



Contribution ID: 454

Type: **Talk**

Distributed management and processing of ALICE monitoring data with Onedata

Thursday 24 October 2024 17:27 (18 minutes)

Onedata [1] platform is a high-performance data management system with a distributed, global infrastructure that enables users to access heterogeneous storage resources worldwide. It supports various use cases ranging from personal data management to data-intensive scientific computations. Onedata has a fully distributed architecture that facilitates the creation of a hybrid cloud infrastructure with private and commercial cloud resources. Users can collaborate, share, and publish data, as well as perform high-performance computations on distributed data using different interfaces.

Within the ALICE project, we are designing an architecture that live-streams monitoring data from MonALISA data and stores it in Onedata, utilising S3 storage. This data is accessible through the POSIX filesystem HPC, cloud infrastructures, and external MLOps systems (via S3 or REST API). When a computational task requires the data, it is seamlessly transferred and cached at the task's location. Onedata's distributed and multi-protocol nature facilitates the creation of a hybrid data processing infrastructure, where Onedata functions as the data plane. The platform also includes robust security features to safeguard data and metadata from unauthorised changes, ensuring the integrity of the datasets during the final preparation stages. Additionally, Onedata enables long-term archiving of datasets, preserving crucial information for future reference. Data can be structured hierarchically within Onedata, and datasets are annotated with metadata, simplifying the organisation and retrieval of specific information.

Currently, Onedata is used in European projects: EUreka3D [3], EuroScienceGateway [4], DOME [5], InterTwin [6] where it provides a data transparency layer for managing large, distributed datasets on dynamic hybrid cloud containerised environments.

Acknowledgements: This work is co-financed in part supported by the Ministry of Science and Higher Education (Agreement Nr 2023/WK/07) and by the program of the Ministry of Science and Higher Education entitled "PMW".

References:

- [1] Onedata project website. <https://onedata.org>.
- [2] ALICE - A Large Ion Collider Experiment. <https://alice-collaboration.web.cern.ch>.
- [3] EUreka3D: European Union's REKconstructed in 3D. <https://eureka3d.eu>.
- [4] EuroScienceGateway project: open infrastructure for data-driven research. <https://galaxyproject.org/projects/esg/>.
- [5] DOME: A Distributed Open Marketplace for Europe Cloud and Edge Services. <https://dome-marketplace.eu>.
- [6] InterTwin: Interdisciplinary Digital Twin Engine for Science. <https://intertwin.eu>.

Author: Dr ORZECHOWSKI, Michał (AGH University of Krakow, Faculty of Computer Science, Poland)

Co-authors: Dr BALIŚ, Bartosz (AGH University of Krakow, Faculty of Computer Science, Poland); Dr DUTKA, Łukasz (Academic Computer Centre Cyfronet AGH, Krakow, Poland); Prof. KITOWSKI, Jacek (AGH University of Krakow, Faculty of Computer Science, Poland)

Presenter: Dr ORZECHOWSKI, Michał (AGH University of Krakow, Faculty of Computer Science, Poland)

Session Classification: Parallel (Track 1)

Track Classification: Track 1 - Data and Metadata Organization, Management and Access