MONIT infrastructure, the sole focus was to make the data for the targeted metrics available in the WLCG Grafana organization. There were some constraints to be navigated: the solution had to be fault-tolerant, simple and had to make use of the existing software and APIs on the ALICE side. A Python script [16][2] to extract the necessary data from the ALICE MonALISA [17] service was developed. This script extracts the data hourly using GitLab [18] CI/CD pipelines. As can be seen in Figure 4, the extracted data is then sent to an OpenSearch index that is managed by MONIT and which is made available in the WLCG Grafana organization.

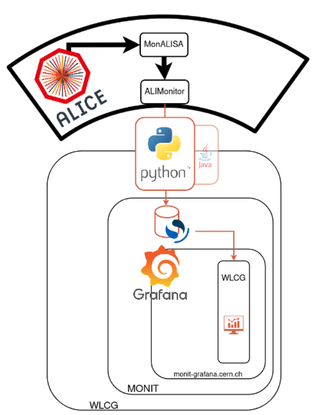

Figure 4. The data pipeline developed by UEM to access the job monitoring data from ALICE. The extracted data is stored in an OpenSearch index that is made available in the WLCG Grafana organization.

[2]CERN credentials are required to access the repository

3.4 Dashboards-as-code with Jsonnet

Creating unified dashboards is technically straightforward once all the data sources are made available to the WLCG Grafana organization. However, creating dashboards manually can be error-prone and, especially in the development phase, repetitive. On top of that, there is no way to store the knowledge that went into the making of a given dashboard. Since the visualizations of Grafana dashboards can be easily edited, this risks that the dashboards, over time, are no longer showing what they were developed for. It is possible to add the JSON representation of a dashboard to a Git repository to counter this, but JSON files are not easily versioned, as changes to dashboards typically lead to large differences in the JSON configuration. Instead, the configuration language Jsonnet [19] and the Grafana extension Grafonnet [20] are used to create the dashboards programmatically. The Grafonnet extension allows specifying every element of a Grafana dashboard, such as data sources, queries to these data sources and style sheets to be used by the dashboard, in a programmatic way. Now, consistency and, with the use of a Git repository, versioning are ensured. The current workflow, pictured in Figure 5, uses GitLab pipelines to build dashboards and push them to the MONIT Grafana instance on every commit and merge request.

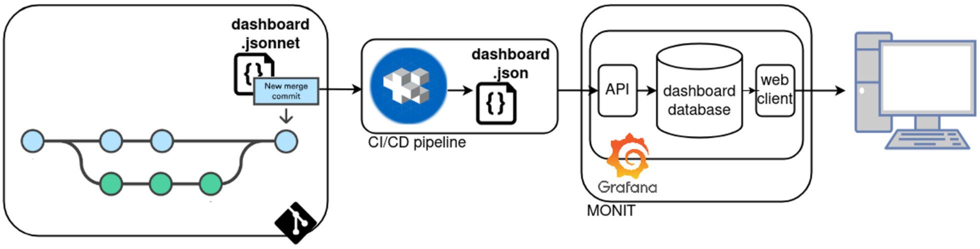

Figure 5. Changes in the dashboard definitions in the GitLab repository trigger a CI/CD pipeline. This pipeline first rebuilds the JSON versions of the dashboards and pushes those JSON files to the Grafana instance using Curl.

4 Conclusion

Merging the job monitoring of the LHC experiments to create unified WLCG level job monitoring dashboards is not easy. It requires that all necessary data sources are made available in the MONIT infrastructure, that the data structures and definitions of all experiments are well understood and that common definitions are created. However, for the first time, it is now possible to have a WLCG-level job monitoring overview in the WLCG Grafana organisation [21]. The infrastructure that makes this possible is modular and leans as much as possible on the existing LHC experiment infrastructures. Hence, creating additional views or adding more metrics does not require more infrastructure to be build. Rather, the existing framework is easy to expand because of its ligth footprint. To speed up the dashboard development process and to improve versioning, the dashboards were made with Jsonnet and its Grafana extension Grafonnet. In the next steps, the existing dashboards, showing the metrics across WLCG tiers, will be validated with the experiments and new views, showing metrics across activities and resource types, will be created. For these new views, challenges in the availability of certain metrics and the matching of the different activity definitions used by the LHC experiments still remain as can be seen in Table 3.

Table 3. Activities across LHC experiments

LHC experiment	Activity label	Proposed UEM label
ALICE	MC production	Monte Carlo
ALICE	Organized analysis	Analysis
ALICE	Raw data reconstruction	Reconstruction
ALICE	Organized analysis run 3	Analysis
ALICE	User	User jobs
ATLAS	Data processing	?
ATLAS	Event index	?
ATLAS	Group analysis	Analysis
ATLAS	Group production	?
ATLAS	MC event generation	Monte Carlo
ATLAS	MC merge	Monte Carlo
ATLAS	MC reconstruction	Monte Carlo
ATLAS	MC resimulation	Monte Carlo
ATLAS	MC simulation	Monte Carlo
ATLAS	testing	?
ATLAS	Userstatus	analysis
CMS	Analysis	Analysis
CMS	Cleanup	?
CMS	Harvesting	?
CMS	LogCollect	?
CMS	Merge	?
CMS	Processing	?
CMS	Production	?
LHCb	MCFastSimulation	Monte Carlo
LHCb	MCMerge	Monte Carlo
LHCb	MCReconstruction	Monte Carlo
LHCb	MCSimulation	Monte Carlo
LHCb	Merge	?
LHCb	Sprucing	?
LHCb	User	User jobs
LHCb	WGProduction	?
LHCb	LbAPI	?

References

[1] WLCG (2021) The Worldwide LHC Computing Grid, Accessed 28.02.2025, http://wlcg-public.web.cern.ch
[2] WLCG-CRIC, Accessed 28.02.2025, http://wlcg-public.web.cern.ch/resources
[3] WLCG tiers, Accessed 28.02.2025, https://wlcg-public.web.cern.ch/tiers
[4] A. Aimar et al (2019) MONIT Monitoring the CERN Data Centres and the WLCG Infrastructure, EPJ Web of Conferences **214**, 08031, https://doi.org/10.1051/epjconf/201921408031
[5] The CMS Collaboration (2008) The CMS experiment at the CERN LHC, JINST 3:S08004, https://doi.org/10.1088/1748-0221/3/08/S08004
[6] The LHCb Collaboration (2008) The LHCb detector at the LHC, JINST 3:S08005, https://doi.org/10.1088/1748-0221/3/08/S08005

[7] The ATLAS Collaboration (2008) The ATLAS experiment at the CERN large hadron collider, JINST 3:S08003, https://doi.org/10.1088/1748-0221/3/08/S08003
[8] Grafana, Accessed 28.02.2025, https://grafana.com
[9] OpenSearch, Accessed 28.02.2025, https://opensearch.org
[10] MySQL, Accessed 28.02.2025, https://mysql.com
[11] LHCb Grafana, Accessed 28.02.2025, https://monit-grafana.cern.ch/d/Q78h6E-nz/home?orgId=46
[12] WLCG Grafana, Accessed 28.02.2025, https://monit-grafana.cern.ch/d/000000425/default?orgId=20
[13] LHC Resources Review Boards (RRB), Accessed 28.02.2025, https://committees.web.cern.ch/LHCRRB
[14] E. Ketele, RRB Grafana dashboard, Accessed 28.02.2025, https://monit-grafana.cern.ch/d/GnbGjuPVz/rrb-work-in-progress?orgId=20
[15] The ALICE Collaboration (2008) The ALICE experiment at the CERN LHC, JINST 3:S08002, doi:10.1088/1748-0221/3/08/S08002.
[16] E. Ketele, uem-alimonitor, GitLab repository, Accessed 28.02.2025, https://gitlab.cern.ch/wlcg/uem-alimonitor
[17] ALICE MonALISA service, Accessed 28.02.2025, http://alimonitor.cern.ch
[18] GitLab, Accessed 28.02.2025, https://gitlab.com
[19] Jsonnet, Accessed 28.02.2025, https://jsonnet.org
[20] Grafonnet, Accessed 28.02.2025, https://github.com/grafana/grafonnet
[21] UEM Unified Experiment Job Monitoring, Accessed 28.02.2025, https://monit-grafana.cern.ch/d/e8b582b5-1c65-430c-96ef-8c9e99ecd2ee1/uem-unified-experiment-job-monitoring